**Proteomes in 3D: *in situ* protein structural states as a readout for proteome functional alterations**

Valentina Cappelletti1#, Thomas Hauser1#, Ilaria Piazza1#, Monika Pepelnjak1, Paul Boersema1, Natalie de Souza1,2 and Paola Picotti1\*

1 Institute of Molecular Systems Biology, Department of Biology, ETH Zürich, Zürich, Switzerland.

2 Department of Quantitative Biomedicine, University of Zürich, Zürich, Switzerland.

# Equal contribution

\* Correspondence: picotti@imsb.biol.ethz.ch (P.P.)

**Abstract**

Many biological processes are regulated via molecular events, such as intermolecular interactions, post-translational modification and conformational changes that do not affect protein levels and that therefore escape detection in classical proteomic screens. Reasoning that these events locally or globally affect the structures of proteins, we propose here that a global readout of protein structure could detect various types of functional alterations simultaneously and *in situ*. We test this idea using limited proteolysis coupled to mass spectrometry (LiP-MS), which monitors structural changes in thousands of proteins within a complex native-like environment. In bacteria adapting to different nutrient sources and in yeast responding to acute heat or osmotic stress, we show that this global structural readout captures enzyme activity changes, allosteric regulation, phosphorylation, and protein complex formation, with sufficiently high resolution to identify alterations of single functional sites. Comparison with prior knowledge, including flux, phosphoproteomics and metabolomics data, shows that LiP-MS captures most known enzyme functional alterations and suggests novel metabolite-protein binding events. The structural readout dramatically increases the coverage of classical protein expression profiling in detecting functional changes and its structural resolution can guide the generation of mechanistic hypotheses, thus better linking holistic and reductionist approaches. In summary, we show that proteome structural data report on multiple functional molecular events, simultaneously and *in situ*, and is thus a powerful approach to detect protein alterations underlying cellular phenotypes.

**Introduction**

Cellular systems are dynamic and complex, and multiple molecular events take place simultaneously at different subcellular sites. Systems biological approaches can provide a holistic view of molecular changes within cells or tissues. At the protein level, quantitative mass spectrometry-based proteomics is routinely used to measure global protein expression profiles across different conditions (Aebersold and Mann, 2003). Sets of proteins that change in abundance can be assigned to functional classes and used to identify altered biological processes. This approach has been applied to identify pathways regulated upon specific cellular perturbations, nutrient shifts, and disease development, and to identify mechanisms of drug action and resistance (Boisvert et al., 2010; Costenoble et al., 2011; Ideker et al., 2001; Kolkman et al., 2006; Mertins et al., 2016; Ressa et al., 2018).

An intrinsic limitation of this approach is that many molecular events that result in protein functional changes do not involve changes in protein abundance. Proteins undergo functional changes upon post-translational modification (PTM) (Ardito et al., 2017); upon binding to other proteins, small molecules, lipids, or nucleic acids (Chubukov et al., 2014; Niphakis et al., 2015; Nussinov et al., 2013; Sahni et al., 2013); upon cleavage (Russell, 2014), or upon conformational changes induced by environmental changes (e.g. pH or temperature) (Damaghi et al., 2013; Robertson and Murphy, 1997). The regulation of many cellular processes, such as signaling cascades (Kolch, 2005; Shaul and Seger, 2007), relies solely on these types of events rather than on altered protein levels. Variants of the proteomics workflow such as phosphoproteomics (Batth et al., 2018; Humphrey et al., 2015), interactomics (Sowa et al., 2009; Wepf et al., 2009), or activity-based proteomics (Cravatt et al., 2008) can capture specific molecular events that affect protein function. Though such approaches are powerful, they typically report on only a single type of molecular mechanism. The high-throughput, simultaneous analysis of diverse regulatory events on a proteome-wide scale is not practically feasible.

We speculated that the global analysis of protein structures could serve as a new type of quantitative readout to capture the majority of events that alter the functional states of proteins. It is dogma that the structure of a protein is intimately linked to its function (Pauling et al., 1949; White and Anfinsen, 1959). Protein structures integrate different types of molecular cues that result in functional alterations: binding of other molecules, protein-protein interactions, PTMs, mutations, aggregation, and conformational alterations due to changes in the cellular matrix all result in local or global structural alterations of proteins. We hypothesized that by measuring altered structural states of proteins on a proteome-wide scale, we could detect protein functional changes of various types that occur simultaneously in the cell. We reasoned that this strategy would yield a more detailed and nuanced picture than that offered by the measurement of abundance changes alone.

We have previously developed a mass spectrometric approach, limited proteolysis-coupled mass spectrometry (LiP-MS), that monitors protein structural changes directly within complex biological extracts and on a proteome-wide scale (Feng et al., 2014). The approach relies on the application of a broad-specificity protease for a limited amount of time to a proteome extracted under native conditions. Mass spectrometry analysis yields detailed structure-specific proteolytic fingerprints for every detectable protein in the extract. Comparison of structural fingerprints from different proteomes can identify proteins that undergo a structural alteration and can pinpoint structurally altered regions.

LiP-MS and other structural proteomics approaches have so far been used in protein structural studies, for example to study protein unfolding, aggregation, ligand-induced structural changes, and protein-protein interaction (Huber et al., 2015; Leuenberger et al., 2017; Liu et al., 2018; Piazza et al., 2018; Savitski et al., 2014), or to provide spatial constraints for protein structure determination (Aebersold and Mann, 2016). Here we test the idea that LiP-MS, and a global structural proteomic readout in general, can monitor functional changes within the cell. We show that this structural approach captures multiple types of known molecular events, including enzyme activity changes, enzymatic substrate site occupancy, allosteric regulation, phosphorylation, and protein-protein interaction, with a resolution that pinpoints single functional sites, thereby driving the generation of new molecular hypotheses. Applying this approach to three well-studied systems, yeast acutely responding to heat or osmotic stress, and bacteria undergoing metabolic adaptation, we show that the structural approach both captured most known functional changes and suggested new molecular events. We compared the information provided by LiP-MS to that obtained from protein abundance measurements alone and showed that LiP-MS detects a greater number of altered biological processes. The structural approach captures partially overlapping but also complementary information to metabolomics, flux analyses and phosphoproteomics, within a single readout. Our data show that a global structural approach reports on many functional events simultaneously and within the complex cellular milieu and thus constitutes a powerful readout to detect molecular events that underlie physiological and pathological phenotypes.

