Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry

Simone Schopper[1,3], Abdullah Kahraman[2], Pascal Leuenberger[1], Yuehan Feng[1,3], Ilaria Piazza[1], Oliver Müller[1], Paul J Boersema[1] & Paola Picotti[1]

[1]Institute of Biochemistry, Department of Biology, ETH Zurich, Zurich, Switzerland. [2]Institute of Molecular Life Sciences, University of Zurich, Zurich, Switzerland. [3]Present addresses: Faculty of Health and Medical Sciences, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark (S.S.); Stanford Genome Technology Center, Stanford University, Palo Alto, California, USA (Y.F.). Correspondence should be addressed to P.P. (paola.picotti@bc.biol.ethz.ch).

Published online 26 October 2017; doi:10.1038/nprot.2017.100

Protein structural changes induced by external perturbations or internal cues can profoundly influence protein activity and thus modulate cellular physiology. A number of biophysical approaches are available to probe protein structural changes, but these are not applicable to a whole proteome in a biological extract. Limited proteolysis-coupled mass spectrometry (LiP-MS) is a recently developed proteomics approach that enables the identification of protein structural changes directly in their complex biological context on a proteome-wide scale. After perturbations of interest, proteome extracts are subjected to a double-protease digestion step with a nonspecific protease applied under native conditions, followed by complete digestion with the sequence-specific protease trypsin under denaturing conditions. This sequential treatment generates structure-specific peptides amenable to bottom-up MS analysis. Next, a proteomics workflow involving shotgun or targeted MS and label-free quantification is applied to measure structure-dependent proteolytic patterns directly in the proteome extract. Possible applications of LiP-MS include discovery of perturbation-induced protein structural alterations, identification of drug targets, detection of disease-associated protein structural states, and analysis of protein aggregates directly in biological samples. The approach also enables identification of the specific protein regions involved in the structural transition or affected by the binding event. Sample preparation takes approximately 2 d, followed by one to several days of MS and data analysis time, depending on the number of samples analyzed. Scientists with basic biochemistry training can implement the sample preparation steps. MS measurement and data analysis require a background in proteomics.

INTRODUCTION

The structures of proteins closely reflect their functional state, as they integrate multilayered cues that regulate protein activity. Measuring dynamic alterations of protein structures on a proteome-wide scale can thus provide functional readouts of biological systems and pathological states. Mass-spectrometry (MS) methods are frequently used to identify and quantify proteins in biological samples, to analyze protein–protein interactions, and to investigate post-translational modifications (PTMs)[1]. Protein structural alterations, for example, alterations as a consequence of binding of other molecules, chemical derivatization, or mutations, have for a long time not been amenable to proteomic analysis on a global scale and at high throughput. To address this limitation, we recently developed a technique called limited proteolysis-coupled MS (LiP-MS)[2]. LiP-MS enables measurement of protein structural transitions directly in biological matrices and on a proteome-wide scale.

LiP-MS can detect subtle alterations in secondary structure content[2]; larger-scale movements, such as those induced by allostery[2]; and more pronounced transitions such as switching between folded and unfolded states[3] or multimerization events[2]. The method can also be used to pinpoint protein regions undergoing structural transition with peptide-level resolution[2,3]. LiP-MS has been used to detect protein structural rearrangements induced by specific perturbations[2–4]. Further, the technique has been applied to the analysis of protein aggregation[2] and to the identification of protein–small molecule interactions[2,5], as the binding event can change the structural properties of the target protein, resulting in altered LiP patterns. As LiP experiments can be coupled to discovery and targeted proteomics analyses[2], the method supports both the structural analysis of specific proteins of interest in a biological sample and unbiased, whole-proteome measurements.

LiP-MS: monitoring protein structural changes in complex biological samples

LiP experiments rely on proteases with broad specificity applied for a short time to a proteome extract under native conditions. This ensures that proteolysis sites are dictated by the structural features of the protein substrates (**Fig. 1**). Detecting differences in LiP patterns enables the identification of protein regions involved in structural rearrangements. To date, LiP has typically been applied to purified proteins[6–11], owing to the challenge of identifying LiP sites in complex backgrounds. The LiP-MS method represents the proteome-wide extension of LiP experiments. To enable the identification and quantification of LiP products on a proteome-wide scale and directly in complex biological extracts, the LiP-MS approach exploits a double-digestion step that generates peptides amenable to bottom-up proteomic analysis and the sensitivity and specificity of advanced MS to probe LiP patterns in complex matrices. The approach can be applied in an unbiased, discovery-driven manner to detect proteins undergoing structural transitions in differently perturbed samples[2,3]. In this case, each proteome extract is subjected to the double-proteolysis step and the resulting peptides are quantified by unbiased MS techniques (e.g., shotgun proteomics). Differences in relative peptide abundances between the samples are indicative of proteins that

change structural properties across the conditions compared. Alternatively, the method has been used for the targeted analysis of structural changes of specific proteins of interest[2,4]. For this purpose, peptide markers along the whole protein sequence are measured using a targeted MS workflow, and regions of the protein of interest that undergo structural changes are identified. Compared with discovery-driven methods, a targeted approach monitors fewer peptides and gains in terms of reproducibility and specificity. The two workflows share the same sample preparation steps and can be applied exclusively or sequentially to analyze a proteome, a protein network, or a single protein in proteome extracts from cells or biological fluids.

Overall, LiP-MS is a powerful technique for probing structural changes and thus functional alterations of proteins and for gaining insight into physiological and pathological rearrangements in a proteome.

Applications of LiP-MS

The LiP-MS method finds applications in biological research as well as in biomedical, biotechnological, and pharmaceutical studies. Several applications of LiP-MS have been demonstrated in previous studies:

Global analysis of protein structural changes upon specific environmental perturbations. LiP-MS can be used to detect proteins in a proteome that undergo structural changes upon differential treatment. For example, by coupling LiP-MS data with protein abundance measurements, we demonstrated that certain pathways involved in carbon metabolism in *Saccharomyces cerevisiae* are regulated by enzyme structural changes, whereas others are regulated at the transcriptional level in response to nutrient availability[2]. In another study, the approach was used to detect pH-induced structural changes in viral proteins[4] during viral entry into mammalian cells. Data from LiP-MS experiments can be integrated with data from other proteomic analyses, for example, to simultaneously capture protein-abundance changes, PTM alterations, and protein structural rearrangements[2,4].

Targeted analysis of proteins of interest. LiP-MS has been used to analyze structural transitions in predefined sets of proteins under different conditions. For example, the structural rearrangements of myoglobin, α-synuclein, pyruvate kinase, viral proteins, and the entire network of central carbon metabolism enzymes in *S. cerevisiae* were studied in detail in targeted applications of the method[2,4].

Identification of the specific regions undergoing structural changes. LiP-MS was used to map regions within a protein structure that undergo structural transitions directly in a biological sample[2]. The resolution currently achieved by the method in determining regions with altered structural properties within a protein is in the range of ~10 amino acids.

Large-scale identification of protein–small molecule interactions. Binding of small molecules can change the proteolytic accessibility of a protein and therefore its LiP pattern. LiP-MS has been used for unbiased detection of proteins in a proteome that binds a small molecule of interest. Protein–small molecule interactions are identified by comparing LiP patterns of proteomes exposed to and in the absence of a specific compound. This strategy enabled identification of proteins in *S. cerevisiae* that bind the metabolite fructose-1,6-bis-phosphate[2] and interactors of the metabolite L-arginine in primary T cells[5]. The approach can also reveal small-molecule binding site(s) or protein regions distally affected by the binding event. Although the small molecule of interest can be applied to cultured cells, addition of the compound to a proteome extract (i.e., after cell or tissue lysis) is preferable, to minimize the detection of

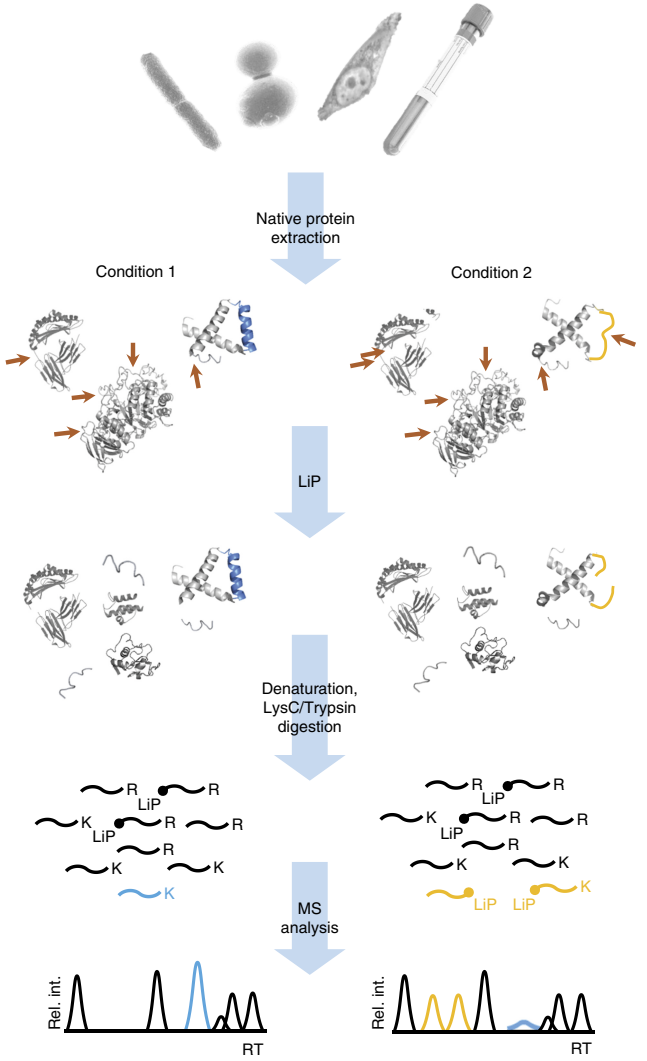

Figure 1 | Experimental LiP-MS workflow. Proteins are extracted from cells or tissues under native conditions. Each proteome extract is split into a control sample and a sample to be subjected to LiP. The workflow applied to LiP samples is visualized. In Step 1, samples for LiP are subjected to pulse proteolysis with a broad-specificity protease, such as PK, to generate structure-specific protein fragments. In Step 2, control samples and samples previously subjected to LiP are denatured and fully digested with LysC/trypsin to generate peptides amenable to bottom-up proteomics. The peptide mixture includes canonical fully tryptic peptides and half-tryptic peptides deriving from the LiP step. Subsequently, all samples are subjected to an unbiased and/or targeted mass-spectrometry analysis. Proteolytic patterns of samples subjected to LiP are compared, after correction for protein abundance changes using control samples. Arrows indicate LiP sites. K and R indicate tryptic termini. K, lysine residue; R, arginine residue; Rel. int., relative intensity.

indirect effects such as protein structural changes in pathways modulated by the compound.

Analysis of protein aggregation. Protein aggregation (including formation of amyloid deposits) substantially reduces the proteolytic accessibility of a protein, which in turn alters LiP patterns. LiP-MS has been used to probe protein aggregation of disease-related amyloidogenic proteins directly in a biological context[2]. This application has particular potential for the study of protein-aggregation diseases, such as Alzheimer's and Parkinson's diseases. Conformation-specific peptides (conformotypic peptides) produced during protease digestion can be quantified to probe the structure of disease-related proteins in clinical samples and have potential as biomarkers. Furthermore, LiP-MS analysis of samples in the presence of drug candidates will reveal their capability to influence the aggregation process directly in complex biological extracts.

The following sections detail additional applications of the approach that can be envisioned but that have not yet been demonstrated.

Protein quality control. LiP-MS could be applied to the quality control of therapeutic proteins (e.g., to monitor the stability of the protein fold over time), an important step for the pharmaceutical industry in the drug production pipeline.

Drug design. As LiP-MS can in principle be used to probe the structure of protein receptors of interest directly within their membrane matrix, the method could be used to aid the design of molecules to target these receptors.

Biomarker discovery. As the LiP-MS approach is compatible with the analysis of human samples, it could be applied to biomarker-discovery studies to identify what we termed 'conformational biomarkers'[2] as opposed to classic, concentration-based biomarkers of disease.

