

## REVIEW

## High throughput strategies for probing the different organizational levels of protein interaction networks

Cite this: *Mol. BioSyst.*, 2013, **9**, 2201

Anders R. Kristensen and Leonard J. Foster\*

Most proteins do not exist as isolated molecules in the cell, but instead serve as nodes of protein interaction networks. A number of techniques have been developed in the last two decades to study protein interaction networks at different levels of detail. Here we describe some of the techniques for characterizing protein interactions and protein complexes on a system-wide scale, focusing especially on newly emerging techniques that use co-migration. These newer approaches have the advantage that no genetic manipulation is necessary, thereby allowing investigation of protein complexes at their endogenous levels in the correct cellular context. Finally, we discuss different approaches for measuring large-scale temporal changes to protein interaction networks, an area that we believe will be one of the frontiers in systems biology in the coming years.

Received 4th April 2013,

Accepted 4th July 2013

DOI: 10.1039/c3mb70135b

[www.rsc.org/molecularbiosystems](http://www.rsc.org/molecularbiosystems)

## 1. Protein interaction networks

Proteins perform most of the structural and catalytic work in the cell but they rarely act alone, instead they form functional units with other proteins that together can do more than the sum of their parts. It is estimated that at least 90% of soluble proteins serve as components of protein complexes in both prokaryotes and eukaryotes, implying that the assembly of protein complexes is a fundamental principle of most, if not all living organisms.<sup>4–6</sup> Protein complexes play central roles in all aspects of cellular functions, including DNA replication, transcription, translation, metabolism, signaling transduction and structural organization.

The study of protein–protein interactions is not a new concept; it has been around for more than a century, even though in the beginning it was far from understood what was being observed. One of the first interactions discovered was trypsin bound to antitrypsin, from which it could also dissociate.<sup>7</sup> Much later came the concept of protein interactions and that these could serve as regulators of metabolism or were critical

elements of signal transduction.<sup>8</sup> In the last two decades it has become apparent that proteins form networks that can be modeled as graphs composed of nodes and edges representing macromolecules and the interactions between them, respectively, where these networks seem to capture or represent phenotypic variation.<sup>9</sup>

When the first protein interaction networks were mapped it was apparent that they were scale free, meaning that many proteins have few interaction partners, whereas few proteins interact with many proteins; proteins that interact with many partners became known as hubs.<sup>10</sup> By further investigating the characteristics of hub proteins based on the correlation of RNA expression among the proteins in complexes, it was realized that they could be divided into two classes: party hubs (also known as intramolecular hubs) and date hubs (also known as intermolecular hubs),<sup>11</sup> although whether this depiction is appropriate is still debated.<sup>12</sup> Party hubs are defined as stable complexes where proteins have higher than average correlating mRNA expression and date hubs were defined as protein complexes with less than average correlating mRNA expression. In other words, party hubs are typically co-expressed with all their interacting partners, whereas date hubs are only co-expressed with a subset of their interacting partners in any

*Department of Biochemistry & Molecular Biology, Centre for High-Throughput Biology, University of British Columbia, Vancouver, BC, Canada.  
E-mail: foster@chibi.ubc.ca*

*Anders R. Kristensen is a PhD student in Biochemistry and Molecular Biology at University of British Columbia. He obtained his Master Degree in 2008 from University of Southern Denmark. His research interest involves how the proteome and interactome of cells changes in response to external perturbation.*

*Leonard J. Foster is an Associate Professor of Biochemistry and Molecular Biology in the Centre for High-Throughput Biology at the University of British Columbia. His research group focuses on developing and applying quantitative proteomic approaches to the study of host–pathogen interactions.*

given condition. It was further realized that date hubs interact with their partners at different times and/or conditions, whereas party hubs are interact with their partners at all times and conditions; put differently, date hubs serve to connect functional modules that are composed of party hubs.<sup>13,14</sup> In addition, it has been observed that date hubs are enriched in signaling domains such as tyrosine kinases, PDZ domains and G $\alpha$  domains, and that mutations in these are more frequently associated with mutations in cancer.<sup>13</sup>

A number of reviews have addressed different aspects of protein interaction networks, including experimental<sup>15,16</sup> and computational approaches<sup>17,18</sup> to measure them, as well as the implications of such networks in disease.<sup>14</sup> The various methods that are commonly used to measure protein interaction networks address different features of the system and understanding these differences is crucial to understanding the data, especially when comparing between datasets. The differences between the features measured are sometimes subtle and are typically not captured in how the data are referred to so here we use a hierarchical nomenclature to describe the protein interaction network, based on the ability different experimental approaches to decipher the levels of the network.

### 1.1. Protein interaction networks and disease

Phenotypic variations in an organism, especially those that result in a disease, can be seen to arise from perturbations of the interaction network. This network view of diseases can therefore often replace the linear causative model that states diseases happen due to failure in one gene. A growing body of evidence suggests that a number of major diseases, *e.g.*, type 2 diabetes and coronary diseases, derive from small defects in a number of genes.<sup>19</sup>

Removal of nodes or edges will result in completely different outcomes of a protein interaction network, since node removal will eliminate not only the function of the protein but also all interactions that the protein participates in, whereas edge removal will only remove some interactions and thereby cause a more subtle effect on the network. This means that mutations in different domains of a protein can affect different edges in a protein interaction network, a principle exemplified by mutations in two domains in TP63 that lead to two different developmental disorders. Mutations in the DNA binding domain cause ectrodactyly ectodermal dysplasia (EEC), whereas mutations in the SAM2 domain, involved in protein interactions, cause ankyloblepharon ectodermal dysplasia (AEC).<sup>20</sup>

Network perturbations are also critical in infectious diseases. One archetypical example is *Salmonella*, a Gram-negative bacterium that secretes a number of virulence proteins through its type 3 secretion systems into the host cell where they subsequently target strategic parts of the host cell's protein interaction network through various means. Among many mechanisms, SopE/E2 and SptP from the bacteria act as guanine nucleotide exchange factors and GTPase activating proteins respectively for the human small G-proteins CDC42 and Rac1; together with another bacterial effector, SopB, these proteins cause a massive actin rearrangement that leads to the

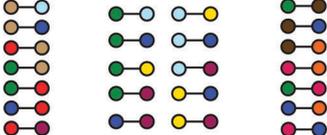
bacteria being engulfed by the host cell.<sup>21–23</sup> Another example is the human immunodeficiency virus (HIV) that blocks part of the host cell antiviral defenses by producing an accessory factor, Vif, which removes a key node of the host protein interaction network. This is accomplished through the hijacking of a host ubiquitin ligase complex that subsequently initiates the degradation of APOBEC3G proteins that normally play a critical role in viral defense by interfering with reverse transcription.<sup>24</sup> More recently, Vif was shown to have a wider impact on the host cell protein interaction network through its recruitment of host CBF- $\beta$  to the ubiquitin ligase complex, disrupting CBF- $\beta$ 's normal interactions with RUNX DNA binding proteins.<sup>25,26</sup> This recruitment helps facilitate Vif folding and is required for binding and poly-ubiquitylation APOBEC3G.

### 1.2. The hierarchy of protein interaction networks

The hub terminology is now widely used to describe the proteins of an interaction network but the nomenclature used to describe how those proteins are organized has not been formalized and is currently quite confusing, with each group inventing their own terms. Defining a new terminology that is universally accepted is challenging but we feel that such an attempt is relevant here in order to appreciate the differences between interactome methods. The idea behind our new terminology is to capture the different features of protein interaction networks that are reported by the various techniques that are commonly used to measure them. The terminology is based on James G. Miller's ideas about general network theory: "The Universe contains a hierarchy of systems, each higher level of system being composed of systems of low levels".<sup>27</sup> Similar to the terms ascribed to the various levels of protein structure, we believe protein interaction networks can also be depicted as having four levels of organization, primary, secondary, tertiary and quaternary (Fig. 1), each of which is addressed by different methodological approaches.

