**Sample preparation and processing for mass spectrometry (MS)**

***E. coli***

All experiments were performed with the *E. coli* BW25113 wild-type in shake-flask cultures. Frozen glycerol stocks were used to inoculate Luria-Bertani (LB) complex medium. After 6 hours of incubation at 37°C under constant shaking at 220 rpm, LB cultures were used to inoculate 25 ml of M9 minimal medium pre-cultures supplemented with 5 g/L of the indicated carbon source (glucose, fructose, sucrose, acetate, gluconate, glycerol, galactose and pyruvate) for over-night culture. The next day, final cultures were inoculated 1:100 (v/v) in 500 ml of M9 minimal medium supplemented with the same carbon source and grown to exponential phase (OD600 = 0.8 ± 0.1) at 37°C under constant shaking at 220 rpm. Cells were then harvested by centrifugation at 4,200 x *g* for 15 min at 4 °C and washed twice with 25 ml ice-cold lysis buffer (LB: 20 mM Hepes, 150 mM KCl, 10 MgCl2, pH 7.5). Cell pellets were resuspended in 500 µl cold LB and mixed with the same volume of acid-washed glass beads (Sigma Aldrich) and disrupted at 4 °C by 4 consecutive rounds of beads-beating at 30 sec with 4 min pause between the runs in a FastPrep-24TM 5G Instrument (MP Biomedicals). *E. coli* lysates were centrifuged at 16,000 x *g* for 15 min at 4 °C to remove cellular debris, the supernatants were collected and transferred to a fresh 1.5 ml tube and the protein concentration was determined with the bicinchoninic acid assay (BCA Protein Assay Kit, Thermo Fisher Scientific). The protein extracts were flash frozen in liquid nitrogen and stored at - 80°C until use.

For the preparation of *E. coli* proteome extracts used as background proteome in the *in vitro* experiments, *E. coli* cells were grown in 500 ml M9 minimal medium supplemented with 5 g/L glucose at 37°C under shaking at 220 rpm and harvested in exponential phase (OD600 = 0.8 ± 0.1). Proteome extracts were prepared as described above for the different carbon sources. Endogenous metabolites and nucleic acids were removed by size-exclusion chromatography (Amicon Desalting Columns 3 kDa MWCO, Merck), protein concentration was determined with the bicinchoninic acid assay (BCA Protein Assay Kit, Thermo Fisher Scientific). The protein extracts were flash frozen in liquid nitrogen and stored at -80°C until use.

***Saccharomyces cerevisiae***

Single colonies of the BY4741 *Saccharomyces cerevisiae* strain picked from a fresh plate were inoculated in synthetic complete (SC, Cold Spring Harbor Protocols, 2016) medium and grown for 6 hours at 30°C under shaking at 180 rpm. The pre-cultures were inoculated into fresh SC medium cultures to a final OD600 of 0.0003 and grown overnight at 30°C under constant shaking. When cultures reached OD600 = 0.8±0.1 the liquid medium was removed by 1 min centrifugation at 1000 x g. For the heat shock experiment, cell pellets were resuspended in the same volume of 42°C pre-warmed SC medium and incubated at 42°C for 3 min under shaking at 180 rpm. As control, the same procedure was followed but resuspending cell pellets with 30°C pre-warmed SC medium and incubating cell cultures at 30°C. For the osmotic stress perturbation, cell pellets were resuspended in SC medium supplemented with 0.4 M NaCl and with an equivalent volume of SC medium in the control samples and cell cultures were incubated for 10 min at 28°C under constant shaking at 180 rpm. Next, the liquid medium was removed by 1 min centrifugation at 1000 x g and cell pellets were resuspended in lysis buffer (100 mM HEPES, 1 mM MgCl2, 150 mM KCl, pH 7.5). Liquid-nitrogen frozen beads of cell suspensions were added to grinding vials and ground in a Freezer Mill (SPEX SamplePrep 6875). To remove cell debris, samples were centrifuged at 800 x g for 5 min at 4°C. The supernatant was collected and protein concentration determined with the bicinchoninic acid assay (Thermo Fisher Scientific).

For the analysis of differentially regulated p-sites during the acute osmotic perturbation, cells were prepared as described above using untreated cells as controls, and peptide mixtures subjected to the enrichment step (see below).