Comparison with other structural methods

Structural techniques, such as X-ray crystallography, NMR spectroscopy, and hydrogen/deuterium exchange, can measure structural changes of purified proteins or simple protein systems reconstructed *in vitro* but are not applicable to complex biological samples[12–15]. Förster resonance energy transfer (FRET) and in-cell NMR[16,17] measure structural changes directly in the cellular environment but cannot be substantially multiplexed and require labeling of the target protein, thus precluding analysis of human samples. Cross-linking MS (CX-MS)[18] is gaining momentum as a tool to derive structural restraints from purified proteins or protein complexes. However, it has limited utility in the quantitative analysis of protein conformational changes in complex biological matrices such as whole-cell lysates, and identification of cross-linked peptides poses analytical and computational challenges. Surface labeling and surface footprinting[19,20] can detect solvent-accessible residues in complex samples, but these methods rely on chemical modifications of proteins that alter the physicochemical properties of the modified regions, such as hydrophobicity, electrostatic interactions, and steric hindrance, which in turn harbor the risk of monitoring non-native protein structures. LiP-MS offers the following advantages with respect to existing structural methods for protein analysis.

Direct applicability to complex biological matrices. LiP-MS can be used to study protein structural changes directly in cell extracts, without manipulation of the target proteins.

Compatibility with hypothesis- and discovery-driven applications. When coupled to targeted proteomics measurements (e.g., selected reaction monitoring (SRM) MS)[21], the LiP-MS approach can be used to probe structural changes of one or several proteins of interest with high sensitivity and reproducibility in their biological matrix. FRET or in-cell NMR could be used for the same purpose, with the above-described limitations. When coupled to discovery-driven MS techniques such as shotgun proteomics, LiP-MS can be applied to a whole proteome to identify proteins that change structural features upon a given perturbation in an unbiased manner. Performing protein structural studies on a proteome-wide scale is a goal of obvious importance that cannot be achieved using other structural methods, with the possible exception of CX-MS[22]. CX-MS could, in principle, be applied to the unbiased detection of conformationally variant proteins in a proteome[22], but it suffers from sensitivity issues in unfractionated cell extracts. As a result, CX-MS experiments are typically performed on immunoprecipitated or *in vitro* reconstituted protein samples[18,23,24].

High throughput. The time required to perform a LiP-MS experiment is comparable with that required to perform a classic quantitative proteomic experiment, and multiple samples can be processed in parallel. For example, analysis of 30 different proteome extracts by LiP-MS requires 2 d of sample preparation. The subsequent discovery-driven or targeted MS analysis requires 3.5 d or 1.5 d of measurement time, respectively. Thus, the experiments can be performed at a relatively high throughput.

Simplicity. The method is based on a simple protocol and can be performed with standard equipment and reagents available in a biochemistry laboratory. One validated application of LiP[2], the unbiased detection of protein–small molecule interactions in cell extracts, has recently been performed with a method involving thermal proteome profiling by quantitative MS[25]. As binding of a small molecule may alter the thermal stability of the target protein, thermal denaturation profiles of proteins along a temperature gradient may enable the detection of proteins undergoing ligand-induced shifts in their thermal stability. Unlike the thermal profiling approach, LiP-MS can pinpoint regions in a protein locally or distally affected by binding of the compound (e.g., binding sites or allosteric paths), and the method relies on a substantially simpler protocol, as it does not require exposure of a proteome to a broad temperature range. In addition, LiP can be used for a range of other applications.

Applicability to various types of samples. A broad range of samples can be processed with the LiP-MS workflow, including microbial or mammalian cell extracts or extracts from human tissues or body fluids. This is an advantage over tagging-based techniques such as FRET and in-cell NMR, which cannot be applied to human samples. LiP-MS can be optionally combined

with enrichment or fractionation steps, as long as the chosen fractionation techniques preserve protein structures.

Limitations

Lack of information on the nature of the structural change. When applied in discovery-based experiments, LiP-MS yields a list of proteins that undergo a structural change upon application of a given chemical or environmental perturbation, and provides information on protein regions involved in the structural rearrangement. However, LiP-MS does not provide information on the nature of the structural change. The change could be due to, for example, a conformational change in the protein of interest, ligand binding, protein aggregation, or altered protein–protein interactions. The nature of the structural change must be addressed in follow-up experiments, as previously shown[2,4]. For example, the trigger of the conformational change (e.g., binding of an allosteric regulator or post-translational modifications) can be suspected based on orthogonal 'omics data sets'[2] or literature knowledge and validated by an *in vitro* reconstitution assay, using LiP[2] or other structural methods as a readout. The effect of the structural change on enzyme activity can be evaluated by activity assays performed in the lysate or with the purified enzymes[2].

Sensitivity. The method is currently most successful for proteins of high-to-medium abundance. Detecting structural changes for low-abundance proteins is challenging, as the corresponding peptides may not be detectable in a complex cell extract. For example, in Feng *et al.* (2014), the method yielded structural data for about one-third of the expressed proteome in the yeast *S. cerevisiae*[2]. Because of subsequent improvements, coverage has now been expanded to 60–70% of the expressed yeast proteome. To further expand the coverage, different fractionation techniques can be coupled to LiP-MS (e.g., at the protein level, size-exclusion chromatography). Fractionation increases detection of low-abundance proteins but, as in all proteomics-based experiments, it reduces the throughput of the method. Coupling of LiP with targeted MS may also increase the likelihood of detecting soluble proteins of interest.

Membrane proteins. The LiP-MS approach is also affected by a bias against membrane proteins due to the native extraction step[3].

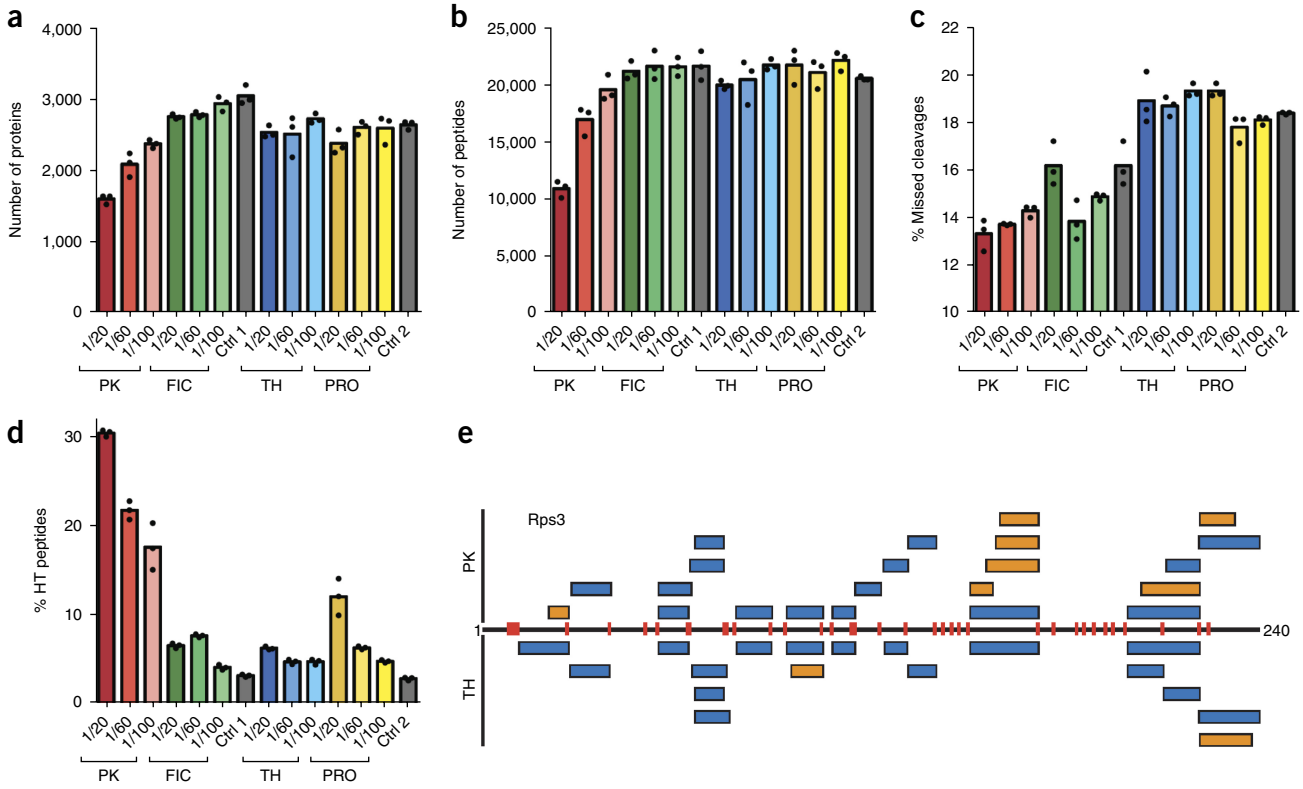

Figure 2 | Effect of different E/S ratios and choice of proteases. (**a**–**d**) Protein extracts from *S. cerevisiae* cells were subjected to the LiP-MS protocol using PK (red), ficin (FIC, green), thermolysin (TH, blue) or pronase (PRO, yellow) at different E/S ratios as LiP proteases. The incubation time with the LiP protease was 1 min for all the samples displayed. The resulting fragments from LiP-treated and control samples were analyzed by LC–MS/MS and can be found in the online repository Dataverse (https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/A3FYXF). Bars indicate average values from three replicated analyses. Dots indicate values of each replicate. (**a**) The total number of proteins and (**b**) peptides, (**c**) the percentage of missed cleavages and (**d**) half-tryptic (HT) peptides out of the total number of peptides identified are reported for the different treatments. For TH and PRO, which bind Ca^{2+} and Zn^{2+} ions as cofactors, respectively, EDTA was added at the protease-quenching time before boiling. Addition of EDTA quenches the activity of thermostable TH, for which boiling would not be sufficient, and it contributes to the efficient quenching of PRO. For consistency, EDTA was added to the respective control samples at the same step. Ctrl 1: control without EDTA (gray); Ctrl 2: control with the addition of EDTA. (**e**) Results for the yeast 40S ribosomal protein S3 (Rps3), using PK or TH as LiP proteases, at an E/S ratio of 1/100. The results of the above samples are described and shown in **a**–**d**. The black line represents the sequence of the protein; red boxes indicate the position of K and R residues. Blue and orange boxes represent fully tryptic and half-tryptic peptides identified during the LiP-MS analysis, respectively. The position of the boxes reflects the position of the associated peptide along the protein sequence.

Box 1 | Chaotropes and surfactants

Preserving native protein structures upon extraction of a proteome from cells and tissues is critical to the success of a LiP experiment. To this end, various detergents commonly added to lysis buffers should be omitted. To improve extraction of specific classes of proteins (e.g., membrane proteins), nondenaturing detergents and nonionic surfactants such as *n*-dodecyl-β-d-maltoside[49] can be used.

Protein fragments generated during the LiP step undergo complete digestion with LysC and trypsin. At this step, the polypeptides are first denatured by the addition of denaturing agents such as ionic[50] or zwitterionic[51] surfactants and chaotropes[52] that disrupt intra- and intermolecular interactions. Under these conditions, disulfide bridges are reduced and free cysteines are alkylated. The concentration of the denaturing agent is then reduced by dilution before the addition of LysC and trypsin to ensure protease activity. Many detergents commonly used in molecular biology are not compatible with MS analysis and therefore cannot be used in this step; among these are SDS and Triton X-100. Compatible agents are, for example, the ionic surfactant DOC and the chaotropes guanidine hydrochloride (Gdn) and urea. DOC is the preferred agent in our workflow because of the high number of fully tryptic and half-tryptic peptide identifications and low number of missed cleavages we observed in validation experiments using this agent (**Supplementary Fig. 1**). DOC was the most effective of the agents tested on a variety of proteins by Proc *et al.*[52] and in our experience. In addition, DOC is stable at high temperature and can be removed by precipitation at acidic pH. Urea (typically at a starting concentration of 8 M) can be used in applications that do not involve heating of the sample to >37 °C; at high temperatures, urea decomposes into isocyanate and ammonia, thereby inducing peptide carbamylation[53]. Gdn at high concentrations (e.g., 7 M) is preferable in applications in which the structural features of insoluble amyloid aggregates are studied, as the efficiency of this chaotrope in denaturing amyloid structures has been demonstrated[54–56]. A drawback of the use of Gdn is a decrease in efficiency of trypsin cleavage, even after dilution to <1 M before addition of the protease[52,57]; this effect can be partly counteracted by digestion with LysC (compatible with 2 M Gdn) before trypsinization[30]. Urea and Gdn are typically removed from the peptide solution during the peptide desalting step before MS analysis.

Recommended use

DOC: Add DOC to a final concentration of 5% after the LiP step (as indicated in Step 8). Formation of a gel can occur, but this does typically not impact subsequent steps. After trypsin digestion, DOC is precipitated by acidification and separated from the peptide samples.

