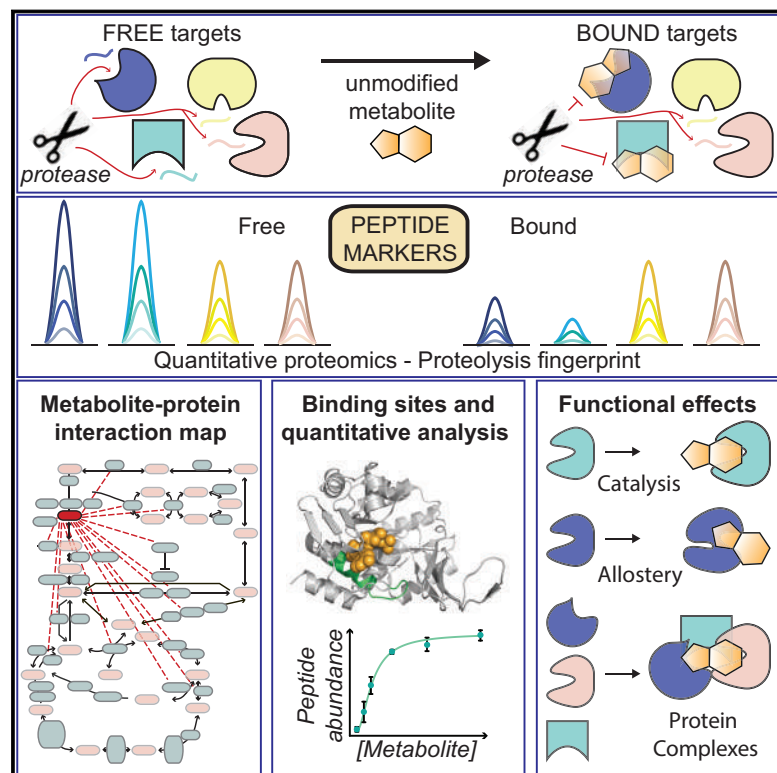# A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication

Graphical Abstract

Authors

Ilaria Piazza, Karl Kochanowski,
Valentina Cappelletti, Tobias Fuhrer,
Elad Noor, Uwe Sauer, Paola Picotti

Correspondence

picotti@imsb.biol.ethz.ch

In Brief

A method to assess proteome-wide binding to metabolites of interest uncovers new allosteric and enzymatic functions and provides a tool to investigate drug targets in the native cellular environment.

Highlights

- Chemical proteomics provides a map of metabolite-protein interactions

- Metabolite binding sites pinpointed on a proteome-wide scale

- New catalytic, allosteric, and metabolite-induced protein-protein interaction events

- Binding of metabolites at binding sites is promiscuous

CellPress

A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication

Ilaria Piazza,[1,2] Karl Kochanowski,[2,3] Valentina Cappelletti,[1,2] Tobias Fuhrer,[2] Elad Noor,[2] Uwe Sauer,[2] and Paola Picotti[1,2,4,*]

[1]Institute of Biochemistry, Department of Biology, ETH Zurich, Zurich, Switzerland
[2]Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland
[3]Present address: Department of Pharmaceutical Chemistry, University of California, San Francisco, San Francisco, CA, USA
[4]Lead Contact
*Correspondence: picotti@imsb.biol.ethz.ch
https://doi.org/10.1016/j.cell.2017.12.006

SUMMARY

Metabolite-protein interactions control a variety of cellular processes, thereby playing a major role in maintaining cellular homeostasis. Metabolites comprise the largest fraction of molecules in cells, but our knowledge of the metabolite-protein interactome lags behind our understanding of protein-protein or protein-DNA interactomes. Here, we present a chemoproteomic workflow for the systematic identification of metabolite-protein interactions directly in their native environment. The approach identified a network of known and novel interactions and binding sites in *Escherichia coli*, and we demonstrated the functional relevance of a number of newly identified interactions. Our data enabled identification of new enzyme-substrate relationships and cases of metabolite-induced remodeling of protein complexes. Our metabolite-protein interactome consists of 1,678 interactions and 7,345 putative binding sites. Our data reveal functional and structural principles of chemical communication, shed light on the prevalence and mechanisms of enzyme promiscuity, and enable extraction of quantitative parameters of metabolite binding on a proteome-wide scale.

INTRODUCTION

Beyond their roles as intermediates in metabolic conversions, metabolites serve as signals that directly or indirectly trigger adaptive responses. Nutritional states, stress, and ecological conditions influence the intracellular levels of hundreds of thousands of metabolites, and the signals these molecules mediate are transmitted through a series of molecular events, including binding of metabolites to proteins. Different types of functional interactions between proteins and metabolites have been reported (Chubukov et al., 2014). The best characterized interactions involve binding of metabolites to the active site of enzymes as substrates, cofactors, or products of enzymatic reactions. As allosteric regulators, metabolites bind to sites that are different from active sites, rapidly and reversibly modifying protein activity (Gerosa and Sauer, 2011). Allosteric interactions with metabolites also influence proteins with non-enzymatic functions such as transmembrane receptors (Changeux and Christopoulos, 2016) and transcription factors (Motlagh et al., 2014). Metabolite binding also regulates the assembly and function of protein complexes (Milroy et al., 2014) and high-molecular weight protein assemblies (O'Connell et al., 2012; Wu, 2013).

There are at least one million protein molecules in a bacterial cell, (Milo, 2013) and metabolites outnumber proteins by about 100-fold (Bennett et al., 2009). Thus, there could be millions of functionally relevant metabolite-protein binding events. Our knowledge of the metabolite-protein interactome is likely very partial, as the transient and low-affinity nature of metabolite-protein interactions (Lindsley and Rutter, 2006) has prevented systematic analyses similar to those performed to identify protein-protein (Hein et al., 2015) or protein-nucleic acid interactions (Castello et al., 2012). Most characterized protein-metabolite interactions have been discovered via hypothesis-driven experiments that rely on *in vitro* activity assays. These assays are laborious, depend on choice of the relevant ligand, and preclude the identification of interactions that do not change an *in vitro* measurable activity. Methods involving chemical modification of metabolites or protein tagging (Diether and Sauer, 2017) have been restricted to studies of lipid-protein interactions (Gallego et al., 2010; Niphakis et al., 2015) and hydrophobic metabolites (Li et al., 2010).

To enable a systematic analysis, unbiased with regard to both metabolites and proteins, we developed a chemoproteomic approach that combines limited proteolysis (LiP) with mass spectrometry (MS) in the presence of unmodified metabolites. We applied this strategy to the identification of known and novel interactions in *Escherichia coli*, the organism with the arguably best-characterized metabolic network (Keseler et al., 2013). We identified an extensive network of known and previously unknown metabolite-protein interactions and binding sites. Our data revealed functional principles of metabolite sensing and shed light on the prevalence and mechanisms of enzyme promiscuity. Our analysis provides a framework to study the effects of metabolite binding on the structure of proteins and protein complexes and to evaluate the effect of metabolite concentration on metabolite-protein interactions in the context of the cellular milieu.
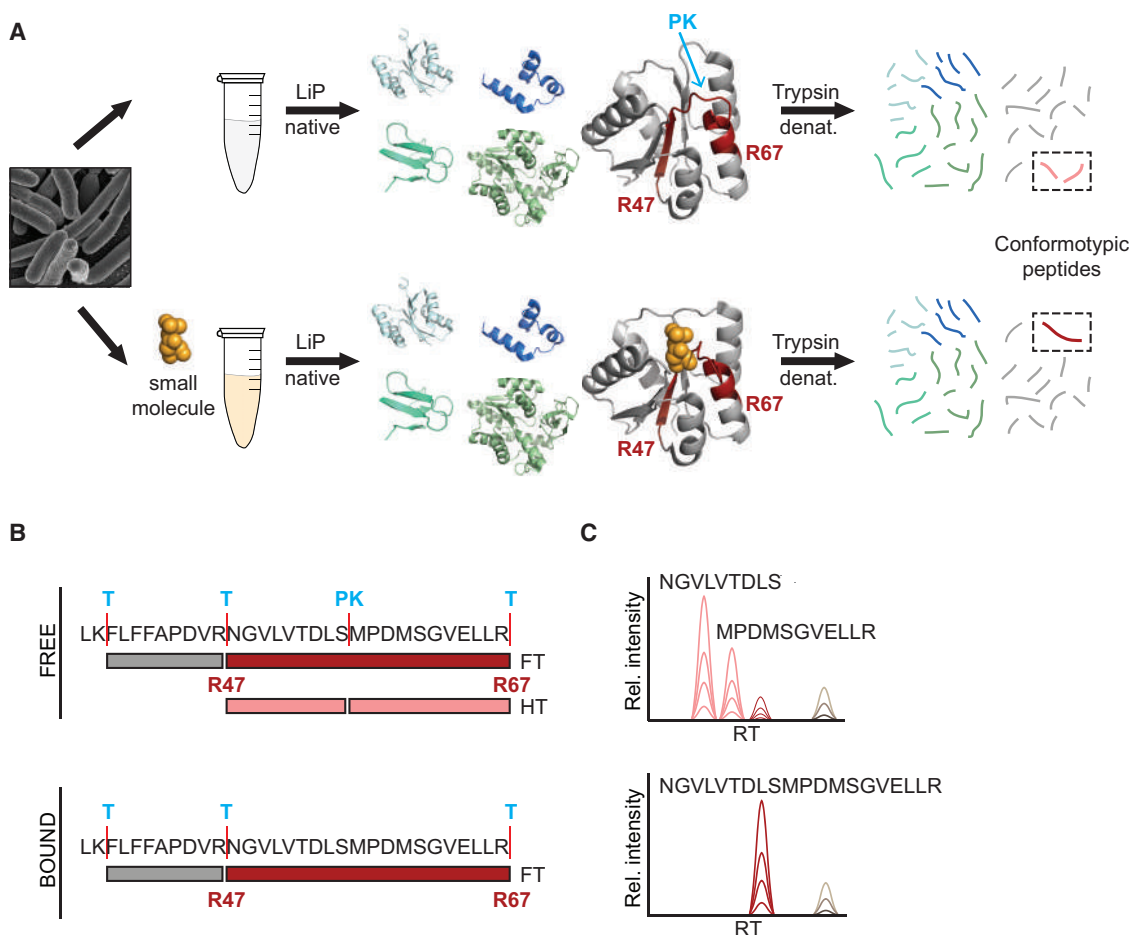
Figure 1. Workflow of the LiP-SMap Approach

(A) Whole-cell lysates extracted under native lysis conditions are treated or not with a metabolite. Metabolite binding to a protein alters local proteolytic susceptibility. A limited proteolysis step is performed with proteinase K (PK) under native conditions, followed by complete digestion with trypsin under denaturing conditions to generate MS-measurable peptides. In the example, structurally informative peptides are produced for the protein FixJ bound to aspartyl phosphate (PDB: 1DBW) and its ligand-free form (PDB: 1D5W). Peptides specific for the bound and unbound conformations (conformotypic peptides) are depicted in red.

(B) Schematic of FixJ peptides generated in the presence and absence of aspartyl phosphate. Binding prevents PK cleavage, resulting in the disappearance of two peptides with non-tryptic ends (HT) and in an increase in concentration of the associated fully tryptic peptide (FT).

(C) The three peptides are detected by MS.

RESULTS

Systematic Identification of Protein-Small Molecule Interactions in Native Proteome Extracts

Metabolite-protein interactions typically result in local or in global alterations of protein structures (Boehr et al., 2009; Nussinov and Tsai, 2013). We reasoned that detecting ligand-induced structural alterations on a proteome-wide scale could provide a universal readout of protein-small molecule interactions. Building on previous work (Feng et al., 2014), we devised a workflow, which we term LiP-small molecule mapping (LiP-SMap), to systematically detect proteins that become differentially susceptible to protease cleavage upon binding of a small molecule added to a proteome extract.

In the LiP-SMap workflow (Figure 1A), proteomes are extracted under conditions that preserve native protein structures, and ex-

tracts are exposed to a small molecule of interest. Samples are subjected to limited proteolysis with the broad-specificity protease proteinase K to generate structure-specific protein fragments (Figure 1B). Fragments are then digested with the sequence-specific protease trypsin to generate peptide mixtures amenable to bottom-up proteomic analysis. Peptides are analyzed by liquid-chromatography-coupled tandem MS, and LiP patterns of proteomes processed in the presence and absence of the small molecule are compared using a label-free quantitative MS approach. Our pipeline includes a quantitative unbiased shotgun proteomic step and a validation step based on data-independent acquisition (Gillet et al., 2016) (Figure 1C; STAR Methods).

Benchmarking the LiP-SMap Approach

To validate our approach, we chose three metabolites with different levels of promiscuity (Figure S1A; Table S1): adenosine

5′-triphosphate (ATP), a highly promiscuous metabolite for which more than 600 interactions are known, L-phenylalanine (L-Phe), for which only 15 interactions are known, and phosphoenolpyruvate (PEP), a metabolite of intermediate promiscuity. We compared proteolytic patterns of the *E. coli* proteome incubated in the absence and upon addition of two physiologically relevant concentrations of metabolite (Table S1). Proteins with altered proteolytic patterns are considered putative metabolite binding proteins (MBPs), and peptides with altered abundance are referred to as conformotypic peptides. The largest number of MBPs, 231, was observed for ATP. For PEP and L-Phe, we detected 129 and 41 MBPs, respectively, in agreement with known binding specificities. Between 30% and 95% more MBPs were observed at the higher concentration of each metabolite than at the lower concentration (Figure S1B).

The LiP-SMap approach captured 104 previously known binding events. Among the 231 MBPs detected for ATP were 92 known ATP binding proteins. For PEP, our approach detected nine of the 58 known MBPs, most notably the PEP producing enolase and the PEP-activated phosphate acetyltransferase Pta and fructose-1,6–bisphosphatase. Among the 41 MBPs detected for L-Phe were three of 15 known interactors (Table S1). At least one-third of the MBPs detected at the lower concentrations for ATP, PEP, and L-Phe involved previously characterized interactions for each metabolite (Figure 2A). More than 90% of MBPs identified at the lower concentration of each metabolite were also detected at the higher concentrations and the fraction of newly discovered interactions increased (Figures 2B and S1B), suggesting that the higher concentrations enable recovery of lower affinity interactions. MBPs of ATP, PEP, and L-Phe did not overlap significantly (Table S1). Of the known interactions recovered, more were allosteric (36.8%) than catalytic (21.2%) (Figure 2C). Detection of known interactions increased with protein sequence coverage achieved with MS (Figures S1C–S1E). These results suggest that, although LiP-SMap is not comprehensive due to proteome undersampling, the method enables the detection of regulatory and catalytic metabolite-protein interactions in an unbiased manner and directly in a complex biological matrix.

To evaluate the likelihood that the method results in false positive identifications, we used LiP-SMap to identify interactors of the antifungal drug cerulenin, which is known to interact with only one protein, Fas2. LiP-SMap analyses of yeast cell extracts treated with cerulenin showed that, of the more than 2,500 proteins identified, only Fas2 had an altered proteolytic pattern in treated relative to untreated extracts (Figure 2D). This experiment demonstrates that LiP-SMap is unlikely to result in false positive identifications.

A Global Map of Metabolite-Protein Interactions and Their Binding Sites

We then applied the LiP-SMap workflow to the unbiased analysis of metabolite-protein interactions in *E. coli*. We focused on 20 metabolites, most of which are intermediates of conserved central carbon metabolism (Figure 3A). We also included four amino acids and seven nucleoside phosphates, and 3′, 5′-cyclic adenosine monophosphate (cAMP) as a prototype of a signaling molecule. These 20 metabolites have a broad range of hydrophobicities, molecular weights, and charges.

We first removed endogenous metabolites from the *E. coli* proteome extracts by gel filtration. This was efficient for all nine metabolites tested (Figure S1F). Analysis was then performed at two physiologically relevant metabolite concentrations (Table S1) (Bennett et al., 2009; Gerosa et al., 2015), with the higher concentration mimicking maximal reported *in vivo* values to maximize the likelihood of detecting interactions spanning a broad range of affinities. Even at the higher concentration, proteolytic patterns remained unchanged for 1,945 of the 2,565 detected proteins, suggesting that the chosen experimental conditions did not result in non-specific binding (Figure S2A).

Overall, we identified 1,678 protein-metabolite interactions. Of these, 1,447 had not been previously reported. Many novel interactions involved organic acids and sugar phosphates (377 and 410, respectively) (Figures 3B and S2B). Of the newly detected metabolite-binding proteins, 76 were proteins without an annotated function; the metabolites these proteins interact with may provide clues about biological activities (Table S1).

We estimated the false discovery rate by comparison of interactions we detected for metabolic proteins with known interactions from the BRENDA database (http://www.brenda-enzymes.org/) (Figure S2C; Table S2). We considered true positives interactions reported in BRENDA that were captured by our approach, and true negatives those not reported in BRENDA and not identified by LiP-SMap. Interactions detected by our method that were not present in BRENDA were considered false positives. The false discovery rate was 5.5%, and the fraction of known interactions recovered increased with increasing evidence of the reported interactions (Figures S2D–S2F; STAR Methods). Importantly, this false discovery rate is an overestimate, since many of what we considered false positives in this exercise are likely to be novel interactions. We also developed a scoring system to estimate the confidence in interactions detected with criteria based on parameters from our experiments and on the degree of supporting literature evidence (STAR Methods). Of the 1,678 interactions detected, 54% were supported by multiple lines of evidence (Figures 3C and S2G; Table S2), indicating that a large fraction of the dataset comprises high confidence interactions.

Conformotypic peptides identified upon treatment of cells with a given metabolite define the regions of a protein that are structurally affected by binding of the metabolite. For example, in the LiP-SMap experiment with cerulenin we detected only one Fas2 peptide with altered abundance upon drug treatment, and this peptide is at the known drug binding site (Figure 3D; Table S2). To determine whether conformotypic peptides pinpoint metabolite binding sites on a proteome-wide scale, we retrieved all experimentally determined structures of protein-metabolite complexes from the Protein Data Bank Ligand Expo (http://ligand-expo.rcsb.org/) and measured the minimal distances between atoms of the metabolite and those of conformotypic peptides. Conformotypic peptides identified by LiP-SMap were most frequently positioned in very close proximity to binding sites (Figures 3E and 3F; Table S3). The median distance between bound metabolite and the closest conformotypic peptide atom was
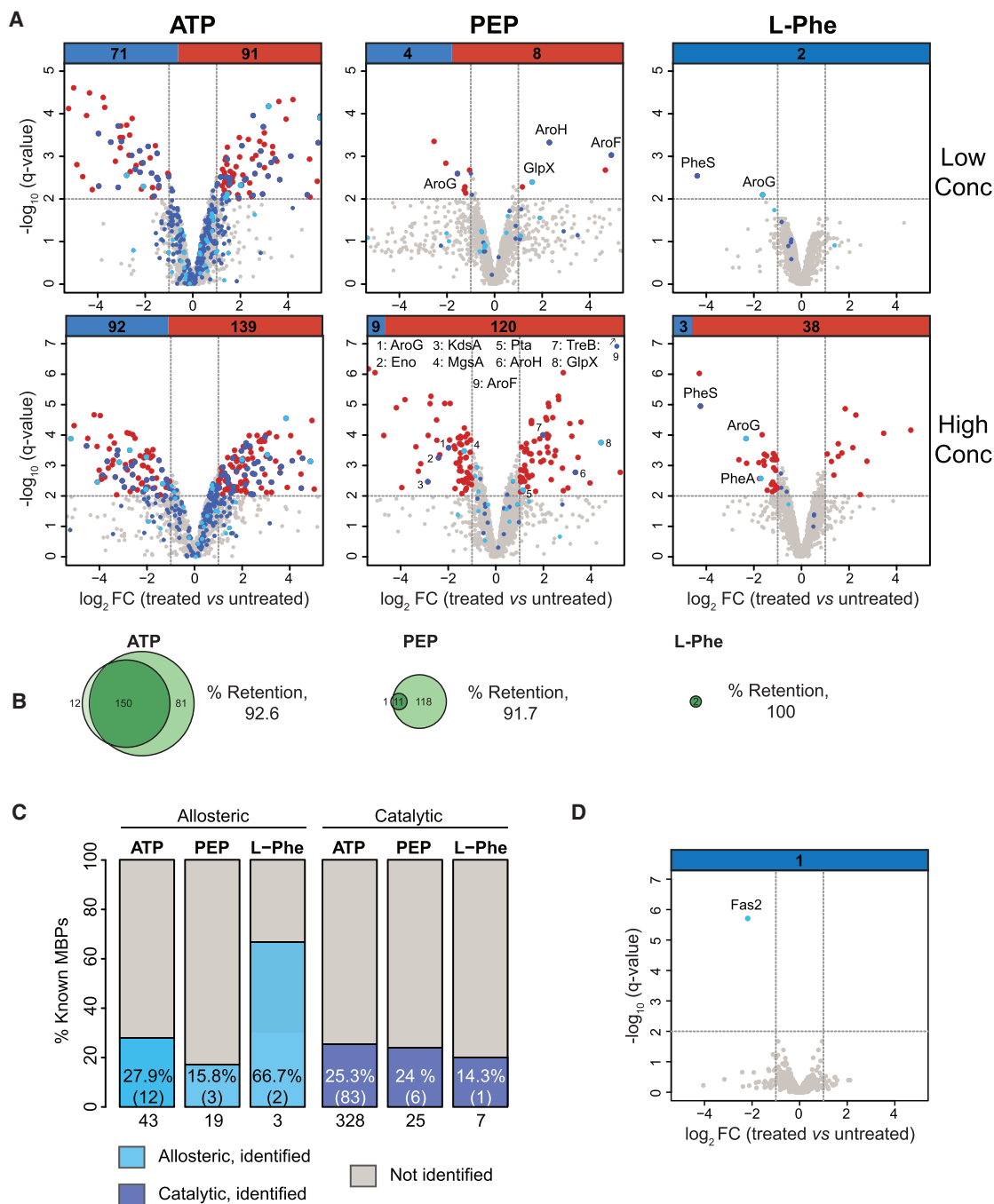
Figure 2. Benchmarking the LiP-SMap Approach

(A) Volcano plots of LiP-SMap experiments with low (top panels) and high (bottom panels) concentrations of ATP, PEP, and L-Phe. Peptide mixtures produced in the presence and absence of each metabolite are compared. Fold changes (FC) in peptide abundance in treated versus untreated samples are shown as a function of significance. Significance cutoffs were $q = 0.01$ (Bayes moderated t tests) and FC = 2 (n = 3). Each protein is represented with a single data point in the graph, corresponding to the peptide with the lowest q value. Known catalytic and allosteric interactions are in blue and cyan, respectively. Novel interactions are in red. Bars indicate the total number of known (blue) and novel (red) MBPs identified.