**Recombinant protein production and purification**

All purification steps were performed at 4°C, and protein concentration was determined spectrophotometrically at 280 nm. Pgk *E. coli* protein was obtained from the ASKA collection (Kitagawa et al., 2005) and expressed as N-terminal His6 fusion proteins. Briefly, 500 mL LB cultures containing 50 µg/mL chloramphenicol were inoculated with an aliquot of an LB overnight culture diluted 1:100. Cells were grown to a final OD600 of 0.5 at 37 °C under constant shaking at 220 rpm followed by induction for 2 h with 0.5 mM isopropyl β-d-thiogalactoside at 37°C. Cells were lysed in 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 5mM phenylmethylsulfonyl, 0.5 mg/ml Lysozyme for 45 min on ice, followed by ultrasonication. The supernatant obtained after high-speed centrifugation (20,000 g, 40 min, 4 °C) was applied to an Ni-IMAC column (GE Biotech) equilibrated in 50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 10 mM imidazole, to capture the His6 fusion protein, followed by washing with 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 25 mM imidazole. The protein was eluted with 50 mM Tris-HCl pH 7.5, 300 mM NaCl, 200 mM imidazole and the protein containing fractions were dialyzed against 20 mM Tris-HCl, pH 7.5, and 50 mM NaCl over-night at 4°C. Proteins were concentrated with 10 kDa MWCO (Milipore) concentrators, 5% glycerol was added, and fractions were flash frozen in liquid nitrogen and stored at -80 °C.

**Limited proteolysis (LiP)**

Each proteome extract was split into a control sample, which was subjected to only tryptic digestion and used to measure protein abundance changes, and a LiP sample, containing information about protein structural changes, which was subjected to a double-protease digestion step with a nonspecific protease followed by complete digestion with trypsin. Both samples contained 100 μg of extracted proteome. Proteinase K from *Tritirachium album* (Sigma Aldrich) was added to the LiP samples at an enzyme/substrate (E:S) ratio of 1:100 (w/w) and incubated for 1 min (*E.coli* experiment) or 3 min (*S. cerevisiae* experiment) at 25°C. A corresponding volume of water was added to the control samples. Digestion reactions were stopped by heating LiP samples for 5 min at 98°C in a thermocycler followed by addition of sodium deoxycholate (Sigma Aldrich) to a final concentration of 5%. The same procedure was applied to control samples. Both LiP and control samples were then subjected to complete tryptic digestion in denaturing conditions as described below. The *in vitro* LiP experiment was performed spiking in 10 µg purified *E. coli* pgk into 100 μg *E. coli* lysate cleared of endogenous metabolites, as described above. Cell lysates were incubated with the pgk substrate 3-phosphoglycerate (3PG) to a final concentration of 5mM, 10mM, 15mM, 20mM, and 25mM for 5 min at 25°C. As control, a cell lysate without metabolite addition was used. The metabolite solutions were freshly prepared from ultra-pure powders in 100mM HEPES, pH 7.5. After solubilization, pH was double-checked with pH strips. LiP experiments were carried out on both the lysate after metabolite addition and on control samples.

**Tryptic digestion**

Proteins fragments generated in the previous step were reduced by incubation of samples with tris(2-carboxyethyl)phosphine (Thermo Fisher Scientific) to a final concentration of 5 mM for 30 min at 37 °C. Next, the alkylation of free cysteine residues was achieved by adding iodoacetamide (Sigma Aldrich) to a final concentration of 40 mM for 30 min at 25°C in the dark. Samples were diluted with freshly prepared 0.1 M ammonium bicarbonate to a final concentration of 1% sodium deoxycholate. Samples were predigested with lysyl endopeptidase LysC (Wako Chemicals) at an enzyme/substrate ratio of 1:100. After 2 hours at 37°C, sequencing-grade porcine trypsin (Promega) was added to a final enzyme/substrate ratio of 1:100, and samples were incubated for 16 h at 37°C under shaking at 800 rpm. Protease digestion was quenched by lowering the reaction pH (< 3) The peptide mixtures were loaded onto Sep-Pak tC18 cartridges or 96 wells elution plates (Waters), desalted, and eluted with 80% acetonitrile, 0.1% formic acid. After elution from the cartridges, peptides were dried in a vacuum centrifuge, resolubilized in 0.1 % formic acid, and analyzed by mass spectrometry.