**RESULTS**

**Protein structural changes during the yeast response to acute stress**

We used multiple experimental systems to test whether global protein structural data can detect functional alterations of proteins and protein networks (Figure 1a). As the first system, we chose to study cellular responses activated on short time scales in yeast. Rapid responses to environmental perturbations are typically independent of gene expression changes and should therefore be less amenable to protein abundance-based screens. We applied a short osmotic or heat stress to exponentially growing yeast cultures, extracted the proteomes under native conditions at time points chosen to precede protein expression changes, and applied the LiP-MS workflow, which monitors in parallel protein abundance changes and protein structural alterations across conditions. We analyzed the resulting peptide mixtures by data-independent acquisition followed by label-free quantification, and corrected data from LiP experiments for protein abundance changes to yield structure-specific proteolytic fingerprints for every detectable protein. We detected LiP-based protein structural fingerprints for more than 2700 proteins and monitored abundance changes for a similar number of proteins (Figure S1b; Table 1). Differential analysis (see Methods) comparing samples before and after each stimulus showed that only 1% or less of the detected proteins varied in abundance (Figure 1b). In contrast, a larger fraction of the detected proteomes underwent structural alterations upon heat shock and osmotic shock (25% and 11%, respectively).

Heat shock in yeast results in protein misfolding, activation of quality control mechanisms, a generalized inhibition of protein translation and the formation of stress granules (Verghese et al., 2012; Wallace et al., 2015). Osmotic stress in contrast activates a different series of molecular events. Plasma membrane osmosensors activate the HOG pathway, a phosphorylation cascade involving sequential MAP kinases (Figure S3a) that leads to activation of the kinase Hog1. During these events, phosphorylation of known regulatory sites by the upstream kinase results in allosteric alterations at catalytic sites and activation of the downstream kinase. Hog1 in turn induces fast cytosolic accumulation of the osmoprotectant glycerol, again via a combination of known phosphorylation (Pfk26, Gpd1 and Ypk1) and allosteric (binding of 2,6-bisphosphate to Pfk1) events that causes increased flux through the glycerol biosynthesis pathway (Brewster and Gustin, 2014; Hohmann, 2015) .

We first asked if the structural readout captured activation of these known processes. We performed a functional enrichment analysis of our LiP-MS hits to identify cellular processes and compartments affected by protein structural alterations in these two yeast stress responses (Figure 1c). For both perturbations, we found an enrichment of glycolytic and gluconeogenic pathways (Figure 1c) and of cytoplasmic stress granules and polysome cellular components (Figure S2). As expected, several categories related to protein translation and folding were enriched specifically in the heat stress dataset (Figure 1c; Table 2, Figure S2). Compatible with the known increase of glycerol production under osmotic stress, we detected enrichment of the “glycerol metabolic process” and the “NADPH regeneration” biological processes only after osmotic stress; both GO categories include glycerol biosynthetic enzymes. Since MAPK signaling is not a specific GO term, we did not capture it in this analysis. Functional enrichment analysis of protein abundance data from the two perturbations did not result in any significantly enriched GO term, likely due to the low number of proteins that change their abundance.

**Structural changes capture multiple regulatory events**

We next assessed whether the structural changes we detect can capture the specific molecular events known to occur in the yeast response to osmotic stress: activation of the kinase cascade, allosteric events and increased flux in the glycerol production pathway (Figure 2; Figure S3a) (Brewster and Gustin, 2014; Hohmann, 2015). We detected structural alterations for most proteins of the HOG1 and glycerol production pathways (Figure 2; Table 1; Table 3). Structurally altered enzymes included the Ste20 and Hog1 protein kinases of the MAPK-HOG1 signaling pathway, enzymes in the glycerol biosynthesis branch of the pathway (Gpd1 and Gpp1/2), and enzymes of both upper (Pfk1, Fba1, Tpi1, Tdh1/2/3) and lower glycolysis (Pgk1, Gpm1, Eno2, Cdc19, Pyk2) (Figure 2).

Interestingly, on average only two LiP peptides were altered per enzyme, suggesting that structural alterations were confined to specific protein regions. We mapped altered LiP peptides to available structures of the relevant proteins in complex with substrates or allosteric regulators, and assessed proximity of LiP peptides to known functional sites. Throughout this work, a peptide was assigned to a known functional site if the minimal distance between its atoms and those of known reactants or regulators was less than 6.4 Å, a threshold value we determined in previous work (Piazza et al., 2018).

Strikingly, most altered LiP peptides exactly reflected the specific molecular events known to occur for these proteins (Figure 2; Table 3). For MAP kinases, regulated LiP peptides mapped either to the protein region embedding the known activating phosphorylation site (Hog1) or to the allosterically regulated catalytic site (Ste20). For the MAP kinase target Gpd1, one LiP peptide mapped to the active site, a second was adjacent to the downregulated phosphosites (Lee et al., 2012), and a third LiP peptide mapped to the C-terminal domain, which undergoes an extensive conformational change upon substrate binding (Mydy et al., 2019). We also detected LiP peptides at sites bound by small molecule allosteric regulators: for the beta subunit of 6-phosphofructokinase Pfk2, one of the two altered LiP peptides mapped exactly to the allosteric binding site of fructose 2,6-bisphosphate (Banaszak et al., 2011; Dihazi et al., 2004; Strater et al., 2011). The increase of Pfk1 activity during osmotic stress increases the flux of upper glycolysis through the two downstream enzymes Fba1 and Tpi1, feeding the demand for intermediates in glycerol production. Consistently, we observed that each of the two regulated LiP peptides for Fba1 and Tpi1 map to their active sites, likely reporting on increased substrate site occupancy of these enzymes. Indeed we previously showed, in experiments with exogenously added metabolites, that LiP peptides at metabolite binding sites are increasingly regulated with increasing occupancy of binding sites (Piazza et al., 2018). Regulated LiP peptides of Gpp1, which generates glycerol, are also in close proximity to its active site (Figure 2).