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

The secondary level of organization, sometimes referred to as core complexes,<sup>5,28</sup> are proteins that are always in the same complex(es), though they might not interact directly. Expression patterns for such proteins would be highly correlated but by themselves would not necessarily form fully functional units. In addition, their biochemical role will be irreplaceable since deletion of any member of this group will typically lead to the loss of functional integrity of the whole complex,<sup>28</sup> similar to party hubs. An illustrative example of a secondary interaction is the set of proteins mTOR, mLST8, DEPTOR and Tti/Tel2, which

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

**Fig. 1** Classification of the interactions in the protein interaction network. Pictographic representations of the protein interaction networks hierarchically levels (see text for detail).

are integral to both the mTOR complex 1 (mTORC1) and mTOR complex 2 (mTORC2);<sup>29</sup> these can be identified as a single unit in interactome studies but they are not actually known to all form a complex in isolation.

The tertiary level of organization within protein interaction networks, sometimes referred to as ‘complex isoforms’, is composed of functional complexes, where different accessory (or halo) proteins are added to the secondary interactions, or core complexes, leading to different protein assemblies in the cell.<sup>5,28</sup> The accessory proteins will therefore typically not share common deletion phenotypes, functional classification or cellular localization with the core proteins.<sup>28</sup> These peripherally attached proteins can, *e.g.*, spatially or temporally affect the complex and its functions. This can again be illustrated with mTORC1 and mTORC2, where in mTORC1 the mTOR/mLST8/DEPTOR/Tti/Tel2 core is joined by RAPTOR and PRAS40, while in mTORC2 the core is joined by RICTOR, mSIN1 and PROTOR1/2. The implications of these different accessory proteins joining the core complex are massive since mTORC1 gets activated by amino acids and ATP, is inhibiting during autophagy and positively stimulates translation, whereas mTORC2 can activate several members of the AGC family of kinases including Akt.<sup>29</sup> Different versions of very similar tertiary assemblies can also exist, such as through alternative stoichiometries. This can be illustrated by the proteasome that can exist as a complex composed of one 20S core complex and two 19S regulatory subunits (*i.e.*, the 26S proteasome) or a complex composed of one 20S core complex and one 19S regulatory subunit. In this case the actual proteins comprising the two different assemblies are the same but there is clearly two discrete structures within the cell.<sup>1</sup>

The quaternary level of interactions represents the union of the tertiary protein interactions and is therefore not necessarily present in any given cell but is nonetheless useful for gaining a bigger-picture view of the system. This view of the protein interaction network is generated computationally through clustering a matrix of primary interactions where 1's represent an interaction between two proteins and 0's represent no interaction. In this way proteins that interact with the same set of other proteins will cluster closely together.<sup>30,31</sup> The quaternary level can also be observed in non-saturated AP-MS experiments where not all proteins are tested as baits, leading to what are actually multiple tertiary complexes being identified as a single quaternary complex. To use the mTOR example again, affinity purification of any member of the mTOR/mLST8/DEPTOR/Tti/Tel2 core brings along with it all components of both the mTORC1 and mTORC2 complexes and from such data it is not clear that there are actually two different structures present. Only by a saturating screen where every component of both mTORC1 and mTORC2 are affinity purified and one sees, *e.g.*, that affinity purification of RAPTOR does not co-purify RICTOR can the composition of the two be resolved.<sup>32</sup>

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

## 2. Primary protein interactions

Primary protein interactions are generally measured using *in vivo* assays, with the very large majority of studies having

been conducted in yeast. Studying interactions *in vivo* is attractive since it can be more physiologically relevant and many of the sources of error involved *in vitro* approaches can potentially be avoided (*e.g.*, interactions forced as a result of breaking the cell). The basic concept of *in vivo* techniques is that engineered fusions of the proteins of interest and reporter proteins are co-expressed and if an interaction occurs a detectable output signal will be generated. The most common applications of mass spectrometry typically generate quaternary-level interaction data (see below) but recent methodological and bioinformatic developments have led to great advances in the analysis of cross-linked proteins by mass spectrometry, which generates primary interaction data.<sup>33,34</sup>

### 2.1. *In vivo* techniques

The Y2H assay, developed by Fields and Song more than 20 years ago, is the original *in vivo* method for studying protein interactions.<sup>35</sup> The core of the approach involves the fusion of the two hypothesized interacting proteins with two domains of the transcription factor GAL4, the DNA binding domain and the activation domain. An interaction between the two target proteins leads to reconstitution of the full transcription factor activity and synthesis of  $\beta$ -galactosidase proceeds; yeast colonies in which this has happened can then be identified colourimetrically since they will turn blue in the presence of X-gal. This approach has been used to test all of the yeast open reading frames (ORF) against each other, thereby generating a complete organism protein interaction network.<sup>36,37</sup>

Several other reporter systems have been developed since Y2H was first developed. One of the most widely applied approaches is to split a reporter protein, *e.g.*, ubiquitin,<sup>38</sup> yellow fluorescence protein (YFP)<sup>39</sup> or dihydrofolate reductase,<sup>40</sup> and then fuse each part to the proteins where an interaction is to be tested. If the two target proteins physically interact, the reporter fragments are brought close together and allowed to fold into their native structure, resulting in reconstitution of the reporter protein activity. For YFP this results in a fluorescent signal in the cells where the interaction takes place and for dihydrofolate reductase this results in a new selectable marker, where survival in the presence of methotrexate can be used to identify colonies in which an interaction is occurring.<sup>31</sup> A significant difference between Y2H and split reporter assays is that in the latter the interactions can presumably take place in the endogenous subcellular localization of the interacting proteins, whereas in Y2H the interaction must occur in the nucleus; additionally, split reporter assays can be used, in principle, in any organism that can be genetically modified, meaning that if they are expressed off their endogenous promoter, the proteins should be present at natural levels and with the correct post-translational modifications.

Another imaging-based method that is widely used in focused<sup>†</sup> studies to examine protein–protein interactions is

<sup>†</sup> We use the term ‘focused’ here to differentiate between ‘omic’-scale studies that measure a more shallow but wider-reaching view of the system from studies that focus on one or a few proteins and try to understand their biology in greater detail.

FRET, where two proteins are tagged with, *e.g.*, cyan (CFP) and yellow fluorescent proteins (YFP), and co-expressed in the same cells; if the two proteins are within 10 nm of each other then excitation of CFP can result in resonance energy transfer to YFP, which then releases a photon at 535 nm, instead of the 480 nm expected for CFP.<sup>41</sup> This apparent shift in the emission wavelength of CFP indicates that the two proteins to which the fluorescence proteins are fused must be in such close proximity that the only reasonable explanation is that they are physically interacting. As mentioned, FRET is widely accepted as an exceptionally accurate method for confirming that two proteins interact but it typically requires significant optimization for any given bait and prey combination and has so far not been used for measuring whole protein interaction networks.

