Identification of Dynamic Protein-Metabolite Complexes in Saccharomyces cerevisiae Using Co-Fractionation Mass Spectrometry

Dennis Schlossarek

Univ.-Diss.

zur Erlangung des akademischen Grades
"doctor rerum naturalium"
(Dr. rer. nat.)
in der Wissenschaftsdisziplin „Biochemie“

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät
Institut für Biochemie und Biologie
der Universität Potsdam

Ort und Tag der Disputation: Potsdam, 19.01.2023

Hauptbetreuer: Prof. Dr. Dr. h.c. Lothar Willmitzer
Betreuer*innen: Prof. Dr. Aleksandra Skirycz
Gutachter*innen: Prof. Dr. Alisdair Fernie
Prof. Dr. Dirk Walther
Prof. Dr. Henning Urlaub

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-58282
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-582826

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbständig und unter Verwendung keiner anderen als den von mir angegebenen Quellen und Hilfsmitteln verfasst habe. Ferner erkläre ich, dass ich bisher weder an der Universität Potsdam noch anderweitig versucht habe, eine Dissertation einzureichen oder mich einer Doktorprüfung zu unterziehen.

Dennis Schlossarek Potsdam, 23.03.2022

Table of Contents

Abbreviations

2’,3’-cAMP	2',3'-Cyclic Adenosine Monophosphate
3’,5’-cAMP	3',5'-Cyclic Adenosine Monophosphate
3PGA	3-Phosphoglycerate
5'-UTR	5'-Untranslated Region
AA	Amino Acid
AD	Activation Domain
AMP	Adenosine Monophosphate
ANOVA	Analysis of Variance
ASV	Alternative Splicing Variant
ATP	Adenosine Triphosphate
BD	Binding Domain
BifC	Bimolecular Fluorescence Complementation
BPGA	Bisphosphoglycerate
CETSA	Cellular Thermal Shift Assay
CF-MS	Co-Fractionation Mass Spectrometry
CN-PAGE	Clear Native Polyacrylamide Gel Electrophoresis
CP	Core Particle (of the Ribosome)
cryo-EM	Cryo-Electron Microscope
CYS3	Cystathionine Gamma-Lyase
DAP	Dihydroxyacetone-Phosphate
DARTS	Drug Affinity Responsive Target Stability
DES	Dis-Elution-Score
DIF-FRAC	Differential Fractionation
DNA	Deoxyribonucleic acid
ESI	Electrospray Ionization
FAD	Flavin Adenine Dinucleotide
FC	Fold Change
FDR	False Discovery Rate

FMN	Flavin Mononucleotide
G3P	Glycerol-3-Phosphate
GAP	Glyceraldehyde-3-Phosphate
Gly/Glu	Glycolysis/Gluconeogenesis
GMP	Guanosine Monophosphate
IEX	Ion Exchange Chromatography
ItSA	Isothermal Shift Assay
K_d	Dissociation Constant
LC	Liquid Chromatography
LiP-SMAP	Limited Proteolysis-Small Molecule Mapping
mRNA	Messenger RNA
MS	Mass Spectrometry
MS/MS	Tandem Mass Spectrometry
MTBE	Methyl Tert-Butyl Ether
NAD	Nicotinamide Adenine Dinucleotide
NADPH	Nicotinamide Adenine Dinucleotide Phosphate
nanoDSF	Nano Differential Scanning Fluometry
NMR	Nuclear Magnetic Resonance Spectroscopy
NPARC	Non-Parametric Analysis of Response Curves
OD_{600}	Optical Denisty at 600nm
PB	Processing Body
PCA	Protein Fragment Complementation Assay
PCC	Pearson Correlation Coefficient
PDB	Proximity Dependent Biotinylation
PISA	Proteome Integral Stability Assay
PLP	Pyridoxal Phosphate
PMC	Protein-Metabolite-Complex
PMI	Protein-Metabolite Interaction
POI	Protein of Interest
PPI	Protein-Protein Interaction

PPP	Pentose Phosphate Pathway
PROMIS	Protein Metabolite Interactions Using Size Separation
PROMISed	Protein Metabolite Interactions Using Size Separation Easy Data Analysis
PTM	Posttranslational Modification
RNA	Ribonucleic Acid
RP	Regulatory Particle (of the Ribosome)
rRNA	Ribosomal RNA
SDG	Sucrose Density Gradient Centrifugation
SDS-PAGE	Sodium Dodecyl Sulfate
SEC	Size Exclusion Chromatography
SG	Stress Granule
siRNA	Short Interfering RNA
SLIMP	Supervised Learning of Metabolite-Protein Interactions
SPROX	Stability of Proteins from Rate of Oxidation
SWATH-MS	Sequential Window Acquisition of All Theoretical MS
TAP	Tandem Affinity Purification
TCA	Tricarboxylic Acid
TPP	Thermal Proteome Profiling
tRNA	Transfer RNA
UMP	Uridine Monophosphate
UPLC	Ultra-Performance Liquid Chromatography
V_{max}	Maximal Reaction Rate
Y2H	Yeast-two-Hybrid
YPD	Yeast Extract Peptone Dextrose

1. Summary

Cells are built from a variety of macromolecules and metabolites. Both, the proteome and the metabolome are highly dynamic and responsive to environmental cues and developmental processes. But it is not their bare numbers, but their interactions that enable life. The protein-protein (PPI) and protein-metabolite interactions (PMI) facilitate and regulate all aspects of cell biology, from metabolism to mitosis. Therefore, the study of PPIs and PMIs and their dynamics in a cell-wide context is of great scientific interest. In this dissertation, I aim to chart a map of the dynamic PPIs and PMIs across metabolic and cellular transitions. As a model system, I study the shift from the fermentative to the respiratory growth, known as the diauxic shift, in the budding yeast *Saccharomyces cerevisiae*. To do so, I am applying a co-fractionation mass spectrometry (CF-MS) based method, dubbed protein metabolite interactions using size separation (PROMIS). PROMIS, as well as comparable methods, will be discussed in detail in chapter 1.

Since PROMIS was developed originally for *Arabidopsis thaliana*, in chapter 2, I will describe the adaptation of PROMIS to *S. cerevisiae*. Here, the obtained results demonstrated a wealth of protein-metabolite interactions, and experimentally validated 225 previously predicted PMIs. Applying orthogonal, targeted approaches to validate the interactions of a proteogenic dipeptide, Ser-Leu, five novel protein-interactors were found. One of those proteins, phosphoglycerate kinase, is inhibited by Ser-Leu, placing the dipeptide at the regulation of glycolysis.

In chapter 3, I am presenting PROMISed, a novel web-tool designed for the analysis of PROMIS- and other CF-MS-datasets. Starting with raw fractionation profiles, PROMISed enables data pre-processing, profile deconvolution, scores differences in fractionation profiles between experimental conditions, and ultimately charts interaction networks. PROMISed comes with a user-friendly graphic interface, and thus enables the routine analysis of CF-MS data by non-computational biologists.

Finally, in chapter 4, I applied PROMIS in combination with the isothermal shift assay to the diauxic shift in *S. cerevisiae* to study changes in the PPI and PMI landscape across this metabolic transition. I found a major rewiring of protein-protein-metabolite complexes, exemplified by the disassembly of the proteasome in the respiratory phase, the loss of interaction of an enzyme involved in amino acid biosynthesis and its cofactor, as well as phase and structure specific interactions between dipeptides and enzymes of central carbon metabolism.

In chapter 5, I am summarizing the presented results, and discuss a strategy to unravel the potential patterns of dipeptide accumulation and binding specificities. Lastly, I recapitulate recently postulated guidelines for CF-MS experiments, and give an outlook of protein interaction studies in the near future.

Zusammenfassung

Die Zelle besteht aus einer Vielzahl von großen und kleinen Molekülen, und sowohl das Proteom als auch das Metabolom passen sich dynamisch den vorherrschenden Umweltbedingungen oder zellulären Anforderungen an. Allerdings ist es nicht die bloße Menge an biologischen Molekülen, sondern deren Interaktionen miteinander, die das Leben erst ermöglichen. Protein-Protein (PPI) und Protein-Metabolit Interaktionen (PMI) vollbringen und regulieren alle Aspekte der Zelle, vom Stoffwechsel bis zur Mitose. Die Studie dieser Interaktionen ist daher von fundamentalem wissenschaftlichem Interesse. In dieser Dissertation strebe ich an, eine Karte der Protein-Protein und Protein-Metabolit Interaktionen zu zeichnen, die den Übergang vom fermentativen zum respiratioschen Stoffwechsel in der Hefe *Saccharomyces cerevisiae* umfasst. Zu diesem Zweck nutze ich PROMIS (egl. protein metabolite interactions using size separation), eine auf der co-Fraktionierungs Massensprektrometrie (CF-MS) aufbauende Methode. PROMIS, und ähnliche Methoden zur Untersuchung von Protein-Interkationen, werden ausgiebig in Kapitel 1 vorgestellt.

Da PROMIS ursprünglich für die Modellpflanze *Arabadopsis thaliana* entwickelt wurde, beschreibe ich in Kapitel 2 zunächst die erste Anwendung der Methode in *S. cerevisiae*. Die Ergebnisse stellen eine Fülle an Protein-Metabolit Interaktionen dar, und 225 zuvor prognostizierte Interaktionen wurden das erste Mal experimentell beschrieben. Mit Hilfe orthogonaler Methoden wurde außerdem eine inhibitorische Interaktion zwischen dem proteinogenen Dipeptid Ser-Leu und einem Enzym der Glykolyse gefunden.

In Kapitel 3 präsentiere ich PROMISed, eine neue Web-Anwendung zur Auswertung von Daten von PROMIS oder anderen CF-MS Experimente. PROMISed kann genutzt werden um in rohen Fraktionierungs-Profile lokale Maxima zu finden, aus denen ein Interaktions-Netzwerk basierend auf Korrelationen erstellt wird. Außerdem kann die Anwendung Unterschiede in den Profilen zwischen verschiedenen experimentellen Bedingungen bewerten. PROMISed umfasst eine benutzerfreundliche grafische Oberfläche und bedarf daher keiner Programmierkenntnisse zur Nutzung.

In Kapitel 4 benutze ich schließlich PROMIS und ItSA (engl. isothermal shift assay) um PPI und PMI während des Übergangs vom fermentativen zum respiratorischen Stoffwechsel in Hefe zu untersuchen. Hier beschreibe ich eine zellweite Umbildung der Protein-Metabolit-Komplexe, bespielhaft beschrieben anhand des Auseinanderfallens des Proteasoms im respiratorischen Stoffwechsel, des Verlustes der Interaktion zwischen einem Enzym des Aminosäure Stoffwechsels mit seinem Cofaktor und spezifischen Interaktionen zwischen Dipeptiden und Enzymen des zentralen Stoffwechsels.

In Kapitel 5 fasse ich die gefundenen Ergebnisse zusammen und stelle eine Strategie zur Untersuchung der Spezifität sowohl der Bildung als auch der Protein-Interaktionen von Dipeptiden vor. Zu aller letzt

rekapituliere ich Richtlinien für CF-MS Experimente und gebe einen Ausblick auf die nahe Zukunft der Studien der Protein-Interkationen.

2. General Introduction

Parts of this chapter will be published in a review by Schlossarek and Skirycz (2021)

2.1 Proteins and Metabolites Act in Collaboration

Metabolites are small molecules with a large variety in their chemical structure, size and elemental composition, resulting in a diversity in charge, polarity and stability. The entirety of metabolites in a given sample is defined as metabolome [1], and yields information on the current state of the biological sample. The metabolome can differ significantly between tissues, cells and even organelles within the same organism. Moreover, it is highly dynamic and can change drastically during developmental transitions (eg. yeast metabolic cycle [2,3]) and in response to the environment. The study of the metabolome and its changes is called metabolomic. Metabolites fulfill crucial roles in all living cells, as they serve as energy sources, cellular building blocks and regulators, as well as intra- and extracellular signaling compounds. To do so, metabolites are part of continual chemical reactions, which convert metabolites into one another or incorporate them in macromolecules. The sum of these chemical reactions within a cell is called metabolism, and is largely carried out by enzymes.

Enzymes are predominantly proteins, although ribonucleic acids with enzymatic activities are known as well [4]. Here, I will focus on proteins. In contrast to metabolites, proteins are large macromolecules, varying in size and structure. Proteins are composed of 20 canonical, proteogenic amino acids, which residues can be post-translationally modified, further contributing to their notable chemical diversity. The proteome is defined as the entirety of a cell's proteins, and the study of the proteome and its changes is called proteomics. The statements made for the metabolome are also true for the proteome, in that it is highly dynamic and shows a great spatial variability (e.g. [5,6]). While some proteins are functional monomers, as it is the case for the yeast glyoxalase 1 (GLO1) [7], most proteins act as oligomeric holoenzymes, or are part of transient multi-protein complexes, and often incorporate metabolites as co-factors. These protein-metabolite-complexes (PMC) facilitate all aspects of live, from metabolism and cellular transport, signaling, protein biosynthesis and degradation, to cell division and mobility. These functions need to be dynamically and tightly regulated, to i) ensure an energy efficient steady state and avoid wasting resources, ii) rapidly adapt to dynamic environmental changes and iii) prepare and execute cellular processes, such as cell division.

The protein regulatory network is controlled at multiple layers, which are closely interconnected. For one, gene transcription and translation in combination with protein degradation regulate the abundance of proteins, in an interplay known as proteostasis [8]. Another layer are post-translational modifications (PTM), covalently attaching chemical groups (e.g., phosphorylation) metabolites (e.g., acetylation) or even

small proteins (e.g., ubiquitinoylation) to an amino acid residue, regulating protein activity independently of their abundance. The most prominent PTM is phosphorylation, which can directly affect enzyme activity or the binding affinities to protein- and/or metabolite interaction partners [9]. These interactions in turn act as an additional regulatory layer: Metabolites, for examples, can serve as allosteric regulators, transiently binding to a distal site from the active center, altering its reaction kinetics [10], and transiently binding to other proteins can result, for example, in altering a proteins subcellular localization, as it is the case for many signaling proteins. Moreover, the first two regulatory layers described above rely on protein-protein interactions as well: Protein degradation is facilitated by the direct interaction of a protein with, for example, the proteasome (which in itself is a protein-complex), and protein-phosphorylation requires physical contact to a kinase. While transcriptional-translational regulation is more prevalent in anticipatory processes, such as the cell cycle or the diurnal cycle, it is a relatively slow process, and on its own insufficient for the adaption to rapid and unpredictable environmental changes. These rapid adaptations are therefore often conducted by the fast assembly or disassembly of protein-metabolite-complexes (PMCs), or regulated by posttranslational regulations.

In this work, I will focus on the non-covalent protein-protein and protein-metabolite interactions (PPI and PMI, respectively) that are the basis of the formation of PMCs, and the dynamics of these interactions across varying metabolic and cellular states. Therefore, I will first give a few examples of known PPIs and PMIs to underlie their ubiquitous and dynamic nature. Then, I will discuss recent advances in untargeted, proteome-wide methods that allow us to study dynamic interactions in a "bird's eye view".

2.2 Interactions Drive All Aspects of Life

2.2.1 Protein Interactions

As introduced above, protein-protein interactions play key roles in all aspects of life, from protein biosynthesis to metabolism and signaling. PPIs can be either transient interactions, or result in the covalent attachment of one protein to another, as is the case for ubiquitination (reviewed in Swatek and Komander 2016 [11]) or sumoylation (reviewed in Celen and Sahin 2020 [12]). While both types of interaction have a profound impact on a protein's properties, here, I will focus on the various transient interactions. Strong, transient interactions can lead to the formation of stable multiprotein-complexes. Two proteins interacting in such a way are called a dimer, or an oligomer in the case of multiple proteins. If the oligomer is composed of multiple subunits, or monomers, of the same polypeptide chain, it is called homomeric. In contrast, a heteromer is an oligomer consisting of different polypeptides. In the following, I will demonstrate that

protein-complexes are dynamic and involved in a variety of cellular processes by giving brief examples of recent findings in the field of protein-interactions.

Transient Interactions Channel Metabolism

The formation of complexes is often observed for proteins that act in the same direction of a biosynthetic pathway. One concept that arose here is the metabolon, first defined by Paul Srere in 1985 [13]. Metabolons have been extensively reviewed in the past [14,15], and I will only give a short overview here. A metabolon is a transient complex of proteins catalyzing subsequent steps of the same metabolic pathway, increasing pathway efficiency by substrate channeling [13]. Many advantages of substrate channeling have been proposed, which include the increase in enzyme efficiency, or avoiding competing pathways and the release of cytotoxic intermediates [14]. While metabolons have been proposed to play part in many pathways, such as the mitochondrial electron transport chain [16], the branched-chain amino acid biosynthetic pathway in human mitochondria [17] or carotenoid-biosynthesis in *A.thaliana* [18], substrate channeling was only observed for a few. Following the definition by Paul Srere, *bona-fide* metabolons have been demonstrated for example for the glycolytic pathway [19,20] and the TCA cycle [21–23], underlining their critical role for central carbon metabolism. Moreover, recently, the substrate channeling between nine enzymes of purine biosynthesis could be shown in HeLa cells, demonstrating the existence of the purinosome in human cells [24].