Interestingly, most (79%) of the LiP peptides mapping to proteins of the glycolysis pathway were in the active sites of these enzymes, including all enzymes of lower glycolysis (Pgk1, Gpm1, Cdc19, Eno2 and Pyk2). An alteration in the occupancy of these sites is consistent with decreased flux through lower glycolysis upon acute osmotic stress, possibly as a result of most upper glycolytic flux being diverted to the glycerol biosynthetic branch during this response.

Taken together, our data show that LiP-MS captures multiple molecular events during the yeast response to acute osmotic stress with the resolution of single functional sites. With a single structural readout, this approach simultaneously reports on altered enzyme activity, altered site occupancy, allostery, and phosphorylation events in a given pathway.

Since several of the known events in MAPK signaling are phosphorylation events, we asked how the coverage of the LiP-MS readout upon acute osmotic stress compared to that of a phosphoproteomic analysis conducted in parallel. Overall, phosphoproteomics identified 11078 phosphopeptides mapping to 2022 proteins (Table 4), a similar number as we detected with the structural readout. We found 610 differentially phosphorylated proteins upon osmotic shock. Of the 316 proteins that we detected in both LiP and phosphoproteomics datasets, 48 were structurally altered upon stimulation. Among these 48 proteins, 38% of them have LiP peptides overlapping or in close proximity (± 10 amino acids) to a phosphopeptide; thus, the structural readout can detect a subset of phosphorylation events reported by phosphoproteomics. As expected, differentially phosphorylated proteins included proteins of the HOG1 pathway, in particular MAPK kinases of the upstream osmotic response and plasma membrane osmosensors. In parallel, LiP-MS identified a set of 20 structurally altered proteins within the HOG1, glycolysis and glycerol biosynthesis pathways (Table 3). Six proteins showed both structural variations and differentially regulated phosphopeptides upon osmotic stress: Hog1, Ste20, Gpd1/2, Tdh3 and Fba1 (Figure S3a). Structural variations included peptides that were differentially phosphorylated (Hog1, Fba1), mapped to a sequence adjacent to the phosphosite (Gpd1) or to a region located in close proximity to the phosphosite in the protein 3D structure (Tdh3) (Figure S3b). In summary, the structural analysis detected a subset of proteins that became phosphorylated upon osmotic shock, as defined by phosphoproteomics, but also detected changes in additional proteins. The structural and phosphoproteomics analyses are thus complementary to each other (Figure S3a; Table 4).

**Functional protein assemblies detected via structural alterations**

As a fourth type of functional change, we probed the ability of LiP-MS to report on protein complex assembly under acute stress, focusing on the heat shock dataset. A previous study based on centrifugation of cell extracts and subsequent MS analysis of the resulting pellets identified about 170 proteins that rapidly become insoluble upon heat shock, likely as a consequence of aggregation, misfolding or formation of protein/RNA complexes (Wallace et al., 2015); these proteins have been referred to as “aggregators”. First, we asked whether structural alterations detected by LiP-MS occurred in this set of proteins. We observe a clear enrichment for aggregators (Fisher’s exact test, p-value < 0.05) in our dataset, with 84 aggregators showing structural changes (Figure 1d; Table 5). Among them, only four (Nug1p, Faa4p, Nog2p and Ett1p) showed an abundance change.

Heat shock is also known to activate molecular chaperones (Mackenzie et al., 2016; O'Connell et al., 2014), which should engage in interactions with their clients (Balchin et al., 2016). We detected a significant enrichment for chaperones among structurally altered proteins after heat shock (Fisher’s exact test, p-value < 0.05), whereas yeast subject to osmotic stress did not show a similarly prominent chaperone structural response (Figure 1d). We next asked if chaperones could be involved in interactions with the aggregators, in an attempt to counteract their misfolding. Indeed, based on curated data from the literature on chaperone interactors (see methods), the set of aggregators was enriched for client-chaperone relationships, with 57 out of 84 structurally-altered aggregators known to physically interact with a chaperone in which we also detected a structural change (Figure 1d; Table 5). It is therefore likely that these structural changes are indicative of chaperone-substrate interaction in response to heat-induced protein misfolding or aggregation.

We observed structural changes in chaperones of several subfamilies, known to localize to various subcellular locations (Figure 1d; Table 5). In addition to chaperone-client interactions, we detected structural changes that likely report on the regulation of chaperones by nucleotides, which occurs after their activation after heat stress. For the protein disaggregase ATPase Hsp104p, for instance, we found two regulated LiP peptides out of 65 detected, both mapping to the high-turnover nucleotide binding domain 1 (NBD1) (Figure 1e). ATP-binding to NBD1 triggers substrate binding and stimulates ATP hydrolysis at nucleotide binding domain 2 (NBD2) (Gates et al., 2017). We mapped LiP peptides to both ATP- and ADP-bound Hsp104 hexameric structures. One of the two peptides mapped to the active site while the other was in close proximity, suggesting that LiP-MS captures conformational changes of the chaperone associated to the ATP hydrolysis cycle (Figure 1e). Similarly, three out of four regulated LiP peptides for the Kar2p chaperone, also an ATPase, mapped to the active site (Figure 1e), further supporting that LiP-MS captures structural alterations in chaperones related to nucleotide binding or hydrolysis.

In summary, in yeast responding to acute heat stress, the structural readout captured misfolding events leading to aggregation, assembly of chaperone-client interactions, as well as potential allosteric regulation of chaperones.