(B) Venn diagrams show the number of MBPs identified at the low and high concentrations. % Retention is the percentage of protein identifications from the lowest metabolite concentration that were retained at the higher concentration.

(C) Relative fraction of recovered known allosteric (cyan) or catalytic (blue) interactions over total. Numbers in brackets and below the bars indicate the number found and total known targets, respectively. All numbers refer to proteins for which at least one peptide was detected by MS.

(D) Volcano plot representation as in (A) of the LiP-SMap experiments with cerulenin.
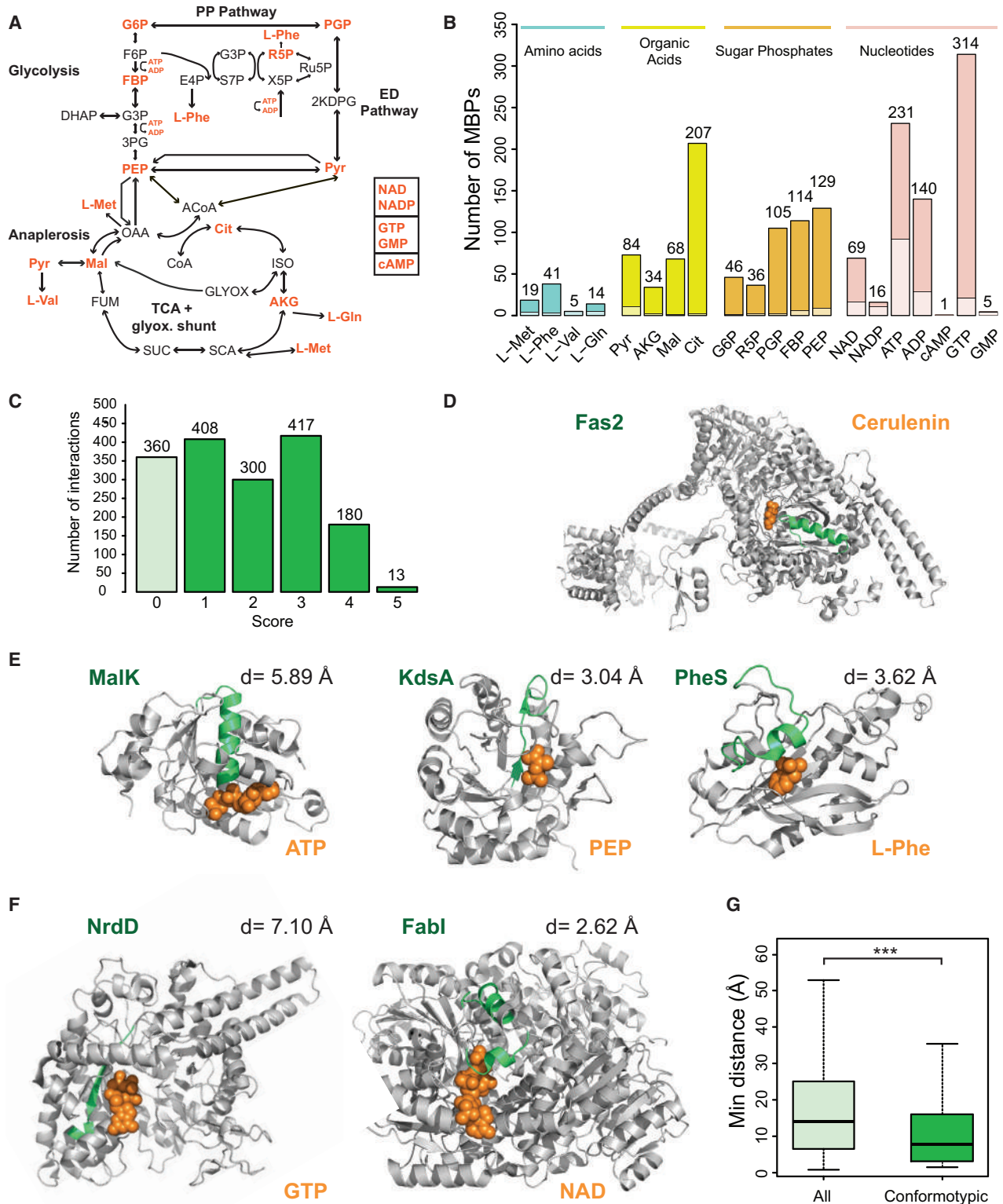
See also Figure S1.

Figure 3. Global Map of Metabolite-Protein Interactions

(A) Representation of the central carbon metabolism network with input, intermediate, and output metabolites as nodes. Metabolites analyzed by LiP-SMap (orange) cover a broad range of reactions. PP = pentose phosphate; ED = Entner-Doudoroff.

(legend continued on next page)

4.41 Å, significantly lower than the 14.07 Å median distance from atoms of all peptides detected (two-sided Wilcoxon test, p < 2.2×10^{-16}). Similar results were obtained when holocomplex structures with only one conformotypic peptide identified were considered (p < 2.93×10^{-6}; median minimal distances of 7.87 and 14.07 Å, respectively; Figure 3G; Table S3). We observed no significant biases in terms of protein secondary structure between conformotypic and non-conformotypic peptides (Figure S3A). Thus, our method provided a peptide-level resolution map of metabolite binding sites on a proteome-wide scale.

Features of the Metabolite-Protein Interactome

The MBPs identified included proteins encompassing a wide range of functions with localization in membranes and cytosol (Figures S3B and S3C). Only half of the discovered metabolite-protein interactions involved metabolic enzymes (Figure S3D). Metabolite binding often preferentially occurred within the metabolic sub-network known to involve the metabolites of interest. For example, MBPs identified for phenylalanine and valine were enriched for tyrosine, tryptophan, and phenylalanine metabolism and valine, leucine, and isoleucine metabolism, respectively (hypergeometric tests: p < 5.65×10^{-4} and p < 3.87×10^{-7}). MBPs identified for glycolysis intermediates fructose-1,6-bisphosphate (FBP) and PEP were both enriched in proteins from "glycolysis-gluconeogenesis," and FBP-binding proteins were also enriched for the pentose phosphate pathway class. Proteins that bound ATP, ADP, guanosine 5′-triphosphate (GTP), guanosine 5′-monophosphate (GMP), and citrate were enriched in purine and pyrimidine biosynthesis pathway (p < 5×10^{-3}). No biases for specific catalytic mechanisms based on analysis of enzyme classification (EC) numbers were observed (Figure S3E).

Next, we assessed the relationship between metabolite-sensitive proteome determined by LiP-SMap and the "core proteome" of *E. coli*, a set of 356 proteins consistently expressed across environmental conditions (Yang et al., 2015). The relative frequency of core proteome proteins was significantly higher among MBPs discovered with LiP-SMap than in the set of detected proteins (Fisher's exact test, p = 2.04×10^{-8}), and several core proteins interacted with multiple metabolites (Figure S4A; Table S4). *E. coli* genes that showed high variability in their expression levels (Schmidt et al., 2016) included only a few MBPs. Additionally, MBPs showed a significantly smaller variation in abundances across growth conditions than the rest of the *E. coli* proteome (Figure S4B). This suggests that many transcriptionally stable and core proteins respond to metabolite-mediated signals.

Novel Regulatory Interactions

The novel interactions we detected could reflect different types of functional events including allosteric or catalytic events or ligand-induced regulation of the assembly of protein complexes. Thus, we next systematically evaluated these three types of functional interactions. To evaluate whether our detected metabolite-protein interactions included novel cases of allosteric interactions, we focused on metabolite interactions with enzymes from central carbon metabolism (Figure 4A; Table S4) for which *in vitro* enzyme assays are available. We focused on the interaction of citrate with phosphoenolpyruvate carboxylase (Ppc), FBP with the PEP synthetase regulatory protein (PpsR), and FBP with glucose-6-phosphate dehydrogenase (G6PDH).

The carboxylation of PEP to oxaloacetate catalyzed by Ppc replenishes carbon removed from the tricarboxylic acid (TCA) cycle (Figure S4C). Allosteric activation of Ppc by acetyl-CoA and FBP (Sanwal and Maeba, 1966), the latter detected in our LiP-SMap experiment (Figure 4A), ensures rapid inactivation of Ppc when glycolytic flux is low (Xu et al., 2012). We identified TCA cycle intermediate citrate as a potential interactor of Ppc (Figure 4A). To test whether citrate influences the catalytic activity of Ppc, we performed a coupled activity assay that recapitulates the reported synergistic activation of Ppc by acetyl-CoA and FBP (Figure S4D). In this assay, citrate acted as a potent inhibitor of Ppc activity (Figure 4B). This inhibition was not significantly affected by FBP or acetyl-CoA (Figure 4B), suggesting that citrate-mediated inhibition of Ppc overrides activation by FBP and acetyl-CoA. Other metabolites that our approach did not identify as Ppc-binding molecules did not show an effect on Ppc activity (Figure S4E).

We also identified FBP as a new regulator of G6PDH activity at concentrations of FBP above 1 mM (Figure S4C). FBP addition reduced both the K_m and V_{max} of G6PDH, suggesting a mixed inhibition mechanism (Figure 4C).

Finally, we confirmed interaction of FBP with PpsR, a dual kinase/phosphatase that catalyzes the ADP-dependent phosphorylation (which inactivates) and the phosphate-dependent dephosphorylation (which activates) of PEP synthetase (PpsA) (Burnell, 2010). PpsR was identified as an FBP binder in the LiP-SMap screen at both FBP concentrations, whereas PpsA was not. Therefore, we hypothesized that FBP regulates PpsA activity indirectly through an allosteric interaction with PpsR. FBP alone did not alter PpsA activity (Figure S4F); however, addition of 5 mM FBP caused a modest (~15%) but significant reduction in PpsA activity in the presence of PpsR (Figure 4D). Thus, the FBP-PpsR binding event detected by LiP-SMap suggests a regulation scheme in which FBP indirectly regulates PpsA through the allosteric regulation of its regulating kinase PpsR.

(B) Interactions detected by LiP-SMap grouped based on the chemical nature of the metabolite involved. Numbers on top of the bars indicate the total number of interactions measured. Stacked bars indicate the number of known (bottom) and newly identified (top) MBPs.

(C) Distribution of the 1,678 LiP-SMap interactions binned according to evidence score.

(D) Structure of Fas2 bound to cerulenin (orange); conformotypic peptide is indicated by green ribbons.

(E) Representative MBP complexes with ATP, PEP, and L-Phe with ligand in orange and conformotypic peptides in green. The minimal Euclidean distances (in angstroms) between peptides and metabolite are reported (PDB: 1Q12, 1G7U, 3PCO).

(F) Representative MBP-ligand complexes for GTP and NAD (PDB: 1HK8, 3GR6).

(G) Distributions of minimal Euclidean distances between metabolite and protein atoms for all detected peptides and for all conformotypic peptides of holocomplex structures with only one conformotypic peptide identified.

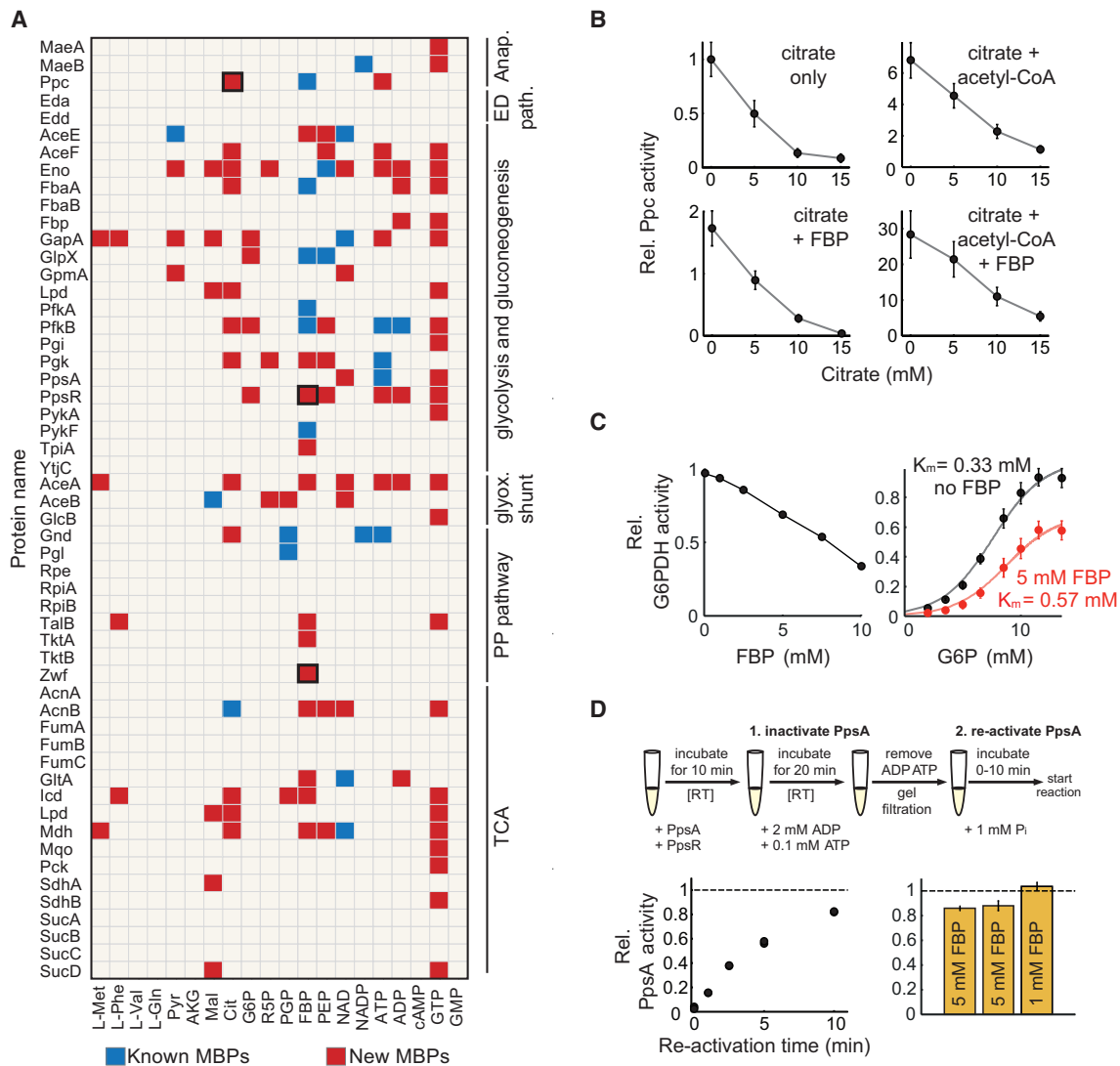See also Figures S2, S3, and S4 and Tables S1, S2, and S3.

Figure 4. Discovery of Novel Regulatory Interactions

(A) Overview of protein-metabolite interactions identified with LiP-SMap within the central carbon metabolism. Blue and red indicate known and novel interactions, respectively. Interactions selected for validation are framed in black. Anap. = anaplerosis.

(B) Activities of purified Ppc as a function of citrate concentration in the absence (upper left) or in the presence of 0.1 mM acetyl-CoA (upper right), 1 mM FBP (bottom-left panel), or both (bottom right) normalized to the maximum activity observed in the absence of the effector. Error bars, SD (n = 4).

(C) G6PDH activity at different FBP concentrations in the presence of 2 mM G6P normalized to the maximum activity observed (left) and relative G6PDH activity at different G6P concentrations in the absence (black) or presence (red) of 5 mM FBP (right). Error bars, SD (n = 4). K_m and V_{max} were estimated by non-linear regression.

(D) Schematic of PpsA activation assay. PpsA was inactivated by treatment with PpsR and ADP and reactivated by adding phosphate in the presence or absence of FBP (upper panel). RT, room temperature. PpsA activity as fraction of the maximal PpsA activity without ADP/ATP (lower left). PpsA activity after 2.5 min. of reactivation in 5 or 1 mM FBP normalized to signal in the absence of FBP (bottom right).

Error bars, SD (n = 4). See also Figure S4 and Table S4.

Novel Interactions at Catalytic Sites

We next asked how many detected interactions occurred at active sites and thus might underlie catalytic events or cases of competitive inhibition. To define an active site, we calculated Euclidean distances between conformotypic peptides and natural reactants at active sites from experimentally derived structural models of *E. coli* enzymes (Table S5). The median distance of this sample, 6.44 ± 0.55 Å, was significantly different from the median value of 11.90 Å for the distribution of the non-conformotypic peptides from the same holocomplexes (two-sided Wilcoxon test, p < 3.47 × 10^{-12}; Figure 5A). Based on these results, a distance of 6.44 Å from the metabolite was used to define the boundaries of an active site measurable with LiP-SMap; binding to a distal
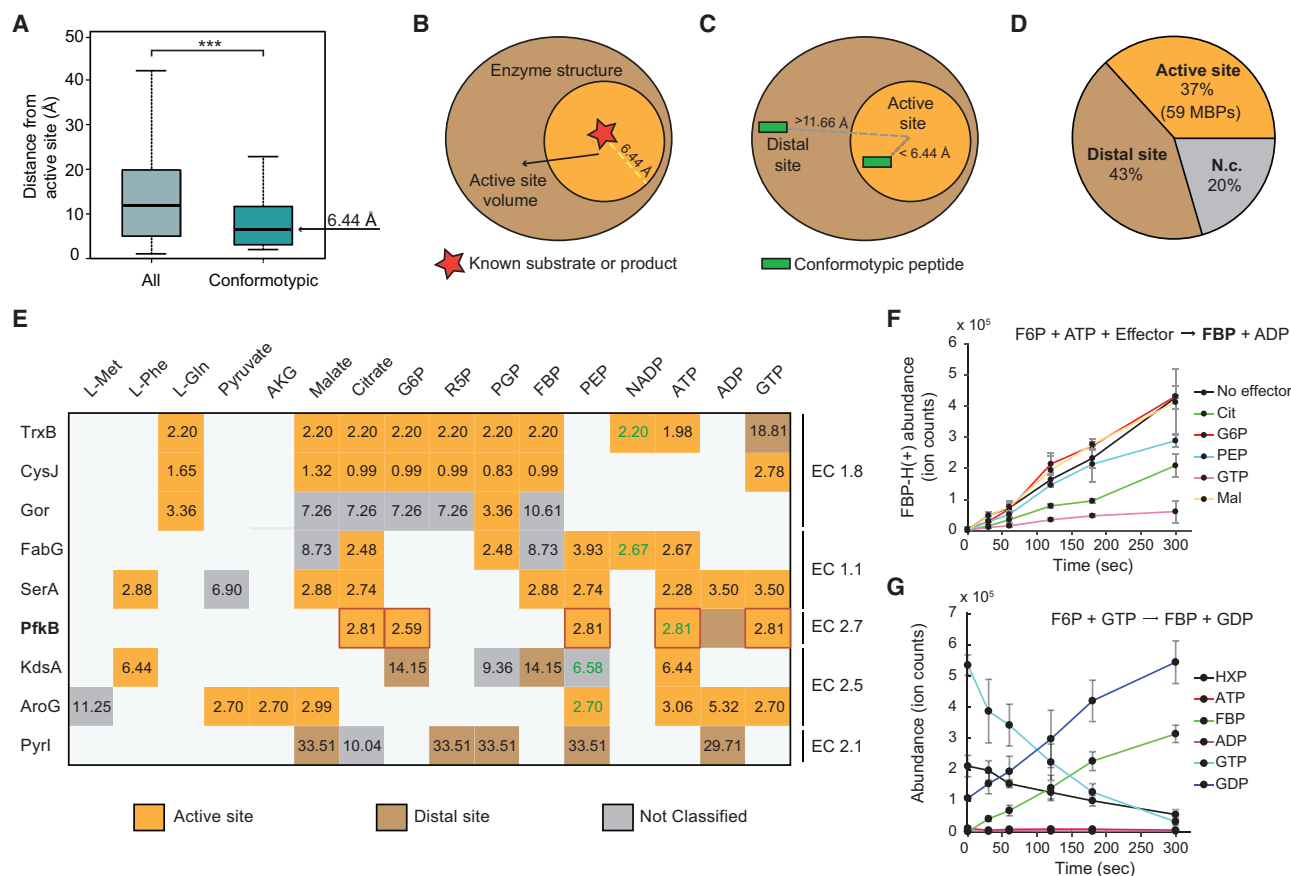
Figure 5. Proteome-wide Map of Metabolite Binding Sites

(A) Distribution of minimal Euclidean distances of all detected peptides and conformotypic peptides from active sites of enzyme-metabolite holocomplex structures.

(B) Operational definition of active site specified by the volume of a sphere of 6.44 ± 0.55-Å radius.

(C) Structural mapping of interactions classified depending on the distance from active site.

(D) Prevalence of LiP-SMap interactions occurring at active or distal sites involving metabolites not previously associated with proteins known to bind a given metabolite. Binding events at distal sites are defined as those with a distance from the active site boundaries of at least 11.66 Å. Non-classified (N.c.) cases are those with distance between 6.44 and 11.66 Å.

(E) Positions of metabolite binding sites of the most promiscuous enzymes. Numbers in the boxes show the measured minimal distances from the active sites. Green labels are relative to known reactant-enzyme pairs. EC are enzyme classification numbers.

(F) Time courses of fructose-1,6-bisphosphate (m/z 338.9884 FBP-H(+)) production by purified PfkB incubated with 2.5 mM fructose-6-phosphate (F6P) and 2.5 mM ATP in the presence of 5 mM potential effectors identified by LiP-SMap citrate (Cit), G6P, PEP, and GTP and in the presence of negative control malate (Mal). Data are means ± SD (n = 3).

(G) Time courses of product formation and substrate consumption in the reaction converting 2.5 mM F6P (m/z 259.0231 HXP -H(+)) and 2.5 mM GTP (m/z 521.9825 GTP -H(+)) into fructose-1,6-bisphosphate (m/z 338.9884 FBP -H(+)) and GDP (m/z 442.0162 GDP -H(+)) catalyzed by purified PfkB.

Data represent mean values ± SD (n = 3). See also Figure S5 and Table S5.

site (Figures 5B, 5C, and S5A) could signify an allosteric site or an additional catalytic site.