### 2.2. Crosslinking combined with mass spectrometry

Crosslinking combined with mass spectrometry (CXMS) can be used to both discover new interactions and to characterize the structure of known interactions. In the latter use, it is more tolerant to impurities and low concentrations than are crystallography and NMR but of course it cannot compete in terms of resolution.<sup>42</sup> CXMS requires bifunctional chemical crosslinkers that can be applied either *in vivo* or *in vitro* to react with and covalently link neighboring proteins. The proteins are then digested to peptides and subjected to liquid chromatography–tandem mass spectrometry (LC-MS/MS) to detect the cross-linked peptides. In recent years, a number of linkers targeting different functional groups on proteins have been developed, with primary amines (mostly on lysine side-chains) being the preferred target as they are plentiful and mostly located at the surface of proteins. Since the abundance of cross-linked peptides are very low compared to linear peptides, the sensitivity of the approach can be boosted significantly by enriching crosslinked peptides prior to LC-MS/MS. A range of enrichment strategies have been proposed, most of which take advantage of the increased size and charge state of the cross-linked peptides using, *e.g.*, strong cation exchange (SCX) chromatography, size exclusion chromatography (SEC) or C4 reverse phase chromatography. The main challenge in CXMS, however, is interpreting the LC-MS/MS data to correctly identify the cross-linked peptides. Even in a clean system with only two proteins, the number of possible species that must be considered is the square of the number of linear peptides, resulting in a vastly enlarged search space. New algorithms that can deal with this issue have been developed<sup>43,44</sup> and these have been instrumental in a number of groundbreaking studies that have employed CXMS to identify primary interactions among proteins in multi-protein complexes. One of the first large complexes analyzed was the RNA polymerase II complex, where Chen *et al.* identified 220 high quality interactions with approximately one third being primary interactions between proteins in the complex.<sup>33</sup> Hereby it was possible to observe the interaction between Pol II and transcription factor IIF at peptide resolution. Very recently the human protein phosphatase 2A complex was subjected to crosslinking followed by mass spectrometry,

revealing an interaction with IGBP2 and the topology of PP2A regulatory subunit binding to chaperonin.<sup>34</sup>

### 3. Identifying secondary and tertiary protein interactions through co-migration

Historically, protein complexes were identified one at a time by purifying them based on enzymatic activity and a complex was inferred when proteins co-migrated through multiple chromatographic methods. This approach provided information about the number of proteins in a complex and the size of the complex from SDS-PAGE and SEC, respectively.<sup>45</sup> However, this approach was really only useful when assays and/or reagents existed for the complex in question and it was not practical until developments in mass spectrometry-based proteomics provided the tools for reconstructing migration profiles of many proteins in parallel.

One class of methods for analyzing profile data, called protein correlation profiling (PCP), was pioneered by Andersen *et al.* in order to characterize the centrosome.<sup>46</sup> Here, profiles of individual proteins were generated across a density gradient by LC-MS/MS, using a label-free, ion intensity-based quantitative approach. PCP was subsequently used to characterize the nucleolus<sup>47</sup> and membrane-enclosed, cytoplasmic organelles.<sup>48</sup> Ion intensity measurements are reasonably accurate but the logical extension of PCP was to marry it to a more accurate approach such as stable isotope labeling with amino acids in cell culture (SILAC), leading to improved accuracy of the profiles and thus enhanced the specificity of assignments made from those profiles.<sup>49–52</sup> A similar approach, called Localization of Organelle Proteins by Isotope Tagging (LOPIT), also involves resolving organelles by density gradient centrifugation, however here the isobaric tag for relative and absolute quantitation (iTRAQ) is used for quantification.<sup>53</sup> Using this approach, a range of different organelles, as well as larger protein complexes like the large and small ribosomal subunits, have been characterized.<sup>54</sup>

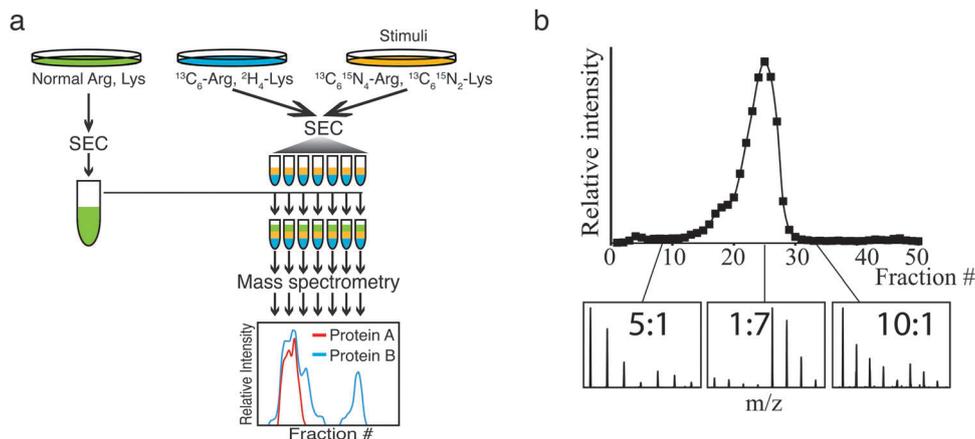
#### 3.1. Characterizing protein complexes using SEC

Protein complexes are poorly resolved by density gradient centrifugation but SEC provides an alternative separation method that can be used for protein complexes in a similar way to how density gradients were used for organelle components above. SEC was developed more than fifty years ago and has been widely used for studying protein complexes since it resolves analytes based on their Stokes radius, which is roughly proportional to molecular weight.<sup>55</sup> It relies on a column packed with a porous matrix that analytes can diffuse in and out of while an isocratic flow drives them through the column. The pore sizes define what range of Stokes radii can be separated, since smaller analytes effectively have a larger volume to diffuse into because they can enter the pores while larger analytes cannot. By comparing the elution times of standards of known molecular weight to measured values of unknowns, the molecular weight of an analyte can be estimated, since the

elution volume decreases with the logarithm of the Stokes radius. This assumes, of course, that all analytes are the same shape, typically globular; while not precisely accurate this is a reasonable first assumption for most protein complexes. The big advantage of SEC over other chromatographic techniques for studying protein complexes is that SEC provides very gentle conditions since no binding and release from the stationary phase is required and the proteins can be maintained in a physiologically relevant buffer the entire time.

The first unbiased large-scale mapping of stable protein complexes using chromatography was done on an *Escherichia coli* lysate by performing three different chromatographic separation steps, followed by iTRAQ to determine the relative migration of the single proteins, which identified 13 protein complexes based on co-migration of component proteins.<sup>56</sup> Around the same time, it was observed that one third of *S. cerevisiae* proteins were eluting at a much higher than expected weight from SEC.<sup>57</sup> Later, SEC was used to follow the co-migration of protein complexes in chloroplast stroma of *Arabidopsis* by first fractionating chloroplast stroma by SEC, followed by mass spectrometry of the individual fractions where adjusted spectral counting was used to estimate levels of individual proteins in each fraction.<sup>58</sup> The spectral count chromatograms could subsequently be hierarchically clustered with the assumption that proteins clustering together should belong to the same complex; several complexes were characterized in this way, including the RuBisCO complex, ribosomes and the chaperone 60 complex. Finally, it was shown by combining SEC with an array of antibody beads that the CDK-cyclin complexes had different profiles in quiescent and proliferating cells.<sup>59</sup>

Recently, we reported an approach in human HeLa cells that combines PCP-SILAC with SEC to map 7209 primary protein interactions and 293 protein complexes based on their co-migration.<sup>1</sup> The basic principle starts with mass-encoding the cells using arginine and lysine isotopologs. The heavy population is subsequently perturbed while the medium and light populations are left untreated and then all three populations are mechanically lysed. An enrichment of the cytosolic protein complexes from each lysate was performed, to ensure irrelevant interactions among proteins in different cellular compartments would be minimized, before each lysate was separated on a SEC column with optimal resolution between 150 kDa and 2 MDa. The fractions from the light sample are then pooled together and aliquots are mixed into each of the medium/heavy fractions prior to tryptic digestion and mass spectrometric analysis (Fig. 2a). In this scheme, the light-labeled proteins act as internal standards and any temporal protein interaction network changes following perturbation of the heavy population are monitored with the medium/heavy ratio (Fig. 2b). To assign interactions among the proteins represented in these chromatograms, we used two types of information: first, secondary interactions among proteins could be deduced by calculating the Euclidian distance between all chromatograms, with the assumption that proteins that always occur together in the same complex(es) would have very similar chromatograms.