**Protein structural changes during nutrient adaptation in *E. coli***

For our second system, we studied *E. coli* grown in eight different carbon sources: acetate, galactose, succinate, glycerol, pyruvate, fructose, glucose and gluconate. We chose this model to leverage a recent systematic analysis in which metabolite levels and fluxes through central carbon metabolism (CCM) were determined in *E. coli* grown under these same conditions (Gerosa et al., 2015) and were shown to be condition-dependent. We reasoned that flux variations could be taken as a proxy for altered functional states of enzymes and thus used to assess the capability of structural changes (as detected by LiP-MS) to report on measurable functional changes in the CCM. We also aimed to further evaluate the information content of the structural readout and its overlap with information extracted from other ‘omic measurements, such as protein expression and metabolomics data.

We cultured *E. coli* in medium containing each of the eight carbon sources, harvested cells in exponential phase (OD = 0.8 ± 0.1), extracted the proteomes under native conditions and applied the LiP-MS workflow. We derived LiP-based structural fingerprints for a minimum of 1895 proteins (growth in galactose) up to a maximum of 1917 proteins (growth in gluconate) (Figure S1a; Table 6). We measured abundance changes for a similar number of proteins, ranging from 2085 proteins (growth in glycerol and galactose) to 2102 proteins (growth in pyruvate) (Figure S1a; Table 6).

Differential analysis of protein structure and abundance across the seven growth conditions, in each case compared to glucose, showed that on average 365 proteins underwent a structural alteration (15-25% of identified proteins) and 190 proteins changed in abundance (3-13% of identified proteins) (Figure 3a). Reassuringly, each growth condition resulted in the up-regulation of the expected nutrient transporters and nutrient uptake regulators (Figure 3b). The exceptions were the acetate transporter and acetyl coenzyme A synthetase, involved in acetate metabolism, which were down-regulated in glucose compared to all the other carbon sources, as previously observed (Schmidt et al., 2016).

With 294 and 313 significantly changing proteins, fructose and gluconate were the two conditions that diverged least from glucose in terms of protein structural states (Figure 3a). Growth in galactose, resulted in the highest number of structurally altered proteins (490), followed by acetate, succinate and glycerol (373, 370 and 350, respectively). Overall, these analyses identified sets of proteins that undergo structural changes, abundance changes, or both, in *E. coli* grown in different carbon sources, and show that a higher number of proteins undergo structural alterations in comparison to abundance changes.

**Global protein structural data are complementary to protein abundance information**

To assess whether the structural readout could identify cellular processes altered during *E. coli* metabolic adaptation and to assess the overlap of information derived from protein abundance and structural measurements, we performed functional enrichment analyses on proteins with altered structure and/or abundance during bacterial growth on the eight carbon sources (Figure 3c; Table 7). The structural and protein abundance readouts captured different sets of biological processes although some Gene Ontology (GO) terms overlapped. We focused on the different branches of central carbon metabolism (CCM) since this network is known to be heavily regulated upon our nutrient shifts.

Among proteins that changed abundance, the tricarboxylic acid cycle (TCA) was enriched in bacteria grown in acetate, galactose and succinate (Figure 3c). The glyoxylate shunt, a variant of the TCA cycle activated during growth in acetate and galactose (Haverkorn van Rijsewijk et al., 2011; Lehning et al., 2017), was enriched among proteins that changed abundance in these specific conditions. We did not observe enrichment of the glycolytic pathway in the set of proteins that changed in abundance (Figure 3c), although this pathway was substantially functionally altered across the eight growth conditions based on flux data (Gerosa et al., 2015). This is consistent with previous observations that glycolytic fluxes are not primarily controlled at the transcriptional level (Gerosa et al., 2015).

In contrast, the set of proteins with structural alterations as detected by LiP-MS, was enriched in enzymes involved in glycolysis in all growth conditions except for gluconate (Figure 3c). Out of 20 identified glycolytic enzymes, 16 showed a structural alteration in multiple conditions, whereas only three changed in abundance in at least one metabolic condition (Figure S4). Similarly, seven out of nine enzymes from the pentose phosphate pathway were altered at the structural level in multiple conditions, whereas none of them changed in abundance. The TCA cycle was regulated at both the abundance and the structural level, with all 16 identified proteins undergoing structural changes in almost all conditions (Figure S4). Finally, we also observed abundance and structural changes in several other processes such as amino acid biosynthesis, response to oxidative stress and fatty acid beta-oxidation (Figure 3c; Table 7). Interestingly, proteins with altered structure in the galactose condition showed significant enrichment in processes associated with amino acid biosynthesis and protein translation. This could be linked to the reduced growth rate observed for this condition (Gerosa et al., 2015).

Overall, these data suggest that different regulatory mechanisms control different branches of the CCM in *E. coli* adapting to different carbon sources. Glycolysis is controlled by regulatory mechanisms that affect protein structure rather than by modulation of gene expression. The TCA cycle, on the other hand, shows both protein abundance and structural changes between growth conditions, suggesting that changes in expression levels are accompanied by other regulatory processes.

The structural readout detected alterations in pathways that were not captured by protein abundance data alone, but also detected changes in proteins with altered expression levels. We reasoned that LiP-MS data that are not corrected for protein abundance changes should reflect regulation at both the structural and expression levels, since peptide intensities depend on both protein concentration and structural accessibility to the LiP protease (Feng et al., 2014; Schopper et al., 2017). Indeed*,* the enrichment profiles of non-normalized LiP data recapitulated most of the functional categories that were enriched when the pure structural and abundance readouts were combined (Figure S5; Table 7). This indicates that raw LiP-MS data capture the majority of the detected regulatory events. Nevertheless, separation of the structural and expression contributions provides more information on the nature of the detected regulation.

**Structural changes reflect functional alterations of metabolic enzymes**

To investigate whether structural signals report on functional changes in the *E. coli* CCM, we used prior 13C-based metabolic flux measurements (Gerosa et al., 2015) performed under the same experimental conditions as our study. Changes in flux for the reaction catalyzed by a given enzyme may occur due to changes in enzyme activity (for example in response to allosteric interactions or PTMs), in reactant concentrations, or in enzyme levels. While altered enzyme levels are captured by protein expression data, we hypothesized that a change in flux due to altered enzyme activity or altered binding of reactants might affect protein structure, resulting in a LiP signal. To test this, we asked whether LiP-MS data, corrected for protein abundance changes, captured structural changes for enzymes known to catalyze the reactions that change flux.