Urea: Add urea to a final concentration of 8 M after the LiP step (before Step 7). Ensure that the temperature of the sample is reduced to <37 °C after the boiling step to prevent peptide carbamylation by urea decomposition byproducts. No dilution is necessary before LysC digestion. Dilute it to <2 M before adding trypsin (Step 13).

Gdn: Add Gdn to a final concentration of 7 M[57,58]. We recommend adding Gdn before transferring LiP samples to boiling water (quenching; Step 7) to facilitate dissolution of the Gdn powder. This can be achieved by pipetting LiP mixtures at the quenching time directly into tubes containing the preweighed amount of Gdn powder. Addition of Gdn will also contribute to denaturing the LiP protease. Dilute to 2 M or 0.5 M Gdn before the addition of LysC or trypsin, respectively.

To target membrane proteins, performing LiP on a suspension of membrane debris or directly on cells permeabilized with suitable detergents may be attempted, although this application has not yet been thoroughly evaluated.

Limited degree of structural characterization when applied to a single condition. Application of LiP-MS to proteins or proteomes in a single condition provides information on protein regions that are most accessible to the protease, i.e., exposed and flexible regions. Based on this, LiP has been used to identify loops and locally unfolded regions, protein segments connecting different domains, aggregation cores[26], and interfaces of proteins and DNA or protein complexes[27]. However, the structural information that can be extracted from a LiP experiment is limited compared with information that can be obtained from high-resolution approaches such as X-ray crystallography, NMR, or in-cell NMR. The unrivaled power of LiP-MS is obvious in comparative experiments conducted on complex biological samples from different conditions and aimed at identifying proteins and protein regions undergoing perturbation-induced structural changes.

Artifactual structural alterations. The protein extraction step could introduce protein structural alterations, e.g., due to detachment of interactors by dilution, loss of post-translational modifications, or degradation. These factors can affect the samples that are being compared to the same extent or to different extents, depending on whether they are dependent on the extraction step alone or also on the specific biological background.

Protocol overview

Protein extraction. LiP can be conducted directly on biological fluids. For other samples, such as cells or tissues, the first step of the LiP-MS protocol is the extraction of proteins under conditions that preserve the structure of proteins (Step 1). For this reason, pH and ionic strength of the lysis buffer must be as close as possible to physiological values, denaturing detergents such as SDS should be omitted, and dilution of the extract should be kept to a minimum. Extraction steps before addition of the LiP protease should be performed as quickly as possible and on ice. To improve the solubilization of specific classes of proteins such as membrane proteins, nondenaturing MS-compatible surfactants can be used (**Box 1**; **Supplementary Fig. 1**). For ease of handling, we recommend using physical methods for cell lysis, such as bead-beating for bacteria and yeast cells, and homogenization with a douncer for mammalian cells and tissues. Optionally, inhibitors of endogenous proteases can be added to the protein extraction

Box 2 | Choice of enzyme to substrate ratio, incubation time and protease

Different broad-specificity proteases can be used for LiP (**Table 1**). Previous studies indicated that under LiP conditions, the sites of initial proteolytic cleavage are dictated by the structural features of the substrate, and different promiscuous proteases cleave at the same regions of a protein structure, even when they display slightly different sequence preferentiality[2,29]. In a LiP-MS workflow, the voracity and sequence preferentiality of the protease can, however, affect the number and type of peptides generated. To illustrate this, we compared the results obtained with the four promiscuous proteases (PK, thermolysin (TH), pronase (PRO), and ficin (FIC)) applied to a yeast proteome extract. We also tested the effect of different E/S ratios. PK was the most aggressive protease, generating the largest fraction of half-tryptic (HT) peptides (**Fig. 2d**). The amount of HT peptides mostly increased at increasing E/S ratios. The voracity of PK resulted in the identification of the lowest number of peptides and proteins (**Fig. 2a**,**b**)—probably because of increased sample complexity and decreased occurrence of fully tryptic (FT) peptides—and in the identification of the largest number of LiP cleavages. This is exemplified by the detailed analysis of yeast Rps3 (**Fig. 2e**), for which the application of PK resulted in a larger number of LiP sites and HT peptides as compared with treatment with TH. This in turn resulted in a higher sample complexity and in a slightly lower sequence coverage for Rps3 in the PK-treated sample as compared with the TH-treated sample. Lower numbers of peptide and protein identifications were also observed at increasing E/S ratios (**Fig. 2a**,**b**). PK and FIC resulted in a slightly lower number of peptides containing missed cleavages as compared with the respective control samples, whereas no difference was observed for the other proteases (**Fig. 2c**). Increasing the incubation time mimicked an increase in the E/S ratio and resulted in lower proteome coverage and in a larger fraction of HT peptides (**Supplementary Fig. 2**). The number of HT peptides decreased again after prolonged incubation, probably because of the occurrence of secondary cleavages.

On the basis of these and other experiments conducted in our laboratory, we recommend the preferential use of PK, because of its broad specificity and efficacy, using an E/S of 1/100 and an incubation time of 1 min. These conditions ensure both the detection of a large number of LiP sites within a proteome and a reasonable proteome coverage (comparable to or only slightly lower than that obtained with other proteases, **Fig. 2**). Long incubation times (>30 min) or high amounts of enzyme (E/S > 1/10) do not seem useful, as they lead to the generation of secondary cleavage products (i.e., initial proteolysis products are further digested to shorter peptides), thus complicating the interpretation of cleavage patterns. An incubation time of 1 min allows for sufficient handling time and minimization of secondary cleavages.

For specific applications, other proteases can be used. For example, TH is a preferable enzyme in LiP analyses at increased temperatures. For TH, we recommend increasing the E/S ratio (e.g., to 1:20), and quenching of protease activity by the addition of 25 mM EDTA. Subsequent boiling of the samples can be performed, but it is not strictly required. To maintain full protease activity, 800 μM $CaCl_2$ is added before trypsin addition. The enzyme pepsin can be used in experiments at low pH, for example, those in which the partly folded or molten globule states of proteins are analyzed[47]. Other proteases that were previously used in LiP experiments on purified proteins are shown in **Table 1**. Although proteases such as trypsin and Lys-C have been used in LiP experiments[29,48], the sequence specificity of these proteases reduces the likelihood that LiP sites are dictated only by the structural features of the substrates in the LiP step. Therefore, we recommend the use of these proteases only in the second, structure-independent proteolysis step of a LiP-MS workflow.

buffer (see Troubleshooting section). Protein extracts and biological fluids can be subjected to fractionation or enrichment steps to increase the likelihood of detecting specific proteins of interest. Fractionation conditions must be chosen in a way that ensures preservation of the protein structure. Recommended extraction methods for bacteria, yeast, and mammalian cells are described in the PROCEDURE (Step 1).

The concentration of total protein in the proteome extracts is determined by a commercial assay (e.g., bicinchoninic acid (BCA) assay) before addition of the LiP protease. Internal standards such as proteins or synthetic peptides can be spiked in at this stage to correct for variations in protease activity and environmental factors. Endogenous proteins with well-documented proteolytic patterns can also be used as internal controls[2,28,29].

The protein extract is split into two samples. One is subjected to the LiP step to generate a digestion pattern that depends on protein structure, followed by denaturation and tryptic digestion; the other is subjected to only tryptic digestion under denaturing conditions. The latter sample is subsequently used as a control to correct for changes in protein abundance across the tested conditions and (optionally) to detect altered activity of endogenous proteases and possible differential PTMs.

Limited proteolysis. In this step, nonspecific proteases, typically proteinase K (PK), are applied to the proteome of interest for a short time to generate proteolytic patterns that depend on the structural features of the proteins (Steps 2–8; **Fig. 1**). The enzyme/substrate (E/S) ratio, incubation time, and available proteases are discussed in **Box 2** (also see **Table 1**; **Fig. 2**; **Supplementary Fig. 2**). After incubation with the nonspecific protease, activity of the protease is quenched. If PK is used, we recommend transferring the sample to a boiling water bath, which irreversibly denatures the enzyme. The following considerations are critical to the LiP step to ensure the reproducibility of proteolytic patterns: first, enzymes must be kept cold, and enzyme aliquots should be discarded after use; second, incubation time with the nonspecific protease must be exactly controlled. Quenching of the protease is a critical step, as immediate and complete inactivation of the protease is necessary to maximize the reproducibility of proteolytic patterns.

Sequential LysC and trypsin digestion, and peptide cleanup. Structure-specific protein fragments generated during LiP are processed through a standard proteomics sample preparation workflow to generate peptides amenable to bottom-up MS analysis (Steps 9–16). Chaotropic agents compatible with MS analysis,

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

TABLE 1 | Proteases that can be used in LiP-MS experiments.

Protease/ species	EC number	Type	Optimum pH	Specificity	Notes	Ref.
Proteinase K/ *Tritirachium album*	3.4.21.64	Ser-P	7.5–8.0	Nonspecific; slight preference for C terminus of hydrophobic aliphatic and aromatic amino acids	Ca^{2+} protects against autolysis and increases thermal stability, but it is not required for activity. A temperature of 50–60 °C increases activity, as does addition of 0.5–1% SDS, 3 M Gdn HCl or 4 M urea. Stable over a broad (4.0–12.5) pH range	2,3,29
Thermolysin/ *Bacillus thermoproteolyticus*	3.4.24.27	Zn-MeP	7–8.5	Preference for N terminus of bulky hydrophobic amino acids, such as L, F, I, V, M and A	Ca^{2+} and Zn^{2+} act as cofactors. Zn^{2+} is required for activity and Ca^{2+} for stability. Zn^{2+} is bound with high affinity; when added in excess it inhibits Th. For optimal stability, Th may be used in the presence of 1–10 mM $CaCl_2$. Thermostable enzyme, active at 25 °C, with an optimal temperature at 65–85 °C. Stable between pH 5.0 and 8.5	3,29,40
Subtilisin/ *Bacillus subtilis*	3.4.21.62	Ser-P	7.0–8.0	Nonspecific; slight preference for cleavage at C terminus of large uncharged amino acids	Stable between pH 4.0 and 11.0	29,40,41
Chymotrypsin/ *Bos taurus*	3.4.21.1	Ser-P	7.0–9.0	Preference for C terminus of F, Y, W, L and I	Activated and stabilized by Ca^{2+}. Addition of 10 mM $CaCl_2$ to the reaction is recommended. Self-digestion may occur at temperatures >37 °C. Autolysis occurs at high pH. Active in the presence of 0.1% SDS and 2 M Gdn HCl	29,42
Papain/ *Carica papaya*	3.4.22.2	Thiol-P	6.0–7.0	Nonspecific; slight preference for peptide bonds adjacent to R, K, Q, H, G, Y and L	The disulfide bond is required for activity. Before use, it should be activated by incubation in 1 mM EDTA and 5 mM cysteine for 30 min	29,43
Elastase/ *Sus scrofa*	3.4.21.36	Ser-P	7.5–8.8	Preference for C terminus of A, V, I, L, G and S	Stable between pH 4.0 and 10.0	29,40,44
Ficin/*Ficus carica*	3.4.22.3	Thiol-P	6.5	Nonspecific; slight preference for the C terminus of G, S, T, M, K, R, Y, A, N and V	Stable between pH 4.0 and 9.5	45
Pronase/ *Streptomyces griseus*	3.4.24.4	Protease mix	7.0–8.0	Nonspecific; mixture of proteases from *S. griseus*	Ca^{2+} is recommended for protection from autolysis. Can be dissolved in 0.01 M Na^+ acetate, 0.005 M Ca^{2+} acetate, pH 7.5 at 37 °C. Stable between pH 6.0 and 9.0	40
Bromelain/ *Ananas comosus*	3.4.22.4	Thiol-P	5.0–6.5	Nonspecific; slight preference for C terminus of K, A, Y and G	A temperature of 45–65 °C enhances activity. Optimum temperature, 62 °C. Above 70 °C, activity is decreased. Activated by cysteine, bisulfite salt, NaCN, H_2S, Na_2S and benzoate. The enzyme is usually sufficiently active without the addition of activators. Stable between pH 3.0 and 9.0	46

(continued)

TABLE 1 | Proteases that can be used in LiP-MS experiments (continued).