Of 1,665 conformotypic peptides mapped to enzymatic holo-complexes, 612 (37%) from 59 MBPs were positioned at active sites known to react with a different compound. Among those, 336 peptides had residues that could theoretically establish hydrogen bonds with the catalytic core, since their distance was less than 3.1 Å (Figures 5D, S5B, and S5C; Table S5). Thus, our data indicate that a substantial fraction of active sites interact with metabolites different from those expected based on the metabolic reactions they catalyzed. Known binding sites of

nicotinamide adenine dinucleotides NADP and NADPH, flavin adenine dinucleotide (FAD), and flavin mononucleotide (FMN) often interacted with compounds of heterogeneous chemistry (Figures S5D and S5E).

Of the most promiscuous 78 MBPs, each interacted with at least six different metabolites (Table S5). For the nine enzymes for which active site structures are available, 41 of 61 conformotypic peptides were within the limits of the active site boundary of 6.44 Å (Figure 5E; Table S5), suggesting that enzyme promiscuity mostly derives from substrate ambiguity. Among the catalytic activities associated with the most promiscuous active sites

were reductase and dehydrogenase activities and reactions that involve transfer of phosphate groups. The regulatory subunit of aspartate carbamoyltransferase (PyrI) was a notable exception to this trend, since none of the newly identified metabolites bound within the catalytic site (Figure 5E; Table S5). These results suggest that enzyme promiscuity is a wide-spread property that mostly derives from binding clefts able to host diverse natural compounds.

To validate our prediction that certain metabolites interact with enzyme active sites leading to novel catalytic events or cases of competitive inhibition, we focused on the enzyme PfkB. Previously unknown interactions with citrate, G6P, PEP, and GTP and the known interaction with ATP were detected by LiP-SMap at the catalytic site of PfkB (Figure 5E). In *in vitro* PfkB activity assays, addition of citrate, PEP, or GTP resulted in a reduction of PfkB activity, suggesting that these three metabolites act as inhibitors of PfkB. No effect was observed for G6P or for malate that was not shown by LiP-SMap to bind PfkB (Figures 5F and S5F).

To determine whether the novel interactions could underlie novel enzyme-substrate relationships, we tested structurally similar compounds glucose-6-phosphate (instead of fructose-6-phosphate) and GTP (instead of ATP) as alternative co-substrates. Glucose-6-phosphate was not used as a substrate by PfkB (Figure S5F). In contrast, PfkB did catalyze the reaction with GTP and fructose-6-phosphate as substrates, and production of GDP and FBP was observed at similar rates as in the control reaction with ATP and fructose-6-phosphate (Figures 5G and S5G). These experiments demonstrate that, as suggested by our initial LiP-SMap analysis, PfkB has a promiscuous active site. Overall, these data indicate that LiP-SMap enables identification of novel enzyme-substrate relationships and cases of competitive inhibition.

High-Order Structural Effects Induced by Metabolite Binding

Binding of metabolites may also regulate formation or dissociation of protein complexes or high-molecular weight protein assemblies (Aughey and Liu, 2015; O'Connell et al., 2012). The latter should be detected in a LiP-SMap experiment as a generalized increase or decrease, respectively, in proteolytic resistance. For a subset of proteins, we detected responses to certain metabolites that were indicative of large structural alterations. We focused on 162 protein-metabolite pairs (involving 69 proteins in total) for which at least 80% of conformotypic peptides indicated increased or decreased resistance to proteolysis upon metabolite binding (Figures S6A and S6B). Of the proteins involved in these interactions, 110 became globally more protease sensitive and 52 more protease resistant. This subset includes 19 nucleotide-binding proteins, 12 oxidoreductases, and six ligases or kinases (Figure 6A; Table S6). Of the 69 proteins that underwent extensive structural rearrangements, 27 are known to form homo-complexes and 42 are subunits of known hetero-complexes (Keseler et al., 2013). To capture additional metabolite-induced alterations to the structure of protein hetero-complexes, we searched for cases in which conformotypic peptides from at least two different subunits of a known hetero-complex consistently indicated increased or decreased resistance to proteolysis upon addition of a metabolite. We found 18 additional cases in which addition of a metabolite to the extract may alter the composition of known protein complexes (Table S6).

We validated our observations using size-exclusion chromatography of cell lysates coupled to MS (SEC-MS) (Kirkwood et al., 2013). In these analyses, the assembly or disassembly of protein complexes is reflected in changes in the elution volume of the proteins involved. We focused on ATP as a test case. Our analysis suggested that 39 proteins underwent high-order structural rearrangements upon ATP addition. For 27 of the 39 proteins, we consistently observed a significant change in elution volume in the presence versus the absence of ATP (Table S6). SEC-MS profiles of non-ATP binding proteins were not affected by ATP addition, and results were independent of the quantification method used (Figures S6C and S6D).

For example, upon addition of ATP we observed a global increase in proteolytic resistance of the protein GalF, which forms a homo-oligomeric complex (Marolda and Valvano, 1996). The SEC-MS profiles of GalF suggest that a transition from monomer to pentamer or hexamer is induced by ATP (Figure 6B). Similarly, the global increase in proteolytic resistance we observed for the subunit IlvH of the acetolactate synthase complex upon ATP addition was reflected by a shift in the SEC-MS profile of IlvH (Figure 6C). This structural rearrangement could indicate formation of previously reported IlvH oligomers or hetero-tetrameric complexes with IlvI (Vyazmensky et al., 1996) (Figure S6E).

We observed massive shifts in the SEC-MS profiles of ribosomal subunits indicative of dissociation of the 50S and 30S complexes in the presence of ATP (Figures 6D and S6F), in line with the global increase in protease sensitivity that we observed by LiP-SMap for subunits of these complexes.

Our data also recapitulated known cases of metabolite-induced formation or dissolution of high-molecular weight protein assemblies. The DNA binding ATPase MinD became more resistant to proteolysis in the presence of ATP and ADP, and FtsZ, which is involved in cell division, became protease resistant in the presence of GTP (Figures 6A and 6E). A change in elution profile was also observed for MinD upon ATP-addition. Interestingly, MinD and FtsZ are components of the division septum that assemble during cytokinesis and have been shown to form ATP- and GTP-dependent filamentous structures, respectively (Mukherjee and Lutkenhaus, 1998; Suefuji et al., 2002). We also measured increased protease sensitivity and a consistent shift in the FtsZ SEC-MS profile for FtsZ in the presence of ATP (Figure S6G), which could imply that GTP and ATP differentially regulate the assembly of FtsZ filaments or oligomeric state.

Overall these data indicate that LiP-SMap detects known and novel cases of metabolite-induced assembly and disassembly of protein complexes and high-molecular weight protein assemblies.

Quantitative Parameters of Metabolite Binding

Last, we asked whether our data could be used to derive quantitative parameters of metabolite binding in the cellular environment. ATP was added to *E. coli* lysates at concentrations ranging
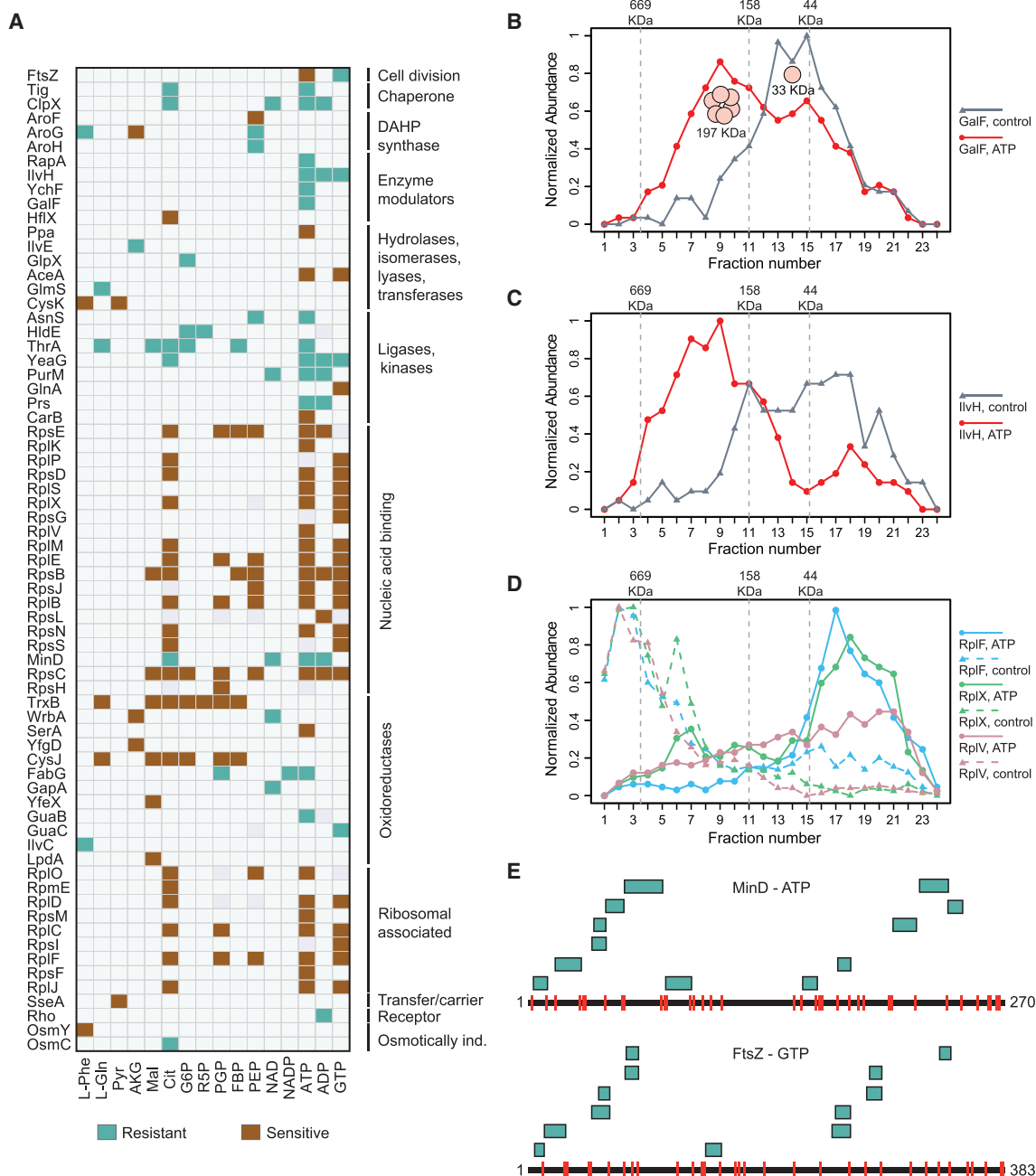
Figure 6. High-Order Structural Changes and Metabolite Binding

(A) MBPs that became globally protease resistant (green) or sensitive (gold) upon addition of metabolites.

(B) SEC-MS elution profiles of GalF when ATP was absent (blue line) or present (red line). Normalized abundance is the sum of peptide counts detected for each SEC fraction relative to the maximum intensity. Monomeric GalF, 33 kDa.

(C) SEC-MS elution profiles of IlvH with (red) or without ATP (blue). Monomeric IlvH, 18 kDa.

(D) SEC-MS elution profiles of RplF, RplX, and RplV subunits of the 50S ribosome complex without (dashed lines) and with (solid lines) ATP.

(E) Abundances of conformotypic peptides for MinD (top) and FtsZ (bottom) are indicative of increased resistance to proteolysis upon binding to ATP and GTP, respectively. Red lines indicate tryptic cleavage sites.

See also Figure S6 and Table S6.

from 1 to 25 mM, and intensities of the 1,051 conformotypic peptides as a function ATP concentration were plotted. In 602 cases, the curves obtained were monotonic and hyperbolic, reminis- cent of binding curves (Figure 7A, see Mendeley data in Data and Software Availability), suggesting that the data could be used to derive measures of affinities in the cell extract.

A

B

C

D

Figure 7. LiP-SMap Determines Metabolite Binding Affinities in Cellular Matrices
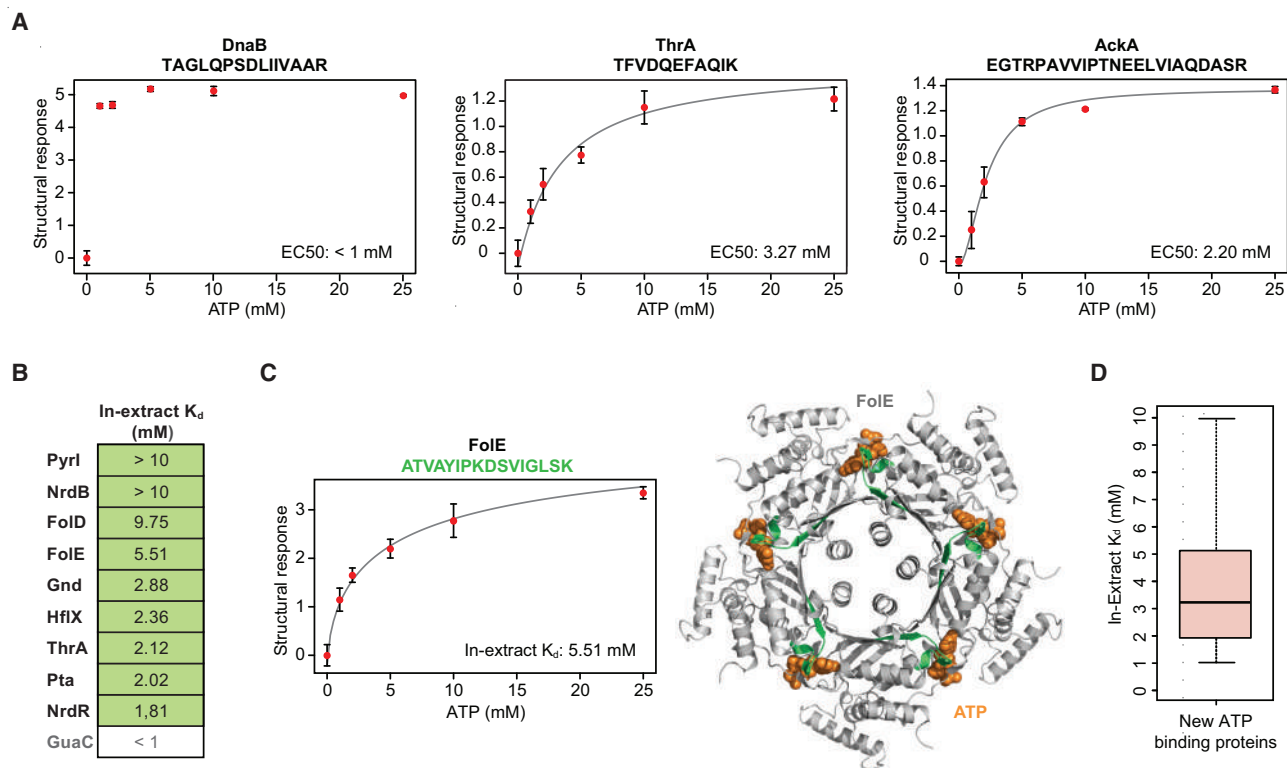
(A) Dose-response curves for three conformotypic peptides from DnaB, ThrA, and AckA. Differences between the $\log_2$ transformed mean intensities of a conformotypic peptides in the treated and untreated conditions were plotted versus metabolite concentration. Error bars, SE (n = 3).

(B) In-extract K_Ds for ATP of proteins for which ATP is an allosteric effector.

(C) Dose-response curve of conformotypic FolE peptide as a function of ATP concentration. Error bars, SE (n = 3). The position of the conformotypic peptide identified is also shown within the structural model of FolE with ATP (PDB: 1A8R).

(D) Distribution of in-extract K_D values extrapolated for newly discovered ATP-binding proteins.

See also Table S7.

ATP binding sites could be divided according to their structural responsiveness to ATP. About 15% of conformotypic peptides showed saturating effects below 1 mM ATP. These peptides mapped to 37 proteins, including 31 enzymes known to use ATP as substrate or to produce ATP, such as the replicative DNA helicase DnaB (Figure 7A, left). The structural responses of 391 peptides could be fit with a sigmoidal dose-response model within the range of 0–25 mM ATP, as exemplified by titration curves of peptides from the ThrA and AckA kinases (Figure 7A, middle and right). Since effects detected by LiP-SMap depend on the degree of occupancy of binding sites, these binding curves are indirect measures of the affinity of the metabolite for a given binding site, allowing extrapolation of the ATP concentration that produced a structural response equivalent to half of the maximum, which we refer to as the in-extract dissociation constant (in-extract K_D) (Table S7).

Binding sites of proteins with in-extract K_D values for ATP in the 1- to 10-mM range would not be saturated at the physiological concentration of ATP of between 1 and 10 mM and may mediate regulatory responses of allosteric nature. To test this hypothesis we analyzed the in-extract K_Ds relative to conformotypic peptides of ATP binding proteins for which

ATP is exclusively an allosteric effector. In nine cases out of 10, in-extract K_Ds were higher than 1 mM and within the physiological concentration range of ATP *in vivo* (Figure 7B). An example is shown for the interaction of ATP with the regulatory sites of the protein GTP cyclohydrolase FolE (Figure 7C). This indicates that in-extract K_Ds could be used to formulate hypotheses on the nature of an ATP-protein interaction. Most (83%) proteins with in-extract K_D values for ATP in the 1- to 10-mM range had not been previously reported to bind ATP, and we discovered 111 new low-affinity ATP binding proteins that have a median in-extract K_D value of 3.23 mM (Figure 7D; Table S7). These experiments show that the structural responses measured with LiP-SMap can be used to determine quantitative parameters of metabolite binding directly in complex biological mixtures.

DISCUSSION

The process of molecular recognition between proteins and metabolites is essential for life. We report here the first systematic analysis of metabolite-protein interactions and binding sites performed directly in the native cellular matrix. The analysis resulted

in a map of the bacterial metabolite-protein interactome, identified a multitude of previously known and novel binding events, and revealed functional principles that control these interactions. Using a combination of biochemical, metabolomics, and proteomics approaches, we validated novel metabolite-protein interactions of allosteric and catalytic nature and interactions that result in high-order structural changes.

Our data suggest that the interaction of FBP with PpsR switches the direction of the glycolytic flux when the intracellular FBP concentration changes. To our knowledge, this is one of the first examples of allosteric regulation of a protein kinase in *E. coli*. Moreover the interaction we validated between G6PDH and FBP provides a potential mechanism to control the flux distribution between glycolysis and pentose phosphate pathway. We also found that citrate is a potent inhibitor of Ppc activity. Interestingly, the intracellular citrate concentration is highest during growth on alternative carbon sources, when the glyoxylate shunt is used for anaplerosis instead of Ppc (Bennett et al., 2009). The novel inhibitory interaction with citrate may explain how Ppc is inactivated when an alternative anaplerotic pathway is used. Further, we found that citrate, PEP, and GTP inhibit activity of PfkB, a member of an important class of kinases in central carbon metabolism, possibly by acting as competitive inhibitors. Moreover, we validated a novel enzyme-substrate relationship, that of GTP with PfkB and identified the reactants and products of the novel reaction. This reveals that PfkB is a promiscuous enzyme that can operate in the absence of ATP.

Although protein complex assembly has been shown to be regulated by exogenous and synthetic small molecules (e.g., rapamycin, Fk506, thalidomide, and auxin) (Fischer et al., 2016), much less is known about effects of endogenous metabolites. Our data show that the alterations in the intracellular concentrations of endogenous metabolites can have profound effects on the assembly of protein complexes and oligomers. Large protein assemblies form upon exposure to specific conditions that alter the composition of the cellular metabolome (Jarosz et al., 2014; Wallace et al., 2015). Our analysis provides a list of proteins and metabolites that might be involved in such events, which can be used for targeted validation experiments.

We observed ribosome dissociation in the presence of ATP. Ribosome dissociation and recycling is known to be ATP dependent *in vivo* in *archaea* (Barthelme et al., 2011). Besides, intracellular ATP and Mg^{2+} levels are tightly regulated (Pontes et al., 2016) and ribosomes may be destabilized due to the chelation of Mg^{2+} by ATP. Other metabolites capable to chelate divalent ions such as citrate may have similar effects on ribosome structures (Figure 6A). Evaluating the likelihood and functional impact of potential chelation events involving metabolites will be an interesting follow up of this study.

Based on our data, about one-quarter of the measured proteome interacted with at least one of the 20 evaluated metabolites, and metabolite-protein interactions impact all known cellular processes, indicating that the size of the metabolite-protein interactome is substantially larger than previously anticipated. We observed, however, that metabolites tend to preferentially bind to those proteins expressed across numerous and varied environmental conditions, the so called "core proteome" (Yang et al., 2015). Thus, a subset of transient metabolite-protein inter-

actions with these stable proteins might dictate the rapid responses of an organism to fluctuating conditions. Genes that are subject to condition-dependent transcriptional regulation are less likely to include MBPs, suggesting that transcriptional control and metabolite-mediated regulation are, at least to a certain extent, independent or even mutually exclusive.

Our approach allowed us to pinpoint putative binding sites at peptide-level resolution, and the data suggest that one-third of enzyme active sites interact with multiple metabolites. This is consistent with genome-scale metabolic models suggesting that one-third of enzymes expressed in *E. coli* and in organisms from the Archaea and Eukarya domains catalyze multiple reactions (Nam et al., 2012). LiP-SMap data indicate that promiscuity is likely due to the capacity of a given binding site to interact with multiple compounds rather than to the coexistence of multiple binding sites on the same protein. Binding promiscuity at metabolite binding sites may reflect active sites that catalyze multiple reactions or competitive inhibition of known substrates. Metabolite binding sites that were not located at active sites (43%) could pinpoint secondary sites of possible allosteric or catalytic nature.

Importantly, LiP-SMap does not require specific reaction chemistries or residue conservation to identify ligand binding domains (Figure S7). This will make the approach particularly useful for the identification of allosteric sites that are generally not conserved (Christopoulos, 2002) and are thus difficult to predict with computational methods. The identification of allosteric sites is in turn particularly relevant for drug discovery, since targeting of these sites increases specificity and modulability relative to targeting of active sites (Nussinov and Tsai, 2013).

We provide an atlas of *in-extract* K_Ds for ATP-binding proteins. These K_Ds reflect physiological events more closely than K_Ds measured from *in vitro* measurements with purified proteins. Our approach identified proteins that can dissociate from ATP within the physiological ATP concentration range and thus are likely to be regulated by this metabolite *in vivo* (Lindsley and Rutter, 2006). The same approach could be used to predict regulatory interactions for other metabolites.