**Fig. 2** Schema of the PCP-SILAC approach to identifying protein interactions. (a) Proteins from differentially treated medium and heavy-labelled cells are fractionated by size exclusion chromatography and then aliquots of light protein are spiked into each fraction to act as an internal standard, prior to LC-MS/MS analysis. The medium/light or heavy/light ratios vs. chromatographic fraction then represent the chromatogram of that protein. (b) Size exclusion chromatogram for the average ratios from all peptides of a protein. The mass spectra of the peptide from which ratios were derived for fractions #10, 25 and 32 are shown below.

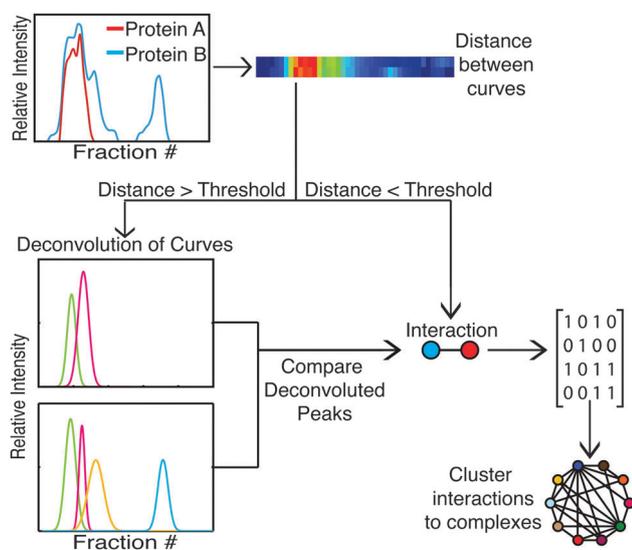
Second, tertiary interactions could be deduced by deconvolving each chromatogram into component Gaussian curves, with the assumption that for large complexes, which are made of independent, stable and observable complexes, the constituent proteins might only show similarities in part of the chromatogram (Fig. 3). This resulted in a protein interaction network with at least the same precision and false positive rate as low throughput techniques when validated using the BioGRID database.<sup>60</sup>

The number of complexes a protein participates in and its relative distribution among them can be important for biological outcomes; *e.g.*, a scaffold protein was shown to either inhibit or activate a cascade depending on its concentration.<sup>61</sup> A great advantage with the SEC-PCP-SILAC approach is it finds all distinct complexes with which the protein associates, since

the complexes elute as separate peaks from the SEC column. We observed 43% of the protein chromatograms measured could be deconvolved into more than one Gaussian peak, suggesting that these proteins participate in multiple complexes. In addition to identifying protein interactions, the SEC-PCP-SILAC approach also allows one to estimate the stoichiometry of a protein between multiple complexes, since the chromatograms are quantitative. The relative stoichiometry of a protein binding to its various partners can be calculated from the areas of the individual Gaussian curves, similar to the information obtained from multisignal sedimentation velocity analytical ultracentrifugation.<sup>62</sup> By investigating the stoichiometry of the individual proteins of the regulatory particle of the proteasome, we observed that the majority of the proteins had similar complex distribution; interestingly, however, we observed the E3 ubiquitin KCMF1 was only binding to the regulatory particle when it, in turn, was bound to the core particle, thereby acting as an accessory protein to the regulatory particle, according to the organizational scheme presented above.

### 3.2. Other co-migration approaches for characterizing protein complexes

Instead of just fractionating protein complexes in one dimension, Havugimana *et al.* took biochemical fractionation to the extreme, using weak anion exchange, isoelectric focusing and density gradients to generate 1163 different fractions from HEK293 and HeLa S3 cells that were then analyzed by mass spectrometry.<sup>2</sup> Using the spectral counts observed for each protein in the 1163 fractions, the results were filtered using machine learning, where additional data from the literature such as evolutionary rates, messenger RNA co-expression, domain co-occurrence and physical associations reported earlier in worms and yeast were taken into account. This resulted in the identification of 13 993 co-complex interactions and 622 protein complexes, which represents the largest experimentally derived catalog of human protein complexes to date.



**Fig. 3** SEC-PCP-SILAC informatics workflow. The general decision tree followed in defining a protein interaction network from SEC-PCP-SILAC data (see text for more details).

The PCP approach has also been successfully applied to membrane proteins by separating them on blue native PAGE, cutting the gel lane into several slices and analyzing each slice by LC-MS/MS; followed this, label free quantitation was used to identify oxidative phosphorylation complexes I–V.<sup>63</sup> Recently, a comprehensive mitochondrial protein interaction network was identified using blue native and large-pore blue native gel electrophoresis combined with label-free protein quantification.<sup>3</sup> Among the identified complexes was a novel complex, given the name ‘mitochondrial complex I assembly complex’, which highlights the advantage of studying protein complexes using non-hypothesis driven techniques.

#### 4. Identifying quaternary protein interactions using affinity purification-mass spectrometry

The AP-MS approach is very widely used in focused studies to identify the interacting partners of a protein of interest. The method is scalable and partially automatable so with substantial resources one can use it to characterize a whole protein interaction network. In the simplest approach, the protein of interest (the bait) is enriched using an antibody and then the proteins bound to it are identified by mass spectrometry; after accounting for potential non-specific interactions using an appropriate negative control, the specific interacting partners can be identified. This approach has the advantage that no genetic manipulation is required and that it can be conducted in all organisms where an antibody is available. On the other hand, the success of this approach depends greatly on the specificity and affinity of the antibody towards the protein of interest, which is often sub-optimal. In addition, antibodies are not available for the majority of proteins, even in well-studied organisms, and those that are available are not of consistent quality. Broad panels of antibodies generated and validated under uniform conditions, such as from the Human Protein Atlas project,<sup>64</sup> will likely make this approach more popular. Recently, this technique was used to identify numerous protein interactions in human cells.<sup>65</sup>

The most popular approach in AP-MS is to fuse the protein of interest to an affinity tag. This can be accomplished relatively easily in yeast by homologous recombination and several whole genome tag collections are available for yeast. Homologous recombination is not possible in human systems though, so there the general approach has been to introduce the tagged version of the gene as a cDNA-based transgene; the problem with this is that these transgenes lack endogenous noncoding regulatory information, such as introns and 3'-untranslated regions, resulting in non-physiological expression levels in most cases. Recently, the bacterial artificial chromosome (BAC) TransgeneOmics approach was introduced that uses BACs to include most of the regulatory elements upstream of the gene of interest in order to achieve near-endogenous expression levels.<sup>66</sup> Beyond expression issues, there are many affinity tags to choose from, ranging from small peptides such as 6xHis,

Hemagglutinin (HA) or FLAG to whole proteins such as GFP and glutathione-S-transferase (GST). Smaller tags seem less likely to disrupt protein complexes than a protein-sized tag, but GFP has the obvious advantage that it can also be used to image the protein of interest to gain additional information such as localization, movement within the cell and co-localization.<sup>67,68</sup> The most common affinity system used in global protein interaction network studies is the tandem affinity purification (TAP)-tag, originally formulated as a fusion of a calmodulin binding peptide (CBP), a Tobacco etch virus (TEV) cleavage site and two IgG units of protein A from *Staphylococcus aureus*,<sup>69</sup> although several variants on this idea now exist.<sup>70</sup> The tagged protein and its interactors are first enriched using IgG-coated beads, then released from those beads by cleaving the tag with TEV protease and then further enriched on calmodulin beads.<sup>69</sup> TAP-MS has been widely used for global screens in yeast,<sup>5,6,71</sup> as well as in more targeted analyses in human systems.<sup>72,73</sup> The great advantage of TAP is that the two rounds of enrichment help to minimize the non-specific proteins that would otherwise co-purify in a single step enrichment; on the flip side, the process still requires tagging and the longer and more complicated procedure to enrich interactors means that only very stable interactions survive to be detected.