Protease/ species	EC number	Type	Optimum pH	Specificity	Notes	Ref.
Pepsin/ *Sus scrofa*	3.4.23.1	Asp-P	1.5–2.5	Nonspecific; slight preference for F, M, L and W adjacent to another hydrophobic amino acids	Recommended for LiP experiments at acidic pH. Still active at pH 4.0; becomes unstable at nonacidic pH. Irreversibly denatured at pH >8.0–8.5. Active in 4 M urea and 3 M guanidine HCl. Stable at 60 °C	47
Trypsin/*Bos Taurus*	3.4.21.4	Ser-P	7.5–9.0	Specific; cleaves at C terminus of R and K	Subject to autolysis, generating pseudotrypsin, which exhibits chymotrypsin-like specificity. MS-grade trypsin is typically modified by reductive methylation of Lys residues to prevent autolysis. Specificity can be further improved by TPCK treatment, which inactivates chymotrypsin. Resistant to 0.1% SDS, 1 M urea, or 10% ACN. Retains 50% activity in 2 M Gdn HCl. Lys-Pro and Arg-Pro bonds are almost completely resistant to cleavage	2,3,29
Endoproteinase Lys-C/*Lysobacter enzymogenes*	3.4.21.50	Ser-P	7.0–9.0	Specific; Cleaves at C terminus of K	Remains active in 8 M urea, 2 M Gdn HCl, 1% SDS, 2% CHAPS and 40% ACN	2,3,48

A, Ala; ACN, acetonitrile; Asp-P, aspartic protease; C, Cys; CHAPS, 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate hydrate; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; Q, Gln; R, Arg; S, Ser; Ser-P, serine protease; Thiol-P, thiol or cysteine protease; TPCK, tosyl phenylalanyl chloromethyl ketone; V, Val; W, Trp; Y, Tyr; Zn-MeP, zinc metalloprotease.

such as urea or guanidinium hydrochloride (**Box 1**), are added to denature proteins; disulfide bridges are reduced; and free cysteines are alkylated to promote accessibility to proteolytic enzymes used in downstream steps. The choice of the denaturing agent must be adapted to the type of sample and temperatures used in the protocol. Suitable chaotropes and surfactants are discussed in **Box 1**. Tryptic digestion is conducted by the sequential addition of LysC and trypsin. Predigestion with LysC, which cleaves at the C-terminal side of lysines, can be conducted at higher concentrations of denaturing agents than those used in the following digestion with trypsin[30]. After tryptic digestion, peptide mixtures are subjected to a desalting step performed using cartridges packed with a C18 resin, the elution solvent is evaporated to dryness, and peptide mixtures are resuspended in an acidic buffer compatible with subsequent LC–MS analysis (Steps 17–19).

Unbiased shotgun proteomics analysis. Peptides can be analyzed with a discovery-driven shotgun proteomics workflow using a high-resolution high-mass accuracy mass spectrometer (e.g., a QExactive Plus (Thermo Fisher Scientific) or a TripleTOF (AB Sciex) instrument), equipped with a nanoelectrospray ion source and coupled to a nano-LC system (Step 20A). Analysis of singly charged precursor ions during the acquisition of LC-MS/MS data is discussed in **Box 3**. After data acquisition, MS/MS spectra are assigned to peptide sequences using well-established database search tools. Label-free quantification of peptide ions is achieved with software packages for quantitative proteomics analyses such as Progenesis (Nonlinear Dynamics), MaxQuant[31], OpenMS[32], or others[33]. These tools perform the alignment of peptide ions across MS runs, attribute peptide identifications based on database search results, integrate peptide peak intensities along the chromatographic time, and perform relative quantification of peptide ion profiles across the different samples (Step 21).

Targeted proteomics analysis. Targeted structural analyses of specific proteins of interest can be performed as an alternative or a follow-up to proteome-wide analyses (Step 20B). A targeted mass-spectrometry technique such as SRM is used to maximize reproducibility, specificity, and sensitivity. In this workflow, a protein of interest is first selected, for which structural changes will be monitored across multiple samples[34]. All possible unique (proteotypic) fully tryptic peptides from this protein are predicted or derived for previous proteomic data sets, for example, using the Skyline software (https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view?; ref. 35). For each proteotypic peptide, an SRM assay is developed, guided by spectra from discovery-driven analyses, where available. A liquid chromatography-coupled triple quadrupole MS (e.g., a 5500QTrap (AB Sciex) or a TSQ Quantiva or Endura (Thermo Fisher Scientific)) is used for the SRM analysis. Relative quantification of peptide abundances can be performed using tools such as Skyline[35] and MSstats (http://msstats.org/)[36]. Abundance changes of fully tryptic peptides can be followed up by monitoring of all their possible N- and C-terminal fragments by SRM to identify exact LiP sites and for confirmatory purposes.

Data analysis. The general idea behind the analysis of LiP-MS data is that protein regions embedding LiP cleavage sites are identified by an abundance decrease of the associated fully tryptic peptides or by the appearance or increased intensity of peaks

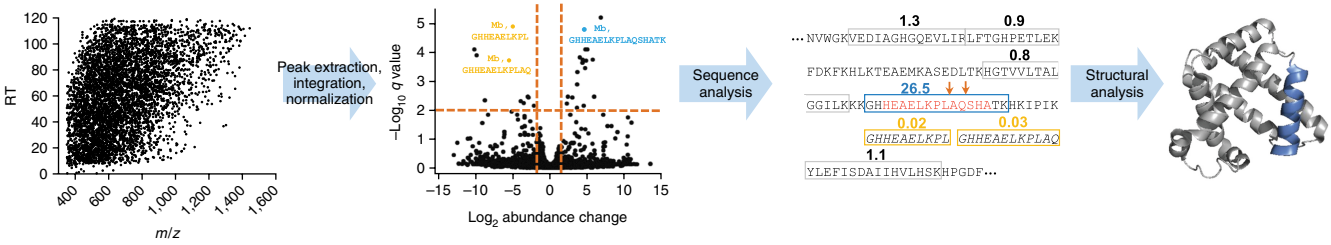

Figure 3 | Analysis of LiP-MS data. Samples in which horse myoglobin (Mb) is in a holo (i.e., bound to the heme group) or apo (i.e., deprived of heme) conformational state are comparatively analyzed with the LiP-MS workflow. (First section) Analysis of LC-MS/MS data. Peptide mixtures from control or LiP-treated samples are analyzed by a label-free LC–MS/MS approach. Peptide ion peaks are detected and integrated over the retention time dimension, using tools such as Progenesis or alternative software. (Second section) Quantitation. Protein abundance changes are calculated from comparison of control samples. Changes in LiP patterns are extracted from comparison of samples subjected to LiP, after normalization for protein abundance changes across conditions. Peptide-abundance changes are plotted against their probability (volcano plot). Significant changes (e.g., Log_2 fold change >2, q value <0.01) appear in the top left and right quadrants of the plot. (Third and fourth sections) Data visualization. Significant peptide abundance fold changes are displayed along the Mb sequence. Software tools such as Pymol are used to map LiP-MS data to the Mb structure. Upon heme removal, Mb undergoes a local unfolding of helix F (sequence marked in red). In the sample containing holoMb, a fully tryptic peptide mapping to helix F shows an abundance increase, as the associated region is more protected from proteolysis because of folding of the helix. Conversely, two half-tryptic peptides deriving from internal cleavage of the same region appear prominently in the sample containing apoMb because of the local unfolding. Peptides outside this region do not significantly change abundance. The experiment enables detection of the Mb region undergoing the structural change.

corresponding to peptides with nontryptic termini (half-tryptic peptides) in the native sample compared with that in the denatured control sample (**Fig. 3**, Steps 22–30). Comparison of LiP patterns from proteomes subjected to different conditions enables the identification of proteins and the respective regions undergoing perturbation-induced structural changes. A change in the intensity of fully tryptic and half-tryptic peptides is interpreted as a change in the probability of proteolytic cleavage under native conditions and, thus, a change in the accessibility or flexibility of the corresponding protein region. We termed these LiP peptides with altered abundances across conditions structure-specific or 'conformotypic' peptides[2].

Peptides generated in a LiP-MS experiment are on average shorter than classic tryptic peptides because of LiP cleavages occurring within the sequence of tryptic peptides. Shorter peptides tend to be more hydrophilic; thus, the distribution of peptide precursor ions in the LiP LC–MS/MS analysis is shifted toward early retention times. This has a few implications: (i) in shotgun analyses, LC–MS/MS runs from samples subjected to the LiP step can be aligned only among themselves. No peak alignment should be attempted against control samples subjected to only trypsin digestion. (ii) LiP-samples are not suitable for the estimation of protein abundance changes across different conditions, as a considerable portion of tryptic peptides will embed structure-specific cleavages. Therefore, unlike standard proteomics analyses, measured peptide intensities are not compiled to infer protein abundances, and each peptide is treated as an independent measurement. (iii) Control samples subjected to only trypsin cleavage are compared, aligned among themselves, and used for the relative quantitation of protein-abundance changes.

In practice, peptide abundances are first extracted from the comparison of LiP samples and protein abundance changes from the comparison of control samples. Protein abundance changes are then used as normalization factors to correct for peptide abundance changes from LiP samples. This prevents the interpretation of alterations in protein concentration across samples as altered structure-dependent proteolytic patterns. Optionally, peptide intensities in control samples can be used to detect or correct for varying efficiency of endogenous proteases or differential PTMs. Fully tryptic and half-tryptic peptides with significant fold changes after the normalization process are indicative of structural changes in the respective protein.

The results can be represented as peptide abundance changes versus their probabilities (volcano plots) (**Fig. 3**) to provide an overview of peptides and proteins that change significantly with the perturbation. The identified conformotypic peptides can be mapped onto the protein sequence and (if available) known 3D structure to evaluate topology and proximity to functionally important sites. Molecular viewers such as PyMOL (https://pymol.org) and VMD (http://www.ks.uiuc.edu/Research/vmd/) can help visualize and interpret the results (**Fig. 3**). Functional annotations from databases such as UniProt (http://www.uniprot.org/), ENSEMBL (http://www.ensembl.org/index.html), PFAM (http://pfam.xfam.org/) and Phosphopep (http://www.phosphopep.org/) can be incorporated as described in PROCEDURE. Functionally important sites may include conserved protein domains; clefts; small molecule–, metal-, or protein-binding sites; protein complex interfaces; PTMs; unstructured regions; alternative splicing sites; and disease-causing mutations.

SRM assays based on conformotypic peptides can be developed and used as markers for specific conformational changes across different conditions, and biochemical follow-up experiments can be designed to verify hypotheses on the nature of the structural change.

MATERIALS

REAGENTS

! CAUTION Even unclassified chemicals should be handled with gloves and protective glasses. Where applicable, classifications according to the Globally Harmonized System of Classification and Labelling of Chemicals (GHS) are indicated. Precautionary statements as specified on suppliers' safety data sheets must be followed.

- *S. cerevisiae* (Euroscarf, cat. no. BY4741)
- Escherichia coli K12 MG1655 (ATCC, cat. no. 700926)