Although powerful in detecting interactions covering a broad range of affinities (from nanomolar to millimolar) based on known interactions, the LiP-SMap approach is not comprehensive. Interactions may be missed due to low MS sequence coverage of the protein involved, which penalizes detection of low-abundance MBPs. This explains why our approach did not identify all known MBPs for the considered metabolites. LiP-SMap requires cell lysis before analysis, which may result in dilution of the cellular medium and loss of compartmentalization. Due to its low degree of subcellular organization, compartmentalization effects should be minimal for *E. coli*, the organism we chose to investigate. Metabolites added to the cell extract could be converted to other compounds by catalytic activities in the extract, resulting in indirect effects. To minimize this issue, we removed endogenous metabolites and cofactors from lysates by gel-filtration, thus minimizing the enzymatic activity of the extract. This removed metabolites with a broad range of chemical properties (Figure S1F). In principle, our approach may be biased against membrane proteins since detergents were not used to ensure protein extraction from membranes prior to LiP. Interestingly, however, 7% of all proteins contained in our dataset (or 33% of the proteins from

our dataset that have an annotated GO subcellular localization) are membrane associated. LiP-SMap could also be applied to total cell extracts containing membrane debris to increase the coverage of membrane proteins.

In summary, LiP-SMap enabled the first systematic mapping of protein-metabolite interactions on a proteome-wide scale and with peptide-level resolution, thus revealing a complex new layer of functional events. Other recently developed chemoproteomics or metabolomics approaches could be applied to the discovery of novel metabolite-protein interactions. These approaches differ from LiP-SMap in that they focus on specific types of interactions (i.e., catalytic) (Sévin et al., 2017), focus on a smaller proteome fraction (Lomenick et al., 2009), require chemical modification of the metabolite or are biased toward particular compound classes (Backus et al., 2016; Niphakis et al., 2015; Parker et al., 2017; Li et al., 2010), or lack peptide-level resolution (Savitski et al., 2014). LiP-SMap is a chemoproteomic approach specifically designed for the discovery of new ligand binding proteins and their binding sites. It does not require any chemical modification of the ligands and is not biased toward compounds with specific properties. We envision that LiP-SMap will set the standard for future studies of protein-ligand interactomes and will be used for drug target deconvolution, as exemplified by the experiment with cerulenin, and for the discovery of orthosteric and allosteric drugs on a cell-wide scale.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
    - *Escherichia coli*
    - *Saccharomyces* cerevisiae
- METHOD DETAILS
    - Whole-proteome preparation for MS analysis
    - Protein concentration determination
    - Preparation of stock solutions of metabolites
    - Quantification of endogenous metabolites with targeted metabolomics
    - Limited proteolysis under native conditions for global analysis of small molecule binding events
    - Proteome preparation in denaturing conditions
    - Recombinant protein production and purification
    - Spectrophotometric assay for Ppc activity
    - Spectrophotometric assay for G6PDH activity
    - Spectrophotometric assay for PpsA-PpsR activity
    - Mass spectrometry-based PfkB activity assays
    - Size exclusion chromatography (SEC) of cell extracts
    - Instrumentation and MS data acquisition for LiP-SMap
    - Peptide identification and quantification of fractionated cell lysates after size-exclusion chromatography
- QUANTIFICATION AND STATISTICAL ANALYSIS
    - Peptide quantification in the LiP-SMap workflow
    - Known protein-metabolite interactions
    - Quality assessment of LiP-SMap results

- Score system for assessment of LiP-SMap results
- Genetic interactions analysis
- Metabolite binding protein analysis
- Volcano plots of candidate MBPs
- Boxplots
- Receiver operating characteristic (ROC) curve
- Evaluation of structural chemical similarity of metabolites
- Structures of protein-metabolite holocomplexes
- Measurements of Euclidean distances
- Secondary structure prediction analysis
- Analysis of the *E. coli* reactome
- Proteins undergoing extensive structural changes
- Metabolite-induced changes in protease resistance
- Isothermal dose-response experiments
- Statistical analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at https://doi.org/10.1016/j.cell.2017.12.006.

AUTHOR CONTRIBUTIONS

I.P. designed and performed the LiP-SMap screen and analyzed all proteomic experiments; K.K. designed in vitro biochemical assays and analyzed data; T.F. designed and performed metabolomics experiments. I.P., V.C., E.N., and K.K. performed bioinformatic analysis of the data. I.P and P.P. wrote the manuscript with contributions from all authors. U.S. supervised part of the project. P.P. conceived and supervised the project.

DECLARATION OF INTERESTS

P.P. is a scientific advisor for the company Biognosys AG (Zurich, Switzerland) and an inventor of a patent licensed by Biognosys AG that covers the LiP-MS method used in this manuscript.

REFERENCES

Aughey, G.N., and Liu, J.-L. (2015). Metabolic regulation via enzyme filamentation. Crit. Rev. Biochem. Mol. Biol. *51*, 282–293.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., and Mori, H. (2006). Construction of

Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. Mol. Syst. Biol. *2*, 2006.0008.

Backus, K.M., Correia, B.E., Lum, K.M., Forli, S., Horning, B.D., González-Páez, G.E., Chatterjee, S., Lanning, B.R., Teijaro, J.R., Olson, A.J., et al. (2016). Proteome-wide covalent ligand discovery in native biological systems. Nature *534*, 570–574.

Barthelme, D., Dinkelaker, S., Albers, S.-V., Londei, P., Ermler, U., and Tampé, R. (2011). Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. Proc Natl Acad Sci U S A *108*, 3228–3233.

Bennett, B.D., Kimball, E.H., Gao, M., Osterhout, R., Van Dien, S.J., and Rabinowitz, J.D. (2009). Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. Nat. Chem. Biol. *5*, 593–599.

Boehr, D.D., Nussinov, R., and Wright, P.E. (2009). The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. *5*, 789–796.

Brunk, E., Mih, N., Monk, J., Zhang, Z., O'Brien, E.J., Bliven, S.E., Chen, K., Chang, R.L., Bourne, P.E., and Palsson, B.O. (2016). Systems biology of the structural proteome. BMC Syst. Biol. *10*, 26.

Burnell, J.N. (2010). Cloning and characterization of Escherichia coli DUF299: A bifunctional ADP-dependent kinase–Pi-dependent pyrophosphorylase from bacteria. BMC Biochem. *11*, 1.

Castello, A., Fischer, B., Eichelbaum, K., Horos, R., Beckmann, B.M., Strein, C., Davey, N.E., Humphreys, D.T., Preiss, T., Steinmetz, L.M., et al. (2012). Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. Cell *149*, 1393–1406.

Changeux, J.-P., and Christopoulos, A. (2016). Allosteric modulation as a unifying mechanism for receptor function and regulation. Cell *166*, 1084–1102.

Christopoulos, A. (2002). Allosteric binding sites on cell-surface receptors: novel targets for drug discovery. Nat. Rev. Drug Discov. *1*, 198–210.

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

Diether, M., and Sauer, U. (2017). Towards detecting regulatory protein-metabolite interactions. Curr. Opin. Microbiol. *39*, 16–23.

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.

Fischer, E.S., Park, E., Eck, M.J., and Thomä, N.H. (2016). SPLINTS: Small-molecule protein ligand interface stabilizers. Curr. Opin. Struct. Biol. *37*, 115–122.

Fuhrer, T., Zampieri, M., Sévin, D.C., Sauer, U., and Zamboni, N. (2017). Genomewide landscape of gene-metabolome associations in Escherichia coli. Mol. Syst. Biol. *13*, 907.

Gallego, O., Betts, M.J., Gvozdenovic-Jeremic, J., Maeda, K., Matetzki, C., Aguilar-Gurrieri, C., Beltran-Alvarez, P., Bonn, S., Fernández-Tornero, C., Jensen, L.J., et al. (2010). A systematic screen for protein-lipid interactions in Saccharomyces cerevisiae. Mol. Syst. Biol. *6*, 430.

Gerosa, L., and Sauer, U. (2011). Regulation and control of metabolic fluxes in microbes. Curr. Opin. Biotechnol. *22*, 566–575.

Gerosa, L., Haverkorn van Rijsewijk, B.R.B., Christodoulou, D., Kochanowski, K., Schmidt, T.S.B., 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.

Gillet, L.C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 11, O111.016717.

Gillet, L.C., Leitner, A., and Aebersold, R. (2016). Mass spectrometry applied to bottom-up proteomics: Entering the high-throughput era for hypothesis testing. Annu. Rev. Anal. Chem. (Palo Alto, Calif.) *9*, 449–472.

Hein, M.Y., Hubner, N.C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I.A., Weisswange, I., Mansfeld, J., Buchholz, F., et al. (2015). A human interactome

in three quantitative dimensions organized by stoichiometries and abundances. Cell *163*, 712–723.

Jarosz, D.F., Brown, J.C.S., Walker, G.A., Datta, M.S., Ung, W.L., Lancaster, A.K., Rotem, A., Chang, A., Newby, G.A., Weitz, D.A., et al. (2014). Cross-kingdom chemical communication drives a heritable, mutually beneficial prion-based transformation of metabolism. Cell *158*, 1083–1093.

Keseler, I.M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martínez, C., Fulcher, C., Huerta, A.M., Kothari, A., Krummenacker, M., et al. (2013). EcoCyc: Fusing model organism databases with systems biology. Nucleic Acids Res. *41*, D605–D612.

Kirkwood, K.J., Ahmad, Y., Larance, M., and Lamond, A.I. (2013). Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. Mol. Cell. Proteomics *12*, 3851–3873.

Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and Mori, H. (2005). Complete set of ORF clones of Escherichia coli ASKA library (a complete set of E. coli K-12 ORF archive): unique resources for biological research. DNA Res. *12*, 291–299.

Li, X., Gianoulis, T.A., Yip, K.Y., Gerstein, M., and Snyder, M. (2010). Extensive in vivo metabolite-protein interactions revealed by large-scale systematic analyses. Cell *143*, 639–650.

Lindsley, J.E., and Rutter, J. (2006). Whence cometh the allosterome? Proc. Natl. Acad. Sci. USA *103*, 10533–10535.

Lomenick, B., Hao, R., Jonai, N., Chin, R.M., Aghajan, M., Warburton, S., Wang, J., Wu, R.P., Gomez, F., Loo, J.A., et al. (2009). Target identification using drug affinity responsive target stability (DARTS). Proc. Natl. Acad. Sci. USA *106*, 21984–21989.

Marolda, C.L., and Valvano, M.A. (1996). The GalF protein of Escherichia coli is not a UDP-glucose pyrophosphorylase but interacts with the GalU protein possibly to regulate cellular levels of UDP-glucose. Mol. Microbiol. *22*, 827–840.

Milo, R. (2013). What is the total number of protein molecules per cell volume? A call to rethink some published values. BioEssays *35*, 1050–1055.

Milroy, L.-G., Grossmann, T.N., Hennig, S., Brunsveld, L., and Ottmann, C. (2014). Modulators of protein-protein interactions. Chem. Rev. *114*, 4695–4748.

Morikawa, M., Izui, K., Taguchi, M., and Katsuki, H. (1980). Regulation of Escherichia coli phosphoenolpyruvate carboxylase by multiple effectors in vivo. Estimation of the activities in the cells grown on various compounds. J. Biochem. *87*, 441–449.

Motlagh, H.N., Wrabl, J.O., Li, J., and Hilser, V.J. (2014). The ensemble nature of allostery. Nature *508*, 331–339.

Mukherjee, A., and Lutkenhaus, J. (1998). Dynamic assembly of FtsZ regulated by GTP hydrolysis. EMBO J. *17*, 462–469.

Nam, H., Lewis, N.E., Lerman, J.A., Lee, D.-H., Chang, R.L., Kim, D., and Palsson, B.O. (2012). Network context and selection in the evolution to enzyme specificity. Science *337*, 1101–1104.

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., and Cravatt, B.F. (2015). A global map of lipid-binding proteins and their ligandability in cells. Cell *161*, 1668–1680.

Nussinov, R., and Tsai, C.-J. (2013). Allostery in disease and in drug discovery. Cell *153*, 293–305.

O'Connell, J.D., Zhao, A., Ellington, A.D., and Marcotte, E.M. (2012). Dynamic reorganization of metabolic enzymes into intracellular bodies. Annu. Rev. Cell Dev. Biol. *28*, 89–111.

Orth, J.D., Conrad, T.M., Na, J., Lerman, J.A., Nam, H., Feist, A.M., and Palsson, B.O. (2011). A comprehensive genome-scale reconstruction of Escherichia coli metabolism—2011. Mol. Syst. Biol. *7*, 535.

Parker, C.G., Galmozzi, A., Wang, Y., Correia, B.E., Sasaki, K., Joslyn, C.M., Kim, A.S., Cavallaro, C.L., Lawrence, R.M., Johnson, S.R., et al. (2017). Ligand and target discovery by fragment-based screening in human cells. Cell *168*, 527–541.

Pontes, M.H., Yeom, J., and Groisman, E.A. (2016). Reducing ribosome biosynthesis promotes translation during low Mg2+ stress. Mol. Cell *64*, 480–492.

Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D., and Noor, E. (2017). Genome-scale architecture of small molecule regulatory networks and the fundamental trade-off between regulation and enzymatic activity. Cell Rep. *20*, 2666–2677.

Sanwal, B.D., and Maeba, P. (1966). Regulation of the activity of phosphoenolypyruvate carboxylase by fructose diphosphate. Biochem. Biophys. Res. Commun. *22*, 194–199.

Savitski, M.M., Reinhard, F.B.M., 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., Ahrné, 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.

Sévin, D.C., Fuhrer, T., Zamboni, N., and Sauer, U. (2017). Nontargeted in vitro metabolomics for high-throughput identification of novel enzymes in Escherichia coli. Nat. Methods *14*, 187–194.

Suefuji, K., Valluzzi, R., and RayChaudhuri, D. (2002). Dynamic assembly of MinD into filament bundles modulated by ATP, phospholipids, and MinE. Proc. Natl. Acad. Sci. USA *99*, 16776–16781.

Vyazmensky, M., Sella, C., Barak, Z., and Chipman, D.M. (1996). Isolation and characterization of subunits of acetohydroxy acid synthase isozyme III and reconstitution of the holoenzyme. Biochemistry *35*, 10339–10346.

Wallace, E.W.J., 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.

Wu, H. (2013). Higher-order assemblies in a new paradigm of signal transduction. Cell *153*, 287–292.

Xu, Y.-F., Amador-Noguez, D., Reaves, M.L., Feng, X.-J., and Rabinowitz, J.D. (2012). Ultrasensitive regulation of anapleurosis via allosteric activation of PEP carboxylase. Nat. Chem. Biol. *8*, 562–568.

Yang, L., Tan, J., O'Brien, E.J., Monk, J.M., Kim, D., Li, H.J., Charusanti, P., Ebrahim, A., Lloyd, C.J., Yurkovich, J.T., et al. (2015). Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. Proc. Natl. Acad. Sci. USA *112*, 10810–10815.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
TCEP (tris(2-carboxyethyl)phosphine hydrochloride)	Pierce	Cat#20490; CAS#51805-45-9
Iodoacetamide	Sigma-Aldrich	Cat#I1149; CAS#144-48-9
Ammonium bicarbonate	Sigma-Aldrich	Cat#09830; CAS#1066-33-7
Formic acid 98-100%	AppliChem	Cat#A38580500
HEPES (4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid, N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid)	Sigma-Aldrich	Cat#H4034
Potassium chloride	Merck	Cat#K41042236-032; CAS#64-18-6
Magnesium chloride hexahydrate	Fluka	Cat#63072; CAS#7791-18-6
Sodium deoxycholate	Sigma-Aldrich	Cat#D6750; CAS #302-95-4
Phenylmethane sulfonyl fluoride	AppliChem	Cat#A0999-0025
DL-Dithiothreitol	Sigma-Aldrich	Cat#D0632; CAS#3483-12-3
Trizma-base	Sigma-Aldrich	Cat#T1503; CAS# 77-86-1
β-Nicotinamide adenine dinucleotide, reduced disodium salt hydrate	Sigma-Aldrich	Cat#N8129; CAS# 606-68-8 (anhydrous)
β-Nicotinamide adenine dinucleotide phosphate hydrate	Sigma-Aldrich	Cat#N5755; CAS# 53-59-8 (anhydrous)
Adenosine 5′-triphosphate disodium salt	Sigma-Aldrich	Cat#A2383; CAS# 34369-07-8
D-Fructose 6-phosphate disodium salt hydrate	Sigma-Aldrich	Cat#F3627; CAS# 26177-86-6 (anhydrous)
L-(-)-Malic acid sodium salt	Sigma-Aldrich	Cat#M1125; CAS# 68303-40-2
L-Phenylalanine	Sigma-Aldrich	Cat#78020; CAS#63-91-2
Phospho(enol)pyruvic acid monopotassium salt	Sigma-Aldrich	Cat#860077; CAS#4265-07-0
Sodium pyruvate	Sigma-Aldrich	Cat#P2256; CAS# 113-24-6
L-Glutamic acid monosodium salt monohydrate	Sigma-Aldrich	Cat#49621; CAS#6106-04-3
Citrate monohydrate	Merck	Cat#K91547044 109
L-Valine	Sigma-Aldrich	Cat#94619; CAS#72-18-4
D-Glucose 6-phosphate disodium salt hydrate	Sigma-Aldrich	Cat#G-7250; CAS# 3671-99-6
α-Ketoglutaric acid	Sigma-Aldrich	Cat#75890; CAS#328-50-7
L-Methionine	Sigma-Aldrich	Cat#M9625; CAS# 63-68-3
D-Ribose 5-phosphate disodium salt dihydrate	Sigma-Aldrich	Cat#83875; CAS#207671-46-3
6-Phosphogluconic acid trisodium salt	Sigma-Aldrich	Cat#P7877; CAS# 53411-70-4
β-Nicotinamide adenine dinucleotide hydrate	Sigma-Aldrich	Cat#N1511; CAS# 53-84-9
β-Nicotinamide adenine dinucleotide phosphate sodium salt hydrate	Sigma-Aldrich	Cat#N0505; CAS# 698999-85-8
Guanosine 5′-triphosphate sodium salt hydrate	Sigma-Aldrich	Cat#G8877; CAS# 36051-31-7
Adenosine 5′-diphosphate sodium salt	Sigma-Aldrich	Cat#A2754; CAS# 20398-34-9
Guanosine 5′-monophosphate disodium salt hydrate	Sigma-Aldrich	Cat#G8377; CAS# 5550-12-9
Adenosine 3′,5′-cyclic monophosphate sodium salt monohydrate	Sigma-Aldrich	Cat#A6885; CAS# 37839-81-9
Proteinase K (PK) from Engyodontium album	Sigma Aldrich	Cat#P2308
Lysyl endopeptidase	Wako Pure Chemical Industries	Cat#125-05061
Trypsin: sequencing-grade modified trypsin	Promega	Cat#V5111

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Heavy-labeled unpurified synthetic peptides	ThermoFisher Scientific; This paper	Table S7
HRM calibration kit	Biognosys AG	Cat#Ki-3003
Malate dehydrogenase from porcine heart	Sigma-Aldrich	Cat#M1567
Cerulenin ((2R,3S,E,E)-2,3-Epoxy-4-oxo-7,10-dodecadienamide)	BASF SE	CAS#17397-89-6
Critical Commercial Assays		
BCA protein assay	Pierce	Cat: 23228
Deposited Data		
Raw and analyzed data	This paper	PRIDE: PXD006543
Differential SEC-MS elution profiles upon ATP binding of ATP binding proteins undergoing large structural changes	This paper; and Mendeley Data	https://doi.org/10.17632/nhsktkcs3d.1
Differential SEC-MS elution profiles upon ATP binding of proteins that do not bind ATP	This paper; and Mendeley Data	https://doi.org/10.17632/nhsktkcs3d.1
Concentration dependent structural response curves relative to single conformotypic peptides	This paper; and Mendeley Data	https://doi.org/10.17632/nhsktkcs3d.1
Experimental Models: Organisms/Strains		
E. coli: Strain background BW25113	Baba et. al., 2006	CGSC#7636
E. coli: 6xHis-tagged G6PDH from ASKA collection	Kitagawa et. al., 2005	EcoCyc: EG11221
E. coli: 6xHis-tagged Ppc from ASKA collection	Kitagawa et. al., 2005	EcoCyc: EG10756
E. coli: 6xHis-tagged PpsA from ASKA collection	Kitagawa et. al., 2005	EcoCyc: EG10756
E. coli: 6xHis-tagged PpsR from ASKA collection	Kitagawa et. al., 2005	EcoCyc: EG11132
E. coli: 6xHis-tagged PfkB from ASKA collection	Kitagawa et. al., 2005	EcoCyc: EG10700
S. cerevisiae: Strain background BY4742: S288C isogenic yeast strain. Genotype: MATα; his3Δ1; leu2Δ0; lys2Δ0; ura3Δ0	Euroscarf	http://www.euroscarf.de/search.php?name=Order
Software and Algorithms		
Rstudio	Rstudio	https://www.rstudio.com
R version v. 3.3.1	The R Foundation	https://www.r-project.org/
Drc package for R	Christian Ritz	CRAN: *Drc.* (https://cran.r-project.org/web/packages/drc/index.html)
Python version v. 2.7	Python Software Foundation	https://www.python.org
Seaborn library for python v. 0.71	Michael Waskom	https://seaborn.pydata.org/
MATLAB v. R2014A	MathWorks	https://ch.mathworks.com
MacPyMol v.1.5.	Schrödinger	https://pymol.org/2/
GraphPad Prism v. 7.02	GraphPad Software	https://www.graphpad.com/
Proteome discoverer v. 1.4	ThermoFisher Scientific	https://www.thermofisher.com/us/en/home.html
Xcalibur v. 3.1	ThermoFisher Scientific	https://www.thermofisher.com/us/en/home.html
Proteome discoverer v. 2.0	ThermoFisher Scientific	https://www.thermofisher.com/us/en/home.html
Analyst v. 1.6.2	AB Sciex	https://sciex.com/
Spectronaut v. 8.0	Biognosys AG	https://biognosys.com/
Other		
Freezer Mill, Sample prep 6875	SPEX	n/a
Sep-Pak Vac, tC18 Cartridges	Waters	Cat: WAT054960
ZebaTM Spin Desalting, 7MWCO	ThermoFisher Scientific	Cat: 89882, 89889
His GraviTrap TALON	GE Healthcare	Cat: 29-0005-94
Orbitrap Q Exactive Plus mass spectrometer	ThermoFisher Scientific	https://www.thermofisher.com/us/en/home.html
5500 QTRAP Triple quadrupole ion trap mass spectrometer	AB Sciex	https://sciex.com/

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed to and will be fulfilled by Lead Contact Paola Picotti (picotti@imsb.biol.ethz.ch).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Escherichia coli

E. coli strain BW25113 (Baba et al., 2006) was cultivated in M9 minimal medium with 2 g/L of glucose at 37°C, 800 rpm. A single colony picked from a fresh plate was used to inoculate Luria-Bertani (LB) complex medium. After 6 hours of incubation at 37°C and constant shaking, LB cultures were diluted 1:100 (v/v) and used to inoculate M9-glucose medium pre-cultures for overnight cultivation. A minimum of three independent final cultures were inoculated in M9-glucose media at a starting OD_{600} of 0.05 and harvested when the OD_{600} reached 0.8. Cells were harvested by centrifugation and carefully washed three times with ice-cold lysis buffer (100 mM HEPES pH 7.5, 150 mM KCl, 1 mM $MgCl_2$). Cell pellets were resuspended in lysis buffer, and cell suspensions were extruded from a gauge needle to produce drops that were immediately flash frozen in liquid nitrogen.