Tagging proteins can be problematic: not only can they be challenging to introduce, they have also been shown to disrupt protein complexes and can potentially lead to change in localization and function. For example, it was only possible to partly recover the chaperonin-containing tailless complex polypeptide 1 (CTT) complex in a study using AP-MS because the ends of the proteins that were tagged were located in the interaction interface.<sup>5</sup> Moving the tag to the opposite terminus can often solve such issues but the majority of high-throughput AP-MS studies only tag one terminus, meaning that many real interactions may be undetectable as a result of the experimental design. Tags can also affect the subcellular localization of proteins, exemplified by a recent report in *Caulobacter* where only one third of proteins localized properly regardless of which end the protein was tagged on, suggesting that the tag disrupts terminal structures or interaction motifs of the protein.<sup>74</sup> Similar, it was observed in yeast that tagging and mild over-expression of a number of mitochondria proteins led to them failing to properly localize to the organelle and instead were delivered to the endoplasmic reticulum (ER), cytosol or nucleus.<sup>75</sup> Finally, in HeLa cells similar observations have been reported using the BAC transgenomics approach where, *e.g.*, Rab5c only localized correctly at the endosomes when tagged at the N-terminus and AURKB showed correct physiological localization dynamics through the cell cycle only when tagged at its C-terminus.<sup>66</sup>

Non-specific interactions can result in a protein being identified as an interactor when it is, in fact, a false positive. TAP minimizes (but does not completely eliminate) such problems through two rounds of enrichment but another approach is to perform the AP-MS in a quantitative fashion. This can be accomplished using stable isotope dilution or label free approaches. In SILAC<sup>76,77</sup> two populations of cells are first

metabolically labeled and then used for the control and specific pull downs respectively, before the samples are combined, digested to peptides and analyzed by MS.<sup>78</sup> This approach has been used in numerous studies and has the great advantage that it is simple to perform and can easily distinguish specific from non-specific interactions.<sup>21,67,79</sup> For organisms/cell lines that are not amenable to SILAC, chemical labeling methods<sup>79</sup> or label free<sup>68,80</sup> approaches can be used to distinguish specific from non-specific interactions since the quantity of a specific interactor should be higher in the specific condition than in the negative control.

Finally, as previously mentioned, one disadvantage to all affinity purification approaches is that they are blind to the arrangement of all the interactors identified. When one identifies interactions in this way, there is no indication from the data whether the interactions are all occurring together (tertiary protein interactions) or whether the average composition of several different complexes is being observed (quaternary protein interactions). If exhaustive reciprocal affinity enrichments are performed with all the interactors then bioinformatic treatment of the entire network can sometimes tease apart such subtleties.<sup>5,6</sup>

## 5. Temporal protein interaction network mapping

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

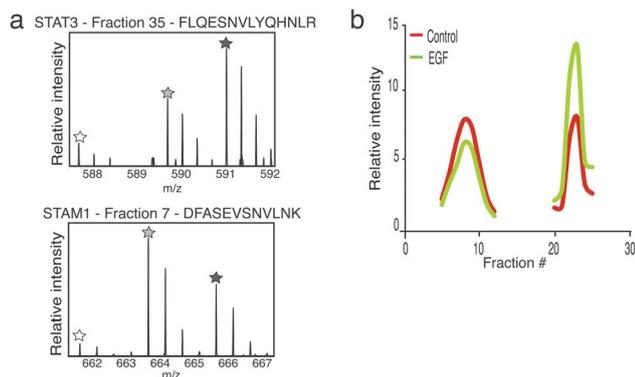
One approach temporal protein interaction network changes can be inferred is to use genetic interaction data, such as epistatic miniarray profiles (E-MAPs), which uses a library of strongly down-regulated genes to screen both essentially and non-essential genes.<sup>81</sup> By conducting E-MAPs under two conditions, it was possible to map 873 differentially genetic interactions during DNA damage response, from which the researchers could identify new roles for Slt2 kinase, Pph3 phosphatase and histone variant Htz1 and observe that protein complexes seem to be stable to perturbation.<sup>82</sup>

Other proteomic methods have also been developed that directly measure changes in physical interactions. One of the earliest approaches combined AP-MS and SILAC, thereby

making it possible to track dynamic changes among the affinity purified proteins.<sup>83</sup> The authors used HeLa cells labeled with triple SILAC and stimulated with EGF for different lengths of time, before antibodies against phospho-tyrosine were used to affinity purify tyrosine phosphorylated proteins and closely related binders. By quantifying the SILAC ratios for these proteins, it was possible to deduce how the stimulation affected a given protein's phosphorylation state. Later, this method was used to distinguish the effect of EGF and PDGF on human mesenchymal stem cells, revealing that while more than 90% of the signaling proteins were used by both pathways, phosphatidylinositol-3-kinase (PI3K) was exclusively activated by PDGF.<sup>84</sup> Finally, by conducting quantitative AP-MS of cyclins E1, A2 and B2 at multiple time points during cell cycle, another study found 295 specific interaction partners, which displayed significant enrichment of proteins with cell-division defects, suggesting a clear link between temporal protein interaction networks and functional roles.<sup>85</sup> These kinds of approaches provide an additional, valuable dimension to interaction data and often reveal novel connections that could not have been predicted otherwise; *e.g.*, the YMER protein, which had similar activation profiles as other proteins involved in early EGF signaling<sup>83</sup> was later shown to function as an inhibitor for the down regulation of the EGF receptor.<sup>86</sup>

On a similar scale, targeted proteomic methods can also be used to measure temporal interaction changes in protein interaction networks. Bisson *et al.*<sup>87</sup> used an approach called AP-selected reaction monitoring (AP-SRM), where GRB2 was affinity purified and its interaction partners assayed by SRM. By integrating the peak areas of the peptides from each partner it was possible to measure the temporal protein dynamics at six different time points after EGF stimulation and additionally follow the changes in GRB2 protein interaction network caused by five other growth factors. SRM is much more sensitive than discovery-driven LC-MS/MS, it has large quantitative dynamic range, it does not require any *in vivo* or chemical labeling and it does not rely on stochastic sampling, leading to more reproducible data. The downsides, however, are that developing the assay is rather time consuming<sup>87,88</sup> and it is not easily scaled to many baits.

Co-migration approaches such as SEC-PCP-SILAC are currently the only method available for large-scale monitoring of temporal changes in the physical protein interaction network.<sup>1</sup> The incorporation of a third SILAC label into the SEC-PCP-SILAC scheme (Fig. 2a) allows for the heavy/medium ratio to be used to quantify how the protein interaction network response to a challenge. In the case of EGF stimulation (Fig. 4), this revealed 351 proteins whose interactions with others changed significantly. These proteins were also highly enriched for proteins known to have altered phosphorylation in response to EGF, highlighting the very well described phenomenon known from more targeted assays that phosphorylation leads to interaction changes or *vice versa*. Since SEC-PCP-SILAC is relative fast and requires only on the order of days of mass spectrometry time, it makes monitoring temporal changes to the protein interaction network across different conditions finally accessible.



**Fig. 4** SEC-PCP-SILAC to distinguish temporal protein interactions. (a) Mass spectra of three peptides that display different spatiotemporal interactions changes following EGF stimulation. The medium:light ratio (M/L) is used to generate chromatograms, whereas the heavy to medium (H/M) ratio represents the impact of EGF stimulation on that protein. In the top panel, the H/M ratio for the peptide FLQESNVLYQHNLNR from STAT3 in fraction 35 indicates that STAT3 gets recruited to a complex in fraction 35 following EGF stimulation. In the bottom panel, the peptide DFASEVSNVNLNK from STAM1 in fraction 7 shows a H/M ratio less than 1.0, indicating that it dissociates from a complex in fraction 7 following EGF stimulation. (b) Chromatogram deriving from STAM1 highlighting the EGF-induced spatiotemporal change. The red line represents the chromatogram from un-stimulated cells (*i.e.*, the plot of M/L as in Fig. 2b), while the green line represents the chromatogram from the EGF-stimulated cells (*i.e.*, the plot of H/L). EGF stimulation causes the complex with STAM1 at fraction 7 to dissociate and STAM1 to associate with a different complex eluding in fraction 22.