- Mammalian cells (HeLa cells; ATCC, cat. no. CCL-2) ! CAUTION The cell lines used in your research should be regularly checked to ensure that they are authentic and that they are not infected with mycoplasma.
- Biological fluid such as cerebrospinal fluid or serum ! CAUTION For the use of human material, ethical approval must be obtained from your institution and use should conform to national regulations. Informed consent is required for the use of human material.
- M9 minimal medium for bacterial growth (Sigma-Aldrich, cat. no. M6030-1KG)
- Synthetic defined (SD) complete medium for yeast growth (Sigma-Aldrich, cat. no. Y1501-20G)
- HEPES BioPerformance, certified 99.5% (Sigma-Aldrich, cat. no. H4034)
- Potassium chloride (Merck, cat. no. 104.936.1000)
- Magnesium chloride hexahydrate, puriss. p.a. (Fluka, cat. no. 63072)
- Glass beads, acid-washed, 425–600 µm (Sigma-Aldrich, cat. no. G8772)
- Albumin standard (Thermo Fisher Scientific, cat. no. 23209)
- Pierce BCA Protein Assay (Thermo Fisher Scientific, cat. no. 23227) ! CAUTION Contains a reagent classified as an aquatic environmental hazard (GHS09).
- Proteinase K from *Tritirachium album,* lyophilized powder, BioUltra (PK; Sigma-Aldrich, cat. no. P2308) ! CAUTION PK is classified as having acute toxicity (GHS07) and as a health hazard (GHS08).
- Ammonium bicarbonate (Ambic; Fluka Analytical, cat. no. 40867) ! CAUTION Ambic is classified as having acute toxicity (GHS07).
- DL- DTT, 98% (Sigma-Aldrich, cat. no. D0632) ! CAUTION DTT poses a risk of acute toxicity (GHS06) and is classified as a health hazard (GHS08).
- Iodoacetamide BioUltra (IAA; Sigma-Aldrich, cat. no. I1149) ! CAUTION IAA is classified as GHS6 and GHS8. ▲ CRITICAL IAA is light-sensitive and can be protected by, e.g., aluminum foil.
- pH-indicator paper for pH 0.5–5.5 and pH 6.4–8.0 (Macherey-Nagel, cat. nos. 90205 and 90210).
- Lysyl Endopeptidase (LysC) for Biochemistry (Wako Pure Chemical Industries, cat. no. 129-02541)
- Sequencing-grade modified trypsin, frozen (Promega, cat. no. V5113)
- Urea puriss. p.a. ACS reagent, ≥99.5% (Sigma-Aldrich, cat. no. 33247)
- Sodium deoxycholate, 97% (DOC; Sigma-Aldrich, cat. no. D6750) ! CAUTION DOC is classified as having acute toxicity (GHS07).
- Guanidinium hydrochloride, ≥99% (Sigma-Aldrich, cat. no. G3272) ! CAUTION Guanidinium is classified as having acute toxicity (GHS07).
- Formic acid, ~98% (Fluka Analytical, cat. no. 94318). ! CAUTION Formic acid is classified as flammable (GHS02), and as GHS05 and GHS06.
- Microsyringe (VWR, cat. no. 549-0523).
- Methanol ROTISOLV, >99.95%, LC–MS grade (Roth, cat. no. AE71.2) ! CAUTION It is classified as GHS2, GHS06, and GHS08.
- Acetonitrile ROTISOLV, >99.98%, ultra LC–MS (Roth, cat. no. HN40.2) ! CAUTION It is classified as GHS02, GHS07, and GHS08.
- HPLC-gradient-grade water (Fisher Scientific, cat. no. W/0106/17)
- Sep-Pak Vac Cartridges, 1-cc (50-mg), tC18 (Waters, cat. no. WAT054960)
- Peptide or protein internal standard (e.g., myoglobin from equine skeletal muscle, Sigma-Aldrich, cat. no. M5696; or human α-synuclein, Sigma-Aldrich, cat. no. S7820)
- PBS

EQUIPMENT

- Eppendorf Safe-Lock micro test tubes, 1.5 ml (Sigma-Aldrich, cat. no. EP0030120086)
- FastPrep 5G cell-disruption device (MP Biomedicals)
- Sterican needle, 26-gauge × 1 inch/0.45 × 25 mm (B. Braun, article no. 4657683)
- Microtube, 2 ml with cap (Sarstedt Nümbrecht, cat. no. 72.693)
- Multifuge 3 S-R (Heraeus)
- Eppendorf concentrator (Eppendorf, model no. 5301)
- Cell disruption device (e.g., MP Biomedicals, model no. FastPrep 5G)
- Dounce homogenizer
- Q Exactive Plus Orbitrap LC–MS/MS system (Thermo Fisher Scientific) equipped with nanoelectrospray ion source and coupled to a 40 cm × 75 µm (inner diameter) HPLC column (New Objective, cat. no. PF360-50-10-N-5) packed with 1.9-µm C18 beads (Dr. Maisch, cat. no. r119.aq)
- Nanoflow HPLC system (Thermo Fisher Scientific, model no. Easy-nLC 1000)
- QTRAP 5500 LC–MS/MS system equipped with nanoelectrospray ion source (AB Sciex, model no. NanoSpray II) and coupled to a 20 cm × 75 µm (i.d.) column (New Objective, cat. no. PF360-50-10-N-5) packed with 5-µm C18 beads (Michrom, cat. no. PM5/61200/00)
- Proteome Discoverer v1.4 (Thermo Fisher Scientific), using the Sequest HT or Mascot database search engines (http://planetorbitrap.com/proteome-discoverer)
- Progenesis v1 (Nonlinear Dynamics; http://www.nonlinear.com/progenesis/qi-for-proteomics/download/; versions up to v2.0.5556 can also be used)
- SafeQuant (https://github.com/eahrne/SafeQuant)
- Skyline (https://skyline.gs.washington.edu/labkey/project/home/software/Skyline/begin.view)
- R (https://R-project.org)
- MSstats R-package (http://msstats.org)
- PyMOL (Schrödinger; https://pymol.org)
- Excel (Microsoft)

REAGENT SETUP

Native lysis buffer Prepare native lysis buffer by mixing 20 mM HEPES, 150 mM KCl, and 10 mM $MgCl_2$ at pH 7.5. The buffer should be freshly prepared.

10× Native lysis buffer Prepare 10× native lysis buffer by mixing 200 mM HEPES, 1,500 mM KCl and 100 mM $MgCl_2$ at pH 7.5. The buffer should be freshly prepared.

DOC solution DOC solution is 10% (wt/vol) DOC in HPLC-grade water. Vortex the solution vigorously to dissolve the DOC. This solution can be stored at room temperature (20–24 °C) for at least 1 month.

Ammonium bicarbonate solution Prepare 100 mM ammonium bicarbonate solution by dissolving ammonium bicarbonate in HPLC-grade water. The solution should be freshly prepared.

DTT solution Prepare 700 mM DTT solution by dissolving DTT in HPLC-grade water. This can be stored at −20 °C for at least 3 months.

IAA solution Prepare 700 mM IAA solution by dissolving IAA in HPLC-grade water. IAA is light-sensitive; the solution should be freshly prepared.

Desalting buffer A/LC solvent A Desalting buffer A/LC solvent A is 0.1% (vol/vol) formic acid in HPLC-grade water. Store at room temperature in the dark for up to 1 year.

Desalting buffer B Desalting buffer B is 80% acetonitrile in HPLC-grade water. Store at room temperature in the dark for up to 1 year.

LC solvent B LC solvent B is 0.1% (vol/vol) formic acid in HPLC-grade acetonitrile. Store at room temperature in the dark for up to 1 year.

EQUIPMENT SETUP

Liquid chromatography Use the following LC gradient at a flow rate of 0.3 ml/min for discovery-driven shotgun proteomics (Step 20A).

Time interval (min)	% Solvent A	% Solvent B
0	95	5
100	75	25
120	60	40
125	10	90
130	10	90

Use the following LC gradient at a flow rate of 0.35 ml/min for targeted proteomic analyses by SRM (Step 20B).

Time interval (min)	% Solvent A	% Solvent B
0	95	5
30	65	35
31	20	90
33	20	90
34	95	5
45	95	5

Mass-spectrometry analysis Settings for shotgun proteomics on a Q Exactive Plus MS instrument are described in the following table.

Method parameter	Value
Polarity	Positive
Full MS	
Microscans	1
Resolution	70,000 at 200 *m/z*
AGC target	3e6
Maximum injection time	64 ms
Scan range	350–1,500 *m/z*
dd-MS^2	
Microscans	1
Resolution	17.500 at 200 *m/z*
AGC target	1e5
Maximum ion time	55 ms
Loop count	1
Isolation window	1.4 *m/z*
Isolation offset	0 *m/z*
Fixed first mass	—
Normalized collision energy	25
dd settings	
Underfill ratio	2
Charge exclusion	Unassigned, 1
Peptide match	Preferred
Exclude isotopes	On
Dynamic exclusion	30 s

Settings for targeted SRM analysis on a QTRAP 5500 MS instrument are described in the following table.

Method parameter	Value
Scan type	MRM
Resolution Q1 and Q3	Unit
Pause between mass ranges	1 ms
CE, precursor charge 2+	CE = 0.044 × (Q1 m/z) + 5.5
CE, precursor charge >2+	CE = 0.051 × (Q1 m/z) + 0.5
Accumulation time	10–20 ms[a]
Declustering potential	80 V
Entrance potential	13 V
Collision cell-exit potential	30 V

CE, collision energy; MRM, multiple reaction monitoring; Q, quadrupole.

[a]Total scan time, including pauses, should not be >2.5 s. Adjust the number of targets and/or accumulation time accordingly.

Computer software/hardware PERL scripts for semiautomated mapping of candidate conformotypic peptide sequences on protein 3D structures are available at https://github.com/kabdullah/LiP-MS_demo. To run the scripts, standard computer hardware with a terminal and PERL v5 is required (type on the terminal: `perl --version`). Windows machines require the latest Windows 10 Anniversary Update and the subsidiary Bash on Ubuntu installation (https://msdn.microsoft.com/commandline/wsl/install_guide). Alternatively, Windows users can install the Unix-like environment Cygwin (https://www.cygwin.com) to execute the scripts.

For the visualization of mapping results, the open-source molecular viewer PyMOL is recommended, which requires Python v2.7 (https://www.python.org/downloads/). A link to the source code of PyMOL is available at http://pymol.org/download. Mac OS X or Linux users should use the package management system MacPorts (https://www.macports.org/) or RPM (http://rpm.org), respectively, for an easy install. Ready-to-run binaries for Windows users are available free of charge for students and teachers at http://pymol.org.

PROCEDURE

Protein extraction ● TIMING 2 h plus cell culture time

1| Proteins for LiP-MS analysis can be extracted from different sources. For protein extraction from Baker's yeast (*S. cerevisiae*), follow option A; for bacteria (*E. coli*), follow option B; for mammalian cells, follow option C; and for biological fluids, follow option D.

(A) Protein extraction from baker's yeast (*S. cerevisiae*)

(i) Grow 50 ml of yeast culture in SD complete medium at 30 °C to OD_{600} ~0.7. Harvest the culture by centrifugation at 3,000*g* for 5 min at 4 °C.
▲ CRITICAL STEP We recommend performing the experiment in triplicate.
? TROUBLESHOOTING

(ii) Transfer the pellet with a small scoop or spatula to a precooled 2-ml screw-cap tube. Make sure that the tube is filled to <20%; if necessary, split the pellet among several tubes. Add 1.5 volumes of acid-washed glass beads.

(iii) Add 500 µl of native lysis buffer (see Reagent Setup).

(iv) Lyse the cells by bead-beating at 4 °C using a cell disruption device at 6 m/s for 30 s at 4 °C. Repeat bead-beating three times with 4-min breaks, resting the sample on ice for 2 min to prevent overheating of the samples.

(v) Centrifuge the sample for 10 min at 16,000*g* at 4 °C to remove cell debris. Transfer the supernatant to a new tube and repeat this centrifugation step. Transfer the supernatant to a new tube.

(vi) Measure the protein concentration using a commercial BCA assay according to the manufacturer's protocol, and dilute the protein extract to 1 mg/ml with native lysis buffer.
■ PAUSE POINT We recommend continuing with the protocol immediately. Alternatively, aliquots can be snap-frozen and stored at –80 °C for at least 6 months.

Box 3 | Singly charged precursor ions

In classic shotgun proteomic analyses based on data-dependent acquisition, fragmentation of singly charged precursors is typically excluded, as such ions often derive from environmental or workflow contaminants[59]. In LiP experiments, short peptides with nontryptic termini may be generated more frequently than in standard proteomic experiments. These peptides are also highly informative, as their detection enables identification of the specific LiP sites. Peptides with a nontryptic C terminus lack a basic arginine or lysine, and may be detected as singly charged peptide precursor ions. Thus, to maximize the information content of a LiP data set, we recommend enabling fragmentation of singly charged precursors during data-dependent acquisition[60].

On the basis of our experience, the number of LiP peptide identifications can be maximized by allowing fragmentation of charge states +1 to +4 within the same MS analysis. This also ensures a higher reliability of false-discovery-rate estimation based on the target-decoy search, as compared with pooling the results from multiple MS runs, in which precursor ions of different charge states are fragmented. For example, from a yeast proteome extract subjected to LiP, ~1% more unique peptide spectral matches can be expected when fragmentation of 1+ precursors is allowed; for about a third of these, the associated peptides are exclusively identified from +1 precursors. On the basis of our experience, identification of multiply charged precursor ions is not significantly affected when including fragmentation of singly charged ions, using a QExactive Plus instrument with a maximum cycle time of 20 × 64 ms (and an effective duty cycle including an MS1 scan with a top-20 method of ~1.8 s) and using a 2-h gradient. This suggests that the analytical setup used allows for this adaptation without compromising on identifications from multiply charged precursors (data not shown).

(B) Protein extraction from bacteria (*E. coli*)

(i) Grow 50 ml of bacterial culture in M9 medium at 37 °C, e.g., to OD_{600} ~0.8. Harvest the culture by centrifugation at 3,000*g* for 5 min at 4 °C.

▲ **CRITICAL STEP** We recommend performing the experiment in triplicate.

(ii) Continue with Step 1A(ii–vi) as described above.

■ **PAUSE POINT** We recommend continuing with the protocol immediately. Alternatively, aliquots can be snap-frozen and stored at –80 °C for at least 6 months.