Saccharomyces cerevisiae

Yeast BY4742 cells were grown at 30°C in YPD media to early log phase from a single colony picked from a fresh YPD plate. Cells were harvested by centrifugation and carefully washed three times with ice-cold lysis buffer (100 mM HEPES pH 7.5, 150 mM KCl, 1 mM $MgCl_2$). Cell pellets were resuspended in lysis buffer, and cell suspensions were extruded from a gauge needle to produce drops that were immediately flash frozen in liquid nitrogen.

METHOD DETAILS

Whole-proteome preparation for MS analysis

Liquid-nitrogen frozen beads of cell suspensions in lysis buffer (100 mM HEPES pH 7.5, 150 mM KCl, 1 mM $MgCl_2$) were mechanically ground in cryogenic conditions with a Freezer Mill (SPEX SamplePrep 6875). Cell debris was removed by centrifugation (5 min, 20000 g, 4°C). Endogenous metabolites and nucleic acids were removed by size-exclusion chromatography (Zeba Spin Desalting Columns 7 MWCO, ThermoFisher Scientific). The sample preparation procedure was performed at 4°C. For proteomics experiments, cell lysis was performed in parallel for a minimum of three biological replicates equivalent to three independent cell cultures. In the LiP-SMap experiment with cerulenin endogenous ligands were not depleted by size-exclusion chromatography.

Protein concentration determination

Protein concentrations in whole proteome samples were determined with the bicinchoninic acid assay (Pierce BCA Protein Assay kit, ThermoFisher Scientific).

Preparation of stock solutions of metabolites

All metabolite solutions were prepared from ultra-pure powders in 100 mM HEPES pH 7.5. After solubilization, pH was measured with a pH meter with micro-electrode and double-checked with pH strips. If necessary, pH was adjusted to pH 7.5 with 12.5 M NaOH. Stock solutions were frozen at −20°C and used for the limited proteolysis or biochemical experiments within a week.

Quantification of endogenous metabolites with targeted metabolomics

Aliquots of cell extracts sampled at different time points were mixed with quenching solution (40:40:20 methanol:acetonitrile:H_2O) containing ^{13}C internal standards. Samples were stored at −20°C, dried, and resuspended in H_2O. Metabolite reactant concentrations were subsequently measured by LC-MS/MS.

Limited proteolysis under native conditions for global analysis of small molecule binding events

Cell lysates from independent biological replicates were aliquoted in equivalent volumes containing 100 μg of proteome sample and incubated for 10 min at 25°C with metabolite of interest. Proteinase K from *Tritirachium album* (Sigma Aldrich) was added simultaneously to all the proteome-metabolite samples at a proteinase K to substrate mass ratio of 1:100 and incubated at 25°C for 5 min. Digestion reactions were stopped by heating samples for 3 min at 98°C in a thermocycler followed by addition of sodium deoxycholate (Sigma Aldrich) to a final concentration of 5%. Samples were then heated again at 98°C for 3 min in a thermocycler. These samples were then subjected to complete digestion in denaturing conditions as described below.

Proteome preparation in denaturing conditions

Protein fragments from the limited proteolysis step were reduced with 5 mM Tris(2-carboxyethyl)phosphine (ThermoFisher Scientific) for 40 min. at 37°C and then alkylated by incubating 30 min at 25°C with 20 mM iodoacetamide (Sigma Aldrich) in the dark. Samples were diluted with 0.1 M ammonium bicarbonate to a final concentration of 1% sodium deoxycholate, and pre-digested with lysyl

endopeptidase (Wako Chemicals) at an enzyme substrate ratio of 1:100 for 4 h at 37°C. Digestions were completed by treatment with sequencing-grade porcine trypsin (Promega) at an enzyme substrate ratio of 1:100 for 16 h at 37°C. Trypsin was inactivated in order to stop peptide digestions by adding a volume of formic acid that lowered the pH to less than 2. Acidified peptide mixtures were loaded onto Sep-Pak tC18 cartridges or into 96-well elution plates (Waters), desalted, and eluted with 50% acetonitrile-0.1% formic acid. Samples were dried in a vacuum centrifuge, solubilized in 0.1% formic acid, and immediately analyzed by mass spectrometry.

Recombinant protein production and purification

6xHis-tagged Ppc and 6xHis-tagged G6PDH were obtained from the ASKA collection (Kitagawa et al., 2005) and purified as described previously (Fuhrer et al., 2017). Briefly, 180 mL LB cultures containing 100 μM IPTG and 20 μg/mL chloramphenicol were inoculated with an aliquot of an LB overnight culture diluted 1:100 and were grown at 37°C with shaking overnight. Cells were harvested by centrifugation and then lysed by three passages through a French pressure cell at 1000 psi (Thermo Fisher Scientific) in 50 mM potassium-phosphate buffer, pH 7.5 with 10 mM $MgCl_2$, 2 mM dithiothreitol, and 4 mM phenylmethylsulfonyl fluoride. Lysates were centrifuged for 30 min at 4°C (14000 g), and purified using GraviTrap TALON Co^{2+} columns. After immobilized metal affinity chromatography (IMAC) purification, protein extracts were re-buffered four times with enzyme assay buffer using filter columns with 10-kD cut-off. The enzyme assay buffers were 100 mM Tris-HCl pH 8.0 with 10 mM $MgCl_2$ for Ppc and 50 mM potassium-phosphate buffer pH 7.5 with 10 mM $MgCl_2$ for G6PDH recombinant proteins.

6xHis-tagged PpsA and PpsR overexpression strains were obtained from the ASKA collection (Kitagawa et al., 2005) and purified following previously reported procedures (Burnell, 2010). After IMAC purification, protein extracts were re-buffered four times with PpsR enzyme assay buffer (50 mM Tris-HCl, pH 8, 10 mM $MgCl_2$, 100 mM NaCl) using filter columns with 10-kD cut-off. Cell lysis, protein purification, and re-buffering were performed in presence of 20% (v/v) glycerol to avoid protein aggregation, as described previously (Burnell, 2010). The strain overexpressing 6xHis-tagged PfkB was obtained from the ASKA collection (Kitagawa et al., 2005) and purified following previously reported procedures (Fuhrer et al., 2017; Sévin et al., 2017). Briefly, cells were pelleted by centrifugation (10 min at 4000 g), resuspended, and lysed in 20 mL of B-PER Reagent (Thermo Scientific) with 1 mM dithiothreitol, DNase I, and lysozyme at room temperature. Debris-free cell lysates were complemented with 20 mM imidazole, loaded onto His GraviTrap Talon columns (1 mL column volume, GE Healthcare), and washed with 10 column volumes of washing buffer (20 mM sodium phosphate, 500 mM NaCl, 20 mM imidazole, pH 7.5). Pure protein was eluted with 20 mM sodium phosphate, 500 mM NaCl, 500 mM imidazole, pH 7.5. Buffer was exchanged to 100 mM HEPES, 1 mM $MgCl_2$, pH 7.5 by three ultrafiltration steps using spin columns with 10-kDa molecular weight cut-off (Millipore), and samples were stored at 4°C.

Spectrophotometric assay for Ppc activity

In vitro activity of 6xHis-tagged purified Ppc was quantified photometrically as previously described (Morikawa et al., 1980). Briefly, Ppc activity was measured as decrease in absorbance at 340 nm due to depletion of NADH in an assay coupling Ppc and malate dehydrogenase: PEP was converted to oxaloacetate by Ppc, and oxaloacetate in turn was consumed by malate dehydrogenase using NADH as a co-factor. Thus, consumption of NADH was stoichiometrically coupled to the formation of oxaloacetate catalyzed by Ppc. Assays were performed in 96-well format in 200 μL reaction volume at 29°C in 100 mM Tris-HCl, pH 8.0; 10 mM $MgCl_2$, 10 mM $NaHCO_3$, 0.2 mM NADH, 1 U malate dehydrogenase. Reagent mixtures were freshly prepared in assay buffer (100 mM Tris-HCl, pH 8.0, 10 mM $MgCl_2$) prior to each experiment to minimize enzyme inactivation. Reactions were initiated by substrate addition, and depletion of NADH was monitored photometrically at 340 nm every 10 s for 30 min. Ppc activity reported in Figure 4B was measured as the variation in absorbance at 340 nm due to changes in NADH concentration normalized to the maximum activity observed. Reaction rates were determined after curve fitting of time course measurements by linear regression. Malate dehydrogenase activity was not affected by citrate (data not shown). The pH of metabolite stock solutions was adjusted to pH 8 for this assay.

Spectrophotometric assay for G6PDH activity

In vitro activity of 6xHis-tagged purified G6PDH was quantified photometrically by measuring the formation of NADPH at 340 nm. Assays were performed in 96-well format in 200 μL reaction volume at 29°C in 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM NADP+, 0.2 μg/mL G6PDH and 2 mM G6P unless stated otherwise. Recombinant enzyme samples were freshly prepared in 100 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ prior to each experiment to minimize inactivation of the enzyme. Reactions were initialized by G6P addition, and formation of NADPH was monitored photometrically every 10 s for 180 s. G6PDH activity reported in Figure 4C was measured as an increase in absorbance at 340 nm due to formation of NADPH normalized to the maximum activity observed. Reaction rates were fitted from time course measurements by linear regression.

Spectrophotometric assay for PpsA-PpsR activity

In vitro 6xHis-tagged purified PpsA enzyme activity was assayed photometrically at 340 nm in 96-well plates using a continuous coupling assay with PEP carboxylase and malate dehydrogenase, as previously described (Burnell, 2010). Briefly, PEP formed by PpsA was first converted to oxaloacetate by PEP carboxylase, and oxaloacetate in turn was consumed by malate dehydrogenase using NADH as a co-factor. Consequently, consumption of NADH was stoichiometrically coupled to the formation of PEP. Reactions were conducted in a 200 μL reaction volume at 29°C in 50 mM Tris-HCl, pH 8; 10 mM $MgCl_2$; 0.2 mM NADH; 1 mM ATP; 1 mM pyruvate; 1 U PEP carboxylase; 1 U malate dehydrogenase; 10 mM $NaHCO_3$; 1 mM G6P.

The Pi-dependent activation of PpsA by PpsR was quantified as follows: first PpsA was completely inactivated under the following inactivation assay conditions: 50 mM Tris-HCl, pH 8, 10 mM MgCl$_2$, 0.1 mM ATP, 2 mM ADP, approximately 5:1 ratio of PpsA and PpsR and incubation for 20 min at room temperature. Any remaining ADP or ATP was removed by size-exclusion chromatography (Zeba Spin Desalting Columns 7 MWCO, ThermoFisher Scientific), and then PpsA was re-activated by adding Pi. Samples of this activation assay were taken and tested for PpsA activity as described above. PpsA activity reported in Figure 4D was measured as the variation in absorbance at 340 nm due to changes in NADH concentration normalized to the maximum activity observed. Within 10 min from the initiation of the reaction approximately 80% of PpsA activity was restored.

Mass spectrometry-based PfkB activity assays

Purified PfkB was incubated at 37°C at a protein concentration of 50 μg/mL in 200 μL buffer containing 100 mM HEPES pH 7.5, 1 mM MgCl$_2$, 2.5 mM of each substrate, and 5 mM predicted effector. At indicated time points, a 10 μL aliquot of the reaction solution was transferred to 140 μL methanol:water (6:1 v/v) pre-cooled on dry ice to quench the reaction by inducing enzyme denaturation. Reactant concentrations were subsequently measured by time-of-flight mass spectrometry. Each experiment was repeated with at least two independent enzyme purifications in experimental triplicates. Negatively charged ions were tentatively annotated based on accurate mass using 0.001 Da tolerance assuming simple deprotonation ([M-H]-).

Size exclusion chromatography (SEC) of cell extracts

Aliquots of fresh, concentrated lysates incubated with a metabolite of interest or vehicle were separated with a Superdex S200 10/300 GL size exclusion column on an AEKTA pure system (GE Healthcare) running at 0.5 mL/min in 100 mM HEPES, 150 mM KCl$_2$, and 1 mM MgCl$_2$ pH 7.5 (Vehicle experiment); 100 mM HEPES, 150 mM KCl$_2$, 1 mM MgCl$_2$ and 10 mM ATP pH 7.5 (ATP SEC experiment). Fractions of 500 μL were collected and denatured in 6 M urea, treated with 5 mM TCEP 35 min at 37°C, and alkylated in 20 mM iodoacetamide for 30 min at room temperature in the dark. Samples were then digested with LysC 1:100 (w/w) at 37°C for 4 hours and then with trypsin 1:100 (w/w) at 37°C for 16 hours after diluting the urea concentration to 2 M with 0.1 M ammonium bicarbonate.

Instrumentation and MS data acquisition for LiP-SMap
LC

Peptide samples were analyzed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) equipped with a nano-electrospray ion source and a nano-flow LC system (Easy-nLC 1000, Thermo Fisher Scientific). Peptides were separated on a 40 cm x 0.75 μm i.d. column (New Objective, PF360-75-10-N-5) packed in house with 1.9 um C18 beads (Dr. Maisch Reprosil-Pur 120). For LC fractionation, buffer A was 0.1% formic acid and buffer B was 0.1% formic acid in 100% acetonitrile and the following gradient was employed: linear from 5% to 25% buffer B over 100 min, linear from 25% to 40% buffer B over 10 min, linear from 40% to 90% buffer B over 5 min and isocratic with buffer B concentration fixed at 90% for 5 min. The flow rate was 300 nL/min and the column was heated to 50°C.

Peptide samples produced from SEC fractions were analyzed on an Orbitrap Fusion Trihybrid mass spectrometer (Thermo Fisher Scientific). Peptides were separated on a 15 cm x 5 μm i.d. column packed in house with 1.9 μm C18 beads (Dr. Maisch Reprosil-Pur C18-AQ) with the same settings as those described in the previous paragraph. The flow rate was 200 nL/min and the column was heated to 50°C.

Data-dependent acquisition

For shotgun LC-MS/MS data dependent acquisition (DDA), 1 μL peptide digests from each biological replicate were injected independently at a concentration of 1 μg/μL. MS1 spectra were acquired from 350 to 1500 m/z at a resolution of 70000. The 20 most intense precursors that exceeded 1300 ion counts were selected for fragmentation at 25 eV normalized collision energy and the corresponding MS2 spectra were acquired at a resolution of 17500 using maximally 100000 ions, collected for maximally 55 ms. All multiply charged ions were used to trigger MS-MS scans followed by a dynamic exclusion for 30 s. Singly charged precursor ions and ions of undefinable charged states were excluded from fragmentation.

Data-independent acquisition

Peptide digests (1 μL aliquots) from each biological replicate were injected independently at a concentration of 1 μg/μL and measured in data-independent acquisition (DIA) mode on an Orbitrap QExactive Plus mass spectrometer (Thermo Fisher Scientific) using DIA settings. The DIA-MS method consisted of a survey MS1 scan from 350 to 1500 m/z at a resolution of 70000 with AGC target of 3×10^6 or 120 ms injection time followed by the acquisition of DIA isolation windows. Twenty variable-width windows were optimized based on previous shotgun measurements of similar samples to equally distribute the number of precursor ions over the DIA isolation windows. The DIA isolation setup included a 1 m/z overlap between windows. The m/z isolation ranges applied on each DIA configuration are listed in Table S7. DIA-MS2 spectra were acquired at a resolution of 35000 with a fixed first mass of 150 m/z and an AGC target of 1×10^6. To mimic DDA fragmentation, normalized collision energy was 25 eV calculated based on the doubly charged center m/z of the DIA window. Maximum injection times were automatically chosen to maximize parallelization resulting in a total duty cycle of approximately 3 s.

Peptide identification and spectral library generation

The collected DDA spectra were searched against the *E. coli* (strain K12) Uniprot fasta database (version June 2015) using the Sorcerer-SEQUEST® database search engine (Thermo Electron). We allowed up to two missed cleavages, excluded cleavage of KP and RP peptide bonds and applied a semi-specific tryptic digestion rule type. Cysteine carboxyamidomethylation (+57.0214 Da) and methionine oxidation (+15.99492) were allowed as fixed and variable modifications, respectively. Monoisotopic peptide tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.02 Da. The identified proteins were filtered using the high peptide confidence setting in Protein Discoverer (version 1.4, Thermo Fisher Scientific), which correspond to a filter for 1% FDR on peptide level.

For generation of spectral libraries the DDA spectra were analyzed with Proteome Discoverer 1.4 as described above and imported to the prepare perspective tool of the software Spectronaut version 8 (Biognosys AG). The spectral libraries were obtained from at least three LC-MS/MS replicates per conditions and contained normalized retention time iRT values for all peptides.

DIA-MS targeted data extraction

Targeted data extraction of DIA-MS acquisitions was performed with Spectronaut version 8 (Biognosys AG) with default settings. Retention time prediction type was set to dynamic iRT with correction factor 1 for determination of XIC extraction windows. Retention time correction was performed with a non-linear calibration strategy, and interference correction on MS2 level was enabled. The false discovery rate (FDR) was estimated with mProphet and set to 1% at peptide precursor level.

Peptide quantification

Both fully and semi tryptic peptides were used for quantification purposes, considering only peptides that were uniquely present in the sequence of one protein of the database. For details regarding DDA-MS and DIA-MS quantification pipelines see section 'Quantification and statistical analysis'.

Selected reaction monitoring (SRM) measurements

Shotgun proteomic data were used to select up to five intense singly charged fragment ions of the y-series from doubly or triply charged precursor ions to perform selected reaction monitoring (SRM) analysis. This set of transitions was then experimentally tested in SRM mode. Matching of retention times and relative fragment ion intensities observed in SRM and shotgun experiments was confirmed after realignment of the gradients used. Transitions associated with obvious interference based on visual inspection using the analysis software Skyline (version 3.6 MacCoss Lab Software) were discarded. In the same experiment, exact peptide retention times from the LC-SRM platform were annotated in order to perform scheduled SRM acquisition using a 360 s retention time window. All SRM analyses were performed on a triple quadrupole/ion trap mass spectrometer (5500 QTrap, ABSciex) equipped with a nano-electrospray ion source and operated in SRM mode. Peptides were separated using an on-line Eksigent 1D-plus Nano liquid chromatography system (Eksigent/ABSciex), equipped with an 18-cm fused silica column with 75-μm inner diameter (New Objective). Columns were packed in-house using Magic C18 AQ 5-μm beads (Michrom Bioresources). A cooled (4°C), autosampler (Eksigent/ABSciex) was used to load the dissolved peptides (∼10 μg), and then 1 μg of the peptide mixture was loaded and separated using a linear gradient from 5 to 35% acetonitrile in water over 30 min. Q1 and Q3 were operated at unit resolution (0.7 m/z half maximum peak width) with a dwell time of at least 50 ms and a cycle time of less than 2 s measuring approximately 200 transitions per run.

Collision energies (CE) were calculated according to the formulas: $CE = 0.044 \cdot m/z + 5.5$ for doubly charged precursor ions or $CE = 0.055 \cdot m/z + 0.55$ for triply charged precursor ions. We confirmed co-elution and peak shape similarity of the transitions monitored for a given peptide and used the respective areas under the curve for quantification. Outlier SRM transitions (e.g., noisy or shouldered transition traces from potentially similar peptides in Q1/Q3 of the background proteome) were not considered in the final calculations.

Peptide identification and quantification of fractionated cell lysates after size-exclusion chromatography

Peptide identifications were performed essentially as described in the 'Peptide identification and spectral library generation' section, with the exception that a tryptic digestion rule type was applied. The number of peptides observed in each fraction (counted as spectral counts) was integrated to quantify the amount of protein present in each fraction in conditions with and without metabolite. The total protein amounts loaded on the SEC column were assumed to be identical since aliquots of the same lysates were used in each condition. SEC chromatograms obtained by monitoring UV absorbance at 280 nm support this assumption. The area under the curve relative to precursor ions (measured with MS1 scans) was also applied as protein quantification method using the Precursor ion area detector node of Proteome Discoverer 2.0 obtaining equivalent results. In both cases the spectral counts profiles (or the equivalent sum of precursor ion intensities) were normalized to the maximum limit of 1 for the representation of the data relative to the chromatograms of single fractionated proteins. We defined a 'shift' in the protein elution volume when a change in the protein elution volume of at least two fractions could be observed in presence versus absence of metabolite.

QUANTIFICATION AND STATISTICAL ANALYSIS

Peptide quantification in the LiP-SMap workflow

A standard workflow in MS design and assay development usually includes a discovery phase followed by a validation phase with targeted proteomics. In the first version of the limited proteolysis combined with mass spectrometry (LiP-MS) protocol a MS label-free shotgun approach based on spectral counts was used to identify and quantify peptides that change in abundance when

comparing conditions. Candidates from shotgun runs were then further validated with Selected Reaction Monitoring (SRM) based targeted proteomics that allowed relative quantification of peptide with better precision and sensitivity than shotgun MS. The SRM approach typically requires several rounds of refinement of the methods, which could be tedious, since they require manual revision of the transition lists and of the MS peaks after data acquisition.

Within the LiP-SMap pipeline the MS data analysis pipeline has been further developed and automated as described below to exclude manual intervention and increase confidence without substantial increase in instrument time albeit reducing data analysis time.

Step 1: MS1-based quantification

In this MS workflow we first compare and quantify MS1 peptide ion maps from DDA-MS runs of the limited proteolysis treated samples in three different conditions (typically, two different concentrations of a metabolite and a control sample with vehicle only) using the alignment software Progenesis 2.0 (Nonlinear Systems). The investigation aims at identifying conformotypic peptides that significantly change their abundance across the different metabolite concentration points. Results acquired from at least three biological replicates are statistically tested for differential conformotypic peptide abundances between conditions applying a fold change cut-off of 2 and a q-value-cut-off of 0.01. Empirical Bayes moderated t tests are applied, as implemented in the R/Bioconductor limma package. The resulting per peptide condition comparisons p values are subsequently adjusted for multiple testing using the Benjamini-Hochberg method. Peptide abundances statistics are obtained by grouping different precursor ions of the same peptide sequences.