We first re-analyzed absolute flux measurements of 25 CCM reactions in *E. coli* grown under the eight nutrient conditions (Gerosa et al., 2015), and calculated flux ratios relative to glucose. Among the 25 reactions for which fluxes were measured, between 18 (for acetate) to 25 (for gluconate) significantly changed in flux in the different growth conditions (t-test, adjusted p-value < 0.05) (Table 8). LiP-MS detected a structural alteration for enzymes associated with the majority of flux changes (87%, Figure 4a), supporting the notion that a structural readout captures alterations in enzyme functional states. Protein abundance data detected changes in enzymes, mostly of the TCA and glyoxylate cycles, associated with only 39% of flux alterations, confirming that only some flux changes are explained by altered concentrations of the associated enzymes (Figure 4a).

Flux associated with an enzyme could be regulated by the same molecular event across growth conditions. Alternatively, different molecular events could regulate flux in different conditions. To distinguish between these two scenarios, we used linear regression to ask for which enzymes structural or abundance changes were quantitatively related to metabolic flux changes across the eight conditions. For three out of eleven glycolytic enzymes (gapA, pgk, eno), we detected one or multiple LiP peptides that linearly correlated (R2 > 0.7 and adjusted p-value < 0.05, Table 8) with flux measurements over the set of growth conditions (Figure 4b), suggesting that for these enzymes structural changes at specific sites are a quantitative predictor of fluxes. For the remaining enzymes in the network, there was no linear correlation between fluxes and a specific LiP peptide changing across at least four conditions, suggesting that fluxes for these enzymes are likely not regulated by the same molecular events in the different conditions. The abundance of enzyme mdh was linearly correlated with flux through the associated reaction, suggesting that, for mdh, fluxes are regulated by enzyme abundance in the conditions studied (R2 > 0.7 and adjusted p-value < 0.05, Table 8).

Of all the LiP peptides detected for glycolytic enzymes gapA, pgK and eno, only 11% correlated with fluxes. We thus asked whether these peptides were located at specific functional sites for each enzyme and could pinpoint the exact molecular events associated with flux changes. We mapped LiP peptides correlating with fluxes to available enzyme structures and assessed their proximity to catalytic sites or to binding sites of known allosteric regulators (Table 8). For all the three glycolytic enzymes (pgk, gapA and eno), a large fraction (pgk 2/2, gapA 11/14 and eno 9/26) of LiP peptides that correlated with flux data also mapped to the enzyme active site and, in the case of gapA, to a known allosteric site (Table 8).

We speculated that LiP peptides at active sites report on substrate occupancy and that they correlate with flux because substrate occupancy integrates different events that affect flux. To assess this, we performed *in vitro* LiP-MS on purified pgk in the presence of different amounts of its substrate 3-phosphoglycerate. To mimic as much as possible the *in vivo* experiment, we spiked purified pgk into an *E. coli* lysate cleared of endogenous metabolites and performed LiP-MS upon addition of increasing amounts of 3-phosphoglycerate to the lysate. We detected 180 peptides for pgk (Table 9). Remarkably, the two peptides that increase with added substrate *in vitro* covered exactly the same active site region we detected as undergoing structural changes by LiP experiments on the endogenous enzyme across the different nutrient conditions (Figure 4b). This suggests that these LiP peptides report on substrate occupancy for pgk *in situ* and that substrate occupancy monitored by LiP correlates with flux changes.

Taken together, our data show that the structural readout captures functional changes of *E. coli* CCM enzymes and can directly capture substrate occupancy at the active site of an enzyme *in situ*.

**The structural readout generates new hypotheses on regulatory events in CCM**

Metabolites can regulate enzyme activity by allosteric interactions (Chubukov et al., 2014). Thus, we asked if some of the structural alterations we detected for CCM enzymes could underlie novel cases of allosteric regulation. We reasoned that, in case of such regulation, levels of metabolite regulators should correlate with structural alterations at the allosteric site of the target enzyme. We used linear regression to test for a correlation between structural changes in enzymes of the CCM and relevant metabolite levels, across the eight growth conditions. We found a linear correlation between metabolite concentration and LiP peptide abundance for 32 enzymes of the CCM (Figure 5a, b; Table 8). Among these metabolite-enzyme pairs, three were well-characterized allosteric (FBP-pykF) or catalytic (FBP-pfkA, FBP-fbaA) interactions (see methods) and five (FBP-ptsI, FBP-pgk, NAD-eno, CIT-pta, NAD-pfkA) were corroborated by recent physical interaction data (Diether et al., 2019; Piazza et al., 2018). Importantly, LiP peptides that correlated with metabolite levels were at (fbaA) or in close proximity to (ptsI, pfkA, pykF, eno) the binding site identified by these previous experiments (Piazza et al., 2018), when this information was available (Figure 5c; Table 8). Overall, our LiP analysis suggests that seven metabolite-protein interactions that had been previously detected in cell lysates (Diether et al., 2019; Piazza et al., 2018) are physiologically relevant under the eight metabolic conditions we analyzed. Further, this structural analysis hypothesizes 12 novel metabolite-protein interactions (Figure 5b), including putative binding sites, that may play regulatory roles under the studied conditions.

**Discussion**

With this study, we demonstrate that detecting dynamic alterations of protein structures on a proteome-wide scale provides a powerful global readout of protein functional alterations *in situ*. Our approach captured protein functional alterations due to different molecular events, including post-translational modification, metabolite-driven allosteric events, protein-protein interactions, and protein misfolding. For enzymes in particular, the structural readout detected activity changes and altered occupancy of active sites, both of which are associated with altered reaction fluxes.