(C) Protein extraction from mammalian cell lines

(i) Begin with 1 × 10^7 HeLa cells.

▲ **CRITICAL STEP** We recommend performing the experiment in triplicate.

(ii) Wash the cells twice with 0.5 ml of 10 mM (or 1×) PBS.

(iii) Resuspend the cells in as little 10 mM PBS as possible (e.g., 500 µl) and transfer the cell suspension to a cooled dounce homogenizer.

(iv) Lyse the cells in PBS with 15 cycles of douncing on ice.

(v) Pass the homogenate through a 26-gauge syringe needle.

(vi) Centrifuge the sample at 1,000*g* for 5 min at 4 °C to remove the cell debris, and transfer the supernatant to a fresh tube.

(vii) Measure the protein concentration using a commercial BCA assay according to the manufacturer's protocol, and dilute the protein extract to 1 mg/ml with 10 mM (or 1×) PBS.

■ **PAUSE POINT** We recommend continuing with the protocol immediately. Alternatively, aliquots can be snap-frozen and stored at –80 °C for at least 6 months.

(D) Protein extraction from biological fluids

(i) Begin with 30–50 µl of biological fluid.

(ii) (Optional) Deplete the biological fluids of abundant proteins (e.g., albumin and immunoglobulins from blood plasma) using commercial depletion kits (e.g., Top 2 Abundant Protein Depletion Spin Columns (Thermo Scientific, cat. no. 85161) or Top 12 Abundant Protein Depletion Spin Columns (Thermo Scientific, cat. no. 85164).

(iii) Measure the protein concentration using a commercial BCA assay. Add 10% (vol/vol) 10× native lysis buffer (see Reagent Setup). Do not dilute it to less than 1 mg/ml of total protein.

■ **PAUSE POINT** We recommend continuing with the protocol immediately. Alternatively, aliquots can be snap-frozen and stored at –80 °C for at least 6 months.

LiP with PK ● TIMING 25 min

▲ **CRITICAL** Perform LiP with PK on only half of the sample. The remaining sample will be digested with trypsin only (see Step 9 and below) and will serve as an internal control. This control will be used to correct for changes in protein abundance across the tested conditions and (optionally) to detect altered activity of endogenous proteases and possible differential PTMs.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

2| Take 100 µg of protein extract per sample.

3| Optionally, spike a peptide or protein internal standard the user is familiar with, for which differential proteolytic patterns are known and ideally which does not occur naturally in the biological extract, into the lysate and mix well. Examples of internal standards are presented in Feng *et al.*[2]. Myoglobin is commercially available; to mimic a medium- to high-abundant protein, it can be spiked in at 1–5 pmol per µg.

4| Place the sample into a heat-block/water bath to bring the temperature to 25 °C.
▲ **CRITICAL STEP** Equilibration time will depend on the sample volume and the original sample temperature. It is important that the sample reach 25 °C before proceeding to the next step.

5| Add PK to the protein extract at a 1:100 enzyme/substrate ratio (wt/wt).

6| Incubate the sample for exactly 1 min at 25 °C.
▲ **CRITICAL STEP** Exact timing of the incubation step is crucial. To increase reproducibility, we strongly advise that the proteolytic reaction be precisely timed.

7| Transfer the sample rapidly to boiling water for 5 min to inactivate the PK.
▲ **CRITICAL STEP** Make sure that the temperature of the water bath is >95 °C to ensure complete inactivation of PK. Consistent heat throughout the vessel is important for proper quenching. We recommend using a device such as the one shown in **Supplementary Figure 3** to hold the microcentrifuge tubes. Make sure that the water level is sufficient to submerge the microcentrifuge tubes above the sample level, but avoid floating of the tubes. Note that the water level in the water bath will decrease over prolonged use with serial quenching of multiple samples or sample batches.

8| Remove the tubes from the water bath and let them cool at room temperature for 5 min. Add DOC to a final concentration of 5% (wt/vol), using the 10% (wt/vol) stock DOC solution.

Sequential LysC–trypsin digestion ● TIMING 6-h handling time, followed by overnight digestion

9| Reduce the cysteine residues by adding DTT to a final concentration of 12 mM, then incubate the tubes for 30 min at 37 °C.

10| Alkylate the reduced cysteine residues by adding IAA to a final concentration of 40 mM. Incubate the tubes for 45 min at room temperature in the dark (IAA is light-sensitive).

11| Ensure that the pH is 7.5–8.5 by placing a drop of the sample on a pH indicator paper. If the pH is too low, adjust the pH to ~8 by adding 100 mM Ambic.

12| Add LysC to a 1:100 enzyme/substrate ratio (wt/wt). Incubate the tubes at 37 °C for 4 h with agitation at 800 rpm.

13| Dilute the sample to 1% (wt/vol) DOC by adding 4 volumes of 100 mM Ambic.

14| Add trypsin to a 1:100 enzyme/substrate ratio (wt/wt). Incubate the tubes at 37 °C overnight under agitation at 800 r.p.m.

15| Stop the digestion and precipitate the DOC by adding 98% (vol/vol) formic acid to a final concentration of 2% (vol/vol). Verify that the final pH of the sample is <3.
▲ **CRITICAL STEP** Should the pH not be at a pH <3, that would point to an error in preparation and a new sample should be freshly prepared.

16| DOC will form a white precipitate upon acidification. To remove the precipitate, centrifuge the sample at 16,000*g* for 10 min at room temperature, carefully transfer the clear liquid phase to a new microcentrifuge tube, and repeat this centrifugation step.
▲ **CRITICAL STEP** When transferring the supernatant, do not disturb the DOC pellet, as DOC may interfere with the MS measurement.
■ **PAUSE POINT** Samples can be stored at –20 °C for at least 6 months.

Peptide cleanup ● TIMING 2–3 h

17| Use a Sep-Pak column packed with 50 mg of C18 resin to desalt the samples. Load, wash, and elute the peptide mixtures, as recommended by the manufacturer using desalting buffers A (for washing) and B (for elution), described in the Reagent Setup. The eluting volume can be 2× 400 µl, for example.

18| Evaporate the solvent from the eluted peptides using a vacuum centrifuge at 45 °C.

19| Resuspend the peptides in 0.1% (vol/vol) formic acid to a final concentration of ~1 mg/ml. Estimate the volume of 0.1% formic acid to add based on the protein concentration determined by the BCA assay in Step 1A(vi) (and the corresponding steps in options B–D), assuming no loss of proteins or peptides during sample preparation. Transfer the sample to an MS vial for subsequent MS analysis.

■ **PAUSE POINT** Samples can be stored at –20 °C for at least 6 months.

Data acquisition and analysis ● TIMING 1–4 h of measurement time per sample plus 1.5 d of data analysis

20| For data acquisition using discovery-driven shotgun proteomics, use option A; for targeted proteomic analyses by SRM, follow option B. The shotgun approach of option A couples LiP to discovery-driven MS techniques, which can be applied to a whole proteome and to unbiasedly identify proteins changing structural features upon different perturbations. Alternatively, the targeted analysis in option B investigates structural changes of one or a few specific protein(s) of interest. The targeted approach measures fewer peptides and outcompetes discovery-driven methods in terms of reproducibility and specificity. The two workflows can be applied exclusively or sequentially.

(A) Discovery-driven shotgun proteomics

(i) Perform shotgun proteomic measurements, as described elsewhere[37,38]. Perform peptide separation with a 2- to 4-h water/acetonitrile gradient, as described in the Equipment Setup. Typically, 1–3 µl of a peptide sample at a concentration of 1 mg/ml is injected.

(ii) Perform a database search of the spectra with search engine tools, as described in Moulder *et al.*[37] using the following settings: precursor and fragment mass tolerance 10 p.p.m., cleavage-specificity trypsin (semispecific), maximally two missed cleavages, carbamidomethylation on cysteines as fixed modification, and oxidation of methionines as variable modification.

(iii) Import the MS raw data files into software designed for relative quantification of label-free MS data, such as Progenesis (Nonlinear Dynamics; http://www.nonlinear.com/progenesis/qi-for-proteomics/download/), OpenMS (https://www.openms.de/downloads/), MaxQuant[31,37], or similar[33]. Perform alignment of peptide ion maps and peak quantification from each sample, as described in the user manuals.

(iv) Perform two separate analyses. (i) Analyze control samples treated with trypsin only as in a standard quantitative proteomics pipeline to obtain protein abundance changes. (ii) Analyze LiP samples by treating every peptide as an independent entity to obtain peptide-level abundance changes.

(v) Perform statistical testing to identify significantly changed peptides and proteins with the R package SafeQuant. This enables calculation of median abundance changes and associated *P* values corrected for multiple testing (*q* values). Abundance changes from Progenesis (or MaxQuant or alternative) can be imported into SafeQuant.

(vi) Filter the protein abundance changes using suitable cutoffs. We recommend using a log2 abundance change cutoff of twofold and *q* values <0.01.

(vii) Use significant protein abundance changes as normalization factors for peptide-level LiP data. This is achieved by dividing peptide-level abundance changes across samples by the (significant) abundance change of the respective protein. For proteins that do not significantly change abundance, use a normalization factor of 1 (i.e., no correction).

(viii) Filter normalized peptide-level abundance changes from LiP data using suitable cutoffs. We recommend using a log2 abundance change cutoff of twofold and *q* values <0.01.

(ix) Plot the results in the form of a volcano plot, representing peptide abundance changes versus the associated *q* values. Proteins associated with peptides that significantly change abundance in the LiP samples are deemed structurally variant. These peptides identify the specific protein region undergoing the structural change.

? TROUBLESHOOTING

(B) Targeted proteomic analyses by SRM

(i) Select the proteins of interest for which a differential structural analysis is to be performed.

(ii) Identify by *in silico* prediction all possible unique tryptic peptides for each target protein. Develop and validate SRM assays for each peptide according to the guidelines described by Feng and Picotti[34].

(iii) Perform SRM measurements as previously described[21,34,39]. Separate peptides using a 30-min water/acetonitrile gradient as described in the Equipment Setup. Typically, 2–3 µl of a peptide sample at a concentration of 1 mg/ml is injected.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(iv) Import the raw SRM files into the interface of the Skyline software[35] and perform relative quantification of peptides and proteins as described in the user manual. Follow the quantification and normalization procedure described in Step 20A(vii) for shotgun proteomics data, referring to the Skyline user manual or to Feng and Picotti[34] for Skyline-specific steps.

(v) Performing statistical testing to identify significantly changed peptides and proteins with the R package MSStats[36] enables calculation of median abundance changes and associated *P* values corrected for multiple testing (*q* values). Abundance changes from Skyline can be imported into MSStats.

(vi) Plot the results in the form of a volcano plot representing peptide abundance changes versus associated *q* values. Fully tryptic peptides that significantly change in abundance in the LiP samples identify proteins that undergo structural transitions across the selected conditions and the associated protein regions. Decreased abundance of a tryptic peptide in a LiP sample indicates a structural change resulting in increased accessibility to the LiP protease and thus increased exposure and/or flexibility of the associated protein region. Conversely, increased abundance of tryptic peptides is suggestive of decreased accessibility to proteolysis of the associated structural segments.

(vii) Confirm the results from fully tryptic peptides by quantifying half-tryptic peptides resulting from LiP cleavage that map to the same protein region. If no information on the cleavage site is available via previous unbiased analyses, half-tryptic peptides can be identified by a follow-up SRM measurement. To this aim, all possible N- and C-terminal peptide fragments resulting from internal cleavage of significantly changed fully-tryptic peptides are identified *in silico*. SRM assays for these peptides are developed and performed on the same set of samples. Quantification is performed as described in Step 20A(viii), yielding half-tryptic peptides from the target protein region that significantly changes abundance. Theoretically, fully tryptic and half-tryptic peptides should display inverse abundance change directions, and this consideration can be used as a confirmatory criterion to support the structural conclusions. However, this is not always the case because of secondary cleavages of half-tryptic peptides and other unanticipated biochemical events. Thus, a significant abundance change of half-tryptic peptides, independent of the direction, should be considered confirmatory of the detected structural change for the target protein region.

(viii) If shotgun proteomics data sets are available, use information contained in the MS spectra from shotgun data to guide the identification of suitable SRM transitions, as previously described[34]. If the targeted SRM experiments follow a shotgun analysis of the same LiP samples, SRM assays for both fully tryptic peptides and half-tryptic peptides detected in the unbiased analysis can be directly developed and simultaneously measured.