MS raw data files from DDA analysis are also used for the spectral library construction for DIA-MS analysis described in step 2.

Step 2: MS2-based quantification

The method then compares time-resolved MS/MS maps measured with the DIA-MS method (Gillet et al., 2012) of the limited proteolysis treated samples in the same three different conditions (see Step 1) using the software Spectronaut version 8 (Biognosys AG).

The DIA method allows the identification and quantification of several thousands of proteins in complex samples, yet maintaining high reproducibility, quantification consistency, broad dynamic range and precision similar to that of SRM, which makes it intrinsically a targeted proteomics method and the subsequent validation stage with SRM assays redundant.

The DIA analysis approach implemented in Spectronaut relies on spectral libraries to assign co-eluting fragment ion traces to peptide precursors and quantify peak intensities. To maximize the confidence of precursor ion identification, the assay libraries used in the LiP-SMap screen are constructed with experiment-specific MS/MS spectra acquired by previously measuring the same peptide samples in DDA mode.

Precursor peptides quantified from at least three biological replicates measured by DIA-MS were statistically tested for differential (conformotypic) peptide abundance between conditions using MSstats based on a linear mixed-effects model (http://msstats.org/). For each metabolite concentration comparison, MS stats provides model-based estimates of fold changes as well as p values that are adjusted for multiple testing using the Benjamini-Hochberg method to control the FDR at the cut-off level of 0.05. Peptide abundance statistics are obtained by grouping different fragment ions that co-elute and are assigned to the same parental precursor. Peptides that present a differential peptide abundance among conditions with a fold change of at least 1.5 are considered for the next step of filtering for the identification of conformotypic peptides.

Step 3: Final candidates

The final conformotypic peptides candidates corresponding to differentially abundant peptides between metabolite treated and not treated samples are selected among those that pass the established cut-off thresholds (see above) with both the DIA-MS and DDA-MS quantification workflows. This is done to increase the confidence in the identification of conformotypic peptides and minimize false positives.

DIA analyses result in permanent digital records of the content of a sample that can be browsed in a targeted manner for multiple queries. In our pipeline, the DIA maps we produced were browsed for the targeted validation of hits from the shotgun-MS screens. However, they can be freely interrogated for the analysis of any peptide of interest in the future.

Step 4: Evaluation of metabolite induced changes in PK activity and endogenous proteolytic activity of cell extracts

We evaluated the possibility that changes in proteolytic patterns measured with LiP-SMap were due to a change in the activity of proteinase K (PK) induced by the addition of metabolites. If it was the case, global changes in proteolytic patterns should be detected on a proteome-wide scale. However, this was not the case for any of the metabolites assayed in this work. We first evaluated the activity of PK in the presence or in the absence of the promiscuous metabolite GTP using a mixture of 39 synthetic peptides (Thermo Scientific) listed in Table S7. Those peptides possess a broad range of propensities to PK cleavage, and are not supposed to acquire structure in aqueous solution; therefore, the extent of PK cleavage should exclusively depend on environmental conditions. The only difference between the experimental conditions we chose is the presence or absence of GTP.

The abundance of the 39 peptides after PK and trypsin digestion and after trypsin digestion only was quantified by SRM-MS using a targeted scheduled method that selectively measured the peptide transitions listed in Table S7. All measurements were evaluated in an experiment with three biological replicates (independent digestions). Overall, there was no significant difference in PK activities measured in presence or in absence of GTP, since the small intensity fluctuations observed for the abundances of the 39 peptides are in the range of the typical error of label-free proteomics experiments (Figure S7A). All other 19 metabolites tested with LiP-SMap did not affect significantly PK activity (Figure S7B). Based on these results, we conclude that metabolite-induced changes in PK activity are negligible in the experimental conditions used in the LiP-SMap screen.

We also tested whether the presence or absence of the metabolites screened by LiP-SMap influenced the endogenous proteolytic activity of the cell lysates. We processed aliquots of the same proteome extracts processed through the LiP-SMap pipeline, but omitting the limited proteolysis step. Under such conditions, differences in proteolytic patterns should reflect the differential activity of endogenous-proteases or other enzymes (e.g., kinases and phosphatases targeting specific proteins in the lysate). Trypsin digestion was performed on three biological replicates of native *E. coli* lysates, and the resulting peptide mixtures were analyzed with the LiP-SMap data analysis pipeline. No tryptic peptides significantly (p value < 0.01) changed abundance in the presence of ATP, PEP or G6P relative to a control with no added metabolite and only two peptides changed abundance in the presence of FBP at the highest concentrations screened in the LiP-SMap assay (Figure S7C). Based on our 1% false discovery rate, as also commonly employed in proteomics experiments, these peptides are likely to be false positives. Comparable results were obtained with the other sixteen metabolites (data not shown). Thus, addition of each of the 20 metabolites does not significantly influence the activity of endogenous proteases in the cell lysates.

Known protein-metabolite interactions

A manually curated list of *E.coli* proteins previously reported to bind to chemical compounds in the 20 metabolite set used in this study was created by mining literature and publicly accessible databases. The research for protein-reactant and allosteric interactions was automated with a MATLAB script that sought for protein-natural compound interactions from the *E. coli* genome scale metabolic model described by *Orth et al.* (Orth et al., 2011) and from the Ecocyc database (Keseler et al., 2013). The resulting list was further expanded to cover all non-metabolic enzymatic reactions not explicitly included in the *E. coli* genome-scale model. For instance, all protein kinases with E.C. number 2.7.1.x or 2.7.10.x, all ligases that bind ATP, and all protein IDs classified as '*ATP-binding*' (KW-0067) or '*GTP-binding*' (KW 0342) in the Uniprot database were manually added to the list. Final curation was completed by manually checking that the automatic annotation truly recovered the right protein IDs.

Quality assessment of LiP-SMap results

In order to assess the quality of our results, we treated LiP-SMap as a classification method. To evaluate the quality of a classification method, one must first identify true and false positive and true and false negative hits (in this case, interactions) generated by the approach. The BRENDA repository (http://www.brenda-enzymes.org/) was mined to identify interactions across species, as a source of known interactions. Uniprot unique protein identifiers were mapped with Enzyme Commission (EC) numbers, which functionally classify enzyme according to catalyzed reactions. We considered the space of interactions that could be identified by LiP-SMap as all potential interactions between the 510 *E. coli* enzymes and our 20 metabolites, excluding 483 enzyme-substrate relationships. Thus, $510 \times 20 - 483 = 9717$ potential enzyme-metabolite interactions were considered overall. Among those, 157 were known enzyme-metabolite interactions supported by experimental evidence specific for *E. coli* in BRENDA. In addition, we included 427 enzyme-metabolite interactions that were reported to exist in other species, and used their Enzyme Commission (EC) numbers to match enzymes to *E. coli* proteins. This approach was recently used to reconstruct the small-molecule regulatory network in *E. coli* (Reznik et al., 2017). Of these additional known interactions, 144 were detected in other bacteria and 283 were only detected in Archaea or eukaryotes. We then restricted our LiP-SMap dataset to those interactions that involved one of the 510 enzymes in *E. coli* and excluded known enzyme-substrate relationships resulting in 592 interactions. We considered true and false positives the subsets of these interactions that were and were not known, respectively, based on the above described criteria. True and false negatives were the subsets of all potential interactions that were not and were detected by LiP-SMap, respectively. Based on this approach, we estimated the false discovery rate (FDR) from the number of false positives using the formula: number of false positives / (number of true negatives + number of false positives).

In order to further assess the quality of our results, we evaluated the confidence in a certain interaction in two ways: first, based on the number of studies providing experimental evidence for the interaction (Figure S2E) and, second, based on the number of species in which the interaction was detected, assuming that the closer the species are to *E. coli*, the more likely the interaction is to occur in *E. coli* (Figure S2F). Based on both criteria, the fraction of interactions recovered by LiP-SMap increased with an increase in the confidence in that interaction, thus supporting the quality of data generated by LiP-SMap.

Score system for assessment of LiP-SMap results

Our scoring system prioritized interactions: i) that were detected at both metabolite concentrations; ii) that were detected with peptide fold changes larger than 4 fold; iii) that were detected via multiple conformotypic peptides from the same protein; iv) for which a metabolite resided in a pathway containing one or more genes sharing a genetic interaction with the protein, since a fraction of genetic networks are known to coincide with physical interaction networks (Gallego et al., 2010); and v) that had been previously reported in other organisms or for which metabolite and protein had been previously associated by other types of literature evidence.

Genetic interactions analysis

To link MBPs to previously reported genetic interaction data, we retrieved curated physical and genetic interactions for *E. coli* (strain K12 / W3110) from the currently released BioGRID database (version 3.4.153, September 25th, 2017). Metabolite-associated genes were extracted from the *E. coli* metabolic network described by Orth and collaborators (Orth et al., 2011), deposited in the BiGG Models database (model id: iJ01366) and accounting for 2251 metabolic reactions, 1366 genes and 1136 metabolites.

Metabolite binding protein analysis

For all metabolites screened in the LiP-SMap screen except GTP, the hits found at the highest concentration were used for the global analysis of metabolite-mediated proteome regulation (data reported in Figures 2, 3B, 4A, and S1–S3). Metabolite binding proteins (MBPs) identified at the high concentration included most hits detected at the low concentration. Concentrations used in the LiP-SMap screen are specified in Table S1.

Volcano plots of candidate MBPs

For the data presented in Figure 2A, the threshold levels of significance necessary to assign conformotypic peptides indicative of a structural rearrangement associated with metabolite binding were: relative fold change between the metabolite exposed and free condition > 2 and q-value < 0.01 of LiP-SMap assays performed with at least three biological replicates. Each point in the plots corresponds to the relative fold change measured between samples with and without metabolite for the peptide with minimal q-value among all peptides quantified for a certain protein. Thus one protein is represented with a single dot in the graph. A protein was defined as an MBP if it had at least one conformotypic peptide in the corresponding LiP-SMap experiment.

Boxplots

In boxplots, horizontal lines define the median and boxes the 25th (Q1) and 75th (Q3) percentiles. The upper whisker is the lower of the maximum observed value and the third quartile plus the interquartile range multiplied by 1.5. The lower whisker is the higher of the minimal observed value and the first quartile minus the interquartile range multiplied by 1.5.

Receiver operating characteristic (ROC) curve

All protein-metabolite interactions were assigned the minimal q-value calculated for each protein-metabolite pair when measuring differential peptide abundances by LiP-SMap. All 9717 pairs were rated by subtracting this value to 1.

Evaluation of structural chemical similarity of metabolites

Chemical structure similarity searches were performed using the advanced searches option in the Chemical Entities of Biological Interest database (ChEBI). Each query metabolite was matched to the collection of PDB ligands available in the PDBdb. OrChem was used in combination with the JChemPaint applet included in the Chemistry Development Kit (CDK) (https://jchempaint. github.io/) to convert chemical structures into molecular fingerprints. Similarity searches were performed by calculating an intramolecular similarity coefficient (Jaccard-Tanimoto) for each structure within the database against the query compound structure. The Jaccard-Tanimoto similarity score (S) was defined as:

$$S = \frac{Nsub}{Na + Nb - Nsub}$$

where N_{sub} was the number of atoms in the maximum common substructure, N_a the number of atoms of molecule 'a' and N_b the number of atoms of molecule 'b'.

Exact matches were given a Tanimoto score of 1, and all random graph matching score were equal to less than 0.5. The Jaccard-Tanimoto coefficient is a measure of how many structural features two chemical structures have in common based on the chosen fingerprint. For this work fingerprints were calculated on a chemical structure path depth of eight. A cut-off of 0.9 was used to ensure selection of significant biologically relevant matches, as, for instance, the correct ligand of an enzyme and its non-cleavable analog. Such non-cognate ligands are often employed in structural and pharmacological studies to aid crystallization, and we wished to recover those cases.

Structures of protein-metabolite holocomplexes

All experimentally determined structures of *E. coli* protein-metabolite holocomplexes were retrieved from the PDB Ligand Expo repository. All structural models containing biologically relevant small molecules analogs with Tanimoto similarity score > 0.9 were considered.

We then expanded this array of *E. coli* structures and searched for homologous structures of protein-metabolite holocomplexes. All PDB entries containing any of the compounds screened or their structural analogs with Tanimoto similarity score > 0.9 were retrieved using the protein-ligand holocomplexes database from PDBsum (https://www.ebi.ac.uk/thornton-srv/databases/ pdbsum). The biological assembly coordinates were then downloaded from the PDB database. When necessary, ChEBI IDs of the metabolite structural analogs were converted to three-letter PDB codes. Protein sequences in FASTA format were then generated from the ATOM coordinates of the selected PDB chains and aligned against the *E. coli* proteome. To this end BLAST databases were constructed from the *E. coli* proteome deposited in the UniProt database. *E. coli* proteins having a sequence identity $\geq$ 40% and an E-value > 0.001 were identified as homologous proteins for the selected PDB chains and considered as valid structural models for the investigated protein-small molecule holocomplexes.

Measurements of Euclidean distances

To determine whether the conformotypic peptides found in LiP-SMap experiments were informative for pinpointing metabolite binding sites, all MS detected peptides were mapped on the homologous PDB chains using the BLAST output described in the previous paragraph. All Euclidean distances in angstroms (Å) between peptide atoms and chemical compound atoms included in the PDB files were calculated with custom-made PyMOL-Python scripts and PERL scripts. The minimum distance found between the MS-detected peptide and the investigated metabolite was reported as the representative distance value for each small molecule-peptide pair.

For the structural models described in Figures 3C–3E we considered the median value of minimal Euclidean distances expressed in Å between metabolites and the conformotypic peptides detected by LiP-SMap for all holocomplexes containing the 20 metabolites screened or their structural analogs with Tanimoto similarity score > 0.9. The average number of amino acid residues was 11 for half-tryptic and 13 for fully tryptic peptides.

Secondary structure prediction analysis

To define the secondary structure of the identified peptides we first built a dataset of secondary structure predictions of the entire *E. coli* (K12) proteome using the Protein Secondary Structure Prediction Server, Jpred 4 (Jpred RESTful API, version 1.5). Next, we mapped the peptides to the protein sequence dataset and retrieved the corresponding secondary structure.

Analysis of the *E. coli* reactome

We created a matrix of 850 metabolites and 1367 *E. coli* enzymes based on all characterized enzymatic reactions as described in the genome-scale models with protein structures (GEM-PRO) for *E. coli* (Brunk et al., 2016). This provided a reference map, here referred to as the *E. coli reactome*, containing all experimentally determined structures of enzyme-substrate and enzyme-product complexes. This information allowed us to establish an operational definition of the boundaries of active sites, based on all experimentally available structural models of *E. coli* enzymes bound to their natural reactants. First, minimal Euclidean distances were measured between amino acid residues of conformotypic peptides (n = 151) or MS detected peptides (n = 6937) and substrates or products of the metabolic reaction catalyzed by the proteins. The difference between the medians of the distances in the conformotypic peptide and MS-detected peptide groups was statistically significant (Figure 5A). Therefore we used the 50th percentile (median) of the distribution of this group of conformotypic peptides (equivalent to 6.44 Å) as the representative radius of the active site volume (Figure 5B). The active site boundaries were therefore defined as the volume of a sphere of radius equal to 6.44 Å.

All candidate binding sites identified by the position of conformotypic peptides with a distance shorter than or equal to 6.44 Å from the active site center were considered as candidate cases for substrate ambiguity or competitive inhibition. When the measured distance was larger than 11.66 Å (a distance equivalent to the third quartile of the distribution of conformotypic peptides), we assumed binding of the metabolite to a secondary site. All intermediate cases with minimal Euclidean distances between 6.44 Å and 11.66 Å were not classified.

To estimate the error associated with the operational definition of active site boundaries we used a non-parametric bootstrap approach. We defined a null distribution generating 10000 bootstrap resamples with replacement of the original dataset and for each of the bootstrap samples we estimated the median. Next, we defined the error associated with active site boundary estimations as the standard deviation of the distribution of active site boundaries (median) across the bootstrap samples. The error calculated was 0.55 Å as discussed in the main text.

Proteins undergoing extensive structural changes

Assignments of half-tryptic (HT) and fully tryptic (FT) peptides were performed with a custom made R script that maps peptides to proteins. Only peptides unambiguously associated with a protein sequence (proteotypic peptides) were considered for the analysis. Fully tryptic (FT) peptides are those originating from two tryptic cleavages at both the N and C termini of the peptide. Half tryptic (HT) cleavages are generated from one sequence unrestricted cleavage and a second tryptic cleavage at the N or the C terminus of the peptide. The mapping of LiP clusters of HT and FT peptides and graphical representations (as in Figure S6) were achieved with a custom made Python script.

Proteins undergoing extreme structural changes were defined as those for which at least 80% of all FT peptides had a change in the same direction (equivalent to the log$_2$ abundance ratio between the condition in presence or in the absence of the metabolite), or at least 80% of all HT peptides had a change in the same direction (equivalent to the log$_2$ abundance ratio between the condition in presence or in the absence of the metabolite). This was quantified with a fully tryptic or half-tryptic *extremity score*:

$$Extremity\ score\ FT = \frac{\sum_{i=1}^{N} sgn(log2(differential\ abundance\ FT\ peptide))}{N}$$

where *differential abundance FT peptide* is the ratio of the median intensity of a fully tryptic peptide *i* measured in presence of a fixed concentration of a metabolite and the median intensity of the fully tryptic peptide *i* measured without metabolite, sgn is the sign function of log$_2$(*differential abundance FT peptide*) and N is the total number of fully tryptic peptides per protein.

$$Extremity\ score\ HT = \frac{\sum_{i=1}^{N} sgn(log2(differential\ abundance\ HT\ peptide))}{N}$$

where *differential abundance HT peptide* is the ratio of the median intensity of a half-tryptic peptide *i* measured in presence of a fixed concentration of a metabolite and the median intensity of the half-tryptic peptide *i* measured without metabolite, sgn is the sign function of log_2(*differential abundance HT peptide*) and N is the total number of half-tryptic peptides per protein.

Protein candidates with an extremity score FT or an extremity score HT higher than 0.8 or lower than −0.8 were further filtered to select those with at least five conformotypic peptides (full or half-tryptic) and a protein sequence coverage of 20%.

Protein hetero-complexes considered for this analysis included assemblies of more than two protein IDs where at least two different MBPs were found with the same metabolite LiP-MS experiment. The list of all manually curated *E. coli* (K-12 substr. MG1655) protein complexes from EcoCyc (1075 entries, latest updated of August 2016) was used as a reference. Protein complexes undergoing extreme structural changes were selected among those that presented at least 80% of all FT peptides assigned to the complex with a change in the same direction (equivalent to the log_2 abundance ratio between the condition in presence or in the absence of the metabolite) or at least 80% of all HT peptides assigned to the complex with a change in the same direction (equivalent to the log_2 abundance ratio between the condition in presence or in the absence of the metabolite). This was quantified with a fully tryptic or half-tryptic protein *community score*:

$$Community\ score\ FT = \frac{\sum_{i=1}^{N} sgn(log2(differential\ abundance\ FT\ peptide))}{N}$$

where *differential abundance FT peptide* is the ratio of the median intensity of a fully tryptic peptide *i* measured in presence of a fixed concentration of a metabolite and the median intensity of the fully tryptic peptide *i* measured without metabolite for a protein hetero-complex, sgn is the sign function of log_2(*differential abundance FT peptide*) and N is the total number of fully tryptic peptides of a protein hetero-complex.

$$Community\ score\ HT = \frac{\sum_{i=1}^{N} sgn(log2(differential\ abundance\ HT\ peptide))}{N}$$

where *differential abundance HT peptide* is the ratio of the median intensity of a half tryptic peptide *i* measured in presence of a fixed concentration of a metabolite and the median intensity of the half tryptic peptide *i* measured without metabolite for a protein hetero-complex, sgn is the sign function of log_2(*differential abundance HT peptide*) and N is the total number of half-tryptic peptides of a protein hetero-complex.

Metabolite-induced changes in protease resistance

Candidates for MBPs that become protease resistant upon metabolite binding were selected from those with 'extremity score FT' $\geq$ 0.8 or an 'extremity score HT' $\leq$ −0.8. Candidates for MBPs that become protease sensitive upon metabolite binding were selected from those that had an 'extremity score HT' $\geq$ 0.8 or an 'extremity score FT' $\leq$ −0.8. Candidates were required to have at least 5 conformotypic peptides (full or half-tryptic) and a protein sequence coverage of 20%.

Isothermal dose-response experiments

Concentration-dependent structural effect curves for single conformotypic peptides were generated by plotting the log_2 transformed mean intensities over the compound concentration range centered on the log_2 transformed mean peptide abundance in the absence of metabolite ($C_0 = 0$ mM, vehicle). These relative abundance values represent the conformotypic peptide local structural response at the corresponding metabolite concentration relative to a control condition. For each peptide the structural response in the absence of metabolite was fixed as basal structural effect level, thus the maximal structural effect as a function of metabolite concentration was equal to zero for all cases where the conformotypic peptide abundance decreased with metabolite concentration. Conversely, the minimal structural effect as a function of metabolite concentration was equal to zero for all cases where the conformotypic peptide abundance increased with metabolite concentration. Error bars show the standard error of the mean (SEM). Means of the C_0-centered log_2 abundances were calculated based on data from at least three biological replicates. Secondary axes of the curves reported in Mendeley data indicate the coefficient of variations of the abundances of peptides measured at a fixed concentration of metabolite calculated over at least three biological replicates.

Fitting of dose-response curves to a sigmoidal dose-response model was performed using the R package drc (https://www.r-project.org). For each peptide, log_2 C_0-centered displacement values relative to the control were fitted to concentrations of metabolite using a 4-parameter logistic equation:

$$f(x(b,c,d,e)) = c + \frac{d-c}{1 + exp\{b(\log(x) - \log(e))\}}$$

where the parameter *b* is the relative slope around the logarithm of the inflection point, *c* is the lower limit, *d* is the upper limit, and *e* corresponds to EC_{50}, that is the dose producing a response half-way between upper and lower limits.