Counter-Enzyme Complex Formation

In contrast to metabolons, interactions between proteins fulfilling antagonistic roles in metabolism are barely known. In fact, the arginase-ornithine transcarbamylase complex was the only known example of a counter-enzyme complex [25,26], until Jayaraman and colleagues described an intriguing mechanism for the regulation of glutamate metabolism in *Bacillus subtilis* [27]. Glutamate lies at the crossroad of carbon- and nitrogen metabolism, and is synthesized by the heterodimeric glutamate synthase GltAB, whereas glutamate dehydrogenase GudB catalyzes glutamate breakdown into α-ketoglutarate and ammonia. Unlike in most organisms, where GudB expression is tightly regulated, in *B. subtilis*, GudB is constitutively expressed. In contrast, GltA and GltB expression is upregulated in glutamate-limited conditions, where they form a heterodimeric holoenzyme. In addition, one GltAB-dimer binds to one of each of the three dimers of the GudB hexamer, creating a complex that is 1.6 mD big, almost half the size of the ribosome. The binding of GltAB effectively blocks access to the active center of GudB, therefore inhibiting glutamate breakdown and preventing a futile loop [27]. With their work, Jayaraman and colleagues demonstrate an intriguing novel mechanism of metabolic regulation.

The Ribosome as a Dynamic Protein-RNA-Complex

Apart from metabolic regulation, dynamic protein complex formations are pivotal for many fundamental cellular processes, such as mRNA translation by ribosomes. Ribosomes are large protein-RNA complexes, which are composed of two multimeric subunits. In *S. cerevisiae*, for example, the 40S subunit contains 32 proteins and one rRNA component, the 60S subunit contains three rRNAs and 46 proteins. For decades, the ribosome was mostly regarded as a homogeneous complex without any regulatory functions, although contradictory evidence is going back to the 1980s (extensively reviewed in Genuth & Barna, 2018 [28] ; Sulima & Dinman, 2019 [29]). By today, it is widely accepted that the composition of "the ribosome" is largely context specific, both in regards to its protein-subunits, as well as its rRNA components [28,29]. Conclusively, this differentially composed ribosomes show selective transcription of subsets of mRNAs, corroborating their regulatory role [30]. Moreover, context specific ribosomal heterogeneity is not randomly distributed across the ribosome, but spatially confined: In a recent cryo-EM study in *S. cerevisiae*, an altered protein composition at the tRNA entrance and exit sites was reported after a switch from a glucose- to a glycerol-containing medium [31]. In *A. thaliana*, using a combination of transcriptomics, proteomics and structural simulations, Martinez-Seidel and colleagues demonstrated a remodeling at the polypeptide exit tunnel, the P-stalk and ribosomal head in cold-acclimated versus control roots [32]. Although ribosomal heterogeneity may be implemented already during ribosome biogenesis, in yeast, changes in composition were observed as early as 30 minutes after switching the medium [31], demonstrating the possibility of transcription independent ribosomal rearrangements. In line, it was shown that ribosomal proteins are able to detach upon phosphorylation, as is the case for human L13a [33]. However, more work needs to be dedicated towards the understanding of the transient nature and dynamics of ribosomal heterogeneity.

2.2.2 Metabolite Interactions

Similar to protein-protein interactions (see above), proteins and metabolites can interact transiently, or a metabolite can be covalently bound to a protein in form of a post-translational modification (PTM). Metabolites used as PTMs vary in size and structure, spanning from phosphate, acetyl- and methyl-groups to larger metabolites such as glutathione or glycosides [34]. Although PTMs have a great regulatory impact, here, I will focus on transient protein-metabolite interactions. As the basic building blocks of the cell, metabolites serve as substrates and products of enzymatic reactions, requiring a transient interaction with their enzyme. Moreover, metabolites are able to bind to enzymes distinct from the active center, affecting their activity in a way known as allosteric regulation. Additionally, many enzymes require a coenzyme or prosthetic group, a metabolite that is directly involved in the chemical reaction catalyzed by the enzyme.

Thus, metabolites are not only subjects of metabolism, but actively play part in its regulation. Besides their involvement in, and regulation of metabolism, metabolites play a major role in intra- and extracellular signaling, and regulate many developmental processes in the form of chemical signals. In the next paragraphs, I will give brief examples of the various roles of metabolites and recent findings in the field of metabolite interactions, and will also discuss modes of actions of metabolites beyond the canonical specific protein-metabolite pair.

Multilevel Allosteric Regulation Provides Metabolic Robustness

Many proteins, including enzymes and transcription factors, can specifically bind metabolites that are not directly involved in the catalyzed reaction, but regulate the proteins activity in a mode known as allosteric regulation. Here, the regulating metabolite binds the protein at the allosteric site, which is distant from the active center, causing a conformational change of the protein, ultimately altering its activity. Allosteric interactions are found regulating many biological processes on both, the enzymatic as well as transcriptional level. In *Escherichia coli*, for example, except for glycine, alanine, aspartate, and glutamate, which are considered cheap to produce concerning energy investment [35], the biosynthetic pathways of amino acids are allosterically feedback-regulated by their end products [36]. For example, tryptophan inhibits the heterotetrameric anthranilate synthase trpDE, which catalyzes the first step of tryptophan [37]. Moreover, the expression of the trp-operon, containing all five enzymes required for the biosynthesis of tryptophan from chorismate, is regulated by TrpR. TrpR is constitutively expressed, and binds the trp-operator upon binding to tryptophan, resulting in transcriptional inhibition [10]. Moreover, expression of the trp-operon is regulated by attenuation, depending on the availability of $tRNA^{trp}$. Thus, the tryptophan biosynthesis, as well as other amino acid biosynthetic pathways, is allosterically regulated on multiple levels. In a recent study, Sander and colleagues explored the effect of removing the allosteric regulation of the rate limiting enzymes of 7 amino acid biosynthetic pathways, namely arginine, threonine, leucine, isoleucine, proline, histidine and tryptophan. The authors showed that, in mutants insensitive to allosteric regulation of enzyme activity, the corresponding amino acids accumulate, while enzymes of the respective pathways decrease in abundance [38]. Moreover, for all mutants, metabolic flux through the manipulated pathways was either unaffected, or higher than wild type. Following on this, by combining modelling-approaches and experimental data, the authors demonstrated that the multi-level allosteric regulation results in an overabundance of enzymes which are operating below maximum capacity [38]. Combined, this results in a robustness of the metabolic pathways to transcriptional fluctuations, e.g caused by transcriptional stochastic or environmental perturbations [38], highlighting the importance of allosteric regulation for cellular homeostasis and the ability to adapt to environmental changes.

Coenzymes Are Essential for Metabolism

While allosteric regulation is mediated by a metabolite interacting at a site distant from the active center, coenzymes bind directly at the active center, and play active parts in the catalyzed reactions. Different enzymes incorporate many different metabolites as coenzymes, including e.g. adenosine triphosphate (ATP), glutathione, biotin or pyridoxal phosphate (PLP), and the reactions they catalyze are as versatile as their structures. PLP alone is involved in a wide range of reaction-types, including decarboxylation, racemization and transamination [39]. For example, PLP is the known cofactor of the yeast cystathionine gamma-lyase (CYS3), which catalyzes the conversion of cystathionine to cysteine [40]. In fact, PLP is predominantly involved in amino acid metabolism and the biosynthesis of other coenzymes, such as NAD or biotin [41]. Moreover, in prokaryotes, 1.5 % of all genes encode PLP-dependent proteins [41], further underlying the prevalent role of coenzyme interactions between proteins and metabolites.

Intracellular Signaling: 2',3'-cAMP Drives Stress Granule Formation

Besides their role in regulation of enzymatic activity, metabolites serve as signals in response to internal and external cues, such as stress. Stress granules (SGs) are highly dynamic protein-RNA associations. Similar to processing bodies (PBs), SGs are membraneless, microscopic organelles within the cytoplasm composed mainly of proteins and mRNAs (reviewed in Riggs et al. 2020 [43]; Youn et al. 2019 [42]). SGs, as well as PBs, are formed by liquid-liquid phase separation and show a gel-like structure. While PBs are constitutively present, SGs become only detectable in certain cellular stress-conditions [42,43]. However, two independently performed proximity-dependent biotinylation studies revealed proximal interactions between SG components even in the absence of stress, suggesting the existence of stable or transient protein complexes that can be rapidly assembled to phase-separate into visible organelles upon stress [44,45]. While, in plants, the occurrence of SGs is well described under different stress conditions, not much is known about the mechanisms of SG formation. In *A. thaliana*, an interaction between 2',3'-cAMP and Rbp47b, a key component of SGs in plants was recently demonstrated [46]. 2',3'-cAMP is a positional isomer of the second messenger 3',5'-cAMP and is a product of RNA breakdown [47], accumulating in conditions of excessive mRNA degradation [48–50]. In addition to the physical interactions with known SG proteins, the addition of 2',3'-cAMP promotes stress granule formation [46], and triggers abiotic stress responses on the transcriptome, proteome and metabolome level [51]. Since 2',3'-cAMP specifically binds to the RNA-binding motifs RRM2 and RRM3 of Rbp47b, the authors hypothesized that 2',3'-cAMP facilitates Rbp47b aggregation similar to the binding of Rbp47b to RNA [46], placing 2',3'-cAMP as the bridge between RNA degradation and stress signaling, however this hypothesis requires further testing.

Metabolites as Unspecific Regulators of Protein Stability

While the vast majority of research has focused on specific PPIs and PMIs, the role of unspecific interactions is gaining more and more attention. The involvement of small molecules, especially osmolytes, on protein folding is long known [52–56]. Many metabolites, such as arginine [57] and proline [56], as well as other amino acids [57–59], glycerol [56,60–62], glycine betaine and trehalose [56] have been shown to affect protein folding kinetics and thermodynamics *in vitro*. In a recent study, Verma and colleagues showed that changes in a cell's metabolic state alters proteostasis and protein functions [63]. Intriguingly, this alteration is due to differences in the chaperoning potential of accumulating metabolites, and independent of the canonical, protein-based chaperone machinery [63]. Moreover, unlike test tubes, in which proteins are usually investigated in heavily diluted, isolated systems, the cell is a crowded place. *In vivo*, the total concentration of macromolecules can reach 50-400 mg/ml, drastically confining the space available to one molecular species in an effect termed molecular crowding (as reviewed in Minton 2001 [64]). Molecular crowding not only affects protein folding kinetics and interactions, but can directly alter enzyme activity as well [64,65]. Thus, the cellular milieu, similarly to specific interactions, affects all aspects of a proteins existence, including proteostasis, folding, interactions and activity.

Beyond Proteins: Metabolic Control of Riboswitches

Before the emergence of proteins, according to the RNA-world theory, it is hypothesized that RNAs fulfilled a dual role as both, information storage as well as biological catalyzers (e.g. reviewed in Bernhardt 2012 [67]; Higgs and Lehman 2015 [66]). Remnants of this RNA-world can still be found in modern organisms, for example in the ribosome, where RNA makes up around 50 % of their mass, or in the presence of riboswitches. Riboswitches are structural elements found in the untranslated-regions (UTR) of mRNAs, forming selective small molecule bindings sites, which confer structural changes to the mRNA upon ligand binding [68–71]. Riboswitches are best known from prokaryotes, where, for example, the cofactors belonging to the group of coenzyme B12 and thiamine pyrophosphate regulate gene expression by binding directly to riboswitches [68–71]. In plants, the first riboswitch was demonstrated regulating thiamine biosynthesis via transient interactions with thiamine pyrophosphate [72,73]. Recently, a riboswitch-mechanism in the regulation of carotenoid biosynthesis in *A.thaliana* was proposed [74]. Phytoene synthase (PSY) catalyzes the rate limiting step in carotenoid biosynthesis, forming phytoene. In contrast to most plant families, which bear multiple PSY genes with differential regulation, Arabidopsis has only one PSY gene. However, AtPSY mRNA is present in two different splice variants, ASV1 and ASV2, which differ only in their 5'-UTR. ASV1 bears a structural confirmation that is predicted to swap between a translation-inhibiting and a translation-permissive confirmation. The ASV2 5'-UTR is constitutively translation-permissive [74].

Moreover, while no proteins were associated with the ASV1 5'-UTR, increased carotenoid-biosynthetic flux downregulated translation of the ASV1 variant. Thus, the authors hypothesize that the ASV1 5'-UTR serves as a flux-sensing riboswitch by interacting with an apocarotenoid to downregulate AtPSY translation in a negative feedback-loop [74]. However, to this date, the identity of this apocarotenoid-signal is still unknown.

2.3 Methods to Untangle the Protein-Interaction Network

As demonstrated, protein-protein and protein-metabolite interactions and their dynamics play major roles in metabolism, proteostasis and responses to external cues, such as stress. Consequently, the goal of research has shifted from the mere abundance-focused study of metabolites and proteins, to addressing functions and mode of actions by asking "who with whom", leading to the creation of large interaction networks. While many methods exist that rely on chemical or genetic screenings, I will focus here on biochemical approaches. Historically, interaction networks have been built from combining the results of targeted interaction studies, meaning experiments restricted to one protein of interest (POI) and its interactors. Therefore, these networks are often biased towards a number of well-studied proteins and metabolites. In contrast, the advent of untargeted, mass spectrometry based proteomics and metabolomics allows for a more holistic view. In the next sections, I will discuss recent advances in the methods that allowed us to uncover those interaction networks. I will first give a brief overview of the traditional, targeted methods for the identification of binary PPIs and PMIs, focusing on protein fragment complementation assays (PFC) and proximity dependent biotinylation (PDB) assays. Then, I will focus on global, untargeted approaches utilizing different biochemical strategies to create a snapshot of the protein-interaction network, namely thermal proteome profiling (TPP) and co-fractionation mass spectrometry (CF-MS). As will be discussed below, PDB, as well as TPP and CF-MS methods allow for a "birds-eye view" on the dynamics of protein-interactions, all with their own strengths and weaknesses.

2.3.1 Targeted Approaches for Studying Interactions

Ligand-Detected Nuclear Magnetic Resonance Spectroscopy

A recent method for the binary identification of protein-metabolite interactions is ligand-detected nuclear magnetic resonance spectroscopy [67,68]. Here, a purified POI is incubated with a complex mixture of putatively interacting metabolites, and interaction is detected by a depletion in the metabolites NMR spectrum. In a recent study, the Sauer lab was able to identify 22 known and 76 new interactions between

28 *E.coli* proteins involved in central carbon metabolism and 55 metabolites [69]. Ligand-detected NMR is an elegant way to identify protein-metabolite interactions, but its binary nature – testing one-protein to one metabolite – requires extensive effort to employ large-scale screenings.

Tandem Affinity Purification

Tags can be small epitopes, enzymatic cleavage-sites, protein-domains, functional proteins, or a combination of all, which are fused to the protein of interest (POI). In tandem affinity purification (TAP), a tag is used to immobilize the POI on a matrix, either by using antibodies or a streptavidin-biotin system. After washing away the unbound lysate, the POI is eluted from the matrix, and molecules bound to the POI are identified. TAP is well established to study protein-protein-interactions, and, for example, has been used to unravel the interplay between the light-signaling pathway and mRNA processing in *A.thaliana* [78]. Moreover, by simultaneously extracting proteins and metabolites from the eluate, TAP has recently been extended to study protein interactions with lipids [79] and metabolites [46,80].

Protein Fragment Complementation Assays

Another wide-spread use of protein-tags are protein fragment complementation assays (PCA). In PCAs, a reporter protein is split in two inactive fragments. To measure protein-protein interactions, the POI ("bait") is tagged with one fragment, and putative interactors ("prey") with the other fragment, respectively. If bait and prey interact, both fragments come in contact, and the reporter-proteins function is reconstituted and can be quantified. The most common PCA methods are bimolecular fluorescence complementation (BiFC), and yeast-two-hybrid (Y2H) assays. BiFC uses a split fluorescent protein, and can be used not only to study interactions *in vivo*, but also determine their subcellular localization [81]. Y2H uses a transcription-factor, which is split into its DNA binding domain and activation domain, and activates the transcription of a reporter gene when reconstituted [82]. The two domains are fused to the POI and a library of putative interaction partners, respectively, to screen for their interactions in yeast cells. Originally limited to soluble proteins, the method has been subsequently extended to cytoplasmic and membrane-bound proteins. The combination of all three methods proves to be a powerful approach towards protein-protein interactions: Recently, Welsch and colleagues used a combination of affinity purification, BiFC and the split-ubiquitin assay – a modified Y2H – to demonstrate that the phytoene synthase (AtPSY) is a target of degradation by the plastidial Clp-protease in *A.thaliana* [83]. Moreover, it was shown that the soluble AtPSY interacts with the membrane-bound orange protein (AtOR) [84], which protects AtPSY from Clp-mediated degradation and presumably recruits it to the plastidial inner-membrane, where carotenogenesis takes place [83].