Regulation of many cellular processes (e.g., signaling) relies solely on these types of molecular events rather than on altered protein levels. Deregulation of such processes is therefore typically not captured by classical protein expression screens, which are instead useful for capturing transcriptionally-regulated cellular processes. In all the systems we examined, the structural readout indeed captured more altered proteins and processes (as assessed by GO term mapping) than the abundance readout and the processes captured by the two approaches were often complementary. This suggests that integrating structural and abundance-based proteomics readouts will maximize detection of altered biological processes. We note that protein abundance information is not lost in a LiP-MS experiment since the control step of the approach, in which proteins are not subject to limited proteolysis, reports on altered protein levels. A LiP-MS experiment can thus capture both structural and abundance changes, allowing us to simultaneously probe most types of molecular events leading to functional alterations of proteins and biological processes.

The LiP-MS structural readout provides data complementary to those obtained with other ‘omic techniques. Integration with metabolomics data enabled the detection of potential novel metabolite-protein interactions, suggesting that the coupling of these two ‘omic analyses may be beneficial for the detection of regulatory events based on small molecule binding. When conducted in parallel to a phosphoproteomic analysis, the structural readout captured multiple types of molecular events known to regulate protein function and that were not captured in the phosphorylation dataset (Figure S2). However, phosphoproteomic analyses additionally revealed phosphorylation-associated functional alterations of low-abundance proteins, since the phosphopeptide enrichment step intrinsic to these analyses increases proteome coverage. For proteins analyzed by both approaches, LiP detected structural alterations for 40% of the regions that changed phosphorylation state. The phosphoproteomic analysis is restricted to the phosphorylated fraction of protein molecules, as it is typically done after enrichment of phosphopeptides, while LiP-MS monitors the average structural state of both the phosphorylated and non-phosphorylated protein pools. Thus LiP-MS may not detect a structural change for a differentially phosphorylated protein region if the overall degree of phosphorylation is low*.*

In bacteria adapting to different nutrient conditions, we detected structural alterations for most enzymes associated with flux changes, suggesting that LiP-MS is a good readout for enzyme functional alterations. In contrast, slightly more than a third of flux changes were explained by abundances of the associated enzymes. The structural readout reports on different molecular events that can affect fluxes, including allosteric and phosphorylation events and changes in the relative levels of enzyme and/or substrate, affecting occupancy of enzyme active sites. For three enzymes, fluxes across metabolic conditions correlated with levels of LiP peptides at active sites and we showed that these peptides typically report on site occupancy. This indicates that, for these enzymes, fluxes in these conditions are regulated by relative levels of enzyme and reactants. For enzymes in which we did not detect a correlation with flux across metabolic conditions, it is likely that different types of molecular events regulate fluxes across these conditions, as previously suggested (Gerosa et al., 2015).

Our data show that LiP peptides located at functional sites can be used as markers to probe specific functional events *in situ*. For example, LiP peptides at the active site of pgk and at the metabolite binding sites of PtsI and eno reported on the occupancy of those sites both *in vitro* and *in situ*. Similarly, the LiP peptide from Hog1 that contains a phosphorylation site changed concomitantly with phosphorylation. It may be possible to extract similar markers for a variety of proteins by mapping structural proteomics data from perturbed proteomes to high-resolution structures and integrating prior knowledge on protein functional states in these conditions.

Our analysis detected a variety of previously characterized molecular events across the eight metabolic conditions we studied in *E. coli* and upon heat or osmotic shock in yeast. Notably, as we have seen before in yeast undergoing a metabolic shift (Feng et al., 2014), cell lysis and native proteome extraction does not in itself disrupt structural changes due to molecular events in the cell. Our present study extends and confirms these observations. In yeast, only a few LiP peptides changed for each protein known to be involved in the response to osmotic shock, and we could map these peptides to known functional sites on the protein structures; this specificity and resolution is useful for generating mechanistic hypotheses. For example, in this study we could map binding of the metabolite FBP to the active site of the enzyme Pts1 in *E. coli* undergoing a metabolic shift, an interaction that we had previously also detected in *E. coli* lysates with exogenously added metabolite. Pts1 catalyses both the phosphorylation and uptake of glucose into the cell. Our data raise the possibility that FBP may act as a competitive inhibitor for both PEP (the phosphate donor in this reaction) and glucose binding to Pts1, such that uptake of glucose is inhibited upon accumulation of FBP within the cell. Finally, we also made several new discoveries. We demonstrated the *in vivo* relevance of protein-metabolite interactions that had been previously detected in experiments in cell lysates, including FBP-ptsI, FBP-pgk, NAD-eno, CIT-pta. Furthermore, we identified 12 potential novel interactions that appear to be involved in the regulation of *E. coli* metabolism.

A limitation of our structural approach is that it does not directly inform on the causes of the detected structural alterations. However, LiP-MS detects structural changes with peptide-level resolution (i.e. the change can be pinpointed to stretches of < 10 amino acids), allowing us to relate effects of perturbations to specific functional sites, such as active or allosteric sites, sites of PTMs, or protein-protein interaction interfaces. This supports the direct generation of testable molecular hypotheses, the design of follow-up biochemical experiments, and the design of mutations for functional studies, thus linking holistic and mechanistic approaches. Our method is particularly useful if a high-resolution structure of a protein is available, but sequence-based information on locations of functionally relevant sites or domains may be sufficient for hypothesis generation. Structurally altered proteins identified in a global analysis could also be followed-up by high-resolution structural studies (e.g., by x-ray crystallography or cryo-electron microscopy) and visualized in cells and tissues (e.g., by cryo-electron tomography), thus potentially leading to new structural biology workflows.

Our approach is adversely affected by the typical bias against low-abundance proteins, which is also an issue with classical proteomic screens, and by the need for extensive coverage of a protein sequence to detect alterations at specific functional sites. This issue could be mitigated in the future by coupling LiP-MS with fractionation or enrichment techniques (e.g., size-exclusion chromatography), with terminal amine isotopic labeling of substrates (Kleifeld et al., 2010), or with a semitryptic peptide enrichment strategy (Ma et al., 2018), all of which could increase coverage at the protein and peptide level.

Other mass spectrometric techniques such as cross-linking mass spectrometry or surface footprinting could in principle also be applied to the *in situ* detection of protein functional changes. Although these approaches struggle with the analysis of complex proteomes and with the comparative analysis of differently treated samples, recent technical developments suggest that they might be a promising direction for the dynamic analysis of structural proteomes.