For upregulated peptides (relative abundance increased with metabolite concentration) the lower limit was fixed at zero, and for downregulated peptides (relative abundance decreased with metabolite concentration) the upper limit was fixed at zero to allow proper fitting. The EC_{50} is the dose that produces a structural response equal to 50% of the upper limit, and it was extrapolated from the 4-parameter logistic equation reported above. If the EC_{50} concentration was below the lowest (not-vehicle) concentration tested (e.g., 1 mM ATP), the EC_{50} was not considered valid, and it was assigned a value < 1 mM, since less than three concentrations were measured between the vehicle control concentration and the extrapolated EC_{50} (three measurable points within the EC_{50} range). In this work we extracted EC_{50} values from LiP-SMap data in order to estimate apparent affinity quantitative parameters in cell extracts. We therefore call these parameters 'in-extract K_d' values.

For the analysis of the allosteric effector proteins for ATP the in-extract K_d values reported in Figure 7B were relative to the conformotypic peptide measured for the listed proteins. In cases where more than one conformotypic peptide was detected, the average in-extract K_ds of all conformotypic peptides is reported.

Statistical analyses

All statistical analyses, string processing, and data visualization after the estimation and error analysis of peptide abundances were performed using R (version 3.3.1) and Python (version 2.7) and the Python library Pandas (version 0.18.1). Significance levels of p values are graphically represented with: *** for p values less than 0.001, ** for p values between 0.01 and 0.001 and * for p values between 0.1 and 0.01.

DATA AND SOFTWARE AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD006543. The complete SEC-MS and dose-response curve datasets are available through Mendeley (https://doi.org/10.17632/nhsktkcs3d.1).

A

Number of metabolite–protein interactions

- Catalytic
- Allosteric

614, 426, 158, 116, 113, 86, 71, 62, 42, 41, 32, 27, 27, 24, 23, 21, 20, 15, 8, 5

ATP, ADP, NAD, Pyr, NADP, GTP, L−Gln, PEP, AKG, L−Met, G6P, GMP, Cit, L−Val, Mal, FBP, R5P, L−Phe, PGP, cAMP

B

Low Conc	ATP	PEP	L-Phe
New	91 (56.2%)	8 (66.7%)	0 (0%)
Known	71 (43.8%)	4 (33.3%)	2 (100%)
Total	162	12	2

High Conc	ATP	PEP	L-Phe
New	139 (60.2%)	120 (93%)	38 (92.7%)
Known	92 (39.8%)	9 (7%)	3 (7.3%)
Total	231	129	41

C

% Sequence coverage

31.4   29.1   29.3

ATP   PEP   L-Phe

D

% Sequence coverage

***

Not identified   Identified

Known MBPs

E

All detected, S.C. > 30%

ATP   PEP   L−Phe

- Identified
- Not identified

40.4% (78)   23.3% (7)   42.9% (3)

193   30   7

F

Residual concentration relative to t_0

Succinate, Aspartate, Glutamate

Malate, Phenylalanine, AKG

PEP, ATP, ADP

(t_0, t_1, t_2)

- No filtration
- Gel Filtration
- 10K MW filter
- 3K MW filter

(legend on next page)

Figure S1. Sensitivity and Specificity of the LiP-SMap Pipeline, Related to Figure 2

(A) Protein binding specificity of *E. coli* metabolites. The characterized interactions are classified as allosteric (cyan) or catalytic (blue). Catalytic interactions include both metabolic and non-metabolic interactions.

(B) Number of known and novel MBPs identified by LiP-SMap at low and high concentrations of ATP, PEP, and L-Phe.

(C) Distribution of protein sequence coverage of MBPs identified with LiP-SMap for ATP, PEP, and L-Phe. Horizontal lines define the median, and boxes the 25th and 75th percentiles; whiskers represent the maximum and minimum values. Median values for all proteins detected in the three LiP-SMap experiments are reported on top of the boxplots. The sequence coverage is a measure of the fraction of the protein sequence for which structural data could be derived.

(D) Distribution of protein sequence coverage between known protein targets identified (dark green) and not identified by LiP-SMap (light green) in the experiments with ATP, PEP, and L-Phe. Known protein targets were derived from literature mining. The median sequence coverage of the known MBPs detected by the assay is significantly higher than the median sequence coverage of the known MBPs not captured by LiP-SMap. Significance determined using two-sided Wilcoxon test, *** p value $< 1.48*10^{-13}$).

(E) Sensitivity of LiP-SMap relative to all proteins with at least 30% protein sequence coverage by MS. S.C. = Sequence coverage.

(F) Residual endogenous metabolite concentration relative to the initial concentration measured in diluted cell lysates at t = t_0 after different time points. t_1: After 10 min. from the lysis procedure while different procedures for removing endogenous metabolites were performed (see below). t_2: After further incubation at 25°C for 10 min. Methods for removal of endogenous metabolites were: no treatment (blue), Filtration with 10K MW cut-off filter (orange), filtration with 3K MW cut-off filter (gray), Gel filtration (yellow). Gel filtration was the most efficient method for metabolite dilution, thus it was implemented in the LiP-SMap protocol. Final Concentrations of the tested metabolites after the gel filtration clean-up were: succinate 17.2 μM, aspartate 3.1 μM, glutamate 4.8 μM, malate 8 μM, phenylalanine 1.7 μM, AKG < 0.1 μM, PEP 4.4 μM, ATP 8.6 μM, ADP 2.3 μM.
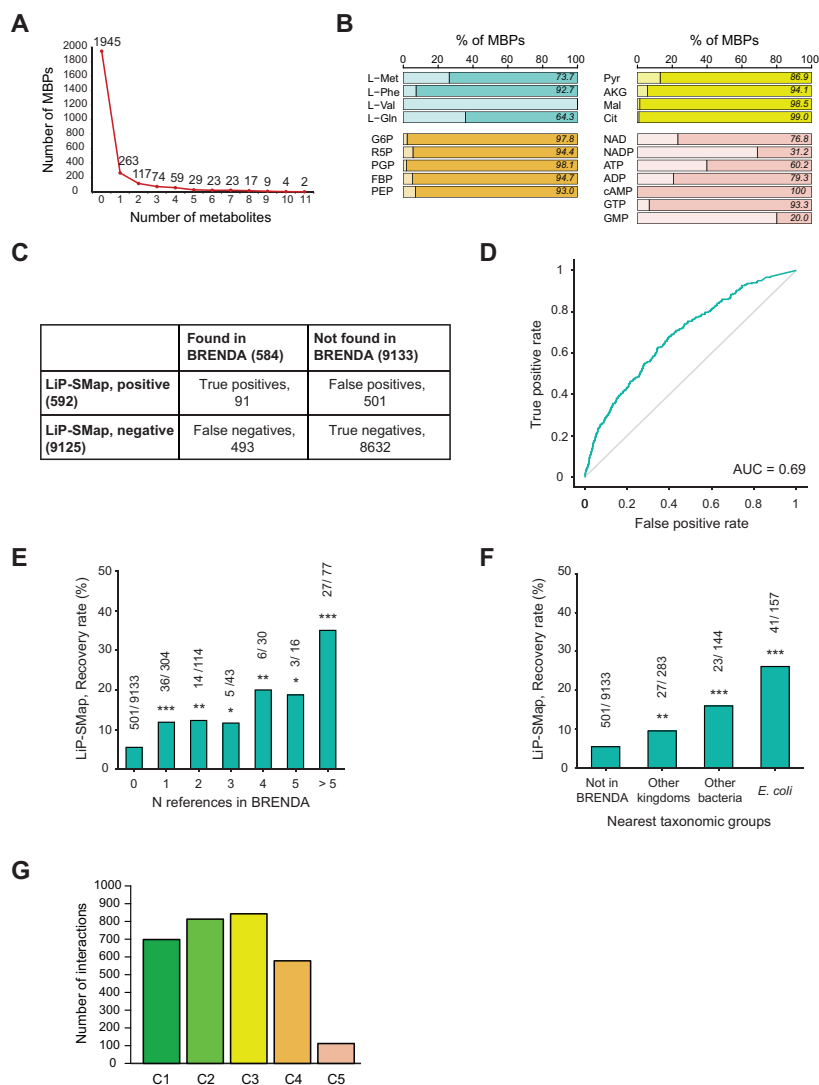
Figure S2. Assessment of Overall Data Quality, Related to Figure 3

(A) Number of distinct metabolites bound by the same protein as detected by LiP-SMap. The fraction of the proteome undergoing structural variations upon metabolite binding corresponds to approximately one third of the total (620 proteins out of 2565).

(B) Fraction over total of previously reported (light hue) and new (dark hue) MBPs identified in the LiP-SMap screen. Metabolites are divided into four classes: amino acids, organic acids, sugar phosphates, and nucleotides. Numbers inside the bars indicate the percentage of new MBPs relative to all MBPs hits.

(C) Contingency table summarizing the overlap between the hits found with the LiP-SMap method and the known interactions from the BRENDA database. Of the 9717 potential pairs that could be mapped to enzyme classification (EC) numbers in BRENDA, 584 were reported in at least one organism (157 from *E. coli*, 144 from other bacteria but not in *E. coli* and 283 from Archaea or eukaryotes but not in any bacterium). The true positive rate (i.e., the percentage of known interaction pairs that were recovered by LiP-SMap) was 15.6%, and the false positive rate was 5.5%. LiP-SMap hits have a much higher probability to be found in the BRENDA database than protein not classified as metabolite binders by LiP-SMap (Fisher exact test p value = 1.7×10^{-17}).

(D) Receiver operating characteristic (ROC) curve of all protein-metabolite interactions rated with the minimal q-value calculated for each protein-metabolite pair when measuring differential peptide abundances by LiP-SMap. The ground truth is represented by the 157 known *E. coli* enzyme-metabolite pairs (Reznik et al., 2017). The gray line represents a random classifier.

(E) Relationship between the number of distinct references relative to metabolite protein interactions found in protein orthologs after mining the BRENDA database and the recovery rate obtained by LiP-SMap. Recovery rate is defined as the percentage of reference enzyme-metabolite pairs in BRENDA that were classified as a hit by Lip-SMap. Numbers on top of the bars indicated the absolute number of recovered interactions found by LiP-SMap and expected interactions from the BRENDA reference set respectively. The number of references found in BRENDA was used to bin each interaction by frequency (0; 1; 2; 3; 4; 5; > 5). For each category, we applied the Fisher exact test to see whether the enrichment of BRENDA interactions among LiP-SMap hits was statistically significant: * indicates a p value between 0.01 and 0.1, ** between 0.001 and 0.01, and *** are p values below 0.001.

(F) Same as in (e), but with each enzyme-metabolite pair found in BRENDA binned by taxonomy (other kingdoms, other bacteria, *E. coli*).

(G) Number of interactions that fulfill each of the criteria we defined to prioritize interactions found by LiP-SMap: C1 = Detected at both concentrations; C2 = detected with peptide fold change > 4; C3 = Multiple conformotypic peptides per protein found; C4 = evidence of genetic interactions; C5 = found in orthologs.
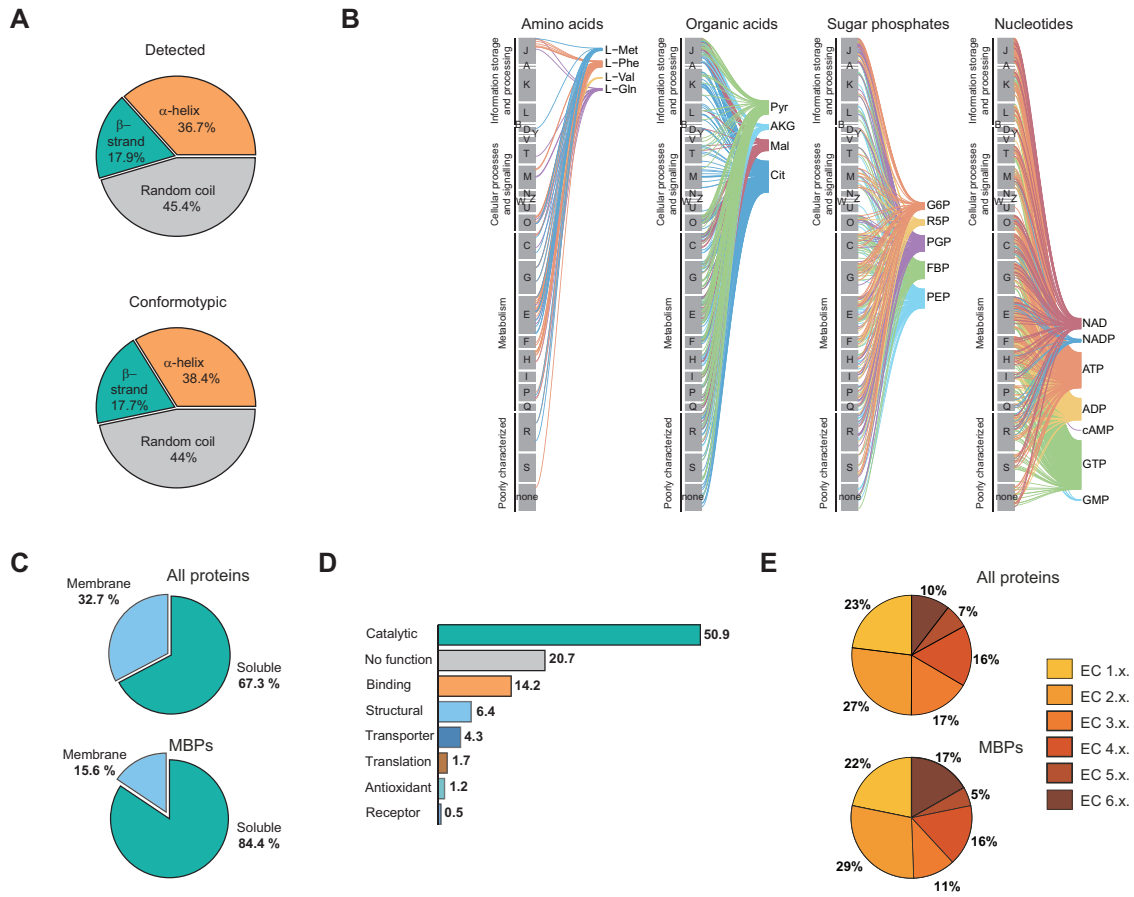
Figure S3. Characterization of the Metabolite-Proteome Interaction Network in *E. coli*, Related to Figure 3

(A) Prevalence of secondary structural elements in conformotypic and detected peptides. Most of the peptides have an average peptide length of 12 amino acids and map to regions devoid of secondary structure or loops.

(B) Comprehensive maps of all *E. coli* metabolite-protein interactions for each metabolite type. MBPs included proteins that encompassed a wide-range of functions as categorized in Clusters of Orthologous Groups (COGs). Each line represents an interaction found with LiP-SMap between one of the 20 metabolites screened and a protein target. Letters in gray boxes correspond to COG categories: A, RNA processing and modification; B, chromatin structure and dynamics; C, energy production and conversion; D, cell cycle control and mitosis; E, amino acid metabolism and transport; F, nucleotide metabolism and transport; G, carbohydrate metabolism and transport; H, coenzyme metabolism; I, lipid metabolism; J, translation; K, transcription; L, replication and repair; M, cell wall/membrane/envelope biogenesis; N, cell motility; O, post-translational modification, protein turnover, chaperone functions; P, inorganic ion transport and metabolism; Q, secondary structure; T, signal transduction; U, intracellular trafficking and secretion; Y, nuclear structure; Z, cytoskeleton; R, general functional prediction only; S, function unknown.

(C) Relative distribution of detected proteins (top) and MBPs (bottom) between the membrane and soluble (i.e., cytosolic) compartment according to the PANTHER GO slim cellular component functional annotation. Percentages refer to the total number of proteins with GO annotation in PANTHER (Protein ANalysis THrough Evolutionary Relationships) version 12.0, released 2017-07-10.

(D) Frequencies of MBPs detected by LiP-SMap that belong to Gene Ontology (GO) functional categories expressed as percentages over the total number of MBPs detected. The categories considered were: Catalytic activity (GO:0003824), no function annotated, binding (GO:0005488), structural molecule activity (GO:0005198), transporter activity (GO:0005215), translation regulator activity (GO:0045182), antioxidant activity (GO:0016209), receptor activity (GO:0004872). GO classification was performed using PANTHER.

(E) Distribution of enzyme classifications (EC) numbers among enzymes analyzed with LiP-SMap. EC numbers of metabolic enzymes were obtained from the genome-scale model of *E. coli* of Orth et al. (Orth et al., 2011). Percent of proteins detected in this study with indicated EC numbers (left). Percent of enzymes shown to bind at least one metabolite with indicated EC numbers (right): EC 1, oxidoreductases; EC 2, transferases; EC 3, hydrolases; EC 4, lyases; EC 5, isomerases; EC 6, ligases. If an enzyme was associated with more than one EC number, each association was treated independently. For all metabolites screened, except GTP, data from the samples treated with the highest concentration of metabolite are reported. Concentrations used are specified in Table S1.

A

Gene name

AlaS
ArgD
AroA
AroB
Asd
AsnS
AspC
AspS
AtpA
AtpD
Cmk
CysS
DeoD
Eno
FabB
FabD
FabG
FolD
FolE
GapA
GlgA
GlmM
GltX
GlyA
GlyS
GpsA
Gsk
HemG
HemL
HisG
HisH
HisI
HisS
IleS
IlvB
IlvC
IlvD
IlvE
IlvN
IscS
IspE
KdsA
KdsC
LeuA
LeuC
LeuD
LeuS
LptB
LuxS
LysC
LysS
MetE
MetF
MetG
MetK
MetL
MtnN
MurF
MurG
NadB
PdxB
PdxJ
Pgk
PheA
PheS
PheT
ProS
Prs
PssA
PurD
PurF
PurL
PurM
RibB
RibC
RibF
SdhA
SdhB
SerC
SerS
ThrS
Tmk
TrpS
TyrA
TyrS
ValS

L-Met, L-Phe, L-Val, L-Gln, Pyr, AKG, Mal, Cit, G6P, R5P, PGP, FBP, PEP, NAD, NADP, ATP, ADP, cAMP, GTP, GMP

■ MBP   □ non-MBP

B

Fraction of proteins vs Coefficient of variation — All
***

Fraction of proteins vs Coefficient of variation — ATP GTP Citr.
**

C

Ppc (Phosphoenolpyruvate carboxylase)
PEP + HCO$_3^-$ ⟶ OAA + Pi

G6PDH (Glucose-6-phosphate dehydrogenase)
G6P + NADP$^+$ ⟶ PGL + NADPH + H$^+$

PpsA (Phosphoenolpyruvate synthetase)
PYR + ATP + H$_2$O ⟶ PEP + AMP + Pi + 2H$^+$
           ⊤
          PpsR

D

Normalized Ppc activity: Ppc only, + Acetyl-CoA, + FBP, + Both

E

Rel. Ppc activity vs AKG (mM)
Rel. Ppc activity vs Malate (mM)

F

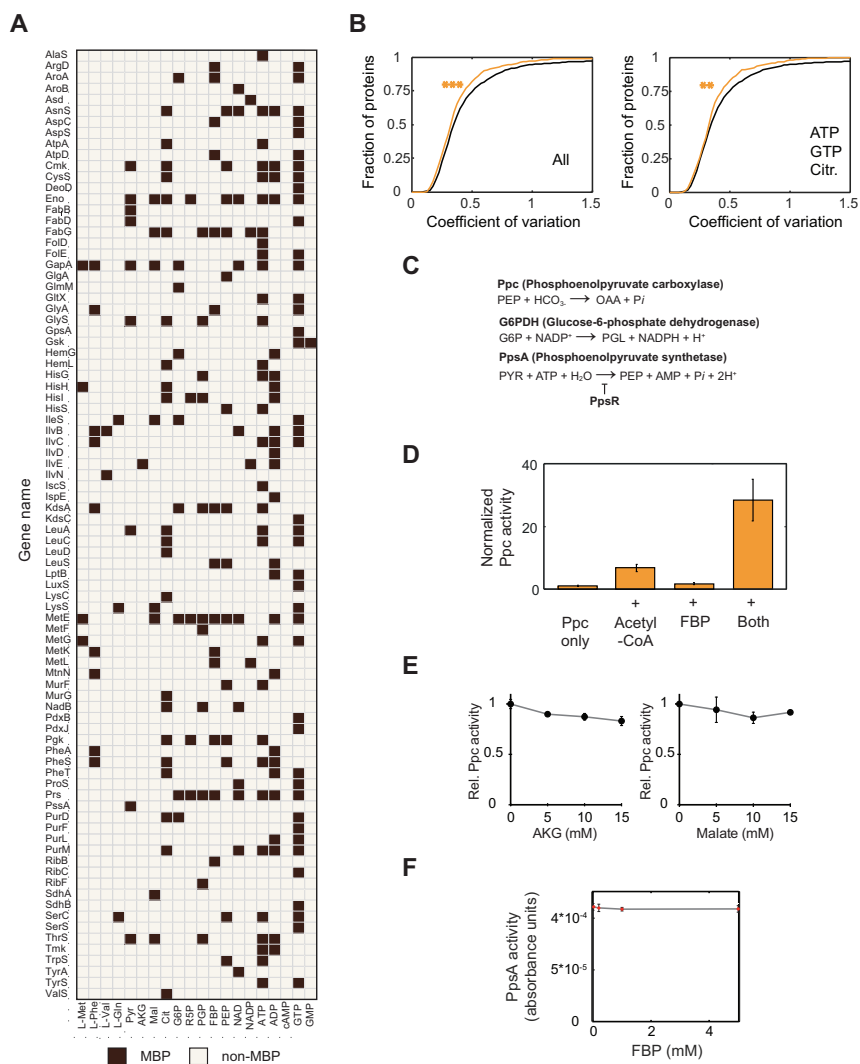PpsA activity (absorbance units) vs FBP (mM)

Figure S4. Further Characterization and Novel Regulatory Processes of the Metabolite-Proteome Interaction Network in *E. coli*, Related to Figures 3 and 4

(A) Interaction map of LiP-SMap-detected protein-metabolite interactions within the family of 'core proteome' proteins as defined in Yang et al. (Yang et al., 2015). Each row shows the core proteome proteins detected by MS in this screen. Dark brown boxes indicate MBPs assigned with LiP-SMap to the metabolite in the corresponding column.