? TROUBLESHOOTING

Analysis of conformotypic peptides ● TIMING 1 h

21| Protein regions that change their structural properties identified based on analyses described here can be visualized along the sequence of the protein and mapped onto its 3D structure, if available, to evaluate the functional and structural relevance of the detected changes. Automated analysis and visualization of LiP data with respect to available structures can be performed (option A) with a master BASH script and two subsidiary PERL scripts that are available at https://github.com/kabdullah/LiP-MS_demo. In the following, we will use a small test data set that is provided along with the scripts in LiP-MS_demo, to guide the reader through the necessary steps to map peptides from a LiP experiment conducted on the 3D structure of the protein myoglobin (**Fig. 3**). Alternatively, manual conformotypic peptide mapping can be performed using option B.

The master BASH script run_demo.sh executes in a serial manner the PERL scripts getPeptidePos.pl and mapPeptideOnPDB.pl.

(A) Automated mapping of conformotypic peptides to protein structures

(i) Download the LiP-MS_demo package from https://github.com/kabdullah/LiP-MS_demo/archive/master.zip.

(ii) Open the command line terminal.

(iii) Assuming that the demo package has been saved on the Desktop, change to the Desktop directory by typing the command: `$>cd ~/Desktop`

(iv) Unzip the demo folder using the operating system's archive manager or the unzip command on the terminal:
`$>unzip LiP-MS_demo-master.zip`

(v) Change to demo directory and run the master BASH script: `$>cd LiP-MS_demo-master` followed by
`$> ./run_demo.sh`

(vi) Reply with `yes` or `y` if you want to execute the scripts on the demo input files. Alternatively, reply with `no` or `n` and rerun the master script with the following explicit parameters:
`$> ./run_demo.sh input/myoglobin_lip_peptides.tsv input/horse_proteome.fasta output/ /opt/local/bin/pymol`

where `input/myoglobin_lip_peptides.tsv` is a simple tab-delimited text file listing LiP peptides in the first column, identified from a LiP-MS analysis of the protein myoglobin (**Fig. 3**). The `input/horse_proteome.fasta` file is a FASTA file of the proteome of interest from the UniProt database (see the README.md file in the demo

directory for the URL). `output` specifies the directory name in which all intermediate and final results will be stored. `/opt/local/bin/pymol` is the path to a local PyMOL installation, which you might need to adjust according to your own PyMOL installation.

Alternatively, run the PERL scripts `getPeptidePos.pl` and `mapPeptideOnPDB.pl` individually without the master script:

```
$>perl scripts/getPeptidePos.pl -in input/myoglobin_lip_peptides.tsv -fasta
input/horse_proteome.fasta -col 2 -header -v >output/myoglobin_lip_peptides_pos.tsv
$>perl scripts/mapPeptideOnPDB.pl -in output/myoglobin_lip_peptides_pos.tsv -
uniprotCol 2 -pepCol 1 -maxDownload 5 -v
```

The `getPeptidePos.pl` script finds the UniProt ID of the proteins from which the candidate conformotypic peptides originate. In addition, it outputs other useful information such as the start and end position of the peptides in the protein sequence, the number of occurrences of the peptide in the protein sequence and in the whole proteome, whether the peptide is fully tryptic, half-tryptic or nontryptic, and an extended peptide sequence. `-in` and `-fasta` options are described in the previous paragraph. The `-col` option represents the column number in which the peptide sequence is found in the input file, whereas the `-header` option skips the first line of the input file. `-v` outputs additional information such as a header for the output file.

The `mapPeptideOnPDB.pl` script maps candidate conformotypic peptides on protein structures. It downloads UniProt entry files for the identified proteins, selects PDB-IDs of high-resolution X-ray structures, or, if nonexistent, downloads MODBASE homology models and aligns the UniProt sequence with all PDB sequences to identify the location of the peptide sequences on the PDB structures. `-in` is the output file of `getPeptidePos.pl`, but can also be any other tab-delimited file with the sequence of the candidate conformotypic peptides and their associated UniProt ID. `-uniprotCol` is the column number of the UniProt ID and `-pepCol` is the column number of the peptide sequence. With `-maxDownload`, the user can specify the maximum number of PDB files for the mapping, which cuts computation time for proteins with many PDB structures, such as horse myoglobin with >80 PDB structures. Please also refer to the `README.md` file at https://github.com/kabdullah/LiP-MS_demo for additional information. To get a list of all command-line options for further customizing the calculations, type:

```
$>perl scripts/getPeptidePos.pl -help
```

or

```
$>perl scripts/mapPeptideOnPDB.pl -help
```

(vii) The mapping script `mapPeptideOnPDB.pl` will output PyMOL.pml scripts into the subdirectory pml. Load a pml script into PyMOL by starting PyMOL on the command line and appending the path to one of the pml scripts: `$>pymol output/pml/MYG_HORSE_2frfA.pml`. The pml scripts show one of the protein structures in a cartoon representation in a gray color with the candidate conformotypic peptide sequences highlighted in yellow and half-tryptic peptide ends emphasized in red. In addition, small molecules are displayed in a ball-and-stick representation with multiple colors.

Alternatively, the script can be loaded into a running PyMOL viewer using the `@` prefix in the PyMOL command bar: `@ output/pml/MYG_HORSE_2frfA.pml`.

(B) Manual mapping of conformotypic peptides to protein structures

(i) Check whether an experimentally determined structure is available for the protein of interest by searching its UniProt Accession number or entry name in the Protein Data Bank (PDB, http://www.rcsb.org/pdb).

? TROUBLESHOOTING

(ii) Download the biologically relevant conformation from PDB as a .pdb1 file.

▲ **CRITICAL STEP** Many proteins are represented by multiple structures in the PDB that differ in sequence coverage, mutations, proteins they are fused to, molecular-binding partners, or applied experimental conditions. This diversity should be kept in mind when selecting the structures for this mapping. Note that the residue numbering in the PDB protein structure might not correspond to the amino acid numbering in the protein sequence.

▲ **CRITICAL STEP** In the case of X-ray crystallography structures, the standard asymmetric unit should be avoided. The asymmetric unit corresponds to the smallest part of a protein crystal that the entire structure can be built from. It is thus ill-suited for protein-biology-based studies. The coordinates for the biological conformation can be found in the 'Download Files' drop-down menu of a PDB entry or in the data/biounit/PDB directory of PDB FTP server.

(iii) Load all available structures downloaded from PDB into a molecular visualization software. For this protocol, PyMOL is used (http://www.pymol.org).

(iv) Choose the most suitable structure and align all the remaining structures to it using the '(A)ction' > 'align' > 'all to this' option in the object list on the right half of the viewer panel.

(v) To display the sequences of the protein structures, type the following command into the PyMOL command line: `set seq_view, 1`.

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

(vi) Mark candidate conformotypic peptide sequences in the PDB sequences using the mouse cursor. Select multiple amino acids by clicking and dragging the mouse cursor over a sequence region.
(vii) The selected candidate conformotypic peptide sequence should appear as a selection object on the right half of the viewer panel named '(sele)'. Give the selected peptide a unique name by renaming the selection object using the '(A)ction' > 'rename selection' option. All candidate conformotypic peptides should have the same name and only differ in a counter number (e.g., conformPeptide_1, conformPeptide_2, and so on).
(viii) Repeat Steps 20B(vi) and (vii) for all candidate conformotypic peptides.
(ix) Highlight all candidate conformotypic peptides with a unique color using the PyMOL command color (e.g., `color yellow, conformPeptide_*`).

Assessment of functional relevance of conformotypic peptides ● TIMING 30 min

22| Binding of small molecules to proteins can be probed in a LiP experiment. The presence of a small molecule in the vicinity of a candidate conformotypic peptide within the 3D structure can be one indication that the LiP pattern was produced because of the binding of the small molecule. Select small molecules and metal ions in PyMOL and name the selection 'ligands' with the `select` command: `select hetatm, ligands`.

23| Highlight all small molecules and metal ions with an atomic sphere representation with the `show` command: `show spheres, ligands`.

24| To investigate the effect of protein–protein interactions in a LiP experiment, check whether candidate conformotypic peptides are adjacent to another protein structure in the 3D structure. Protein–protein interactions can impact the LiP pattern. Highlight different proteins in PyMOL by coloring each of them with a different color using the `util.cbc` (cbc = color by chain) command: `util.cbc`.

25| In less investigated proteins, missing coordinates or alignment gaps can occur in the protein structure. These can be indicative of flexible protein regions. To search for chain breaks in the protein structure or gaps in an alignment of PDB sequence—and FASTA protein sequence—retrieve the UniProt–PDB sequence alignment from the PDBsum webserver via the link:
http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/pdbsum/GetPage.pl?template=align.html&l=1&pdbcode=XXXX
where XXXX corresponds to the PDB accession code of your structure (e.g., 1uuh). Alternatively, visit PDBsum via the URL http://www.ebi.ac.uk/pdbsum, and search for your protein by its PDB accession ID. On the 'Top page' tab of your structure, click on the secondary structure representation of the protein sequence in the 'Protein Chain' information boxes to retrieve the UniProt–PDB sequence alignment. Look for dash symbols in the PDB sequence to identify gaps in the PDB sequence.

26| To support the possibility that gaps in the PDB sequence are due to disordered protein segments, run a protein disorder prediction web service tool such as DISOPRED3 (http://bioinf.cs.ucl.ac.uk/psipred/?disopred=1) or IUPred (http://iupred.enzim.hu) on your protein sequence.

27| Check whether candidate conformotypic peptides are located next to PTM sites. The proximity to PTMs might indicate conformational changes on the protein due to changes in the PTMs. Known PTMs can be retrieved from UniProt and mapped on a protein structure using the PyMOL command `uniprot_features`: `uniprot_features MYG_HORSE, 2frfA`.
? TROUBLESHOOTING

28| Highlight known PTMs using an atomic sphere representation using the PyMOL command `show`: `show spheres, feature_modified_residue`.

29| Create images of the conformotypic peptides and the protein structure. For high-quality images, ray-trace (computation of light reflection and shadows) your protein structure using the PyMOL command `ray`: `ray`.

30| Create a PNG image using the `png` command. For example:
`png MYG_HORSE_2frfA_conformotypicPeptides.png`.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1	Low yield of protein in the extract (as determined in the BCA assay)	Incomplete extraction (Step 1)	Adjust the disruption method. For example, 1 liter of *S. cerevisiae* cells (BY4741 strain) grown in SD medium to an OD_{600} value of 0.4 should yield ~0.35 g of pellet
20	Too many or too few LiP cleavages in the proteome, fully tryptic peptides are low in abundance or LiP digestion is irreproducible	Too aggressive or too weak LiP digestion, incomplete quenching of the LiP protease (Steps 2–7)	Adjust the E/S ratio and/or protease incubation time; check that quenching of the LiP protease is efficient (the temperature of the water bath may have been too low). As a reference, the total number of peptides identified in the sample subjected to LiP should not decrease more than 25% compared to the sample subjected only to trypsin digestion, and half-tryptic peptides should be up to 40% of the identified peptides
	Detection of polymers or background noise is too high during LC–MS analysis	Buffer, salt, detergent, or plasticware interference (Steps 1–19)	Use only MS-compatible plasticware. Transfer highly concentrated acids with and into acid-stable materials. Make sure that no leftovers from dishwashing or autoclaving processes contaminate the sample. Omit non-MS-compatible detergents or remove them before the MS-step, if suitable removal protocols exist. If polymeric materials are used during sample preparation (e.g. beads), make sure that they are MS-compatible
	Low yield of identified peptides or large amount of half-tryptic peptides in the control sample	Endogenous proteases (Step 1)	Extract proteins quickly and at 4 °C to avoid degradation. Certain protease inhibitors can be added at the concentration recommended by the vendor and will not significantly affect the activity of PK added in a 1:100 E/S ratio. For mammalian cells, we recommend using one tablet of Roche cOmplete Protease Inhibitor Cocktail (EDTA-free) for 50 ml of lysis buffer. For yeast, we recommend using 1 ml of Sigma-Aldrich Protease Inhibitor Cocktail for use with fungal and yeast extracts for 100 ml of lysis buffer. If other inhibitors are used, we suggest performing a test for the activity of PK in the presence of the chosen inhibitors
	No (or too many) significant changes found	Experimental design	Depending on experimental design and expected fold changes, the number of required replicates must be statistically determined
21	No protein structure available	Experimental determination of protein structure has failed or has not been attempted (Steps 21)	Download homology models from SWISS-MODEL (http://www.swissmodel.expasy.org) or MODBASE (http://modbase.compbio.ucsf.edu). Choose the model with the highest sequence identity to that of the target protein and the highest sequence coverage. Only homology models with sequence identity >50% should be considered reliable. Models with <50% sequence identity should be used with caution. Models with <30% sequence identity should be discarded due to low quality and errors in their structural coordinates
27	Mapping of UniProt features is shifted on the protein structure	The PDB residue numbering does not correspond to the UniProt residue numbering (Step 27)	Calculate the residue number difference from the UniProt–PDB sequence alignment from Step 25. Adjust the PDB residue number with the PyMOL command 'alter XXXX, resi=resv+YYY', where 'XXXX' is the object name of your target protein structure (Step 24) and 'YYY' is the residue number difference