**Phosphopeptide enrichment**

After the peptide clean-up step, each sample was diluted in 280 μl phtalic acid (PA) solution (86.7 mg/ml PA, 80% acetonitrile, 3.5% trifluoroacetic acid (Thermo Scientific)) by vortexing and sonicating for 5 min. The peptide solution was centrifuged at 16,000 x g for 5 min to remove solid debris. The peptide solution was transferred to Mobicol spin columns containing titaniumdioxide (TiO2) beads (GL Science) that had been washed with 280 μl of methanol by vortexing in short pulses and centrifuged at 800 x g, and equilibrated with 280 μl of PA solution by vortexing in short pulses and centrifuged at 800 x g. Transfer of peptide solution to the spin columns was achieved as follows: the bottom of the Mobicol columns was closed with a small plug and the columns were vortexed to mix the peptide solution with the TiO2 beads and incubated for 30 min at room temperature under end-over-end rotation. After incubation, the beads were washed twice with 280 μl of PA solution (load, vortex, spin down at 800 g) followed by two washing steps with 280 μl of 80% acetonitrile, 0.1% trifluoroacetic acid, two washing steps with 280 μl of 40% acetonitrile, 0.1% trifluoroacetic acid and two washing steps with 280 μl of 0.1% trifluoroacetic acid. For the elution, 280 μl of 0.3 M NH4OH were added to the beads, which were then incubated for 3 min at room temperature and centrifuged at 800 x g. The elution was performed a second time with the same parameters. The solution was acidified immediately after elution with 40 μl of 25% trifluoroacetic acid. The cleanup of phosphopeptides was performed as described in the “tryptic digestion” step using Sep-Pak tC18 cartridges (Waters). Dried phosphopeptides were re-solubilized with 30 µl 0.1% formic acid prior to analysis by mass spectrometry.

**LC-MS/MS data acquisition**

Peptide digests for both LiP, control and phospho-enriched samples were analyzed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher) equipped with a nanoelectrospray ion source and a nano-flow LC system (Easy-nLC 1000, Thermo Fisher) while peptide digests of the *in vitro* samples were analyzed on an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher) equipped with a nanoelectrospray ion source and an UPLC system (ACQUITY UPLC M-Class, Waters).

For shotgun LC-MS/MS data dependent acquisition (DDA), 1 μl peptide digests from each biological replicate of LiP, control and phospho-enriched samples were injected independently at a concentration of 1 μg/μl. 1 μl of the same samples were also measured in data-independent acquisition (DIA) mode. Peptides were separated on a 40 cm x 0.75 µm i.d. column packed in-house with 1.9 µm 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 using a linear LC gradient from 5% to 25% or 5% to 35% acetonitrile, respectively, over 120 min and a flowrate of 300 nL/min and the column was heated to 50°C.

For DDA measurement on the Orbitrap Q Exactive Plus, MS1 scans were acquired over a mass range of 350-1500 m/z with a resolution of 70,000. The 20 most intense precursors that exceeded 1300 ion counts were selected for collision induced dissociation and the corresponding MS2 spectra were acquired at a resolution of 35000, 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.

For DIA measurements 20 variable-width DIA isolation windows were recursively acquired. The DIA isolation setup included a 1 m/z overlap between windows, as described in Piazza et al. 2018. DIA-MS2 spectra were acquired at a resolution of 17500 with a fixed first mass of 150 m/z and an AGC target of 1 x 106. To mimic DDA fragmentation, normalized collision energy was 25, 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. A survey MS1 scan from 350 to 1500 m/z at a resolution of 70,000, with AGC target of 3 x 106 or 120 ms injection time was acquired in between the acquisitions of the full DIA isolation window sets.

*In vitro* LiP-MS experiments were acquired with an Orbitrap Fusion tribrid mass spectrometer. For DDA measurements, MS1 spectra were acquired from 300 to 1500 m/z at a resolution of 120,000. Survey spectra were scheduled for execution at least every 3 s, with the embedded control system determining the number of MS/MS acquisitions executed during this period. Precursors were selected for higher-energy collision dissociation and the corresponding MS2 spectra were acquired at a resolution of 30,000, collected for maximally 54 ms. All multiply charged ions were used to trigger MS-MS scans followed by a dynamic exclusion for 25 s. Singly charged precursor ions and ions of undefinable charged states were excluded from fragmentation.