Pioneering computational biology studies have exploited information from static protein structures to assess the properties of specific biological systems, thus illustrating the potential of global structural data. We propose that the incorporation of dynamic *in situ* structural data obtained for proteomes under different conditions, as those generated by our LiP-MS approach, will be critical to fully realise the potential of structural systems biology and to link systemic and reductionist approaches. The quantitative measurement of molecular events such as active site occupancy may support the development of novel frameworks for the modeling of biological systems. By linking dynamic and high-resolution structural data, our global structural approach may bring us one step closer to a 3D model of the functioning of a cell.

REFERENCES

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

Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. Nature *537*, 347-355.

Ardito, F., Giuliani, M., Perrone, D., Troiano, G., and Lo Muzio, L. (2017). The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). Int J Mol Med *40*, 271-280.

Balchin, D., Hayer-Hartl, M., and Hartl, F.U. (2016). In vivo aspects of protein folding and quality control. Science *353*, aac4354.

Banaszak, K., Mechin, I., Obmolova, G., Oldham, M., Chang, S.H., Ruiz, T., Radermacher, M., Kopperschlager, G., and Rypniewski, W. (2011). The crystal structures of eukaryotic phosphofructokinases from baker's yeast and rabbit skeletal muscle. J Mol Biol *407*, 284-297.

Batth, T.S., Papetti, M., Pfeiffer, A., Tollenaere, M.A.X., Francavilla, C., and Olsen, J.V. (2018). Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling. Cell Rep *22*, 2784-2796.

Boisvert, F.M., Lam, Y.W., Lamont, D., and Lamond, A.I. (2010). A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. Mol Cell Proteomics *9*, 457-470.

Brewster, J.L., and Gustin, M.C. (2014). Hog1: 20 years of discovery and impact. Sci Signal *7*, re7.

Chubukov, V., Gerosa, L., Kochanowski, K., and Sauer, U. (2014). Coordination of microbial metabolism. Nat Rev Microbiol *12*, 327-340.

Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., and Aebersold, R. (2011). Comprehensive quantitative analysis of central carbon and amino-acid metabolism in Saccharomyces cerevisiae under multiple conditions by targeted proteomics. Mol Syst Biol *7*, 464.

Cravatt, B.F., Wright, A.T., and Kozarich, J.W. (2008). Activity-based protein profiling: from enzyme chemistry to proteomic chemistry. Annu Rev Biochem *77*, 383-414.

Damaghi, M., Wojtkowiak, J.W., and Gillies, R.J. (2013). pH sensing and regulation in cancer. Front Physiol *4*, 370.

Diether, M., Nikolaev, Y., Allain, F.H., and Sauer, U. (2019). Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli. Mol Syst Biol *15*, e9008.

Dihazi, H., Kessler, R., and Eschrich, K. (2004). High osmolarity glycerol (HOG) pathway-induced phosphorylation and activation of 6-phosphofructo-2-kinase are essential for glycerol accumulation and yeast cell proliferation under hyperosmotic stress. J Biol Chem *279*, 23961-23968.

Feng, Y., De Franceschi, G., Kahraman, A., Soste, M., Melnik, A., Boersema, P.J., de Laureto, P.P., Nikolaev, Y., Oliveira, A.P., and Picotti, P. (2014). Global analysis of protein structural changes in complex proteomes. Nat Biotechnol *32*, 1036-1044.

Gates, S.N., Yokom, A.L., Lin, J., Jackrel, M.E., Rizo, A.N., Kendsersky, N.M., Buell, C.E., Sweeny, E.A., Mack, K.L., Chuang, E.*, et al.* (2017). Ratchet-like polypeptide translocation mechanism of the AAA+ disaggregase Hsp104. Science *357*, 273-279.

Gerosa, L., Haverkorn van Rijsewijk, B.R., Christodoulou, D., Kochanowski, K., Schmidt, T.S., Noor, E., and Sauer, U. (2015). Pseudo-transition Analysis Identifies the Key Regulators of Dynamic Metabolic Adaptations from Steady-State Data. Cell Syst *1*, 270-282.

Haverkorn van Rijsewijk, B.R., Nanchen, A., Nallet, S., Kleijn, R.J., and Sauer, U. (2011). Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli. Mol Syst Biol *7*, 477.

Hohmann, S. (2015). An integrated view on a eukaryotic osmoregulation system. Curr Genet *61*, 373-382.

Huber, K.V., Olek, K.M., Muller, A.C., Tan, C.S., Bennett, K.L., Colinge, J., and Superti-Furga, G. (2015). Proteome-wide drug and metabolite interaction mapping by thermal-stability profiling. Nat Methods *12*, 1055-1057.

Humphrey, S.J., Azimifar, S.B., and Mann, M. (2015). High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat Biotechnol *33*, 990-995.

Ideker, T., Thorsson, V., Ranish, J.A., Christmas, R., Buhler, J., Eng, J.K., Bumgarner, R., Goodlett, D.R., Aebersold, R., and Hood, L. (2001). Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. Science *292*, 929-934.

Kleifeld, O., Doucet, A., auf dem Keller, U., Prudova, A., Schilling, O., Kainthan, R.K., Starr, A.E., Foster, L.J., Kizhakkedathu, J.N., and Overall, C.M. (2010). Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. Nat Biotechnol *28*, 281-288.

Kolch, W. (2005). Coordinating ERK/MAPK signalling through scaffolds and inhibitors. Nat Rev Mol Cell Biol *6*, 827-837.

Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M.M., Pronk, J.T., Slijper, M., and Heck, A.J. (2006). Proteome analysis of yeast response to various nutrient limitations. Mol Syst Biol *2*, 2006.0026.

Lee, Y.J., Jeschke, G.R., Roelants, F.M., Thorner, J., and Turk, B.E. (2012). Reciprocal phosphorylation of yeast glycerol-3-phosphate dehydrogenases in adaptation to distinct types of stress. Mol Cell Biol *32*, 4705-4717.