However, all of these methods come with certain drawbacks: Since NMR-based methods analyze proteins in isolation, experiments are limited by the possibility to purify functional proteins, which also reduces the potential for high-throughput applications. The same is true for TAP. Also, both TAP and Y2H are known to create a high number of false positive readouts through unspecific binding with the matrix (TAP) or the tendency of the fragmented protein-tags to complement (Y2H). Moreover, TAP only detects high affinity and stable interactions, and is not suitable for insoluble or membrane proteins. While these can be measured in Y2H, the use of ADs and BDs does not allow to measure interactions of transcription factors. In addition to these limitations, targeted approaches can only always give insight into the binary interactions of one protein or metabolite at a time, although these methods can be upscaled to achieve high-throughput, screening against whole libraries of potential interaction partners (as discussed in Suter and Walker 2009 [85]).

Probing the Neighborhood – Proximity Dependent Biotinylation

A powerful approach to study protein-protein interactions *in-vivo* are proximity labeling methods, such as proximity dependent biotinylation (PDB). PDB was originally developed for the use in animal cells [86], but has recently been adapted to several model plants as well [87]. The method has been extensively reviewed in the past [88]. In short, using a flexible linker, the bait protein is fused to a biotin ligase, which is able to covalently attach biotin to proteins that come in close proximity to the bait-protein. After cell lysis, biotinylated proteins are selectively enriched using high affinity streptavidin-biotin pairing, similar to TAP, and identified using LC-MS/MS based proteomics. Since the method has first been reported, several different biotin ligases have been developed, showing different strengths and weaknesses. APEX2 for example is able to perform labeling within a minute, but requires the toxic compounds biotin-phenol and H_2O_2 as substrates [89]. In contrast, BioID2 uses non-toxic biotin as a substrate, but labeling takes up to 18-24 hours [90]. Moreover, both enzymes work optimally at 37 °C, making them unsuitable for many applications outside animal cells and microbiology. Two recently developed ligases, TurboID and miniTurbiID, combine fast labeling time and low toxicity, and are fully active at room temperature (25 °C), opening the field of PDB to a broader range of researchers, such as plant scientists [91].

PDB is able to identify stable, as well as weak, transient and even hydrophobic protein-protein interactions. Moreover, biotinylation occurs *in vivo* and at physiological conditions, allowing for labeling in the proteins subcellular localization without the disruption by cell-lysis [88]. Thus, PDB can create a snapshot of the bait's neighboring proteins and can chart interaction networks for otherwise hard to probe subcellular localizations, such as stress granules [44,45]. However, as with TAP and PCA approaches, one major drawback of PDB is the need to introduce a functinal protein-tag, in this case an active biotin ligase.

The tagging itself may disrupt the folding, function and localization of the tagged protein, so that the functionality of the bait protein as well as the ligase has to be confirmed experimentally. Moreover, biotin is an endogenous metabolite and serves various functions, for example as a co-factor in gluconeogenesis, fatty acid synthesis and amino acid metabolism (e.g. reviewed in León-Del-Río 2019 [92]). Hence, the external addition or depletion of biotin may cause major physiological problems [88]. Additionally, biotinylation requires negatively charged amino acid residues or lysine, depending on the ligase. If those residues are not available on the prey-protein's surface, labeling will not occur, leading to false positives [88].

Lastly, while PDB has successfully been implemented to study protein-protein interactions, to my knowledge, it has not been applied to protein-metabolite interactions to this date. After lysis, biotinylated proteins are specifically enriched using the high-affinity streptavidin-biotin interaction, similar to TAP, and eluted using high concentrations of free biotin or digested directly on the beads. In a similar approach to previous TAP experiments identifying protein-metabolite interactions from native cell lysates [79,80]; protein-bound metabolites could be extracted from the eluate using MTBE extraction [80]. To avoid overloading of the metabolomic samples with biotin, elution could be achieved by enzymatically cleaving streptavidin at a previously introduced cleavage site, similar to the TAP-tag. This would allow for the identification of not only the protein-neighborhood of a POI, but also their metabolite partners, and could provide an insight into subcellular protein-metabolite communities.

2.3.2 Untargeted Methods to Chart Interaction-Networks

Stronger together – Thermal Proteome Profiling

For a long time, chemical genetic screenings have guided the discovery of novel small molecule ligands. However, chemical screenings are limited to a measurable phenotype, followed by forward genetic approaches to unravel the true protein partner of the bioactive compound. Instead of measuring a phenotypic response on the cellular or organismal level, a recently developed set of methods measures the changes in protein properties upon ligand addition to infer protein-ligand pairs. For example, SPROX utilizes the differences in the rate of oxidation [93–95], while in DARTS and LiP-SMAP [96,97], the susceptibility to proteolysis is measured in the presence and absence of a small molecule. Additionally, LiP-SMAP also allows the identification of putative metabolite binding-sites [97,98]. Another commonly tested protein property is thermal stability (TS). In the cellular thermal shift assay (CETSA) [99] and the thermal proteome profiling (TPP) [100–102] a cell culture or lysate is subjected to a range of denaturing temperatures in the presence or absence of an external ligand. After temperature treatment, samples are centrifuged to sediment denatured and aggregated proteins. A protein melting curve is obtained by measuring the protein abundance

in the supernatant using LC-MS/MS based proteomics (TPP) or western-blotting (CETSA), and melting temperatures are compared between treated and control samples.

In their initial work, Savitski and colleagues demonstrated the proteome-wide effect of several drugs on protein thermal stability. For example, they showed that staurosporine and GSK3182571, two broad-spectrum kinase inhibitors, affect the thermal stability of many kinases, as well as proteins interacting with kinases [100]. Moreover, ATP dependent proteins are less stable in cellular extracts compared to intact cells, and the addition of ATP to extracts increases thermal stability of those proteins [100]. In a follow-up study, the Savitski Lab applied a 2D-TPP approach, utilizing different concentrations of ATP in a physiological relevant range. Here, they demonstrated diverse regulatory roles of ATP in a concentration-dependent manner: In sub-millimolar concentrations, ATP primarily serves as a substrate and allosteric regulator to ATP-dependent proteins, while it modulates protein-complexes and protein-DNA associations, as well as the solubility of disordered proteins within membraneless organelles in low millimolar concentrations [103]. Besides ATP, novel functions for one of its derivates, the second messenger 3'5'-cAMP, have been described in plants, where its role is still enigmatic: In a recent preprint, using TPP, an involvement of 3',5'-cAMP in the regulation of the actin cytoskeleton in *A.thaliana* was reported, presumably mediated by calcium signaling [104].

Although initially developed to identify protein targets of small molecules, TPP has been rapidly adapted to investigate protein-protein complexes as well: Proteins that are part of the same stable complex tend to coaggregate upon heat denaturation, resulting in similar thermal stabilities, and putative complexes can be inferred from the protein's melting curves [105]. Thus, in the past years, a growing number of "meltomes" across many species, including model organism such as humans [106,107], mouse [107], fruitfly [107], zebrafish [107,108], *C. elegans* [107], *S. cerevisiae* [107,109], *A. thaliana* [107,110] and *E. coli* [107,111,112], as well as pathogens, such as *Toxoplasma gondii* [113], *Trypanosoma cruzi* [114], and more [107] have been created to be mined for novel protein complexes. These datasets also span multiple conditions, such as cell cycle stages [106], or even responses to infection with the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) [115]. They can also serve as a baseline of proteome thermal stability of an organism, and can be used to address the susceptibility to thermal degradation of a POI in an early stage of a project [107]. Moreover, TPP experiments have been applied to mutant libraries of both, *E. coli* [112] and *S. cerevisiae* [109]. In yeast, mutantTPP (mTPP) was used to investigate the mode of action of missense mutations of two proteasomal subunits, the core-subunit PUP2 (*pup2-ts*) and the regulatory subunit RPN5 (*rpn5-ts*). Intriguingly, mTPP revealed different modes of action for the two mutants: In *rpn5-ts*, the proteasomes thermal stability is not affected, hinting to perturbed interactions outside of the proteasome, putatively causing a mislocalization of the regulatory subunit [109]. In *pup2-ts*, the proteasomal core-particle, but not the regulatory particle, shows

significantly reduced thermal stability, suggesting a disruption of the complex by the incorporation of the mutant PUP2 protein [109], pointing out the strength of TPP to capture protein-protein interactions.

While most proteins thermal stability is easily accessible by TPP, Jarzab and colleagues found large differences in the average melting temperature between cell types, tissues and species in their "meltome-atlas", spanning 13 eukaryotic and prokaryotic species and several different human cell lines and tissues [107]. Indeed, many factors have been identified to affect the susceptibility to thermal proteome profiling. Not surprisingly, the amino acid sequence has the largest influence on thermal stability, notable the frequency of polar amino acids [107], as well as protein structure, especially the presence of disordered regions [116], although the exact relationship is still unknown [107]. Moreover, a weak, but statistically significant negative correlation between protein size and thermal stability has been observed in many studies [107,111,116], in addition to a weak, but again significant, correlation to protein concentration [107,116]. Furthermore, protein phosphorylation affects thermal stability [117], especially at those phosphosites that are likely to be biologically relevant [118,119]. Lastly, while proteins may be susceptible to denaturation in higher temperatures, they may still be soluble and resist aggregation, as has been discussed for human body-fluid proteins, and therefore appear to be non-melters in a classical TPP experiment [107].

Since its introduction in 2014, the thermal proteome profiling has seen a great number of applications and advancements, as reviewed in [102]. For example, ultracentrifugation was initially used to remove aggregated proteins, which has been replaced by filter separation using a benchtop centrifuge. Most recently, the application of vacuum separation for this purpose [115] leads to a decrease in sample preparation time and allows for the full protocol to be performed in 96-well filter plates, which makes it possible to automate [102]. However, as for many proteomic studies, LC-MS/MS measurements are often limiting, both in time and resources. Therefore, two recent methods, namely proteome integral stability alteration (PISA [120,121]) and the isothermal shift assay (ItSA [122]), aim to reduce the sample size for LC-MS/MS measurements by simplifying the classical TPP approach. In PISA, a sample is subjected to a range of temperatures, similar to TPP, but is pooled before labeling, digestion and LC-MS/MS measurements. Thus, instead of measuring a melting curve, the curves integral is measured as the total protein abundance in one pooled sample [120]. Moreover, Li and colleagues reported that the sensitivity of PISA can be further increased by restricting the sampling to a range of temperatures around the average melting temperature of the sample [121]. The independently developed ItSA is taking this approach one step further: Instead of pooling multiple samples, the differences in the fraction of soluble proteins is measured in a single elevated temperature close to the average melting temperature [122]. Both approaches avoid the need of curve fitting, and allow for higher number of replications, thereby increasing statistical power. In their work, Ball and colleagues demonstrated a similar sensitivity of ItSA compared to Savitskis initial TPP experiment in identifying

staurosporine targets with a significantly reduced time investment [122]. Although, to my knowledge, neither PISA nor ItSA have been used to study protein-protein interactions, I expect that thermal proximity coaggregation [105] will be also observable in a correlated decrease of soluble proteins in these single-sample approaches. Furthermore, I envision that the application of PISA/ITSA to a 96-well plate format by utilizing vacuum separation, will allow for high-throughput, automated assays for protein-protein as well as protein-metabolite interactions. Therefore, perturbations in thermal stability could be readily studied in response to drug libraries, in mutant populations or different growth conditions.

Don't let go – Co-Fractionation Mass Spectrometry

As discussed above, thermal proteome profiling and its derivations are powerful methods for the investigation of PPIs and PMIs. Another untargeted method for this purpose is co-fractionation mass spectrometry (CF-MS). Unlike TPP, CF-MS has been originally developed to study protein-protein interactions [123], and was subsequently adapted to include protein-metabolite interactions as well [124–127]. Co-fractionation mass spectrometry relies on the chromatographic separation of native protein-metabolite-complexes (PMCs) and the identification of proteins and metabolites in the resulting fractions by mass-spectrometric analysis, therefore creating fractionation profiles (also called elution profiles) of all identified molecules. Proteins and metabolites of the same complex elute in the same fractions, referred to as co-elution or co-fractionation, and therefore putative interactions can be derived from profile-similarity. Different methods have been applied to fractionate PMCs prior to MS-analysis, most commonly size-exclusion chromatography (SEC) and ion-exchange chromatography (IEX), but also colorless native polyacrylamide gel electrophoresis [128], sucrose density gradient centrifugation [129] and size filtration [125,126] have been applied.

Since its first implementation in 2008, CF-MS has been used to create large protein-interaction networks in a number of organism, including humans [130], *M. musculus* [131], *S. cerevisiae* [123,124,132,133], *C. thermophilum* [134], *A. thaliana* [128,129,135] and *Synechocystis* [136], also comparing different cell-lines and tissues [130,131] and cellular states [137]. In addition, comparing co-fractionation data of different organism has led to uncover evolutionary conserved protein-complexes in plants [138] and metazoans [139]. Moreover, by reanalyzing 206 published CF-MS datasets, Skinnider and colleagues created an evolutionary tree of protein interactions across 27 eukaryotic species [140] and proclaimed a set of guidelines for future CF-MS experiments (see chapter 6). Moreover, in an elegant study using CF-MS on RNAse treated and control cell lysates, Mallam and colleagues showed that 20 % of known protein complexes are associated with RNA in human [141].

Besides protein-protein interactions, CF-MS has quickly been expanded to protein-metabolite interactions: In 2012, Chan and colleagues first applied IEX separation on protein extracts in the presence and absence of a small-molecule, showing that proteins and their bound small-molecules co-elute [124]. In 2015, it was shown that external small molecules binding to a recombinant protein can be separated from unbound molecules by single step size-filtration [125]. Finally, Veyel and colleagues applied this principle to the untargeted study of protein-metabolite interactions in native *A.thaliana* cell-lysates using size filtration and SEC [126,127] in an approach dubbed protein-metabolite interactions using size separation (PROMIS). PROMIS has recently also been applied to the fungi *S. cerevisiae* [133] and *C. thermophilum* [134]. In the original study, PROMIS identified hundreds of protein-bound metabolites, such as the proteogenic dipeptide Tyr-Asp, which was identified co-eluting with a number of glycolytic proteins [127]. Downstream analysis using affinity purification and nano differential scanning fluometry (nanoDSF) showed that Tyr-Asp interacts with glyceraldehyde-3-phosphate dehydrogenases (GAPC) from *A.thaliana* and humans [127]. Moreover, Tyr-Asp was shown to inhibit GAPC in *A. thaliana* and tobacco seedlings, leading to a rewiring of metabolism towards NADPH production, increasing the plants tolerance to oxidative stress [142]. Therefore, by applying CF-MS, novel regulatory mechanism for poorly studied small molecules, such as dipeptides, can be unraveled. However, PROMIS, as most untargeted metabolomics studies, is limited by the highly incomplete characterization of metabolomes [143], and, by judging the wealth of unknown metabolic features found co-eluting with proteins, many more small molecule regulators are expected to be found in the future. Moreover, a single protein or metabolite can co-elute with hundreds of other molecules, of which only one might be a true interaction partner. While the problem of "chance-co-elution" can be minimized by combining different separation techniques [132], co-elution is only an indication, rather than a proof of interaction and orthogonal approaches must the applied to validate any potential interaction.

2.4 Aim of this Thesis

As shown above, protein-protein and protein-metabolite interactions play major roles in all cellular processes, including metabolism, translation or the response to external cues. However, although much effort has been put into unraveling those interactions in the last decades, there is still a lot to be uncovered. This is especially true for protein-metabolite interactions. With the development of TPP and CF-MS, especially PROMIS, we now have tools at hand that enable us to investigate protein-protein-metabolite interactions in a global, label-free and untargeted manner.

The aim of this study was to adapt PROMIS, which has been previously developed in our group, to the model organism *Saccharomyces cerevisiae*, and to generate a protein-metabolite interaction map, including poorly functionally characterized small molecules, such as dipeptides. Following on this, we aimed to track the dynamic changes in the protein-protein-metabolite interactome across the diauxic shift in *S. cerevisiae*, which is characterized by a major metabolic rewiring from the fermentative growth on glucose, to the respiratory utilization of previously produced ethanol. Moreover, to assist these investigations and following studies in the future, I developed and implemented a web-based software tool for the analysis of CF-MS data, which also allows to compare the fractionation profiles between different experimental conditions, cell lines or developmental stages.