The DIA acquisition method for the samples acquired on the Orbitrap Fusion consisted of a survey MS1 scan from 300 to 1500 m/z at a resolution of 120,000, with AGC target of 4 x 105 or 50 ms injection time, followed by the acquisition of 20 variable-width DIA isolation windows. The DIA isolation setup included a 1 m/z overlap between windows.DIA-MS2 spectra were acquired at a resolution of 30,000 with a fixed first mass of 150 m/z and an AGC target of 5 x 104. To mimic DDA fragmentation, normalized collision energy was 28, 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 and protein identification and spectral library generation**

The collected DDA spectra were searched against the *E. coli* (strain K12) Uniprot fasta database (version October 2017) using the SEQUEST HT® database search engine (Thermo Fisher Scientific). Up to two missed cleavages were allowed, cleavage of KP and RP peptide bonds were excluded. For LiP samples a semi-specific tryptic digestion rule type was applied. Cysteine carbamidomethylation (+57.0214 Da) and methionine oxidation (+15.99492) were allowed as fixed and variable modifications, respectively. In case of phosphopeptide search, the phosphorylation of serines, threonines and tyrosines (+79.966 Da) was defined as a variable modification. 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 Proteom Discoverer (version 2.2, Thermo Fisher Scientific), which correspond to a filter for 1% FDR on peptide level.

DDA spectra were searched with Proteome Discoverer 2.2 as described above and the software Spectronaut (Biognosys AG, version 13) was used for spectral library generation. The spectral libraries contained normalized retention time iRT values for all peptides.

**LiP-MS data analysis: peptide quantification and statistical analysis**

Targeted data extraction of DIA-MS acquisitions was performed with Spectronaut (Biognosys AG, version 13) with default settings, using spectral libraries generated as described above. Briefly, the dynamic mass tolerance strategy was applied to calculate the ideal mass tolerances for data extraction and no correction factor was applied (correction factor = 1). The local (non-linear) regression method was used for iRT calibration using the iRT kit peptides. The mutated decoy method was used to generate label-free decoys. Interference correction was enabled to exclude fragment ions with interferences from quantification across all runs but keeping at least three fragments for quantification. In the LiP samples only fully and semi tryptic peptides that were uniquely present in the sequence of one protein of the database (proteotypic peptides) were used for quantification while in control samples only fully tryptic proteotypic peptides were used for protein quantity. The false discovery rate (FDR) was estimated with the mProphet approach (Reiter et al., 2011) and set to 1 % on peptide and protein level. Protein inference was performed using the implemented IDPicker algorithm to define protein groups (Zhang et al. 2007). Comparison analysis of protein (control samples) and peptide (LiP samples) levels was performed with the MSstats package (Choi et al. 2014). Spectronaut normalized peak areas were used as intensity values. Data were then processed with the “dataProcess” function which includes logarithm transformation with base 2 of intensities, median normalization, feature selection (all fragment ions in the dataset were selected) and imputation of missing values by AFT (accelerated failure time model). The “groupComparison” function using linear mixed-effects model was finally used to compare peptide and protein abundances between conditions. For each conditional comparison, MS stats provides model-based estimates of fold changes as well as p-values that are adjusted for multiple testing (q-values) using the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Significant protein abundance changes were used to correct LiP-peptide abundance changes (LiP samples) by dividing peptide-level abundance ratios by the significant abundance ratio of the respective protein. For proteins that did not significantly change abundance, a normalization factor of 1 was used (i.e., no correction). In the *E. coli* dataset, generated by exposing cells to the long-term metabolic perturbation which largely altered protein and peptide abundance levels, we used stringent cutoffs (|log2FC| >2, q-value < 0.05) to select for significant changes. In the *S. cerevisiae* datasets in contrast, where few proteins changed abundance due to the short perturbation time, the following cutoffs were applied: |log2FC| >1, q-value < 0.05).