(B) For each protein, the coefficient of variation (CV), calculated as the relative standard deviation in protein expression across conditions was plotted versus the cumulative frequency of proteins (fraction of proteins). Only data from conditions in which a protein was reliably quantified (relative error of quantification < 30%) was used, and only proteins for which more than 50% of the conditions yielded reliable protein quantification were used. Protein concentrations were calculated from protein copy numbers and cell volumes and were obtained for 22 steady-state *E. coli* growth conditions as described in Schmidt et al. (2016). The left panel shows the cumulative distribution of CV for proteins with at least one conformotypic peptide measured in LiP-SMap experiments with any of the 20 metabolites assayed (orange) and proteins with no conformational changes measured in LiP-SMap experiments with any of the 20 metabolites assayed (black). The right panel shows the cumulative distributions of CVs for the two classes excluding the LiP-SMap experiments with the highest number of protein hits (GTP, ATP, citrate). MBPs have a lower median CV than the rest of the proteome (two-sided Wilcoxon test: *** p value < $6.54*10^{-5}$ for all metabolites, ** p value = $5.8*10^{-3}$ for all metabolites excluding ATP, GTP and Citrate).

(C) Enzymatic reactions catalyzed by Ppc, G6PDH, and PpsA.

(D) Ppc activity in presence of 0.1 mM acetyl-CoA, 1 mM FBP, or both metabolites normalized to Ppc activity in absence of the effector. Error bars indicate SD (n = 4). Addition of 0.1 mM acetyl-CoA strongly enhanced Ppc activity ∼8-fold, whereas 1 mM FBP enhanced activation 2-fold.

(E) *In vitro* activity of purified Ppc at different 2-oxoglutarate (AKG) or malate concentrations. Error bars indicate SD (n = 4).

(F) PpsA activity at different FBP concentrations measured as a decrease in absorbance at 340 nm in coupled assay with PEP carboxylase and malate dehydrogenase.
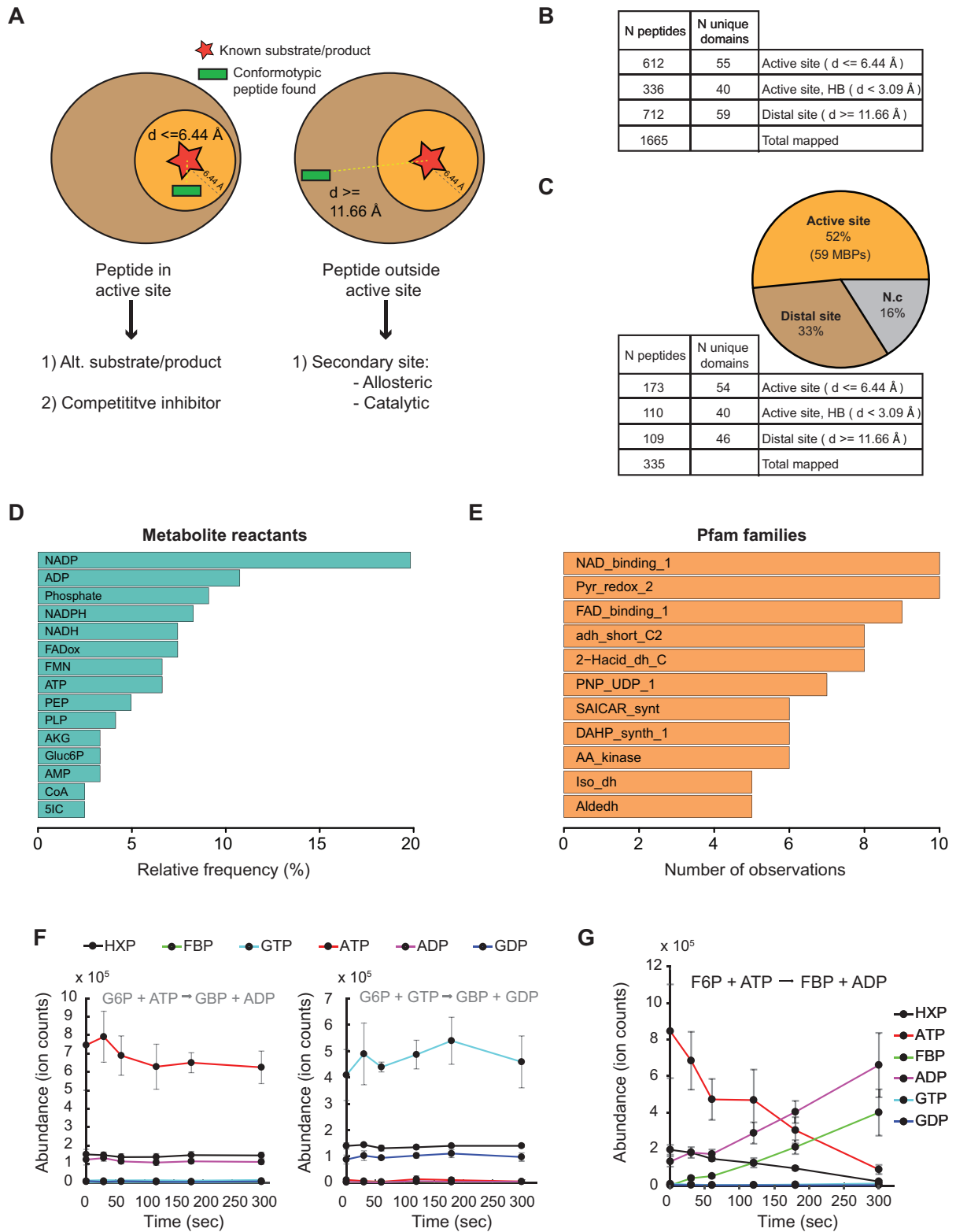
A

★ Known substrate/product

▬ Conformotypic peptide found

Peptide in active site
d <=6.44 Å

Peptide outside active site
d >= 11.66 Å

1) Alt. substrate/product
2) Competititve inhibitor

1) Secondary site:
- Allosteric
- Catalytic

B

N peptides	N unique domains	
612	55	Active site (d <= 6.44 Å)
336	40	Active site, HB (d < 3.09 Å)
712	59	Distal site (d >= 11.66 Å)
1665		Total mapped

C

Active site 52% (59 MBPs)

N.c 16%

Distal site 33%

N peptides	N unique domains	
173	54	Active site (d <= 6.44 Å)
110	40	Active site, HB (d < 3.09 Å)
109	46	Distal site (d >= 11.66 Å)
335		Total mapped

D

Metabolite reactants

NADP, ADP, Phosphate, NADPH, NADH, FADox, FMN, ATP, PEP, PLP, AKG, Gluc6P, AMP, CoA, 5IC

Relative frequency (%)

E

Pfam families

NAD_binding_1, Pyr_redox_2, FAD_binding_1, adh_short_C2, 2−Hacid_dh_C, PNP_UDP_1, SAICAR_synt, DAHP_synth_1, AA_kinase, Iso_dh, Aldedh

Number of observations

F

●HXP ●FBP ●GTP ●ATP ●ADP ●GDP

x 10⁵

G6P + ATP ➝ GBP + ADP

Abundance (ion counts)
Time (sec)

x 10⁵

G6P + GTP ➝ GBP + GDP

Abundance (ion counts)
Time (sec)

G

x 10⁵

F6P + ATP — FBP + ADP

●HXP ●ATP ●FBP ●ADP ●GTP ●GDP

Abundance (ion counts)
Time (sec)

Figure S5. Novel Interactions at Catalytic Sites, Related to Figure 5

(A) Schematics of positions of conformotypic peptides (green bars) that map within the active site boundaries (distance below 6.44 Å; left) and outside the active site neighborhood (distance greater than 11.66 Å; right). In both cases the center of the active site is approximated as the position of known substrates or products

(legend continued on next page)

of the reaction catalyzed by the active site in an experimentally determined structural model of the protein enzyme (red star). Depending on the relative position of the conformotypic peptide from the active site different types of functional protein-metabolite interactions can be hypothesized.

(B) Summary of the number of conformotypic peptides and unique domains mapping inside an active site, inside an active site closely enough to form hydrogen bonds (HB), and at a distance outside the active site neighborhood. See STAR Methods for details.

(C) Summary of the percentages, numbers of conformotypic peptides, and numbers of unique domains mapping inside an active site, inside an active site closely enough to form hydrogen bonds (HB), and at a distance outside the active site neighborhood. Relative frequencies and percentages were calculated considering a single peptide and one metabolite holocomplex for individual proteins. For the pie chart of Figure 5C, relative frequencies and percentages were calculated for multiple holocomplexes involving the same protein.

(D) Metabolite reactants most frequently found within the active site boundaries of promiscuous binding clefts, as defined by the LiP-SMap analysis. Those metabolite reactants are the real substrate or products of known metabolic reactions catalyzed in the active sites. The percentages of metabolite reactants observed relative to the total number of metabolites mapped to an active site are plotted. Abbreviations not explicitly cited in the main text refer to: flavin adenine dinucleotide oxidized form (FADox), pyridoxal phosphate (PLP), 2-ketoglutarate (AKG), D-glucosamine 6-phosphate (Gluc6P), adenosine monophosphate (AMP), coenzyme A (CoA), 5−amino−1−(5−phospho−D−ribosyl)imidazole−4−carboxylate (5IC).

(E) Number of Pfam protein domain families most often found within the active site boundaries of promiscuous binding clefts as defined by the LiP-SMap analysis.

(F) Mass spectrometry-based PfkB assays. No consumption of two alternative substrate pairs: i) glucose-6-phosphate (G6P) and ATP (left panel) or ii) glucose-6-phosphate and guanosine triphosphate (GTP) (right panel) were observed. The expected product pairs were glucose-1,6-bisphosphate (GBP) and adenosine diphosphate (ADP) (left panel) and glucose-1,6-bisphosphate (GBP) and guanosine diphosphate (GDP) (right panel).

(G) Mass spectrometry-based PfkB activity assay: Time courses of product formation and substrate consumption by purified PfkB in the reaction that converts 2.5 mM fructose-6-phosphate (m/z 259.0231 HXP -H(+)) and 2.5 mM ATP (m/z 505.9882 ATP -H(+)) into fructose-1,6-bisphosphate (m/z 338.9884 FBP -H(+)) and ADP (m/z 426.0216 ADP -H(+)). Data represent mean values ± SD (n = 3).
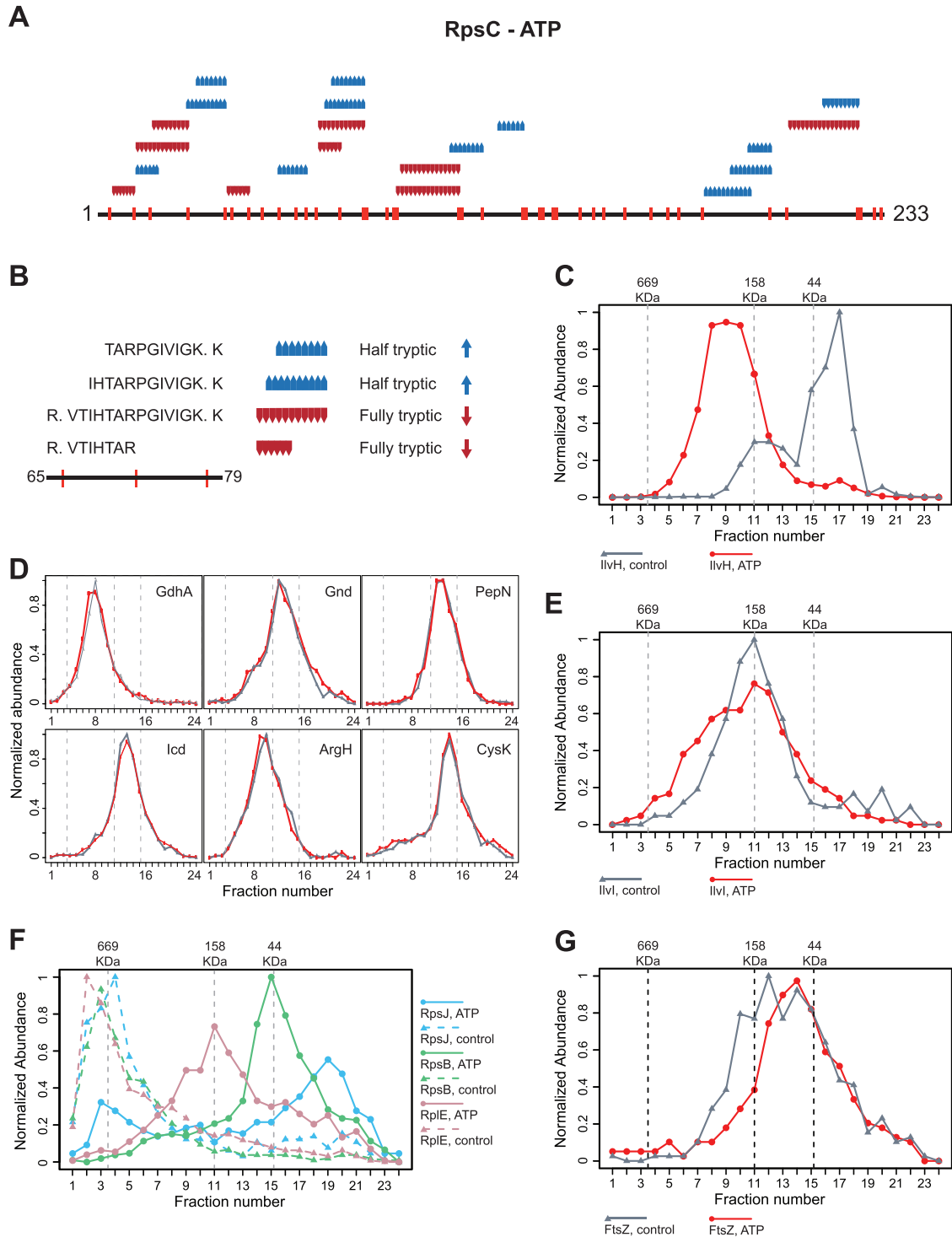
A RpsC - ATP

1 ——————————————— 233

B

TARPGIVIGK. K	Half tryptic ↑
IHTARPGIVIGK. K	Half tryptic ↑
R. VTIHTARPGIVIGK. K	Fully tryptic ↓
R. VTIHTAR	Fully tryptic ↓

65 ———————— 79

C
IlvH, control   IlvH, ATP

D
GdhA   Gnd   PepN
Icd   ArgH   CysK

E
IlvI, control   IlvI, ATP

F
RpsJ, ATP
RpsJ, control
RpsB, ATP
RpsB, control
RplE, ATP
RplE, control

G
FtsZ, control   FtsZ, ATP

Figure S6. Metabolite-Induced High-Order Structural Changes Analysis, Related to Figure 6

(A) Proteins with at least 80% of their conformotypic peptides that indicated resistance or susceptibility to protease activity upon metabolite binding were selected by analyzing separately peptides with non-tryptic termini (half-tryptic, HT) and peptides with both tryptic ends (fully tryptic, FT) (Figure 1B). Since HT peptides result from the proteolytic cleavage of FT peptides at structurally accessible regions after LiP, general upregulation of HT combined with vast down-regulation of FT peptides indicate a broad increase in cleavage sensitivity. Conversely, general upregulation of FT and downregulation of HT signal extensive protease resistance. The plot shows the positions of FT (red) and HT (blue) peptides over the amino acid sequence of RpsC (UniProt: P0A7V3). Vertical red lines indicate the position of tryptic cleavage sites along the RpsC amino acid sequence. Peptides that are present at higher levels in the presence than in the absence

(legend continued on next page)

of ATP are illustrated with triangles with vertices pointing upwards; downregulated peptides are indicated by triangles with vertices pointing downwards. In this example, all (100%) FT peptides were downregulated, and 92% of the HT peptides (11 peptides of 12) were upregulated, therefore RpsC becomes generally more protease sensitive upon binding to ATP (for more details see STAR Methods).

(B) Details of the peptide cluster family of the RpsC sequence between amino acids 65 and 79, which includes 2 upregulated HT peptides and 2 downregulated FT peptides.

(C) SEC elution profile equivalent to that shown in Figure 6C obtained by summing the MS1 intensities of peptide precursors of IlvH for each SEC fractions. Signal intensities were normalized to the maximum signal measured in the 24 SEC fractions collected with and without ATP.

(D) Co-elution of 6 non-ATP binders by SEC upon addition of metabolite in the running buffer. More than 300 profiles showing similar co-elution behaviors are available in Mendeley data (see Data and Software Availability for link).

(E) SEC elution profiles of IlvI. The sums of peptide counts detected for each SEC fraction relative to the maximum intensity measured were plotted when ATP was absent (blue line) or present (red line) in the elution buffer. The molecular weight of monomeric IlvI predicted from its sequence is 63 KDa. The peak shoulder toward lower elution volumes present in the SEC separation performed in presence of ATP may suggest the formation the acetolactate synthase hetero-tetrameric complex with IlvH with stoichiometry $[IlvI]_2[IlvH]_2$ and molecular mass of 155 KDa, which is the active form of the enzyme (Vyazmensky et al., 1996).
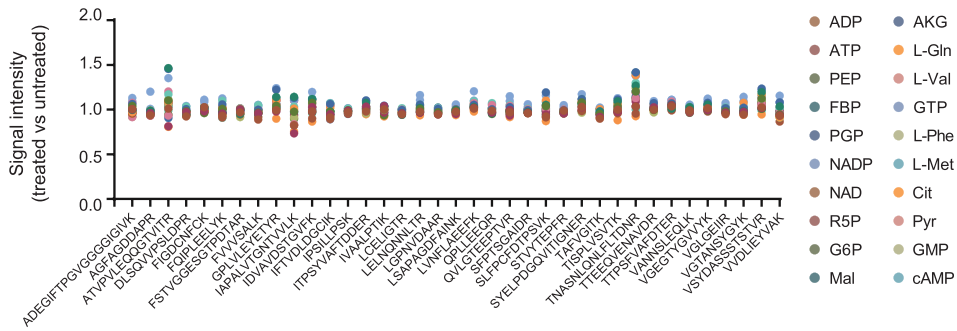
(F) Differential SEC elution profiles of RpsJ, RpsB, and RplE subunits of the 30S ribosome complex. These three subunits were among those seven ribosomal proteins that changed their elution volume when ATP was added to the running buffer.

(G) Differential SEC elution profiles of FtsZ as described in (e).
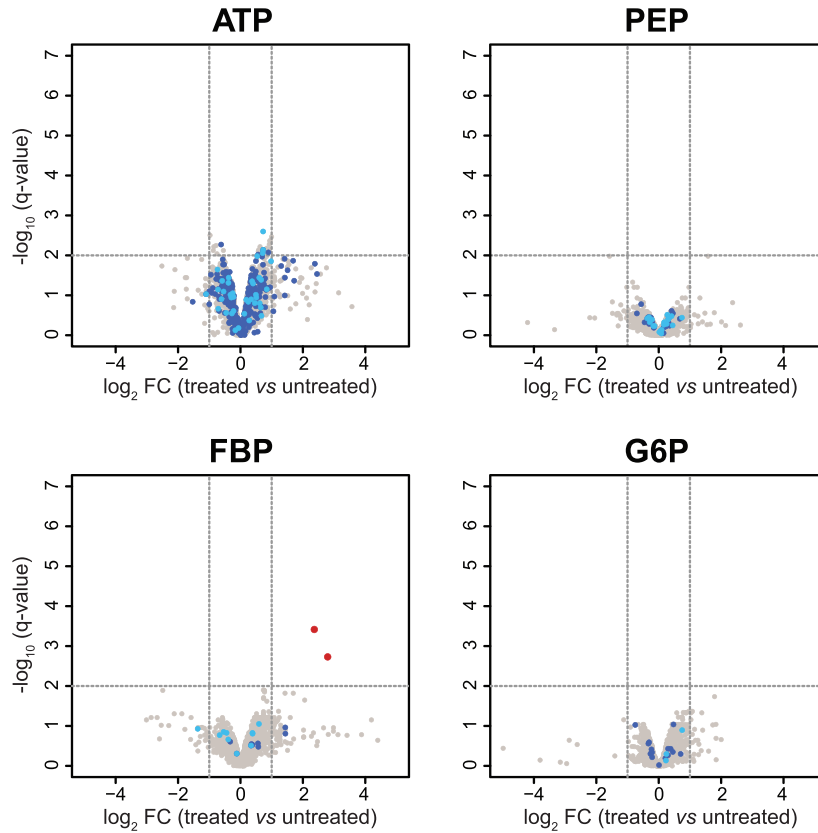
(legend on next page)
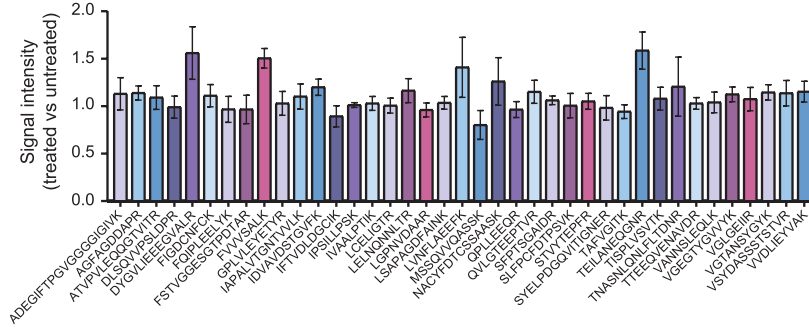
Figure S7. Metabolite Influence on Proteinase K and on Endogenous Proteolytic Activities, Related to STAR Methods

(A) Influence of metabolites on proteinase K activity. We spiked 39 synthetic peptides into *E. coli* lysates. After limited proteolysis with proteinase K, samples were analyzed by SRM-MS in presence of 25 mM GTP or vehicle, and peptides were quantified using a label-free MS approach. For each peptide the ratio between signal intensities measured in presence of 25 mM GTP (Signal intensity treated) or in presence of vehicle solution (Signal intensity untreated) after limited proteolysis are shown. All LiP intensity values were normalized to intensities observed in samples digested only with trypsin. Data collected on GTP are shown here as GTP bound the highest number of proteins among those ligands evaluated with LiP-SMap. Error bars indicate SE (n = 3).

(B) Same as (a) where each dot represents the mean of three biological replicates for each metabolite (n = 3).

(C) Differential abundance analysis of peptides generated from tryptic digestions of *E. coli* cell lysates in presence of ATP, PEP, FBP, or glucose-6-phosphate (G6P). Lysates treated with metabolite or only vehicle were analyzed in at least three biological replicates. Peptides mapped to proteins that are known interactors of individual metabolites are reported in blue (enzymatic) or cyan (allosteric). Peptides that pass the fold change (FC) cut-off ($|FC| > 2$) and q-value cut-off (q-value < 0.01) are shown in red. No peptides of the lysates changed abundance in presence of any of the metabolites with the exception of KGGPLADGIVITPSHNPPEDGGIK (mapping to phosphoglucomutase, UniProt P36938) and DKSLHALEK (mapping to γ-glutamyl-γ-aminobutyraldehyde dehydrogenase. UniProt: P23883).