2.5 List of Publications and Personal Contributions

*Studies marked with an * are part of this cumulative thesis*

PROMIS: Global Analysis of PROtein-Metabolite Interactions

Authors:	Ewelina M. Sokolowska[1], **Dennis Schlossarek[1]**, Marcin Luzarowski[1], Aleksandra Skirycz [1]: These authors contributed equally
Year of Publication:	2019
Journal:	Current Protocols in Plant Biology, 4, e20101
DOI:	https://doi.org/10.1002/cppb.20101
Personal Contribution:	I wrote this method description together with my colleagues. I wrote the protocols for size separation of native protein-metabolite complexes, protein and metabolite extraction as well as data analysis aspects of the PROMIS method.

Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity *

Authors:	Marcin Luzarowski, Rubén Vicente, Andrei Kiselev, Mateusz Wagner, **Dennis Schlossarek**, Alexander Erban, Leonardo Perez de Souza, Dorothee Childs, Izabela Wojciechowska, Urszula Luzarowska, Michał Górka, Ewelina M. Sokołowska, Monika Kosmacz, Juan C. Moreno, Aleksandra Brzezińska, Bhavana Vegesna, Joachim Kopka, Alisdair R. Fernie, Lothar Willmitzer, Jennifer C. Ewald & Aleksandra Skirycz
Year of Publication:	2021
Journal:	Communications Biology volume 4, Article number: 181
DOI:	https://doi.org/10.1038/s42003-021-01684-3
Personal Contribution:	I helped with data analysis, designed and implemented the shiny app for the interactive visualization of the presented data.

PROMISed: A novel web-based tool to facilitate analysis and visualization of the molecular interaction networks from co-fractionation mass spectrometry (CF-MS) experiments *

Authors:	**Dennis Schlossarek**, Marcin Luzarowski, Ewelina Sokołowska, Michał Górka, Lothar Willmitzer, Aleksandra Skirycz
Year of Publication:	2021
Journal:	Computational and Structural Biotechnology Journal, Volume 19, Pages 5117-5125
DOI:	https://doi.org/10.1016/j.csbj.2021.08.042
Personal Contribution:	I streamlined and expended the previously used data-analysis pipeline, and, from there, conceptualized, implemented and tested the software-tool. I benchmarked the software using an available data set. I wrote the manuscript together with my supervisor.

SLIMP: Supervised learning of metabolite-protein interactions from co-fractionation mass spectrometry data

Authors:	Boris M. Zühlke[1], Ewelina M. Sokołowska[1], Marcin Luzarowski[1], **Dennis Schlossarek[1]**, Monika Chodasiewicz, Ewa Leniak, Aleksandra Skirycz, Zoran Nikoloski [1]: These authors contributed equally
Year of Publication:	2021 (preprint, not peer-reviewed)
Journal:	bioRxiv
DOI:	https://doi.org/10.1101/2021.06.16.448636
Personal Contribution:	I advised on the data analysis and was involved in writing the manuscript.

Rewiring of the protein-protein-metabolite interactome during the diauxic shift in yeast *

Authors:	**Dennis Schlossarek[1]**, Marcin Luzarowski[1], Ewelina M. Sokołowska, Venkatesh P. Thirumalaikumar, Lothar Willmitzer, Jennifer Ewald, Aleksandra Skirycz [1]: These authors contributed equally
Year of Publication:	Submitted to PLOS Biology, 16.03.2022
Journal:	-
DOI:	-
Personal Contribution:	I analyzed the data and provided data visualization. I, together with Marcin Luzarowski and my supervisor, wrote the manuscript. I assisted in the experiments.

Plant Journal (Introduction and Discussion) *

Authors:	**Dennis Schlossarek**, Aleksandra Skirycz
Year of Publication:	In preparation
Journal:	-
DOI:	-
Personal Contribution:	I wrote the review together with my supervisor.

Confirmation of supervisor:

I hereby confirm the contributions made by Dennis Schlossarek to the studies listed above.

Prof. Dr. Lothar Willmitzer

2.6 References – General Introduction

1. Patti, G. J., Yanes, O. & Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* **13**, 263–269 (2012).

2. Tu, B. P., Kudlicki, A., Rowicka, M. & McKnight, S. L. Cell biology: Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science (80-.).* **310**, 1152–1158 (2005).

3. Tu, B. P. *et al.* Cyclic changes in metabolic state during the life of a yeast cell. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16886–16891 (2007).

4. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).

5. Murphy, J. P., Stepanova, E., Everley, R. A., Paulo, J. A. & Gygi, S. P. Comprehensive temporal protein dynamics during the diauxic shift in Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **14**, 2454–2465 (2015).

6. Ball, D. A. *et al.* Oscillatory dynamics of cell cycle proteins in single yeast cells analyzed by imaging cytometry. *PLoS One* **6**, (2011).

7. Frickel, E. M., Jemth, P., Widersten, M. & Mannervik, B. Yeast glyoxalase I is a monomeric enzyme with two active sites. *J. Biol. Chem.* **276**, 1845–1849 (2001).

8. Hipp, M. S., Kasturi, P. & Hartl, F. U. The proteostasis network and its decline in ageing. *Nat. Rev. Mol. Cell Biol.* **20**, 421–435 (2019).

9. Korkuć, P. & Walther, D. Spatial proximity statistics suggest a regulatory role of protein phosphorylation on compound binding. *Proteins Struct. Funct. Bioinforma.* **84**, 565–579 (2016).

10. Berg, J. M., Tymoczko, J. L., Gatto, G. J. J. & Stryer. *Biochemistry*. (2019).

11. Swatek, K. N. & Komander, D. Ubiquitin modifications. *Cell Res.* **26**, 399–422 (2016).

12. Celen, A. B. & Sahin, U. Sumoylation on its 25th anniversary: mechanisms, pathology, and emerging concepts. *FEBS J.* **287**, 3110–3140 (2020).

13. Srere, P. A. The metabolon. *Trends Biochem. Sci.* **10**, 109–110 (1985).

14. Fernie, A. R., Zhang, Y. & Sweetlove, L. J. Passing the baton: Substrate channelling in respiratory metabolism. *Research* **2018**, (2018).

15. Zhang, Y. & Fernie, A. R. Metabolons, enzyme–enzyme assemblies that mediate substrate channeling, and their roles in plant metabolism. *Plant Commun.* **2**, 100081 (2021).

16. Ramírez-Aguilar, S. J. *et al.* The composition of plant mitochondrial supercomplexes changes with oxygen availability. *J. Biol. Chem.* **286**, 43045–43053 (2011).

17. Islam, M. M. *et al.* Branched-chain amino acid metabolon: Interaction of glutamate dehydrogenase with the mitochondrial branched-chain aminotransferase (BCATm). *J. Biol. Chem.* **285**, 265–276 (2010).

18. Camagna, M. *et al.* Enzyme fusion removes competition for geranylgeranyl diphosphate in carotenogenesis. *Plant Physiol.* **179**, 1013–1027 (2019).

19. Giegé, P. *et al.* Enzymes of glycolysis are functionally associated with the mitochondrion in arabidopsis cells. *Plant Cell* **15**, 2140–2151 (2003).

20. Zhu, G. *et al.* Rewiring of the Fruit Metabolome in Tomato Breeding. *Cell* **172**, 249-261.e12 (2018).

21. Wu, F., Pelster, L. N. & Minteer, S. D. Krebs cycle metabolon formation: Metabolite concentration gradient enhanced compartmentation of sequential enzymes. *Chem. Commun.* **51**, 1244–1247 (2015).

22. Wu, F. & Minteer, S. Krebs cycle metabolon: Structural evidence of substrate channeling revealed by cross-

linking and mass spectrometry. *Angew. Chemie - Int. Ed.* **54**, 1851–1854 (2015).

23. Zhang, Y. *et al.* Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat. Commun.* **8**, (2017).

24. Pareek, V., Tian, H., Winograd, N. & Benkovic, S. J. Metabolomics and mass spectrometry imaging reveal channeled de novo purine synthesis in cells. *Science (80-.).* **368**, 283–290 (2020).

25. Messenguy, F. & Wia. The Control of Ornithinetranscarbamylase Activity by Arginase in Saccharomyces Cerevisiae. *FEBS Lett.* **3**, 6–8 (1969).

26. Messenguy, F., Pennickx, M. & Wiame, J.-M. Interaction between Arginase and Ornithine Carbamoyltransferase in Saccharomyces cerevisiae - The Regulatory Sites of Arginase. *Eur. J. Biochem.* **22**, 277–286 (1971).

27. Jayaraman, V. *et al.* A counter-enzyme complex regulates glutamate metabolism in Bacillus subtilis. *Nat. Chem. Biol.* (2021) doi:10.1038/s41589-021-00919-y.

28. Genuth, N. R. & Barna, M. The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. *Mol. Cell* **71**, 364–374 (2018).

29. Sulima, S. O. & Dinman, J. D. The Expanding Riboverse. *Cells* **8**, 1–13 (2019).

30. Shi, Z. *et al.* Heterogeneous Ribosomes Preferentially Translate Distinct Subpools of mRNAs Genome-wide. *Mol. Cell* **67**, 71-83.e7 (2017).

31. Sun, M. *et al.* A Time-Resolved Cryo-EM Study of Saccharomyces cerevisiae 80S Ribosome Protein Composition in Response to a Change in Carbon Source. *Proteomics* **21**, 1–9 (2021).

32. Martinez-Seidel, F. *et al.* Spatially enriched paralog rearrangements argue functionally diverse ribosomes arise during cold acclimation in arabidopsis. *Int. J. Mol. Sci.* **22**, (2021).

33. Mazumder, B. *et al.* Regulated release of L13a from the 60S ribosomal subunit as a mechanism of transcript-specific translational control. *Cell* **115**, 187–198 (2003).

34. Ramazi, S., Allahverdi, A. & Zahiri, J. Evaluation of post-translational modifications in histone proteins: A review on histone modification defects in developmental and neurological disorders. *J. Biosci.* **45**, 3–9 (2020).

35. Link, H., Fuhrer, T., Gerosa, L., Zamboni, N. & Sauer, U. Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nat. Methods* **12**, 1091–1097 (2015).

36. Reznik, E. *et al.* Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. *Cell Rep.* **20**, 2666–2677 (2017).

37. Pabst, M. J., Kuhn, J. C. & Somerville, R. L. Feedback regulation in the anthranilate aggregate from wild type and mutant strains of Escherichia coli. *J. Biol. Chem.* **248**, 901–914 (1973).

38. Sander, T. *et al.* Allosteric Feedback Inhibition Enables Robust Amino Acid Biosynthesis in E. coli by Enforcing Enzyme Overabundance. *Cell Syst.* **8**, 66-75.e8 (2019).

39. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1407–1418 (2011).

40. Messerchmidt, A. *et al.* Determinants of enzymatic specificity in the cys-met-metabolism PLP-dependent enzymes family: Crystal structure of cystathionine γ-lyase from yeast and intrafamiliar structure comparison. *Biol. Chem.* **384**, 373–386 (2003).

41. Richts, B., Rosenberg, J. & Commichau, F. M. A survey of pyridoxal 5'-phosphate-dependent proteins in the gram-positive model bacterium Bacillus subtilis. *Front. Mol. Biosci.* **6**, (2019).

42. Youn, J. Y. *et al.* Properties of Stress Granule and P-Body Proteomes. *Mol. Cell* **76**, 286–294 (2019).

43. Riggs, C. L., Kedersha, N., Ivanov, P. & Anderson, P. Mammalian stress granules and P bodies at a glance. *J. Cell Sci.* **133**, 1–9 (2020).

44. Markmiller, S. *et al.* Context-Dependent and Disease-Specific Diversity in Protein Interactions within Stress Granules. *Cell* **172**, 590-604.e13 (2018).

45. Youn, J. Y. *et al.* High-Density Proximity Mapping Reveals the Subcellular Organization of mRNA-Associated Granules and Bodies. *Mol. Cell* **69**, 517-532.e11 (2018).

46. Kosmacz, M. *et al.* Interaction of 2',3'-cAMP with Rbp47b plays a role in stress granule formation. *Plant Physiol.* **177**, pp.00285.2018 (2018).

47. Thompson, J. E., Venegas, F. D. & Raines, R. T. Energetics of Catalysis by Ribonucleases: Fate of the 2′,3′-Cyclic Phosphodiester Intermediate. *Biochemistry* **33**, 7408–7414 (1994).

48. Jackson, E. K., Ren, J. & Mi, Z. Extracellular 2′,3′-cAMP is a source of adenosine. *J. Biol. Chem.* **284**, 33097–33106 (2009).

49. Verrier, J. D. *et al.* The brain in vivo expresses the 2',3'-cAMP-adenosine pathway. *J. Neurochem.* **122**, 115–125 (2012).

50. Van Damme, T. *et al.* Wounding stress causes rapid increase in concentration of the naturally occurring 2′,3′-isomers of cyclic guanosine- and cyclic adenosine monophosphate (cGMP and cAMP) in plant tissues. *Phytochemistry* **103**, 59–66 (2014).

51. Chodasiewicz, M. *et al.* 2′,3′-cAMP treatment mimics abiotic stress response. *bioRxiv* 2021.07.12.452129 (2021).

52. Tatzelt, J., Prusiner, S. B. & Welch, W. J. Chemical chaperones interfere with the formation of scrapie prion protein. *EMBO J.* **15**, 6363–6373 (1996).

53. Samuel, D. *et al.* Proline inhibits aggregation during protein refolding. *Protein Sci.* **9**, 344–352 (2008).

54. Yang, D. S., Yip, C. M., Huang, T. H. J., Chakrabartty, A. & Fraser, P. E. Manipulating the amyloid-β aggregation pathway with chemical chaperones. *J. Biol. Chem.* **274**, 32970–32974 (1999).

55. Voziyan, P. A. & Fisher, M. T. Chaperonin-assisted folding of glutamine synthetase under nonpermissive conditions: Off-pathway aggregation propensity does not determine the co-chaperonin requirement. *Protein Sci.* **9**, 2405–2412 (2000).

56. Diamant, S., Eliahu, N., Rosenthal, D. & Goloubinoff, P. Chemical Chaperones Regulate Molecular Chaperones in Vitro and in Cells under Combined Salt and Heat Stresses. *J. Biol. Chem.* **276**, 39586–39591 (2001).

57. Dandage, R. *et al.* Classification of chemical chaperones based on their effect on protein folding landscapes. *ACS Chem. Biol.* **10**, 813–820 (2015).

58. Gorensek-Benitez, A. H., Smith, A. E., Stadmiller, S. S., Perez Goncalves, G. M. & Pielak, G. J. Cosolutes, Crowding, and Protein Folding Kinetics. *J. Phys. Chem. B* **121**, 6527–6537 (2017).

59. Pal, S., Pyne, P., Samanta, N., Ebbinghaus, S. & Mitra, R. K. Thermal stability modulation of the native and chemically-unfolded state of bovine serum albumin by amino acids. *Phys. Chem. Chem. Phys.* **22**, 179–188 (2019).

60. Gekko, K. & Timasheff, S. N. Mechanism of Protein Stabilization by Glycerol: Preferential Hydration in Glycerol-Water Mixtures. *Biochemistry* **20**, 4667–4676 (1981).

61. Gekko, K. & Timasheff, S. N. Thermodynamic and Kinetic Examination of Protein Stabilization by Glycerol. *Biochemistry* **20**, 4677–4686 (1981).

62. Blomberg, A. Metabolic surprises in Saccharomyces cerevisiae during adaptation to saline conditions: Questions, some answers and a model. *FEMS Microbiol. Lett.* **182**, 1–8 (2000).

63. Verma, K. *et al.* Distinct metabolic states of a cell guide alternate fates of mutational buffering through altered proteostasis. *Nat. Commun.* **11**, (2020).

64. Minton, A. P. The Influence of Macromolecular Crowding and Macromolecular Confinement on Biochemical Reactions in Physiological Media. *J. Biol. Chem.* **276**, 10577–10580 (2001).

65. Minton, A. P. & Wilf, J. Effect of Macromolecular Crowding upon the Structure and Function of an Enzyme: Glyceraldehyde-3-phosphate Dehydrogenase. *Biochemistry* **20**, 4821–4826 (1981).

66. Higgs, P. G. & Lehman, N. The RNA World: Molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16**, 7–17 (2015).

67. Bernhardt, H. S. The RNA world hypothesis: the worst theory of the early evolution of life (except for all the others)a. *Biol. Direct* **7**, 1–10 (2012).

68. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).

69. Winkler, W. C., Cohen-Chalamish, S. & Breaker, R. R. An mRNA structure that controls gene expression by binding FMN. *PNAS* **99**, 2688–2697 (2002).

70. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043–1049 (2002).

71. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell* **111**, 747–756 (2002).

72. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *Rna* **9**, 644–647 (2003).

73. Bocobza, S. E. & Aharoni, A. Small molecules that interact with RNA: Riboswitch-based gene control and its involvement in metabolic regulation in plants and algae. *Plant J.* **79**, 693–703 (2014).

74. Álvarez, D. *et al.* Carotenogenesis is regulated by 5‘UTR-mediated translation of phytoene synthase splice variants. *Plant Physiol.* **172**, 2314–2326 (2016).