A great benefit of applying SILAC and other stable isotope based quantitation methods to map temporal protein interaction networks is that the control and perturbed samples are mixed prior to mass spectrometric analysis, thereby minimizing technical variation. This also allows a quantitative measure of how much the protein interactions change in response to the stimuli, making the data more informative than simply knowing whether an interaction is changing or not. Because of the higher accuracy of SILAC and other stable isotope-based quantitation methods, smaller but still significant changes in the protein interaction network can be detected.

## 6. Outlook

The protein interaction networks that have been reported in the last few years are still largely coming from model organisms that are more malleable than human cells and that are amenable to high-throughput screens. With an estimated 10 000–12 000 proteins per human cell<sup>89</sup> versus 4400 for yeast,<sup>90</sup> there is a limit to how much can be inferred about human protein interaction networks from model organisms. The human body contains hundreds of different cell types, each with its own unique proteome and, thus, its own protein interaction networks, so the complexity of the complete human protein interaction network is vastly underexplored. The methods used over the past ten years for mapping the static protein interaction network are reasonably accurate with acceptable false positive rates. They are still not as sensitive as they need to be and while approaches like AP-MS have shown themselves to

be accurate, they are not broadly useful for global monitoring the protein interaction network because of their cost and confounding factors (*e.g.*, tagging artifacts). Co-migration approaches have reached a point where they have at least the same throughput (in terms of interactions identified), precision and false positive rate as many low throughput approaches at a small fraction of the cost of AP-MS-based mapping efforts. The added benefit of co-migration is that SILAC can be used to encode how the protein interaction network responds to stimuli, something that had previously been completely inaccessible. As the technology matures over the next few years and becomes more accessible, we expect a revitalized interest in protein interaction network mapping as a tool for understanding biology. Of particular interest may be the use of protein interaction network monitoring late in the discovery phases of drug development, to understand the impact of the drug on the protein interaction network.

## Acknowledgements

The authors thank the other members of our group in the Centre for High-Throughput Biology for fruitful discussions and advice, particular Karina T. Simonsen, Nick E. Scott and Lyda M. Brown for reading the manuscript. Protein interaction network research in our group is supported by an operating grant from the Canadian Institutes for Health Research to LJF (MOP-77688). LJF is the Canada Research Chair in Quantitative Proteomics and ARK was supported by the Danish Agency for Science Technology and Innovation.

## References

- 1 A. R. Kristensen, J. Gsponer and L. J. Foster, *Nat. Methods*, 2012, **9**, 907–909.
- 2 P. C. Havugimana, T. G. Hart, T. Nepusz, H. Yang, A. L. Turinsky, Z. Li, P. I. Wang, D. R. Boutz, V. Fong, S. Phanse, M. Babu, S. A. Craig, P. Hu, C. Wan, J. Vlasblom, V.-U.-N. Dar, A. Bezginov, G. W. Clark, G. C. Wu, S. J. Wodak, E. R. M. Tillier, A. Paccanaro, E. M. Marcotte and A. Emili, *Cell*, 2012, **150**, 1068–1081.
- 3 H. Heide, L. Bleier, M. Steger, J. Ackermann, S. Dröse, B. Schwamb, M. Zörnig, A. S. Reichert, I. Koch, I. Wittig and U. Brandt, *Cell Metab.*, 2012, **16**, 538–549.
- 4 S. Kuhner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castano-Diez, W. H. Chen, D. Devos, M. Guell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Bottcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork and A. C. Gavin, *Science*, 2009, **326**, 1235–1240.
- 5 A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer,

- J. M. Rick, B. Kuster, P. Bork, R. B. Russell and G. Superti-Furga, *Nature*, 2006, **440**, 631–636.
- 6 N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Y. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt, *Nature*, 2006, **440**, 637–643.
- 7 S. G. Hedin, *Biochem. J.*, 1906, **1**, 474–483.
- 8 S. S. Taylor, *J. Biol. Chem.*, 1989, **264**, 8443–8446.
- 9 K.-I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal and A.-L. Barabási, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 8685–8690.
- 10 R. Pastor-Satorras, E. Smith and R. V. Solé, *J. Theor. Biol.*, 2003, **222**, 199–210.
- 11 J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V. Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth and M. Vidal, *Nature*, 2004, **430**, 88–93.
- 12 S. Agarwal, C. M. Deane, M. A. Porter and N. S. Jones, *PLoS Comput. Biol.*, 2010, **6**, e1000817.
- 13 I. W. Taylor, R. Linding, D. Warde-Farley, Y. Liu, C. Pesquita, D. Faria, S. Bull, T. Pawson, Q. Morris and J. L. Wrana, *Nat. Biotechnol.*, 2009, **27**, 199–204.
- 14 M. Vidal, M. E. Cusick and A.-L. Barabási, *Cell*, 2011, **144**, 986–998.
- 15 A.-C. Gingras, M. Gstaiger, B. Raught and R. Aebersold, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 645–654.
- 16 P. Braun, *Proteomics*, 2012, **12**, 1499–1518.
- 17 K. Raman, *Autom. Exp.*, 2010, **2**, 2.
- 18 A.-L. Barabási and Z. N. Oltvai, *Nat. Rev. Genet.*, 2004, **5**, 101–113.
- 19 P. R. Burton, *et al.*, *Nature*, 2007, **447**, 661–678.
- 20 Q. Zhong, N. Simonis, Q.-R. Li, B. Charletoaux, F. Heuze, N. Klitgord, S. Tam, H. Yu, K. Venkatesan, D. Mou, V. Swearingen, M. A. Yildirim, H. Yan, A. E. L. Dricot, D. Szeto, C. Lin, T. Hao, C. Fan, S. Milstein, D. Dupuy, R. Brasseur, D. E. Hill, M. E. Cusick and M. Vidal, *Mol. Syst. Biol.*, 2009, **5**, 1–10.
- 21 L. D. Rogers, A. R. Kristensen, E. C. Boyle, D. P. Robinson, R. T. Ly, B. B. Finlay and L. J. Foster, *J. Proteomics*, 2008, **71**, 97–108.
- 22 J. C. Patel and J. E. Galan, *J. Cell Biol.*, 2006, **175**, 453–463.
- 23 D. Zhou, L. M. Chen, L. Hernandez, S. B. Shears and J. E. Galan, *Mol. Microbiol.*, 2001, **39**, 248–259.
- 24 X. Yu, Y. Yu, B. Liu, K. Luo, W. Kong, P. Mao and X.-F. Yu, *Science*, 2003, **302**, 1056–1060.
- 25 S. Jäger, P. Cimermanic, N. Gulbahce, J. R. Johnson, K. E. McGovern, S. C. Clarke, M. Shales, G. Mercenne, L. Pache, K. Li, H. Hernandez, G. M. Jang, S. L. Roth, E. Akiva, J. Marlett, M. Stephens, I. D'Orso, J. Fernandes, M. Fahey, C. Mahon, A. J. O'Donoghue, A. Todorovic, J. H. Morris, D. A. Maltby, T. Alber, G. Cagney, F. D. Bushman, J. A. Young, S. K. Chanda, W. I. Sundquist, T. Kortemme, R. D. Hernandez, C. S. Craik, A. Burlingame, A. Sali, A. D. Frankel and N. J. Krogan, *Nature*, 2012, **481**, 365–370.
- 26 S. Jäger, D. Y. Kim, J. F. Hultquist, K. Shindo, R. S. LaRue, E. Kwon, M. Li, B. D. Anderson, L. Yen, D. Stanley, C. Mahon, J. Kane, K. Franks-Skiba, P. Cimermanic, A. Burlingame, A. Sali, C. S. Craik, R. S. Harris, J. D. Gross and N. J. Krogan, *Nature*, 2012, **481**, 371–375.
- 27 J. G. Miller, *Behav. Sci.*, 1965, **10**, 193–237.
- 28 Z. Dezso, Z. N. Oltvai and A.-L. Barabási, *Genome Res.*, 2003, **13**, 2450–2454.
- 29 M. Laplante and D. M. Sabatini, *Cell*, 2012, **149**, 274–293.
- 30 A. W. Rives and T. Galitski, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1128–1133.
- 31 K. Tarassov, V. Messier, C. R. Landry, S. Radinovic, M. M. S. Molina, I. Shames, Y. Malitskaya, J. Vogel, H. Bussey and S. W. Michnick, *Science*, 2008, **320**, 1465–1470.
- 32 E. V. Haar, S.-I. Lee, S. Bandhakavi, T. J. Griffin and D.-H. Kim, *Nature*, 2007, **9**, 316–323.
- 33 Z. A. Chen, A. Jawhari, L. Fischer, C. Buchen, S. Tahir, T. Kamenski, M. Rasmussen, L. Lariviere, J.-C. Bukowski-Wills, M. Nilges, P. Cramer and J. Rappsilber, *EMBO J.*, 2010, **29**, 717–726.
- 34 F. Herzog, A. Kahraman, D. Boehringer, R. Mak, A. Bracher, T. Walzthoeni, A. Leitner, M. Beck, F. U. Hartl, N. Ban, L. Malmstrom and R. Aebersold, *Science*, 2012, **337**, 1348–1352.
- 35 S. Fields and O. Song, *Nature*, 1989, **340**, 245–246.
- 36 T. Ito, T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 4569–4574.
- 37 P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg, *Nature*, 2000, **403**, 623–627.
- 38 N. Johnsson and A. Varshavsky, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10340–10344.
- 39 C.-D. Hu, Y. Chinenov and T. K. Kerppola, *Mol. Cell*, 2002, **9**, 789–798.
- 40 I. Remy and S. W. Michnick, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 5394–5399.
- 41 J. Lippincott-Schwartz, E. Snapp and A. Kenworthy, *Nat. Rev. Mol. Cell Biol.*, 2001, **2**, 444–456.
- 42 P. Singh, A. Panchoad and D. R. Goodlett, *Anal. Chem.*, 2010, **82**, 2636–2642.
- 43 T. Walzthoeni, M. Claassen, A. Leitner, F. Herzog, S. Bohn, F. Förster, M. Beck and R. Aebersold, *Nat. Methods*, 2012, **9**, 901–903.
- 44 B. Yang, Y.-J. Wu, M. Zhu, S.-B. Fan, J. Lin, K. Zhang, S. Li, H. Chi, Y.-X. Li, H.-F. Chen, S.-K. Luo, Y.-H. Ding, L.-H. Wang, Z. Hao, L.-Y. Xiu, S. Chen, K. Ye, S.-M. He and M.-Q. Dong, *Nat. Methods*, 2012, **9**, 904–906.