**Phosphopeptide quantification and statistical analysis**

DDA data relative to phosphopeptide-enriched samples were analysed with Progenesis QI (Nonlinear Dynamics, version 2.0). Raw LC-MS/MS files were imported into Progenesis for MS1 feature alignment using the automatic alignment algorithm followed by manual revision and adjustment of the aligned chromatograms. Peak picking was then performed setting the maximum ions charge to 5. Peptide ion abundances were normalized using an automatically selected run as normalisation reference to allow comparisons across the different samples. An analysis of variance (ANOVA) was applied to all peptide ions. MS/MS spectra were exported in the .mgf format and searched against a yeast database using Proteome Discoverer (Thermo Fisher Scientific, version 2.2), as described before. The resulted pepXML files containing peptides identified in the search were filtered with a false discovery rate (FDR) of 1% and imported into Progenesis. Peptide identifications from MS/MS spectra were mapped to the corresponding peptide ions detected in MS1 spectra, according to their accurate m/z and retention time and areas under the extracted ion chromatograms. The list of quantified peptide was exported in the .csv format. The R-framework based analysis tool SafeQuant (version 2.3.1) was used for statistical validation of differentially expressed phosphopeptides during osmotic stress. Peptide abundance values were used for statistical testing of differentially abundant phospho-peptides using an empirical Bayes moderated t-test as implemented in the R/Bioconductor limma package (Smyth et al., 2004). Resulting p-values were finally adjusted by multiple testing using the Benjamin-Hochberg method (Benjamini and Hochberg, 1995). The output of this statistical analysis was filtered using the following cutoffs: q-value < 0.05 and |log2FC|>1. Moreover, phosphopeptides mapping to proteins for which we detected a significant change in abundance during osmotic stress (q-value < 0.05 and |log2FC|>1) were excluded from the analysis. Phosphorylation site abundances were calculated grouping peptides reporting the same phospho modification and calculating fold change as the mean of all the peptide fold changes relative to the same modification. p-sites with a coefficient of variation ([standard deviation](https://en.wikipedia.org/wiki/Standard_deviation) to the [mean](https://en.wikipedia.org/wiki/Mean)) higher than 0.2 were filtered out. To statistically assess the combination of the qvalues, a Fisher's combined probability test has been applied to combine q-values using the “combine\_pvalues” function of the open-source python-based Scipy library.

**Gene Ontology Enrichment Analysis**

The gene ontology (GO) enrichment analysis was performed using the GOATOOLS python-based library (Klopfenstein et al., 2018). The background set correspond to all *E. coli* (Uniprot reviewed proteome, downloaded on August 20th, 2018) or *S. cerevisiae* (NCBI protein table, downloaded on June 12th, 2018) proteins. Proteins significantly changing at abundance or structural levels were selected using the cutoffs reported in the “Peptide and protein identification, quantification and statistical analysis” session. The option "propagate counts" was set to “False” to avoid propagating the annotations of a gene from the assigned GO category to all parent GO terms. The p-values were calculated using Fisher's exact test and then adjusted for multiple testing using the Benjamin-Hochberg correction method (Benjamini and Hochberg, 1995). We next calculated the specificity of the enriched GO terms by computing their information content as follow: IC = −log (frequency), where frequency is the number of genes annotated to the current GO term divided by the total number of associations between genes and GO terms in the full branch. The semantic similarity between all the enriched GO terms was then calculated as the inverse of the semantic distance (number of branches separating the terms). The IC and semantic similarity values were finally used to filter the list of GO enriched categories selecting the more informative GO term (highest IC) among pairs of terms showing a semantic similarity higher than 0.5. To visualize the results of the functional enrichment analysis, GO terms with a number of associated genes bigger than 500 were excluded to reduce the complexity and the redundancy of the plot but preserve the biological outcome.