75. Cala, O., Guillière, F. & Krimm, I. NMR-based analysis of protein-ligand interactions. *Anal. Bioanal. Chem.* **406**, 943–956 (2014).

76. Becker, W., Bhattiprolu, K. C., Gubensäk, N. & Zangger, K. Investigating Protein–Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *ChemPhysChem* **19**, 895–906 (2018).

77. Diether, M., Nikolaev, Y., Allain, F. H. & Sauer, U. Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli . *Mol. Syst. Biol.* **15**, 1–16 (2019).

78. Schwenk, P. *et al.* Uncovering a novel function of the ccr4-not complex in phytochrome a-mediated light signalling in plants. *Elife* **10**, 1–33 (2021).

79. Maeda, K., Poletto, M., Chiapparino, A. & Gavin, A. C. A generic protocol for the purification and characterization of water-soluble complexes of affinity-tagged proteins and lipids. *Nat. Protoc.* **9**, 2256–2266 (2014).

80. Luzarowski, M. *et al.* Affinity purification with metabolomic and proteomic analysis unravels diverse roles of nucleoside diphosphate kinases. *J. Exp. Bot.* **68**, 3487–3499 (2017).

81. Miller, K. E., Kim, Y., Huh, W. K. & Park, H. O. Bimolecular fluorescence complementation (BiFC) analysis: Advances and recent applications for Genome-Wide interaction studies. *J. Mol. Biol.* **427**, 2039–2055 (2015).

82. Brückner, A., Polge, C., Lentze, N., Auerbach, D. & Schlattner, U. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.* **10**, 2763–2788 (2009).

83. Welsch, R. *et al.* Clp Protease and OR Directly Control the Proteostasis of Phytoene Synthase, the Crucial Enzyme for Carotenoid Biosynthesis in Arabidopsis. *Mol. Plant* **11**, 149–162 (2018).

84. Zhou, X. *et al.* Arabidopsis OR proteins are the major posttranscriptional regulators of phytoene synthase in controlling carotenoid biosynthesis. *Proc. Natl. Acad. Sci.* **112**, 201420831 (2015).

85. Suter, B. & Walker, J. M. *Two Hybrid Technologies - Methods and Protocols*. *Life Sciences* vol. 531 (2009).

86. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).

87. Arora, D. *et al.* Establishment of proximity-dependent biotinylation approaches in different plant model systems. *Plant Cell* **32**, 3388–3407 (2020).

88. Yang, X. *et al.* Proximity labeling: an emerging tool for probing in planta molecular interactions. *Plant Commun.* **2**, 100137 (2021).

89. Lam, S. S. *et al.* Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat. Methods* **12**, 51–54 (2014).

90. Kim, D. I. *et al.* An improved smaller biotin ligase for BioID proximity labeling. *Mol. Biol. Cell* **27**, 1188–1196 (2016).

91. Branon, T. C. *et al.* Efficient proximity labeling in living cells and organisms with TurboID. *Nat. Biotechnol.* **36**, 880–898 (2018).

92. León-Del-Río, A. Biotin in metabolism, gene expression, and human disease. *J. Inherit. Metab. Dis.* **42**, 647–654 (2019).

93. West, G. M., Tang, L. & Fitzgerald, M. C. Thermodynamic Analysis of Protein Stability and Ligand Binding Using a Chemical Modification- and Mass Spectrometry-Based Strategy. **80**, 4175–4185 (2008).

94. Strickland, E. C. *et al.* Thermodynamic Analysis of Protein-Ligand Binding Interactions in Complex Biological Mixtures using the Stability of Proteins from Rates of Oxidation (SPROX) Method. *Nat. Protoc.* **8**, 148–161 (2013).

95. Geer, M. A. & Fitzgerald, M. C. Characterization of the Saccharomyces cerevisiae ATP-Interactome using the iTRAQ-SPROX Technique. *J. Am. Soc. Mass Spectrom.* **27**, 233–243 (2016).

96. Lomenick, B. *et al.* Target identification using drug affinity responsive target stability (DARTS). *PNAS* **106**, 21984–21989 (2009).

97. Piazza, I. *et al.* Resource A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication Resource A Map of Protein-Metabolite Interactions Reveals Principles of Chemical Communication. *Cell* **172**, 358-361.e23 (2018).

98. Piazza, I. *et al.* A machine learning-based chemoproteomic approach to identify drug targets and binding sites in complex proteomes. *Nat. Commun.* **11**, 1–13 (2020).

99. Molina, D. M., Jafari, R., Ignatushchenko, M. & Seki, T. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science (80-.).* **341**, 84–88 (2013).

100. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science (80-.).* **346**, (2014).

101. Mateus, A. *et al.* Thermal proteome profiling for interrogating protein interactions. *Mol. Syst. Biol.* **16**, 1–11 (2020).

102. Mateus, A., Kurzawa, N., Perrin, J., Bergamini, G. & Savitski, M. M. Drug Target Identification in Tissues by Thermal Proteome Profiling. *Annu. Rev. Pharmacol. Toxicol.* **62**, 465–482 (2022).

103. Sridharan, S. *et al.* Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat. Commun.* **10**, 1–13 (2019).

104. Chodasiewicz, M. *et al.* A novel role of 3',5'-cAMP in the regulation of actin cytoskeleton in Arabidopsis.

bioRxiv 1–36 (2022).

105. Tan, C. S. H. *et al.* Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. *Science (80-.).* **359**, 1170–1177 (2018).

106. Becher, I. *et al.* Pervasive Protein Thermal Stability Variation during the Cell Cycle. *Cell* **173**, 1495-1507.e18 (2018).

107. Jarzab, A. *et al.* Meltome atlas—thermal proteome stability across the tree of life. *Nat. Methods* **17**, 495–503 (2020).

108. Leijten, N. M., Bakker, P., Spaink, H. P., den Hertog, J. & Lemeer, S. Thermal Proteome Profiling in Zebrafish Reveals Effects of Napabucasin on Retinoic Acid Metabolism. *Mol. Cell. Proteomics* **20**, 0–10 (2021).

109. Peck Justice, S. A. *et al.* Mutant thermal proteome profiling for characterization of missense protein variants and their associated phenotypes within the proteome. *J. Biol. Chem.* **295**, 16219–16238 (2020).

110. Volkening, J. D., Stecker, K. E. & Sussman, M. R. Proteome-wide analysis of protein thermal stability in the model higher plant arabidopsis thaliana. *Mol. Cell. Proteomics* **18**, 308–319 (2019).

111. Mateus, A. *et al.* Thermal proteome profiling in bacteria: probing protein state in vivo . *Mol. Syst. Biol.* **14**, 1–15 (2018).

112. Mateus, A. *et al.* The functional proteome landscape of Escherichia coli. *Nature* **588**, 473–478 (2020).

113. Herneisen, A. L. & Lourido, S. Thermal proteome profiling to identify protein-ligand interactions in the apicomplexan parasite Toxoplasma gondii. *Bio-protocol* **11**, 21769 (2021).

114. Coutinho, J. V. P. *et al.* The thermal proteome stability profile of Trypanosoma cruzi in epimastigote and trypomastigote life stages. *J. Proteomics* **248**, 104339 (2021).

115. Selkrig, J. *et al.* SARS-CoV-2 infection remodels the host protein thermal stability landscape. *Mol. Syst. Biol.* **17**, 1–13 (2021).

116. Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science (80-.).* **355**, (2017).

117. Huang, J. X. *et al.* High throughput discovery of functional protein modifications by Hotspot Thermal Profiling. *Nat. Methods* **16**, 894–901 (2019).

118. Smith, I. R. *et al.* Erratum to: High throughput discovery of functional protein modifications by Hotspot Thermal Profiling (Nature Methods, (2019), 16, 9, (894-901), 10.1038/s41592-019-0499-3). *Nat. Methods* **18**, 760–762 (2021).

119. Potel, C. M. *et al.* Impact of phosphorylation on thermal stability of proteins. *Nat. Methods* **18**, 757–759 (2021).

120. Gaetani, M. *et al.* Proteome Integral Solubility Alteration: A High-Throughput Proteomics Assay for Target Deconvolution. *J. Proteome Res.* **18**, 4027–4037 (2019).

121. Li, J., Van Vranken, J. G., Paulo, J. A., Huttlin, E. L. & Gygi, S. P. Selection of Heating Temperatures Improves the Sensitivity of the Proteome Integral Solubility Alteration Assay. *J. Proteome Res.* **19**, 2159–2166 (2020).

122. Ball, K. A. *et al.* An isothermal shift assay for proteome scale drug-target identification. *Commun. Biol.* **3**, 1–10 (2020).

123. Liu, X., Yang, W. chu, Gao, Q. & Regnier, F. Toward chromatographic analysis of interacting protein networks. *J. Chromatogr. A* **1178**, 24–32 (2008).

124. Chan, J. N. Y. *et al.* Target identification by chromatographic co-elution: Monitoring of drug-protein interactions without immobilization or chemical derivatization. *Mol. Cell. Proteomics* **11**, (2012).

125. Chen, X. *et al.* A Ligand-observed Mass Spectrometry Approach Integrated into the Fragment Based Lead Discovery Pipeline. *Sci. Rep.* **5**, 1–8 (2015).

126. Veyel, D. *et al.* System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. *Sci. Rep.* **7**, 1–8 (2017).

127. Veyel, D. *et al.* PROMIS, global analysis of PROtein-metabolite interactions using size separation in Arabidopsis thaliana. *J. Biol. Chem.* **293**, 12440–12453 (2018).

128. Gorka, M. *et al.* Protein Complex Identification and quantitative complexome by CN-PAGE. *Sci. Rep.* **9**, 1–14 (2019).

129. Firmino, A. A. P. *et al.* Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes. *Plants* **9**, 1–29 (2020).

130. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).

131. Skinnider, M. A. *et al.* An atlas of protein-protein interactions across mouse tissues. *Cell* **184**, 4073-4089.e17 (2021).

132. Pang, C. N. I. *et al.* Analytical Guidelines for co-fractionation Mass Spectrometry Obtained through Global Profiling of Gold Standard Saccharomyces cerevisiae Protein Complexes. *Mol. Cell. Proteomics* **19**, 1876–1895 (2020).

133. Luzarowski, M. *et al.* Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. *Commun. Biol.* **4**, 1–15 (2021).

134. Li, Y. *et al.* Coupling proteomics and metabolomics for the unsupervised identification of protein – metabolite interactions in Chaetomium thermophilum. *PLoS One* 1–13 (2021) doi:10.1371/journal.pone.0254429.

135. Gilbert, M. & Schulze, W. X. Global Identification of Protein Complexes within the Membrane Proteome of Arabidopsis Roots Using a SEC-MS Approach. *J. Proteome Res.* **18**, 107–119 (2019).

136. Xu, C. *et al.* Global Landscape of Native Protein Complexes in Synechocystis sp. PCC 6803. *Genomics. Proteomics Bioinformatics* (2021) doi:10.1016/j.gpb.2020.06.020.

137. Heusel, M. *et al.* A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133-155.e6 (2020).

138. McWhite, C. D. *et al.* A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* **181**, P460-474.E14 (2020).

139. Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).

140. Skinnider, M. A. & Foster, L. J. Meta-analysis defines principles for the design and analysis of co-fractionation mass spectrometry experiments. *Nat. Methods* **18**, 806–815 (2021).

141. Mallam, A. L. *et al.* Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. *Cell Rep.* **29**, 1351-1368.e5 (2019).

142. Moreno, J. C. *et al.* Tyr-Asp inhibition of glyceraldehyde 3-phosphate dehydrogenase affects plant redox metabolism. *EMBO J.* **40**, 1–16 (2021).

143. Wagner, M., Zhang, B., Tauffenberger, A., Schroeder, F. C. & Skirycz, A. Experimental methods for dissecting the terraincognita of protein-metabolite interactomes. *Curr. Opin. Syst. Biol.* **28**, 100403 (2021).

3. Publication #1 – Protein-Metabolite Interactions in Yeast

Global mapping of protein–metabolite interactions in *Saccharomyces cerevisiae* reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity

Authors:	Marcin Luzarowski, Rubén Vicente, Andrei Kiselev, Mateusz Wagner, **Dennis Schlossarek**, Alexander Erban, Leonardo Perez de Souza, Dorothee Childs, Izabela Wojciechowska, Urszula Luzarowska, Michał Górka, Ewelina M. Sokołowska, Monika Kosmacz, Juan C. Moreno, Aleksandra Brzezińska, Bhavana Vegesna, Joachim Kopka, Alisdair R. Fernie, Lothar Willmitzer, Jennifer C. Ewald & Aleksandra Skirycz
Year of Publication:	2021
Journal:	Communications Biology volume 4, Article number: 181
DOI:	https://doi.org/10.1038/s42003-021-01684-3
Personal Contribution:	I helped with data analysis, designed and implemented the shiny app for the interactive visualization of the presented data.

3.1 Abstract

Protein-metabolite interactions are of crucial importance for all cellular processes but remain understudied. Here, we applied a biochemical approach named PROMIS, to address the complexity of the protein-small molecule interactome in the model yeast *Saccharomyces cerevisiae*. By doing so, we provide a unique dataset, which can be queried for interactions between 74 small molecules and 3982 proteins using a user-friendly interface available at https://promis.mpimp-golm.mpg.de/yeastpmi/. By interpolating PROMIS with the list of predicted protein-metabolite interactions, we provided experimental validation for 225 binding events. Remarkably, of the 74 small molecules co-eluting with proteins, 36 were proteogenic dipeptides. Targeted analysis of a representative dipeptide, Ser-Leu, revealed numerous protein interactors comprising chaperones, proteasomal subunits, and metabolic enzymes. We could further demonstrate that Ser-Leu binding increases activity of a glycolytic enzyme phosphoglycerate kinase (*Pgk1*). Consistent with the binding analysis, Ser-Leu supplementation leads to the acute metabolic changes and delays timing of a diauxic shift. Supported by the dipeptide accumulation analysis our work attests to the role of Ser-Leu as a metabolic regulator at the interface of protein degradation and central metabolism.

3.2 Introduction

Metabolism is a complex system of chemical reactions that converts external nutrients to cellular building blocks and energy, as well as signalling molecules, defence agents, and means of communication. In response to perturbations of nutrient supply or intracellular demands, metabolite concentrations and their conversion rates can change by orders of magnitude within seconds [1]. These timescales are too fast for transcriptional regulation, and thus cells have evolved more direct means of regulation: for example, metabolites themselves can act as regulators. Metabolites can regulate their pathways, balance competing pathways, and coordinate metabolism with the physiology of the cell by interacting with and regulating proteins [2]. Examples of protein–metabolite interactions (PMIs) can be found in virtually all protein functional classes, ranging from metabolic enzymes to structural proteins to signalling components, such as transcription factors and kinases [3-7].

Regulation by PMI can be especially important for single-cell organisms that face constant changes in their environment and nutrient supply [2]. The yeast *S. cerevisiae* is a well-established single-cell model organism, and its metabolism has been extensively studied in the context of biotechnology, biomedicine, and ecology. A recent study suggested that 29 out of 56 reactions in central yeast metabolism were at least partially regulated by allosteric interaction [8]. Identifying allosteric interactions can significantly improve the predictive power of metabolic models and the success of bioengineering approaches [9]. In addition to regulating the activity of metabolic enzymes, PMIs can also have global regulatory functions in coordinating metabolic fluxes with the physiology of the cell [10,11].

Despite their significant role in regulating metabolism and coordinating physiology, PMIs have remained understudied. For *S. cerevisiae*, there are approximately six times fewer reports of experimentally validated PMIs than protein-protein interactions (PPIs) in the STITCH database, which is a comprehensive resource integrating PMI for 430 000 chemicals [12,13]. Therefore, we expect to find many more regulatory functions of metabolites in the yeast cell. However, discovering these functions of metabolites requires suitable methods for globally capturing PMIs.

Powerful approaches that enable PMI studies at the cell-wide scale have been recently reported [14]. These technologies include affinity purification [15], thermal proteome profiling [16], drug affinity responsive target stability [17], small molecule limited proteolysis [18], tandem affinity purification [19,20], and capture compound mass spectrometry [21]. These are conceptually very different strategies, but they all share a common characteristic: namely they require a predefined protein or metabolite as a bait. Consequently, they are ideal for studying interactions of a single metabolite or protein. However, they cannot capture the global overview of the interactome in an unbiased way.

To address this limitation, we have developed an approach, termed PROMIS, which enables a cell-wide analysis of the protein–metabolite and protein-protein interactomes [22,23]. Similar to the previously

mentioned approaches, PROMIS starts with a native cellular lysate and thus operates in close to *in vivo* conditions. In brief, PROMIS combines size separation of complexes with proteomics and metabolomics analysis of the obtained fractions and exploits co-elution to define putative interactors. Thereby, needs neither a specific protein nor a specific metabolite as a bait. While this approach may not allow direct identification of binding partners, PROMIS is an ideal method of testing the complexity of an interactome and obtaining leads for targeted studies.