- 45 S. Wilk and M. Orlowski, *J. Neurochem.*, 1980, **35**, 1172–1182.
- 46 J. S. Andersen, C. J. Wilkinson, T. Mayor, P. Mortensen, E. A. Nigg and M. Mann, *Nature*, 2003, **426**, 570–574.
- 47 J. S. Andersen, Y. W. Lam, A. K. L. Leung, S.-E. Ong, C. E. Lyon, A. I. Lamond and M. Mann, *Nature*, 2005, **433**, 77–83.
- 48 L. J. Foster, C. L. de Hoog, Y. Zhang, Y. Zhang, X. Xie, V. K. Mootha and M. Mann, *Cell*, 2006, **125**, 187–199.
- 49 L. Jakobsen, K. Vanselow, M. Skogs, Y. Toyoda, E. Lundberg, I. Poser, L. G. Falkenby, M. Bennetzen, J. Westendorf, E. A. Nigg, M. Uhlen, A. A. Hyman and J. S. Andersen, *EMBO J.*, 2011, **30**, 1520–1535.
- 50 J. Dengjel, M. Hoyer-Hansen, M. O. Nielsen, T. Eisenberg, L. M. Harder, S. Schandorff, T. Farkas, T. Kirkegaard, A. C. Becker, S. Schroeder, K. Vanselow, E. Lundberg, M. M. Nielsen, A. R. Kristensen, V. Akimov, J. Bunkenborg, F. Madeo, M. Jaattela and J. S. Andersen, *Mol. Cell. Proteomics*, 2012, **11**, M111.014035.
- 51 J. Dengjel, L. Jakobsen and J. S. Andersen, *Methods in Molecular Biology*, Humana Press, Totowa, NJ, 2010, vol. 658.
- 52 Y. Z. Zheng, K. B. Berg and L. J. Foster, *J. Lipid Res.*, 2008, **50**, 988–998.
- 53 T. P. J. Dunkley, R. Watson, J. L. Griffin, P. Dupree and K. S. Lilley, *Mol. Cell. Proteomics*, 2004, **3**, 1128–1134.
- 54 D. J. L. Tan, H. Dvinge, A. Christoforou, P. Bertone, A. Martinez Arias and K. S. Lilley, *J. Proteome Res.*, 2009, **8**, 2667–2678.
- 55 J. Porath and P. Flodin, *Nature*, 1959, **183**, 1657–1659.
- 56 M. Dong, L. L. Yang, K. Williams, S. J. Fisher, S. C. Hall, M. D. Biggin, J. Jin and H. E. Witkowska, *J. Proteome Res.*, 2008, **7**, 1836–1849.
- 57 X. Liu, W.-C. Yang, Q. Gao and F. Regnier, *J. Chromatogr., A*, 2008, **1178**, 24–32.
- 58 P. D. B. Olinares, L. Ponnala and K. J. van Wijk, *Mol. Cell. Proteomics*, 2010, **9**, 1594–1615.
- 59 W. Wu, H. Slastad, D. de la Rosa Carrillo, T. Frey, G. Tjonnfjord, E. Boretti, H. C. Aasheim, V. Horejsi and F. Lund-Johansen, *Mol. Cell. Proteomics*, 2008, **8**, 245–257.
- 60 C. Stark, B.-J. Breitkreutz, A. Chatr-aryamontri, L. Boucher, R. Oughtred, M. S. Livstone, J. Nixon, K. Van Auken, X. Wang, X. Shi, T. Reguly, J. M. Rust, A. Winter, K. Dolinski and M. Tyers, *Nucleic Acids Res.*, 2011, **39**, D698–D704.
- 61 A. Levchenko, J. Bruck and P. W. Sternberg, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 5818–5823.
- 62 A. Balbo, K. H. Minor, C. A. Velikovskiy, R. A. Mariuzza, C. B. Peterson and P. Schuck, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 81–86.
- 63 H. J. C. T. Wessels, R. O. Vogel, L. van den Heuvel, J. A. Smeitink, R. J. Rodenburg, L. G. Nijtmans and M. H. Farhoud, *Proteomics*, 2009, **9**, 4221–4228.
- 64 M. Uhlen, P. Oksvold, L. Fagerberg, E. Lundberg, K. Jonasson, M. Forsberg, M. Zwahlen, C. Kampf, K. Wester, S. Hober, H. Wernerus, L. Björling and F. Ponten, *Nat. Biotechnol.*, 2010, **28**, 1248–1250.
- 65 A. Malovannaya, R. B. Lanz, S. Y. Jung, Y. Bulynko, N. T. Le, D. W. Chan, C. Ding, Y. Shi, N. Yucer, G. Krenciute, B.-J. Kim, C. Li, R. Chen, W. Li, Y. Wang, B. W. O'Malley and J. Qin, *Cell*, 2011, **145**, 787–799.
- 66 I. Poser, M. Sarov, J. R. A. Hutchins, J.-K. Hériché, Y. Toyoda, A. Pozniakovskiy, D. Weigl, A. Nitzsche, B. Hegemann, A. W. Bird, L. Pelletier, R. Kittler, S. Hua, R. Naumann, M. Augsburg, M. M. Sykora, H. Hofemeister, Y. Zhang, K. Nasmyth, K. P. White, S. Dietzel, K. Mechtler, R. Durbin, A. F. Stewart, J.-M. Peters, F. Buchholz and A. A. Hyman, *Nat. Methods*, 2008, **5**, 409–415.
- 67 L. Trinkle-Mulcahy, S. Boulon, Y. W. Lam, R. Urcia, F. M. Boisvert, F. Vandermoere, N. A. Morrice, S. Swift, U. Rothbauer, H. Leonhardt and A. Lamond, *J. Cell Biol.*, 2008, **183**, 223–239.
- 68 N. C. Hubner, A. W. Bird, J. Cox, B. Splettstoesser, P. Bandilla, I. Poser, A. Hyman and M. Mann, *J. Cell Biol.*, 2010, **189**, 739–754.
- 69 G. Rigaut, A. Shevchenko, B. Rutz, M. Wilm, M. Mann and B. Séraphin, *Nat. Biotechnol.*, 1999, **17**, 1030–1032.
- 70 M. O. Collins and J. S. Choudhary, *Curr. Opin. Biotechnol.*, 2008, **19**, 324–330.
- 71 A.-C. Gavin, M. Bösch, R. Krause, P. Grandi, M. Marzioch, A. Bauer, J. Schultz, J. M. Rick, A.-M. Michon, C.-M. Cruciat, M. Remor, C. Höfert, M. Schelder, M. Brajenovic, H. Ruffner, A. Merino, K. Klein, M. Hudak, D. Dickson, T. Rudi, V. Gnau, A. Bauch, S. Bastuck, B. Huhse, C. Leutwein, M.-A. Heurtier, R. R. Copley, A. Edelmann, E. Querfurth, V. Rybin, G. Drewes, M. Raida, T. Bouwmeester, P. Bork, B. Seraphin, B. Kuster, G. Neubauer and G. Superti-Furga, *Nature*, 2002, **415**, 141–147.
- 72 R. M. Ewing, P. Chu, F. Elisma, H. Li, P. Taylor, S. Climie, L. McBroom-Cerajewski, M. D. Robinson, L. O'Connor, M. Li, R. Taylor, M. Dharsee, Y. Ho, A. Heilbut, L. Moore, S. Zhang, O. Ornatsky, Y. V. Bukhman, M. Ethier, Y. Sheng, J. Vasilescu, M. Abu-Farha, J.-P. Lambert, H. S. Duetzel, I. I. Stewart, B. Kuehl, K. Hogue, K. Colwill, K. Gladwish, B. Muskat, R. Kinach, S.-L. Adams, M. F. Moran, G. B. Morin, T. Topaloglou and D. Figeys, *Mol. Syst. Biol.*, 2007, **3**, 89.
- 73 T. Bouwmeester, A. Bauch, H. Ruffner, P.-O. Angrand, G. Bergamini, K. Croughton, C. Cruciat, D. Eberhard, J. Gagneur, S. Ghidelli, C. Hopf, B. Huhse, R. Mangano, A.-M. Michon, M. Schirle, J. Schlegl, M. Schwab, M. A. Stein, A. Bauer, G. Casari, G. Drewes, A.-C. Gavin, D. B. Jackson, G. Joberty, G. Neubauer, J. Rick, B. Kuster and G. Superti-Furga, *Nature*, 2004, **6**, 97–105.
- 74 J. N. Werner, E. Y. Chen, J. M. Guberman, A. R. Zippilli, J. J. Irgon and Z. Gitai, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 7858–7863.
- 75 A. Sickmann, J. Reinders, Y. Wagner, C. Joppich, R. Zahedi, H. E. Meyer, B. Schönfisch, I. Perschil, A. Chacinska, B. Guiard, P. Rehling, N. Pfanner and C. Meisinger, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 13207–13212.
- 76 S.-E. Ong, B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann, *Mol. Cell. Proteomics*, 2002, **1**, 376–386.