Lehning, C.E., Siedler, S., Ellabaan, M.M.H., and Sommer, M.O.A. (2017). Assessing glycolytic flux alterations resulting from genetic perturbations in E. coli using a biosensor. Metab Eng *42*, 194-202.

Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., Claassen, M., and Picotti, P. (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. Science *355*.

Liu, F., Lossl, P., Rabbitts, B.M., Balaban, R.S., and Heck, A.J.R. (2018). The interactome of intact mitochondria by cross-linking mass spectrometry provides evidence for coexisting respiratory supercomplexes. Mol Cell Proteomics *17*, 216-232.

Ma, R., Meng, H., Wiebelhaus, N., and Fitzgerald, M.C. (2018). Chemo-Selection Strategy for Limited Proteolysis Experiments on the Proteomic Scale. Anal Chem *90*, 14039-14047.

Mackenzie, R.J., Lawless, C., Holman, S.W., Lanthaler, K., Beynon, R.J., Grant, C.M., Hubbard, S.J., and Eyers, C.E. (2016). Absolute protein quantification of the yeast chaperome under conditions of heat shock. Proteomics *16*, 2128-2140.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F.*, et al.* (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55-62.

Mydy, L.S., Cristobal, J.R., Katigbak, R.D., Bauer, P., Reyes, A.C., Kamerlin, S.C.L., Richard, J.P., and Gulick, A.M. (2019). Human Glycerol 3-Phosphate Dehydrogenase: X-ray Crystal Structures That Guide the Interpretation of Mutagenesis Studies. Biochemistry *58*, 1061-1073.

Niphakis, M.J., Lum, K.M., Cognetta, A.B., 3rd, Correia, B.E., Ichu, T.A., Olucha, J., Brown, S.J., Kundu, S., Piscitelli, F., Rosen, H.*, et al.* (2015). A Global Map of Lipid-Binding Proteins and Their Ligandability in Cells. Cell *161*, 1668-1680.

Nussinov, R., Tsai, C.J., and Ma, B. (2013). The underappreciated role of allostery in the cellular network. Annu Rev Biophys *42*, 169-189.

O'Connell, J.D., Tsechansky, M., Royal, A., Boutz, D.R., Ellington, A.D., and Marcotte, E.M. (2014). A proteomic survey of widespread protein aggregation in yeast. Mol Biosyst *10*, 851-861.

Pauling, L., Itano, H.A., and et al. (1949). Sickle cell anemia a molecular disease. Science *110*, 543-548.

Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., and Picotti, P. (2018). A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. Cell *172*, 358-372.e323.

Ressa, A., Bosdriesz, E., de Ligt, J., Mainardi, S., Maddalo, G., Prahallad, A., Jager, M., de la Fonteijne, L., Fitzpatrick, M., Groten, S.*, et al.* (2018). A System-wide Approach to Monitor Responses to Synergistic BRAF and EGFR Inhibition in Colorectal Cancer Cells. Mol Cell Proteomics *17*, 1892-1908.

Robertson, A.D., and Murphy, K.P. (1997). Protein Structure and the Energetics of Protein Stability. Chem Rev *97*, 1251-1268.

Russell, C.J. (2014). Acid-induced membrane fusion by the hemagglutinin protein and its role in influenza virus biology. Curr Top Microbiol Immunol *385*, 93-116.

Sahni, N., Yi, S., Zhong, Q., Jailkhani, N., Charloteaux, B., Cusick, M.E., and Vidal, M. (2013). Edgotype: a fundamental link between genotype and phenotype. Curr Opin Genet Dev *23*, 649-657.

Savitski, M.M., Reinhard, F.B., Franken, H., Werner, T., Savitski, M.F., Eberhard, D., Martinez Molina, D., Jafari, R., Dovega, R.B., Klaeger, S.*, et al.* (2014). Tracking cancer drugs in living cells by thermal profiling of the proteome. Science *346*, 1255784.

Schmidt, A., Kochanowski, K., Vedelaar, S., Ahrne, E., Volkmer, B., Callipo, L., Knoops, K., Bauer, M., Aebersold, R., and Heinemann, M. (2016). The quantitative and condition-dependent Escherichia coli proteome. Nat Biotechnol *34*, 104-110.

Schopper, S., Kahraman, A., Leuenberger, P., Feng, Y., Piazza, I., Muller, O., Boersema, P.J., and Picotti, P. (2017). Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. Nat Protoc *12*, 2391-2410.

Shaul, Y.D., and Seger, R. (2007). The MEK/ERK cascade: from signaling specificity to diverse functions. Biochim Biophys Acta *1773*, 1213-1226.

Sowa, M.E., Bennett, E.J., Gygi, S.P., and Harper, J.W. (2009). Defining the human deubiquitinating enzyme interaction landscape. Cell *138*, 389-403.

Strater, N., Marek, S., Kuettner, E.B., Kloos, M., Keim, A., Bruser, A., Kirchberger, J., and Schoneberg, T. (2011). Molecular architecture and structural basis of allosteric regulation of eukaryotic phosphofructokinases. Faseb j *25*, 89-98.

Verghese, J., Abrams, J., Wang, Y., and Morano, K.A. (2012). Biology of the heat shock response and protein chaperones: budding yeast (Saccharomyces cerevisiae) as a model system. Microbiol Mol Biol Rev *76*, 115-158.

Wallace, E.W., Kear-Scott, J.L., Pilipenko, E.V., Schwartz, M.H., Laskowski, P.R., Rojek, A.E., Katanski, C.D., Riback, J.A., Dion, M.F., Franks, A.M.*, et al.* (2015). Reversible, Specific, Active Aggregates of Endogenous Proteins Assemble upon Heat Stress. Cell *162*, 1286-1298.

Wepf, A., Glatter, T., Schmidt, A., Aebersold, R., and Gstaiger, M. (2009). Quantitative interaction proteomics using mass spectrometry. Nat Methods *6*, 203-205.

White, F.H., Jr., and Anfinsen, C.B. (1959). Some relationships of structure to function in ribonuclease. Ann N Y Acad Sci *81*, 515-523.