In the current study, we use PROMIS for systematic analysis of protein-small molecule interactions in *Saccharomyces cerevisiae.* We assayed interactions between 74 small molecules and 3982 proteins in the native cell lysate and recovered 16% of the previously reported binding events. We provide a unique data set of 225 interactions for 22 individual metabolites and explore specific examples of metabolite regulators. Most excitingly, our results point to role of proteogenic dipeptides as metabolic regulators at the interface of protein degradation and central metabolism.

3.3 Results

PROMIS detects hundreds of candidate protein–metabolite interactions

The goal of this work was to generate a proteome- and metabolome-wide map of protein-metabolite complexes of actively dividing and metabolically active *S. cerevisiae*. The diploid, prototrophic YSBN2 strain in the logarithmic phase of growth was used as starting material. The overall experimental strategy included: (i) preparation of the native, soluble lysate, (ii) size fractionation of complexes using a size exclusion chromatography (SEC), and (iii) untargeted analysis of the complex components using mass spectrometry-based metabolomics and proteomics [22,23] (Figure 1a). In total, we collected 48 fractions from three biological replicates. Thirty-eight of the 48 fractions contained proteins and protein complexes spanning from 5.2 MDa to 20 kDa.

Metabolomics analysis identified 1016 small molecules of the mass between 100 and 1500 Da (Data S1) that separated together with protein complexes and were, therefore, classified as protein-bound. A protein-free small-molecule extract was used as a negative control to exclude the unlikely possibility that free metabolites would elute together with the high-molecular-weight, protein-containing fractions. Indeed, the negative control tests confirmed this was not the case. Overall, 74 of the identified small molecules could be annotated to a specific compound using chemical standards, and included: purines and pyrimidines, amino acids, dipeptides, and cofactors, as well as signalling molecules (3', 5'-cAMP), and transporters (carnitine) (Data S2). Fifty of these 74 small molecules are known or were predicted to be a part of a yeast protein-metabolite complex (STITCH database), but binding was previously experimentally

confirmed for only 15 of these small molecules in yeast.

Proteomic analysis identified 3982 proteins (Data S3), which accounted for almost 90% of all yeast proteins expressed during the log phase of growth [24] and around 60% of the yeast proteome [25]. Twenty-seven percent of the identified proteins were annotated as subunits of protein complexes (21% of the proteome), 7.5% were involved in molecular transport (8.3% of the proteome), 5% were kinases (3.6% of the proteome), and 8% had putative or unknown functions (17% of the proteome). Proteins integral to membranes or associated with the plasma membrane were significantly underrepresented (0.71-fold enrichment), whereas cytoplasmic and nucleolar proteins were overrepresented (1.28 and 1.48-fold enrichment, respectively) (PANTHER database, Data S4).

Given that the majority of the proteins and metabolites had complex elution patterns characterised by more than one elution maximum, we split the data profiles into single peaks; this is referred to as deconvolution [26]. By doing so, we obtained 1320 and 125 peaks for unknown and annotated metabolites, respectively, and 5834 protein peaks. These were used for further analysis (Data S5-7).

We also determined whether the protein-protein complexes remained intact during the PROMIS separation by examining 5834 protein peaks and calculating the apparent mass of a protein complex based on its elution maximum. We then calculated the ratio between the apparent mass and the theoretical monomeric mass of a protein. This ratio reflects the oligomerisation state of a protein (referred to as its oligomeric state ratio). Assuming that an oligomeric state ratio above 1.5 would indicate an interaction with another protein, we identified 4981 protein peaks, corresponding to 3408 proteins, as part of a multimeric complex (Data S8). We further validated the effectiveness of fractionation by examining the elution profiles of 14 known protein macro-complexes, such as the proteasome [27]. As anticipated, the respective components of the analysed complexes shared an elution profile, validating the suitability of PROMIS for the isolation, fractionation, and identification of native protein complexes (Figure 1b, Data S7).

We estimated co-elution, which we use to define putative interactors, by calculating the Pearson correlation coefficient (PCC) between all annotated metabolite and all protein peaks present in our data set. We then determined the influence of using the PCC threshold on the number of detected true PMIs, which we retrieved from the STITCH database. For this purpose, we created the list of reported true interactions comprising 87 PMIs, including only proteins and metabolites identified in our data. Due to the lack of experimental evidences confirming that given protein-metabolite pair does not interact, we constructed a list of false positives by calculating the Pearson correlation coefficient of 87 randomly picked protein-metabolite pairs present in our dataset (100 iterations). Next, we compared PCC values obtained for true interactions retrieved from STITCH database (Data S9) with randomly picked values (for more detailed description see Supplementary Methods). We calculated the receiver operating characteristic curve, which showed the trade-off between specificity (a low number of false-positive hits) and sensitivity (the number

of retrieved true-positive interactions) (Figure 1c) (Data S9 and S10). To assure specificity but to not reduce sensitivity, we applied a PCC threshold of 0.7 to determine the PMIs (false discovery rate = 17.6%). We recovered 14 of the 87 true PMIs, which is five times greater than the number expected by chance and achieving a true-positive rate comparable to a recent MS-based proteome-wide PMI study in *Escherichia coli*, reporting protein binders for 20 different metabolites (Figure 1d) [18]. Moreover, the correlation coefficients calculated for the true protein-metabolite pairs were higher than the permuted values (Figure S1).

It is important to note that when interpreting PROMIS results, PCC should not be used to rank the interactions. However, we anticipate that many of the small molecules will have few specific protein partners, and so a single protein peak is expected to correspond to a single metabolite peak, equally there will be metabolites for which a single elution peak will correspond to the multiple protein partners, obscuring the PCC. In other words and in the latter case it is the co-elution alone, rather than PCC that is indicative of the interaction. Equally, because metabolite binding may vary depending on a protein oligomeric state or presence in a particular protein complex, multiple correlated protein-metabolite peaks will not always reflect confidence. Taken together, PROMIS results should be seen more as qualitative rather than quantitative, and we recommend that the choice of a PCC threshold should be govern by the best compromise between specificity and sensitivity estimated from the receiver operating characteristic curve (Figure 1c, Data S10).

Finally, we created a user-friendly interface, which can be mined for elution profiles of all measured metabolites and proteins, and for the PMIs. The interface is available at https://promis.mpimp-golm.mpg.de/yeastpmi/.

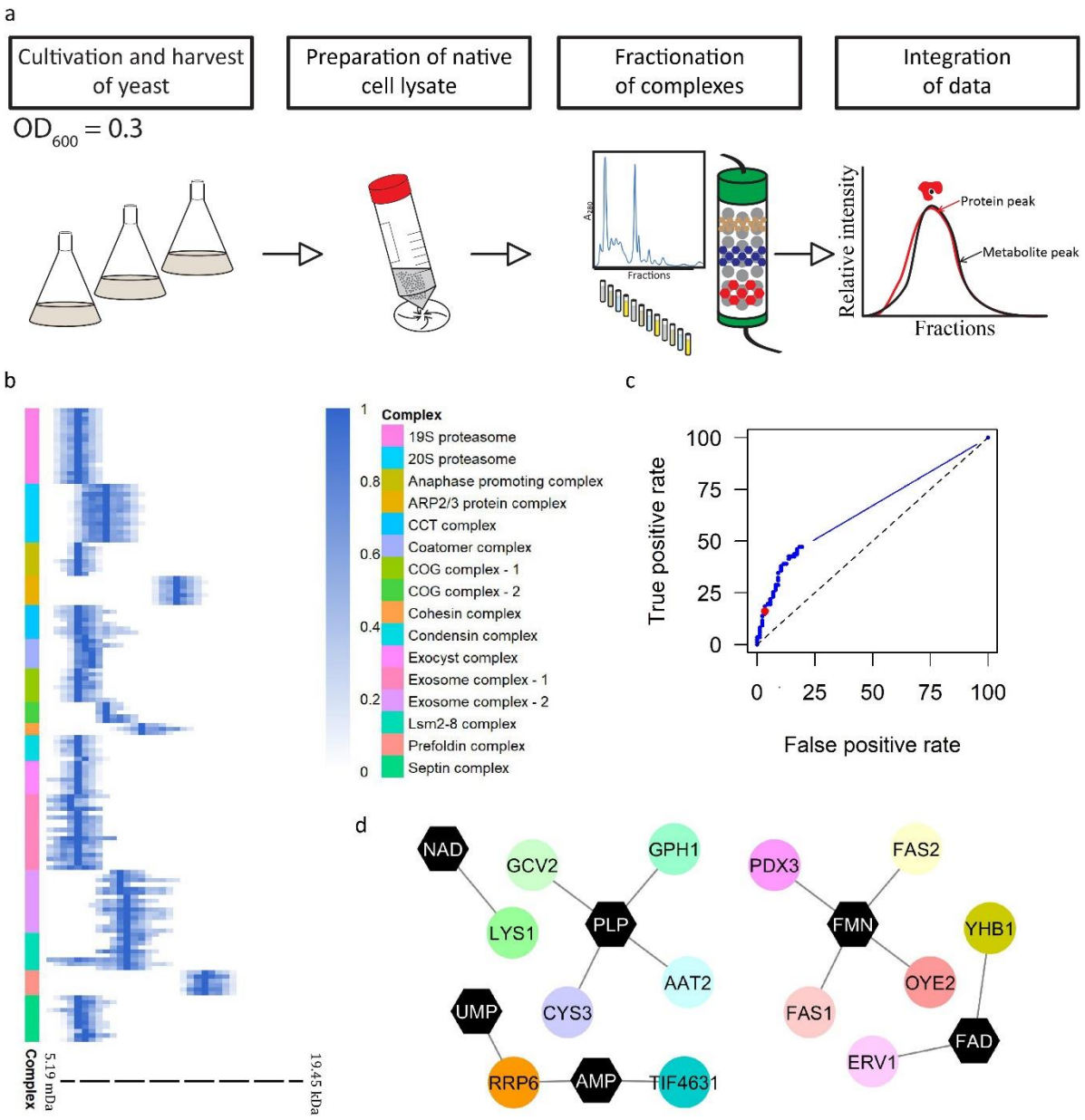


Figure 1: PROMIS allows for system-wide detection of protein–small molecule and protein–protein complexes using size exclusion chromatography. a Dividing yeast cells were harvested in the logarithmic phase of growth and were used as a source of endogenous protein-protein and protein-metabolite complexes. Complexes were fractionated using size exclusion chromatography, lyophilised, and subjected to methyl tert-butyl ether-methanol-water extraction. Polar metabolites and proteins were analysed by liquid chromatography-mass spectrometry. b Known yeast protein macro-complexes remain intact. Multiple subunits of known protein macro-complexes co-elute together. The peak elution profiles of the components of 14 known protein macro-complexes are depicted. The intensity was calculated relative to the maximum intensity of a given protein measured across size exclusion chromatography separation range. Distinct colours are used to mark different protein macro-complexes. c The receiver operating characteristic curve represents a trade-off between numbers of captured true-positive and false-positive protein-metabolite interactions by varying the Pearson correlation coefficient (PCC). The red dot indicates the chosen threshold (PCC ≥ 0.7) used for determining complexes. d Interaction network of captured known protein-metabolite complexes. Overall, 14 of the 87 known protein-metabolite interactions were re-captured in the PROMIS experiment.

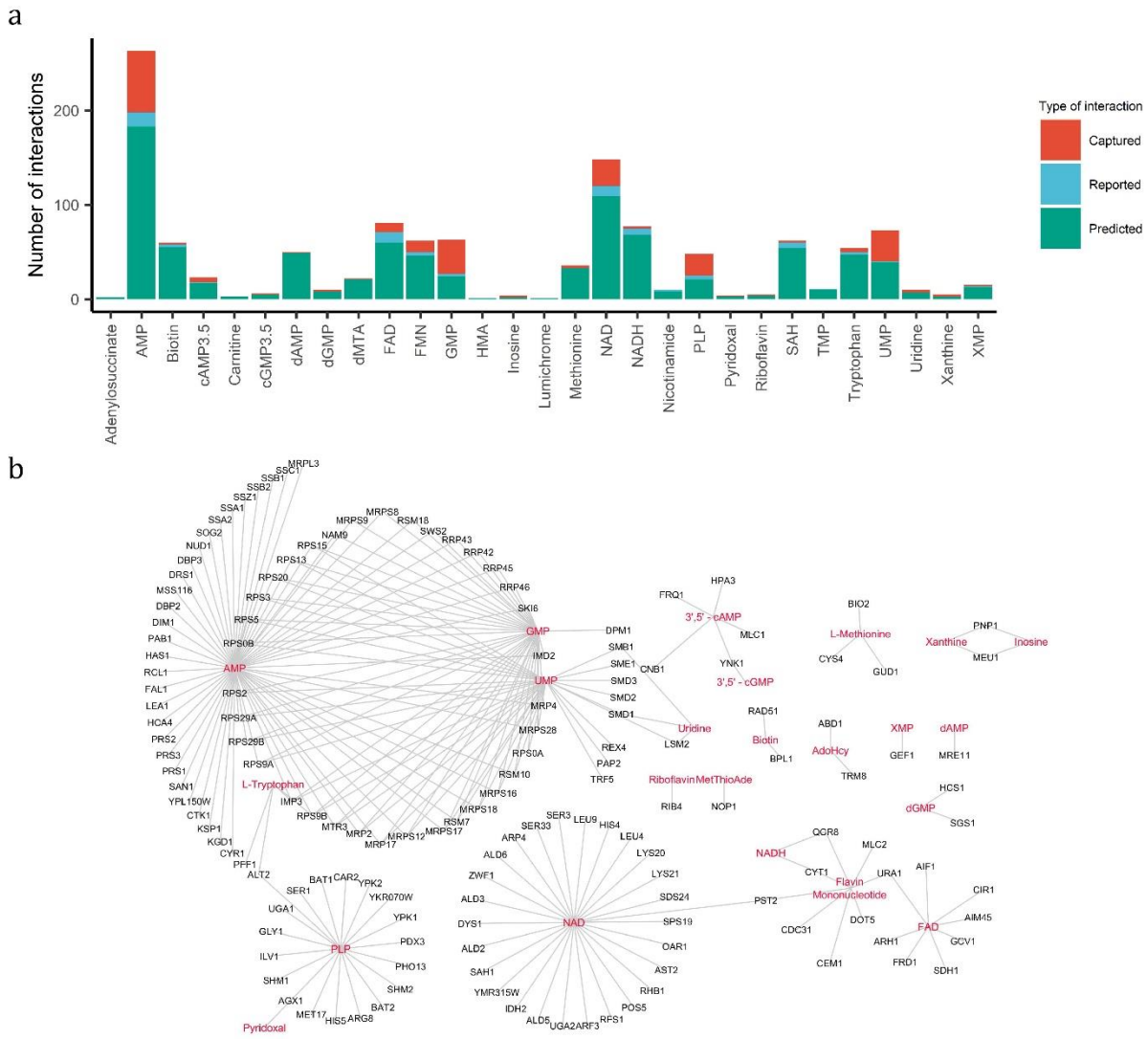


Figure 2: PROMIS provides experimental validation for multiple predicted protein–small molecule complexes. a Number of captured protein-small molecules interaction in relation to previously reported and predicted interactions for each metabolite validated by our data set. b The interaction network of 225 STITCH predicted protein-small molecule interactions validated in this study. Edges represent protein-small molecule interactions and were imported from STITCH, based on the experimental evidence (score ≥ 0.4). Proteins and metabolites are marked as black and red, respectively. Metabolite abbreviations: AdoHcy – adenosyl homocysteine, AMP – adenosine monophosphate, dMTA/MetThioAde – methylthioadenosine, HMA – hydroxy methylglutaric acid, PLP – pyridoxal phosphate, SAH – S-adenosyl-homocysteine, TMP – thymidine monophosphate, UMP – uridine monophosphate, XMP – xanthine monophosphate.

The PROMIS data set captures 225 of the previously predicted yeast PMIs.

In addition to known PMIs, the STITCH database can be mined for predicted PMIs, where prediction is made based on the binding data available for the orthologous proteins, and assuming evolutionary conservation of the interactions [27,28]. We queried lists of predicted PMIs against the yeast

PROMIS data set to provide experimental validation for the previously predicted complexes.

Of the 1122 predicted PMIs, we found experimental evidence for 225 interactions, engaging 22 unique metabolites (Figure 2a, Data S11). A majority of these interactions were between nucleoside monophosphates (NMPs)—such as (deoxy)-AMP, (deoxy)-GMP, and UMP—and DNA-binding and RNA-binding proteins (Figure 2b and Figure S2). After the NMPs, the second largest group was comprised of interactions between enzymes and cofactors (e.g. FMN, FAD, NAD(H), and PLP). Most notably, our dataset validated 19 of the predicted PLP binders, 14 of these were enzymes associated with amino acid metabolism.