● TIMING

Step 1, protein extraction: 2 h plus cell culture time, where applicable
Steps 2–8, LiP with PK: 25 min
Steps 9–16, sequential LysC–trypsin digestion: 6 h of handling time, followed by overnight digestion
Steps 17–19, peptide cleanup: 2–3 h
Step 20, data acquisition and analysis: 1–4 h of measurement time per sample, followed by 1.5 d of data analysis time
Step 21, analysis of conformotypic peptides: 1 h
Steps 22–30, assessment of the functional relevance of conformotypic peptides, 30 min

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

ANTICIPATED RESULTS

The pipeline results in the identification of proteins (from a proteome or a target protein list) that undergo structural alterations in response to the experimental condition of interest, as well as regions within a protein structure involved in the structural change. These changes can be visualized on a protein sequence and, if available, on the 3D structure of the protein (**Supplementary Fig. 4a**) to formulate hypotheses on their structural and biological implications. Analysis of control samples also results in the relative quantification of protein levels across the different conditions, as in a standard proteomic experiment. Protein abundance and structural changes can optionally be overlaid to highlight possible cases of post-translational regulation involving altered protein structures (**Supplementary Fig. 4b**). Structural changes detected by unbiased experiments can be validated and analyzed in more detail by the targeted analysis of all possible fully tryptic peptides for the protein and half-tryptic peptides suspected based on the location of LiP sites. The exact nature of the structural change can be addressed by targeted biochemical experiments. If available, proteome-wide measurements of protein phosphorylation or other PTMs and transcriptomics data can be integrated with LiP-MS data.

If the experiment involves application of small molecules to the biological extract and comparison of proteolytic patterns in the presence and absence of the compound, the approach yields candidate protein targets of the compound and potential binding sites. Identification of novel protein–small molecule interactions can be validated by *in vitro* experiments.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS ACKNOWLEDGMENTS

We thank O. Schubert (UCLA) for insightful discussions. We also thank T. Lehmann for building the device to hold the sample tubes in the water bath and G. de Franceschi for contributing to the setup of the original LiP-MS protocol. P.P. is supported by an EU FP7-ERC Starting Grant (FP7-ERC-StG-337965), a 'Foerderungsprofessur' grant from the Swiss National Science Foundation (grant PP00P3_133670), and Promedica Stiftung (grant 2-70669-11). Y.F. is supported by an ETH Research Grant (grant no. 4412-1); I.P. is supported by an EMBO long-term fellowship (EMBO ALTF2014); and A.K. acknowledges SystemsX.ch for funding.

AUTHOR CONTRIBUTIONS S.S., A.K. and P.P. wrote the paper. Y.F. developed and optimized the original version of the protocol. S.S., P.L., O.M. and I.P. contributed to protocol optimization. I.P. and P.L. generated figures. P.J.B. contributed to editing of the manuscript and setup of the MS pipeline. A.K. developed and optimized algorithms and tools for protein structural analyses and prepared the user guide. P.P. supervised the project.

COMPETING FINANCIAL INTERESTS The authors declare competing financial interests: details are available in the online version of the paper.

1. Walther, T.C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* **190**, 491–500 (2010).
2. Feng, Y. *et al.* Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).
3. Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825 (2017).
4. Stauffer, S. *et al.* Stepwise priming by acidic pH and a high K+ concentration is required for efficient uncoating of influenza A virus cores after penetration. *J. Virol.* **88**, 13029–13046 (2014).
5. Geiger, R. *et al.* L-arginine modulates T cell metabolism and enhances survival and anti-tumor activity. *Cell* **167**, 829–842 (2016).
6. Aceto, A. *et al.* Analysis by limited proteolysis of domain organization and GSH-site arrangement of bacterial glutathione transferase B1-1. *Int. J. Biochem. Cell Biol.* **27**, 1033–1041 (1995).
7. Dieckmann, R. *et al.* Probing the domain structure and ligand-induced conformational changes by limited proteolysis of tyrocidine synthetase 1. *J. Mol. Biol.* **288**, 129–140 (1999).
8. Nouwen, N. *et al.* Domain structure of secretin PulD revealed by limited proteolysis and electron microscopy. *EMBO J.* **19**, 2229–2236 (2000).
9. Pagnozzi, D. *et al.* Stoichiometry and topology of the complex of the endogenous ATP synthase inhibitor protein IF(1) with calmodulin. *Biochemistry* **49**, 7542–7552 (2010).
10. Peng, Z.Y. & Kim, P.S. A protein dissection study of a molten globule. *Biochemistry* **33**, 2136–2141 (1994).
11. Polverino de Laureto, P. *et al.* Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. *J. Mol. Biol.* **334**, 129–141 (2003).
12. Eftink, M.R. Fluorescence techniques for studying protein structure. *Methods Biochem. Anal.* **35**, 127–205 (1991).
13. Ilari, A. & Savino, C. Protein structure determination by x-ray crystallography. *Methods Mol. Biol.* **452**, 63–87 (2008).
14. Pelton, J.T. & McLean, L.R. Spectroscopic methods for analysis of protein secondary structure. *Anal. Biochem.* **277**, 167–176 (2000).
15. Konermann, L., Pan, J. & Liu, Y.H. Hydrogen exchange mass spectrometry for studying protein structure and dynamics. *Chem. Soc. Rev.* **40**, 1224–1234 (2011).
16. Heyduk, T. Measuring protein conformational changes by FRET/LRET. *Curr. Opin. Biotechnol.* **13**, 292–296 (2002).
17. Sakakibara, D. *et al.* Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **458**, 102–105 (2009).
18. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
19. Roeser, J. *et al.* Oxidative protein labeling in mass-spectrometry-based proteomics. *Anal. Bioanal. Chem.* **397**, 3441–3455 (2010).
20. Mendoza, V.L. & Vachet, R.W. Probing protein structure by amino acid-specific covalent labeling and mass spectrometry. *Mass Spectrom. Rev.* **28**, 785–815 (2009).
21. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nat. Methods* **9**, 555–566 (2012).
22. Liu, F. *et al.* Proteome-wide profiling of protein assemblies by cross-linking mass spectrometry. *Nat. Methods* **12**, 1179–1184 (2015).
23. Kosinski, J. *et al.* Molecular architecture of the inner ring scaffold of the human nuclear pore complex. *Science* **352**, 363–365 (2016).
24. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nat. Protoc.* **9**, 120–137 (2014).
25. Savitski, M.M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).
26. Fontana, A. *et al.* Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **51**, 299–321 (2004).
27. Monti, M. & Pucci, P. Limited proteolysis mass spectrometry of protein complexes. *Mass Spectrometry of Protein Interactions* (ed. Downard, K.) Ch. 4 (Wiley, 2007).
28. Polverino de Laureto, P. *et al.* Probing the structure of human growth hormone by limited proteolysis. *Int. J. Pept. Protein Res.* **45**, 200–208 (1995).
29. Picotti, P. *et al.* Modulation of the structural integrity of helix F in apomyoglobin by single amino acid replacements. *Protein Sci.* **13**, 1572–1585 (2004).
30. Glatter, T. *et al.* Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *J. Proteome Res.* **11**, 5145–5156 (2012).

© 2017 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

31. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 (2008).
32. Rost, H.L. *et al.* OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **13**, 741–748 (2016).
33. Perez-Riverol, Y. *et al.* Open source libraries and frameworks for mass spectrometry based proteomics: a developer's perspective. *Biochim. Biophys. Acta* **1844**, 63–76 (2014).
34. Feng, Y. & Picotti, P. Selected reaction monitoring to measure proteins of interest in complex samples: a practical guide. *Methods Mol. Biol.* **1394**, 43–56 (2016).
35. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
36. Choi, M. *et al.* MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526 (2014).
37. Moulder, R., Goo, Y.A. & Goodlett, D.R. Label-free quantitation for clinical proteomics. *Methods Mol. Biol.* **1410**, 65–76 (2016).
38. Richards, A.L. *et al.* One-hour proteome analysis in yeast. *Nat. Protoc.* **10**, 701–714 (2015).
39. Surinova, S. *et al.* Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. *Nat. Protoc.* **8**, 1602–1619 (2013).
40. Wyss, M. *et al.* Limited proteolysis of creatine kinase. Implications for three-dimensional structure and for conformational substrates. *Biochemistry* **32**, 10727–10735 (1993).
41. Vita, C., Dalzoppo, D. & Fontana, A. Limited proteolysis of thermolysin by subtilisin: isolation and characterization of a partially active enzyme derivative. *Biochemistry* **24**, 1798–1806 (1985).
42. Martini, F. *et al.* Investigation of intra-domain and inter-domain interactions of glutathione transferase P1-1 by limited chymotryptic cleavage. *Eur. J. Biochem.* **218**, 845–851 (1993).
43. Hatori, Y. *et al.* Domain organization and movements in heavy metal ion pumps: papain digestion of CopA, a Cu+-transporting ATPase. *J. Biol. Chem.* **282**, 25213–25221 (2007).
44. Griese, M. *et al.* Limited proteolysis of surfactant protein D causes a loss of its calcium-dependent lectin functions. *Biochim. Biophys. Acta* **1638**, 157–163 (2003).
45. Scolaro, L.A., Mersich, S.E. & Damonte, E.B. A mouse attenuated mutant of Junin virus with an altered envelope glycoprotein. *Arch. Virol.* **111**, 257–362 (1990).
46. Kordyukova, L.V. *et al.* Influenza A virus M1 protein structure probed by in situ limited proteolysis with bromelain. *Protein Pept. Lett.* **15**, 922–930 (2008).
47. Polverino de Laureto, P. *et al.* Limited proteolysis of bovine alpha-lactalbumin: isolation and characterization of protein domains. *Protein Sci.* **8**, 2290–2303 (1999).
48. Kim, J. *et al.* Characterization of a unique IgG1 mAb CEX profile by limited Lys-C proteolysis/CEX separation coupled with mass spectrometry and structural analysis. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **878**, 1973–1981 (2010).
49. Laganowsky, A. *et al.* Mass spectrometry of intact membrane protein complexes. *Nat. Protoc.* **8**, 639–651 (2013).
50. Becher, I. *et al.* Chemoproteomics reveals time-dependent binding of histone deacetylase inhibitors to endogenous repressor complexes. *ACS Chem. Biol.* **9**, 1736–1746 (2014).
51. Cadene, M. & Chait, B.T. A robust, detergent-friendly method for mass spectrometric analysis of integral membrane proteins. *Anal. Chem.* **72**, 5655–5658 (2000).
52. Proc, J.L. *et al.* A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J. Proteome Res.* **9**, 5422–5437 (2010).
53. Stark, G.R., Stein, W.H. & Moore, S. Reactions of the cyanate present in aqueous urea with amino acids and proteins. *J. Biol. Chem.* **235**, 3177–3181 (1960).
54. Kang, S.C. *et al.* Guanidine hydrochloride extraction and detection of prion proteins in mouse and hamster prion diseases by ELISA. *J. Pathol.* **199**, 534–541 (2003).
55. Barron, R.M. *et al.* High titers of transmissible spongiform encephalopathy infectivity associated with extremely low levels of PrPSc *in vivo*. *J. Biol. Chem.* **282**, 35878–35886 (2007).
56. Meyer, R.K. *et al.* Detection of bovine spongiform encephalopathy-specific PrP(Sc) by treatment with heat and guanidine thiocyanate. *J. Virol.* **73**, 9386–9392 (1999).
57. Ren, D. *et al.* An improved trypsin digestion method minimizes digestion-induced modifications on proteins. *Anal. Biochem.* **392**, 12–21 (2009).
58. Rashid, F., Sharma, S. & Bano, B. Comparison of guanidine hydrochloride (GdnHCl) and urea denaturation on inactivation and unfolding of human placental cystatin (HPC). *Protein J.* **24**, 283–292 (2005).
59. Michalski, A., Cox, J. & Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10**, 1785–1793 (2011).
60. Biniossek, M.L. & Schilling, O. Enhanced identification of peptides lacking basic residues by LC-ESI-MS/MS analysis of singly charged peptides. *Proteomics* **12**, 1303–1309 (2012).