- 77 S.-E. Ong and M. Mann, *Nat. Protocols*, 2007, **1**, 2650–2660.
- 78 B. Blagoev, I. Kratchmarova, S.-E. Ong, M. Nielsen, L. J. Foster and M. Mann, *Nat. Biotechnol.*, 2003, **21**, 315–318.
- 79 A. Cherkasov, M. Hsing, R. Zoraghi, L. J. Foster, R. H. See, N. Stoyanov, J. Jiang, S. Kaur, T. Lian, L. Jackson, H. Gong, R. Swayze, E. Amandoron, F. Hormozdiari, P. Dao, C. Sahinalp, O. Santos-Filho, P. Axerio-Cilies, K. Byler, W. R. McMaster, R. C. Brunham, B. B. Finlay and N. E. Reiner, *J. Proteome Res.*, 2011, **10**, 1139–1150.
- 80 M. E. Sowa, E. J. Bennett, S. P. Gygi and J. W. Harper, *Cell*, 2009, **138**, 389–403.
- 81 M. Schuldiner, S. R. Collins, N. J. Thompson, V. Denic, A. Bhamidipati, T. Punna, J. Ihmels, B. Andrews, C. Boone, J. F. Greenblatt, J. S. Weissman and N. J. Krogan, *Cell*, 2005, **123**, 507–519.
- 82 S. Bandyopadhyay, M. Mehta, D. Kuo, M. K. Sung, R. Chuang, E. J. Jaehnig, B. Bodenmiller, K. Licon, W. Copeland, M. Shales, D. Fiedler, J. Dutkowski, A. Guenole, H. van Attikum, K. M. Shokat, R. D. Kolodner, W. K. Huh, R. Aebersold, M. C. Keogh, N. J. Krogan and T. Ideker, *Science*, 2010, **330**, 1385–1389.
- 83 B. Blagoev, S.-E. Ong, I. Kratchmarova and M. Mann, *Nat. Biotechnol.*, 2004, **22**, 1139–1145.
- 84 I. Kratchmarova, B. Blagoev, M. Haack-Sorensen, M. Kassem and M. Mann, *Science*, 2005, **308**, 1472–1477.
- 85 F. W. Pagliuca, M. O. Collins, A. Lichawska, P. Zegerman, J. S. Choudhary and J. Pines, *Mol. Cell*, 2011, **43**, 406–417.
- 86 K. Tashiro, H. Konishi, E. Sano, H. Nabeshi, E. Yamauchi and H. Taniguchi, *J. Biol. Chem.*, 2006, **281**, 24612–24622.
- 87 N. Bisson, D. A. James, G. Ivosev, S. A. Tate, R. Bonner, L. Taylor and T. Pawson, *Nat. Biotechnol.*, 2011, **29**, 653–658.
- 88 V. Lange, P. Picotti, B. Domon and R. Aebersold, *Mol. Syst. Biol.*, 2008, **4**, 222.
- 89 N. Nagaraj, J. R. Wisniewski, T. Geiger, J. Cox, M. Kircher, J. Kelso, S. P. A. A. bo and M. Mann, *Mol. Syst. Biol.*, 2011, **7**, 1–8.
- 90 L. M. F. de Godoy, J. V. Olsen, J. Cox, M. L. Nielsen, N. C. Hubner, F. Fröhlich, T. C. Walther and M. Mann, *Nature*, 2008, **455**, 1251–1254.