In the next step, we decided to explore the list of 225 validated PMIs for those of potential regulatory nature. Herein, we will highlight a representative example, which we followed up and validated experimentally. Purines and pyrimidines are pivotal for multiple cellular processes. Perturbation of their homeostasis leads to metabolic dysfunctions and has a serious impact on yeast growth [29-32]. Considering the importance of purine metabolism, we were intrigued by the interaction between xanthine and purine nucleoside phosphorylase (Pnp1), present in the list of 225 PMIs validated by PROMIS.

Pnp1 catalyses the conversion of guanosine and inosine to guanine and hypoxanthine, respectively. In the PROMIS data set, Pnp1 (monomeric mass 33kDa) separated as two distinct elution peaks with maximum intensity in fractions corresponding to 138 kDa and 88 kDa. This indicates that, *in vivo*, Pnp1 exists in two different oligomeric forms or is part of a protein complex. Pnp1 co-eluted with its known substrate, inosine (Figure 3a and 3b). In addition, Pnp1 co-fractionated with xanthine (PCC > 0.95) (Figure 3c). As others enzyme-metabolite pairs identified in this work, the interaction between Pnp1 and xanthine represents a potential catalytic interaction. However, similar to human purine nucleoside phosphorylase, *Sc*Pnp1 is unable to metabolise adenosine and xanthosine. Thus Pnp1-xanthine binding is more likely a putative regulatory interaction [33].

To test this hypothesis, we investigated whether xanthine affects Pnp1 activity. To this end, we purified recombinant Pnp1 from *S. cerevisiae* and used it in an enzymatic assay that measures the conversion rate of inosine to hypoxanthine [34]. The amount of hypoxanthine produced was measured over time using an LC-ESI-MS assay in the presence or absence of 100 µM xanthine. The addition of 100 µM xanthine lowered the total Pnp1 activity by up to 32% (Figure 3d). The accumulation of xanthine in yeast may, therefore, lower Pnp1 activity and slow the conversion of inosine and guanosine to hypoxanthine and guanine, respectively, which subsequently would lead to the reduction of hypoxanthine and guanine level in yeast cells (Figure 3e).

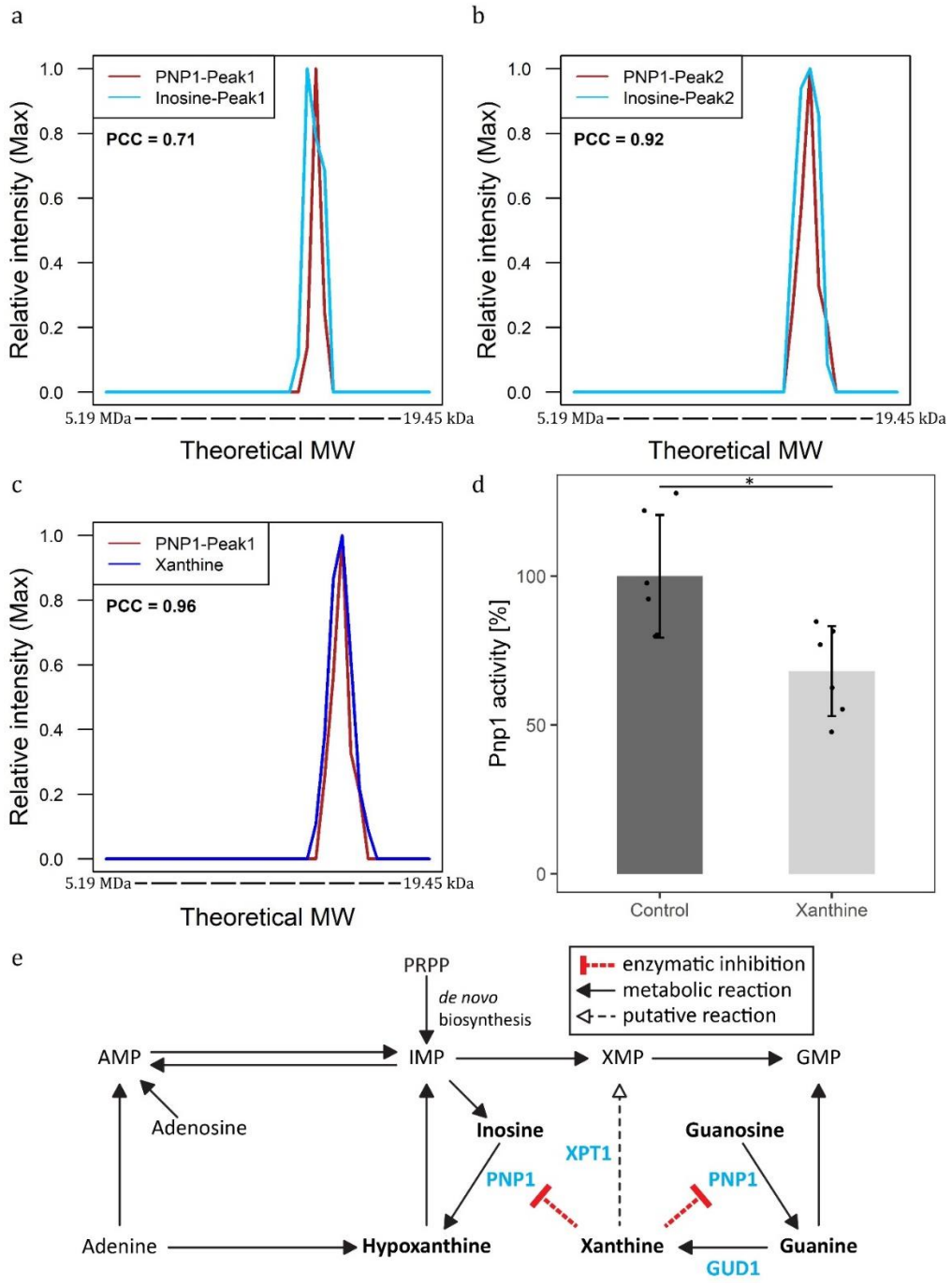


Figure 3: Functional validation of the Pnp1–xanthine interaction. a,b,c) Elution profiles of Pnp1, with its known substrate inosine (a,b) and putative ligand xanthine (c). The intensity was calculated relative to the maximum intensity of the molecule measured across size exclusion chromatography fractions. The theoretical molecular weight (MW) was calculated using reference proteins. d Xanthine inhibits Pnp1 activity. Total activity of recombinant Pnp1 in the presence of 100 µM xanthine was measured using an liquid chromatography-mass spectrometry-based assay. Inhibition was calculated in relation to Pnp1 activity in the absence of xanthine. ± SD, n = 6 independent samples. Asterisks denote significant difference (non-paired, two-tailed t-test P-value < 0.05). e Scheme of purine degradation pathway with predicted regulatory interaction between Pnp1 and xanthine. Molecules discussed in this study are depicted in bold. Enzymes are additionally marked in blue.

The dipeptide Ser-Leu interactome comprises numerous proteins involved in protein and amino acid metabolism.

Of the 74 annotated metabolites that co-eluted with proteins, 36 were proteogenic dipeptides. In yeast the sole reported dipeptide-protein interaction is between dipeptides with the basic N-terminal residue (Arg, Lys, or His) and site-1, and between dipeptides with the bulky hydrophobic N-terminal residue and site-2 (Trp, Phe, Tyr, Leu, or Ile) of the ubiquitin ligase, Ubr1 [35]. Importantly, two of the type 1 dipeptides (Arg-Phe and Lys-Phe) also co-elute with Ubr1 in our data set (Figure S3). Encouraged that we could recapitulate known binding, we decided to determine the precise identity of the protein interactors of a single selected and representative dipeptide, namely Serinyl-Leucine (Ser-Leu). The Ser-Leu elution profile spans reproducibly across a PROMIS protein separation range in all three replicates and is characterised by three local maxima, indicating the presence of a multitude of protein partners. The three Ser-Leu peaks co-elute (PCC $\geq$ 0.7) with 239, 376, and 182 proteins.

To validate the predicted partners, we performed affinity purification experiments starting with Ser-Leu as the bait (c.f. Methods). We used agarose beads coupled to Ser-Leu by the NH_2 group of serine (N-Ser-Leu) or the COOH group of leucine (Ser-Leu-C). We found 162 proteins that were significantly enriched in eluates from the N-Ser-Leu and Ser-Leu-C beads, constituting putative Ser-Leu targets (Figure S4 and Data S12). Proteins involved in protein metabolism (amino acid biosynthesis, protein folding, proteasome, proteins involved in translation and protein targeting) were significantly overrepresented (false discovery rate < 0.05) (Figure S5).

To complement the affinity purification experiments, we used an independent biochemical method for the identification of protein partners of small-molecule ligands, namely thermal proteome profiling. Thermal proteome profiling monitors changes in protein thermal stability caused by ligand binding [16]. We analysed our obtained data by applying the non-parametric analysis of response curves method [36]. The method is independent of melting temperature estimation and tests the differences in curves rather than the differences in melting temperature. We found 94 potential targets that had melting profiles significantly affected by Ser-Leu treatment (Benjamini-Hochberg p-value $\leq$ 0.05) (Data S13 and Figure S6). Again, proteins involved in protein metabolism were significantly enriched (false discovery rate < 0.05) (Figure S7).

In total, 86 proteins, assigned as Ser-Leu binding proteins based on at least two of the three experimental strategies, were queried against a STRING database (Figure 4a). Seventy-seven of the 86 proteins were part of the resulting PPI network (Figure 4b and Figure S8). Functional and enrichment analyses showed a significant overrepresentation of proteins involved in amino acid biosynthesis, translation, and protein folding, degradation, and targeting (Figure S9). Five proteins, identified by all three independent approaches (PROMIS, affinity purification, and thermal proteome profiling), were assigned as

high-confidence Ser-Leu binding proteins (Figure 4a). These five proteins were two subunits of the T-complex (Cct3 and Cct8) [37], the regulatory subunit of acetolactate synthase complex (Ilv6) [38], polyamine acetyltransferase (Paa1) [39], and the yeast prion protein (New1) [40].

Particularly intriguing was appearance of Ilv6, which is involved in the biosynthesis of branched-chain amino acids (valine, leucine, and isoleucine) and feedback inhibited by the binding of valine [38]. Prompted by the published data, we investigated the elution profiles of Ilv6, catalytic subunit of acetolactate synthase complex (Ilv2), and dipeptides containing branched-chain amino acids. We found that, in addition to Ser-Leu, Val-Leu, Leu-Leu, Thr-Leu, Ile-Leu, Asn-Ile, and Thr-Val also co-migrated with the subunits of acetolactate synthase complex (Figure 4c).

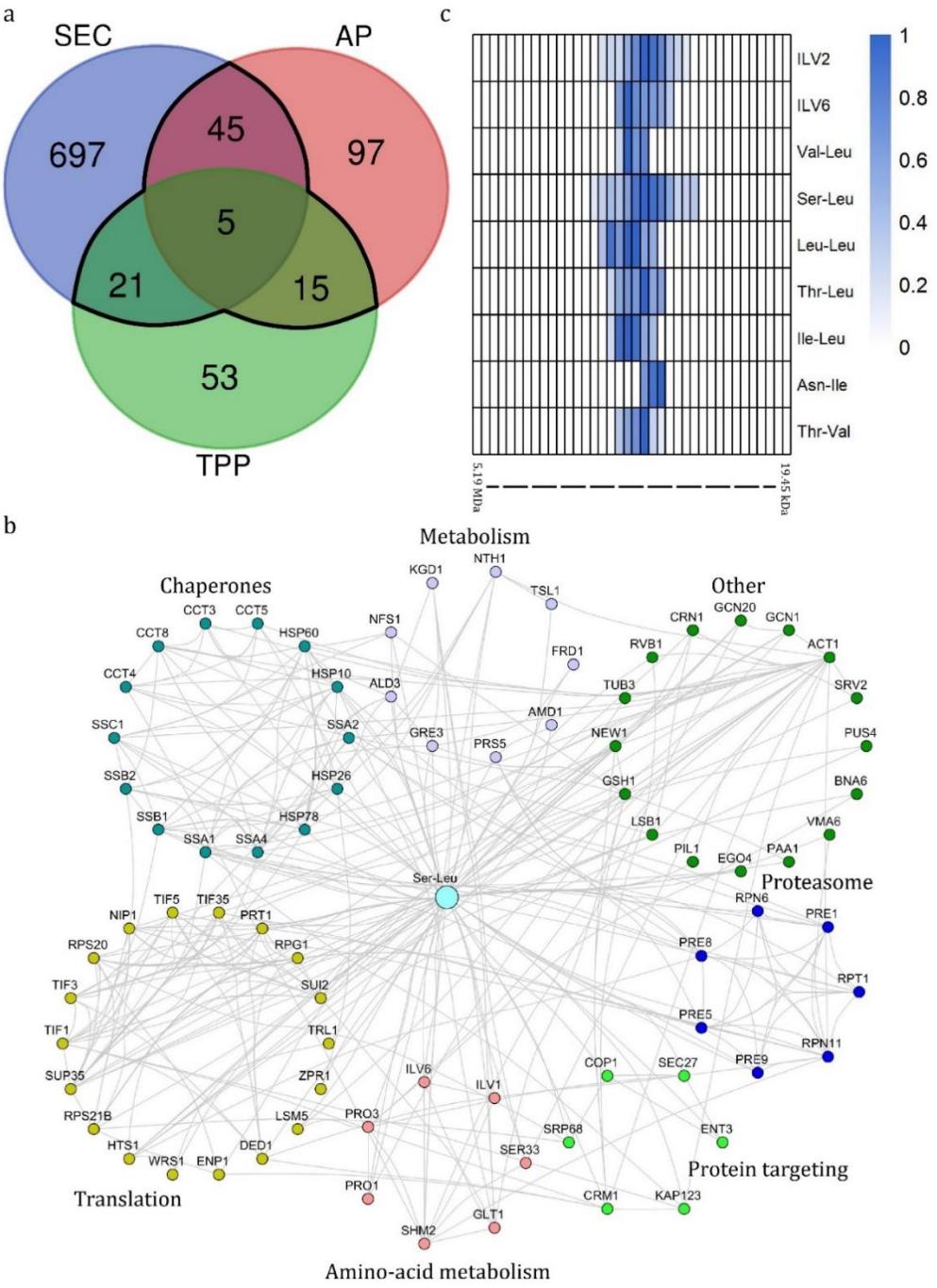


Figure 4: Characterisation of Ser-Leu interactome. a Venn diagram showing the number of putative Ser-Leu targets identified using size exclusion chromatography (SEC), affinity purification (AP), and thermal proteome profiling (TPP). An overlap between at least two orthogonal approaches (86 proteins) was considered to represent the Ser-Leu interactome and is marked in black. b The Ser-Leu interactome network. Edges represent protein-protein interactions and were imported from STRING, based on the experimental evidence (score ≥ 0.4). Functionally related proteins are grouped together. Distinct colours are used to mark different protein groups. c Heatmap showing co-elution of catalytic (Ilv2) and regulatory (Ilv6) subunits of the acetolactate synthase complex with dipeptides containing branched-chain amino acids. The intensity was calculated relative to the maximum intensity measured across the SEC fractions. The theoretical molecular weight was calculated using reference proteins.

Ser-Leu is a regulator of the glycolytic enzyme Pgk1.

The Ser-Leu elution profile is characterised by three local maxima, indicating co-presence of interacting proteins in respective fractions. However, when we checked the PROMIS data set, we found that all five of the high-confidence protein targets (identified simultaneously by affinity purification, thermal proteome profiling, and PROMIS) corresponded to either the first or second Ser-Leu peaks, but none co-fractionated with the third peak. Prompted by our earlier observation that Tyr-Asp binds to plant glyceraldehyde-3-P dehydrogenase (GAPDH) [22], we searched for glycolytic enzymes among the 182 proteins co-eluting with the third peak of Ser-Leu and identified phosphoglycerate kinase (Pgk1) as a putative target of Ser-Leu (Figure 5a).

We validated the direct interaction between Pgk1 and Ser-Leu using microscale thermophoresis with a determined K_d of 416 µM (Figure 5b) [41]. In comparison, no interaction could be measured between Pgk1 and Tyr-Asp, (Figure S10), and between Pgk1 and serine, which was used as a negative control (Figure S10). We decided for serine, as analysis of dipeptide uptake in yeast showed that an amino acid residue at the N-terminus has a more significant role in dipeptide recognition than one on the C-terminus [42]. In line with our results, recent systematic analysis of PMIs in central metabolism using nuclear magnetic resonance showed that Pgk in *Escherichia coli* does not bind to either serine or leucine [43].