**Curve fitting with linear and non linear regression**

A linear regression analysis was performed to investigate the relationship between metabolic fluxes and LiP peptide changes for all enzymes of the CCM for which 13C-based metabolic fluxes were measured by Gerosa and collaborators (Gerosa et al., 2015). First, we converted the estimated absolute fluxes to logarithmic fold changes, comparing each of the seven growth conditions to glucose, the carbon source we used as reference in the proteomics analysis. To identify significantly changing fluxes, absolute fluxes and their standard deviations were used to perform a t-test comparing each condition to glucose. A cutoff of 0.05 was applied to p-values adjusted for multiple testing (Benjamini and Hochberg, 1995). LiP-MS data were used for the regression analysis if q-value < 0.05 and |log2FC| > 1 in at least 4 conditions. A least-squares regression analysis was performed with the “stats.linregress” function of the open-source python-based Scipy library. The Wald test was finally applied to calculate the p-value for a hypothesis test whose null hypothesis is that the slope is zero. Calculated p-values were then adjusted for multiple test correction using the using the Benjamin-Hochberg correction method (Benjamini and Hochberg, 1995). Regression models were selected using a cutoff of 0.05 for the adjusted p-value and 0.7 for R-squared.

The same analysis was repeated to investigate the relationship between protein structural changes (LiP fold-changes) and the concentration of 26 metabolites (Table 8) known to regulate the CCM. Measurements of metabolites concentration over the 8 growth conditions was previously reported in Gerosa et al., 2015. For the regression analysis ratios to glucose were used. Peptides showing a linearity with metabolite concentrations where finally classified as “known” if the metabolite-protein (to which the peptide map) was reported in the EcoCyc database (https://ecocyc.org/) or in the BRENDA database (http://www.brenda-enzymes.org/). Moreover, we reported if the same interaction was identified through the LiP-SMAP approach (Piazza et al., 2018) and/or ligand-detected NMR (Diether et al., 2019). Concentration-dependent structural effect curves for LiP peptides identified in the in vitro experiment were generated by plotting peptide abundance changes (log2FC) over the substrate concentration range. To investigate if LiP peptides followed a dose-response curve we selected peptides significantly changing (|log2FC| >2, q-value < 0.01) over at least 4 substrate concentrations. We then fit a higher-order polynomial regression using the “stats.linregress” function of the open-source python-based Scipy library with the parameter “order” set to 2 and select peptides following a hyperbolic curve. The lower limit of the peptide abundance change was set to zero to allow proper fitting.

**Enrichment analysis of chaperone proteins and aggregators**

We tested if chaperones proteins were over-represented in the *S. cerevisiae* datasets by performing a Fisher’s exact test using the “fisher\_exact” function of the open-source python-based Scipy library, with the alternative parameter set to “greater”. The list of 63 yeast chaperones reported in Gong et al., 2009 was used as reference set. The same analysis was repeated to test for the enrichment of aggregating or misfolded proteins in the heat shock dataset. The reference list of proteins aggregating or misfolding during heat shock was obtained from Wallace et al. We defined as “aggregator” or “misfolded” a protein which was pelletable after heat shock (Wallace et al., 2015). We finally assess the presence of chaperone-client relationships between the two groups of enriched proteins, chaperones and aggregators/misfolded proteins, using the list of yeast chaperone physical interactors retrieved from the BioGRID database (v3.5, <https://thebiogrid.org/>). We restricted the analysis to those interactions detected either by “affinity capture-MS” or by “affinity capture-western” experimental methods and performed the enrichment analysis as described above.

**3D analysis of protein structural alterations**

Selected LiP peptides from the different experiments were mapped to representative protein structures to investigate which regions of the different proteins (e.g. active site, allosteric site) were affected by the structural alterations. Structures of holocomplexes between protein and ligands were selected to position allosteric and active sites. For the yeast data dataset when only the apo- version of a protein was available homology models were built using holocomplexes from homologous proteins that have protein sequence similarity higher or equal to 30%. For those cases where experimental data was not available, we use the predicted active site residues as annotated in Uniprot for positioning metabolite binding. Using a custom-made PyMOL-Python script, we measured the minimal Euclidean distance in angstroms (Å) between all the atoms of the LiP peptide and those of the substrate or allosteric regulator, if present, or alternatively all the atoms of the peptide or amino acid defining the predicted active site (as reported in the Uniprot database). A peptide was assigned to a known functional site if the measured minimal distance was less than 6.4 Å, based on a previous study (Piazza et al., 2018).

**Data visualization and processing**

Data visualization, exploratory data analysis and processing were performed using Python (version 2.7 and 3.0) and the Python library Pandas (version 0.18.1). Heat-map diagrams, bar plots and regression model plots were created using the python data visualization library Seaborn (version 0.9.0).