We characterised the effect of the interaction between Pgk1 and Ser-Leu by testing whether Ser-Leu affects the activity of recombinant Pgk1 [44]. We used a stopped enzymatic assay (Figure S11), which measures the conversion of 3-phosphoglycerate (3PGA) to bisphosphoglycerate (BPGA) and subsequently to glyceraldehyde-3-P (GAP), dihydroxyacetone-P (DAP), and finally glycerol-3-P (G3P) [45,46]. Micromolar concentrations of Ser-Leu significantly increased the activity of Pgk1; however, the activating effect was observable only at relatively low concentrations of the ATP used in the assay (below V_{max}) (Figure 5c, Figure S12). Since high concentrations of ATP diminished the activating effect of Ser-Leu, we hypothesized that Ser-Leu may increase affinity of Pgk1 towards ATP. To test this assumption, we used microscale thermophoresis to determine the K_d of the interaction between Pgk1 and ATP in the presence of a saturating concentration of Ser-Leu (Figure 5d). We first validated the interaction between Pgk1 and ATP (K_d of 122 µM). Next, we demonstrated that Ser-Leu lowered K_d of ATP binding by 40-fold, effectively increasing the affinity of Pgk1 for ATP (K_d of 3 µM).

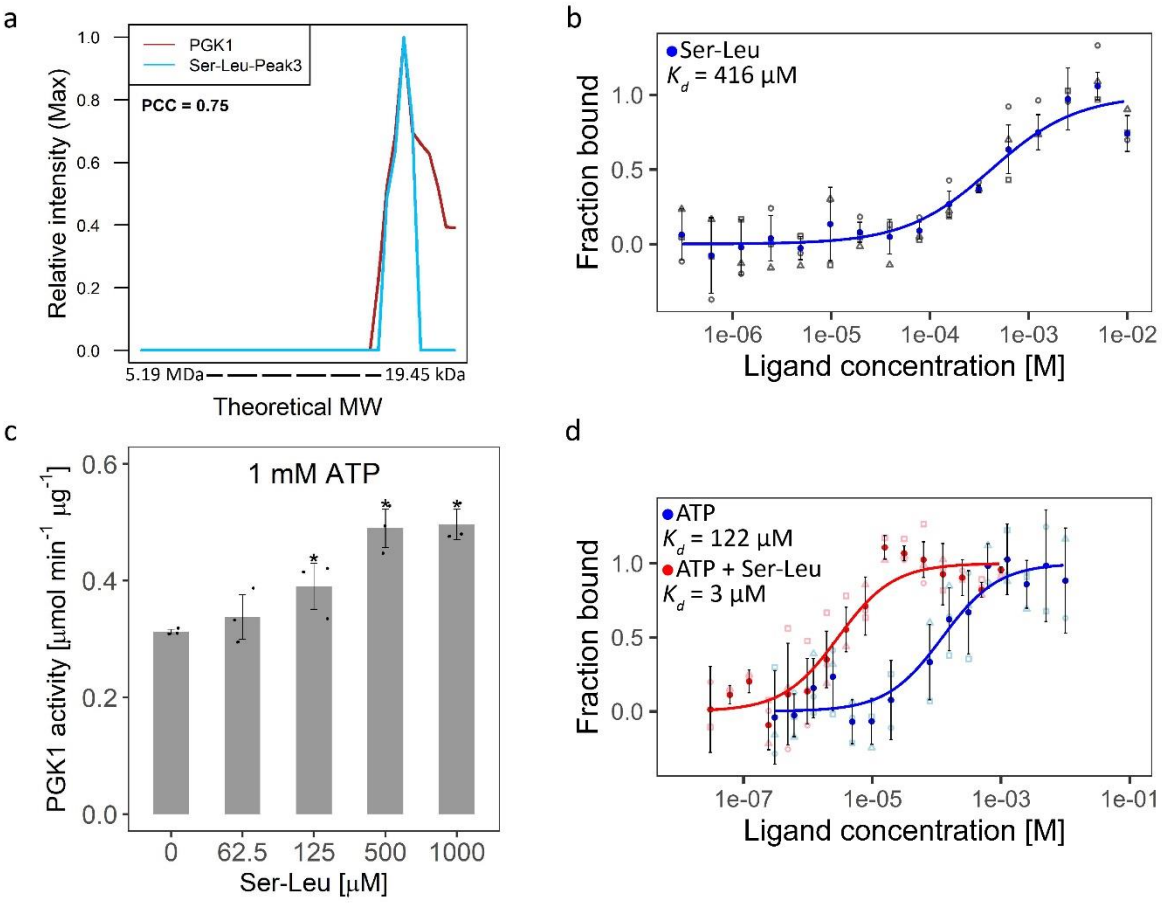


Figure 5: Characterisation of the Pgk1–Ser-Leu interaction. a The elution profile of Pgk1 and its putative ligand Ser-Leu. The intensity was calculated relative to the maximum intensity of the molecule measured across size exclusion chromatography fractions. The theoretical molecular weight (MW) was estimated using reference proteins. The Pearson correlation coefficient (PCC) indicates a correlation coefficient calculated between depicted elution profiles. b Microscale thermophoresis analysis of Pgk1 and Ser-Leu binding. Kd indicates dissociation constant. Data represents the means ± SD, n=3 independent samples. c Functional validation of the interaction between Ser-Leu and Pgk1. Ser-Leu significantly increases Pgk1 activity. Data represents the means ± SD, n=3 independent samples. Asterisks denote significant difference (non-paired, two-tailed t-test P-value < 0.05). d Microscale thermophoresis analysis of Pgk1 and ATP binding in the presence of saturating concentrations (4 mM) of Ser-Leu. Kd indicates dissociation constant. Data represents the means ± SD, n=3 independent samples.

Dipeptide accumulation is associated with glucose depletion.

To learn more about the biological context of Ser-Leu action, we decided to investigate dipeptide and amino acid accumulation during growth on glucose at optimal conditions (30 °C). For this purpose, yeast culture was grown to the stationary phase, followed by transfer to a fresh pre-treated medium (see Methods section). Samples were harvested immediately after transfer to fresh medium and at multiple time-points ranging from 15 to 1440 minutes, quenched in methanol and analysed by liquid chromatography-

mass spectrometry. With few exceptions, all of the measured dipeptides accumulated after 180 minutes of growth (Figure 6a), which corresponds to glucose depletion (Figure S13a). When compared with dipeptides, amino acids displayed a different accumulation pattern, characterised by an increase after 360 and 1440 minutes of growth (Figure 6b). More specifically, level of Ser-Leu and leucine decreased after 30 minutes of growth and started to accumulate after 180 and 1440 minutes of cultivation, respectively (Figure 6c). In contrast, level of serine undergoes fewer fluctuations. Similarly to leucine, it accumulates after 1440 minutes of growth.

Ser-Leu feeding affects both central metabolism and yeast growth

In addition to Pgk1, Ser-Leu protein interactome comprised numerous other enzymes, from amino acids biosynthesis (Prs5, Ser33, Shm2, Ilv6, Glt1, Pro1, Pro3), the TCA cycle (Kgd1), purine (Amd1) and NAD metabolism (Bna6). To examine whether Ser-Leu binding translates into a metabolic effect we followed changes in relative metabolite levels (here described as total intensity) and redistribution of carbon isotope (enrichment level [%] multiplied by relative metabolite level, here described as 13C fraction intensity) in yeast cells upon Ser-Leu supplementation. Specifically, yeast cells at stationary phase were fed with ^{13}C glucose together with either mock, 100 µM Ser-Leu or a mix of 100 µM serine and 100 µM leucine. Samples were harvested at multiple time-points ranging from 5 to 240 minutes following treatment, quenched in methanol and analysed by gas chromatography- and liquid chromatography-mass spectrometry. Over time, the ^{13}C glucose is taken up and metabolized by the cell and metabolites become enriched for ^{13}C until the steady state enrichment is reached. While changes in metabolites levels are valuable information to describe the metabolic state of an organism, they are limited in providing information regarding the flow of mass through the system. ^{13}C enrichment provides further information to access the conversion rates of labelled substrates through metabolism, which can be used to estimate production rate of a given metabolite [47].

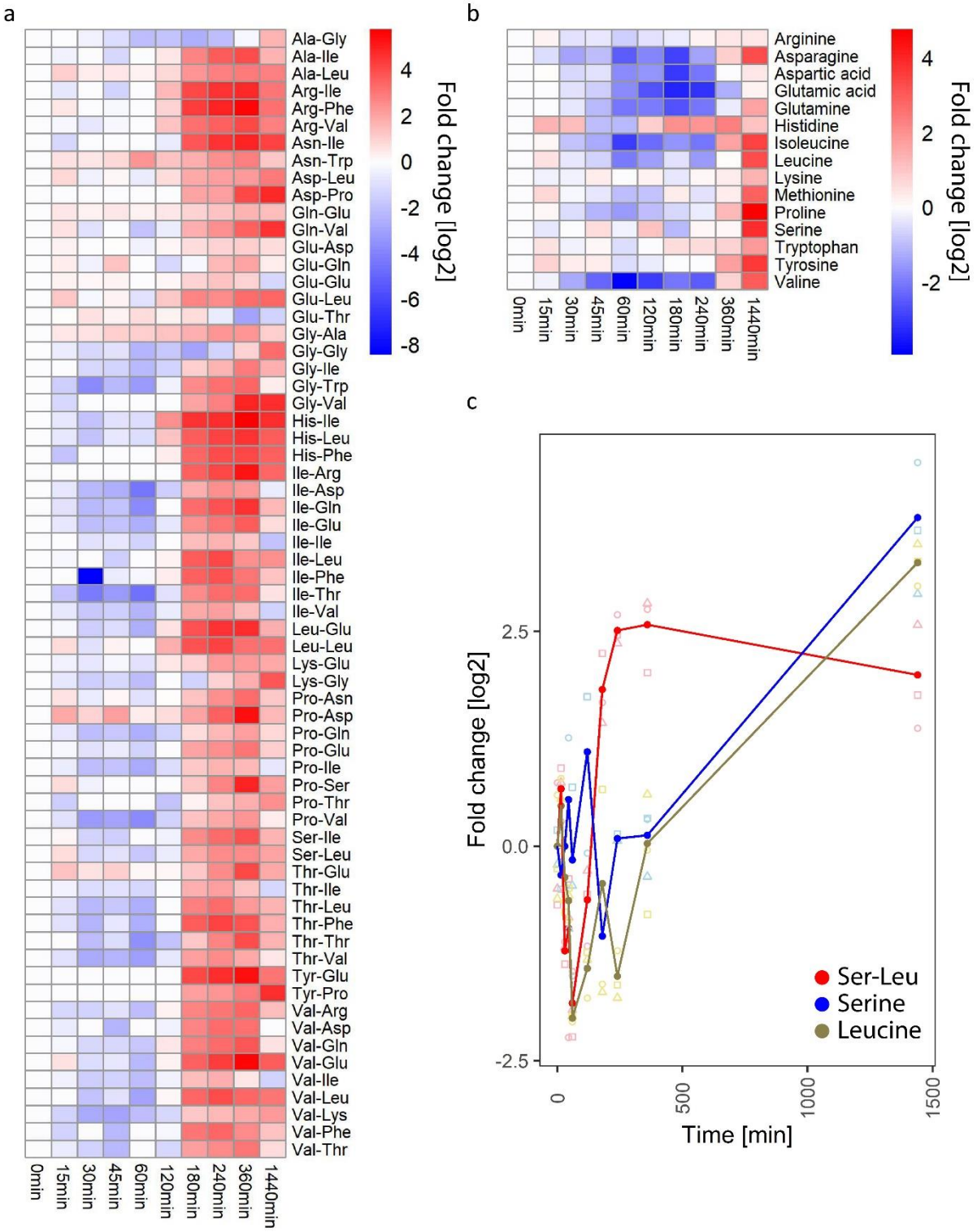


Figure 6: Analysis of dipeptide and amino acids fluctuations in yeast under control growth conditions (30 °C). a Heatmap showing fluctuation of dipeptides level in yeast. b Heatmap showing fluctuation of amino acids level in yeast. c Plot showing fluctuation of Ser-Leu, serine and leucine level in yeast. Shown are relative changes to time point 0. Data represents the means n=3 independent samples. Ratios were log transformed (log base 2).

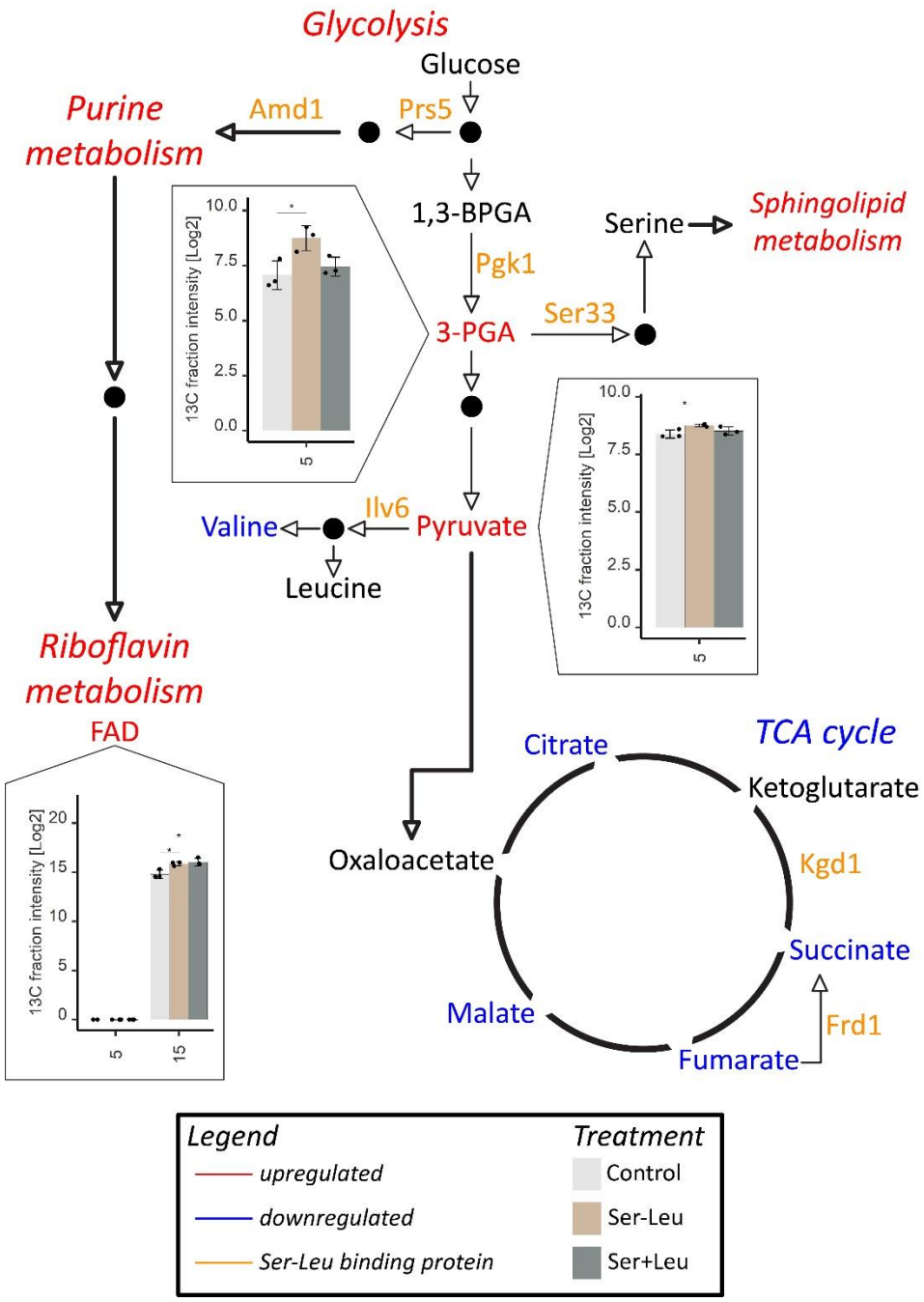


Figure 7: YSBN2 response to Ser-Leu supplementation. Liquid chromatography- and gas chromatograhy-mass spectrometry analysis of metabolomic changes caused upon supplementation with 100 µM Ser-Leu or mixture of 100 µM serine and 100 µM leucine. Presented are changes in metabolite levels (here described as total intensity) and redistribution of carbon isotope (enrichment level [%] multiplied by metabolite level, here described as 13C fraction intensity) in yeast cells. 13C enrichment in combination with metabolite levels provides information regarding the conversion rate of labelled substrate to the metabolite. X-axis represents time [min] upon treatment. Data represents the means ± SD, n=3 independent samples. Asterisks denote significant difference (Tukey's test, * P-value < 0.05, ** P-value < 0.01). Ser-Leu binding proteins are marked orange. Level of metabolites marked red was significantly increased at least in one time-point comparing to other treatment. Level of metabolites marked blue was significantly decreased at least in one time-point comparing to other treatment. Presented are cropped images (see Figures S14-S21).

The choice of Ser-Leu concentration was guided by the absolute cellular levels of Ser-Leu, which we estimated to approximate 6 µM by spiking different amounts of Ser-Leu (from 100 nM to 100 µM) into metabolic extract prepared from ^{13}C labelled S288c yeast culture corresponding to the stationary phase of growth.

GC- and LC-MS analysis of the Ser-Leu, serine and leucine concentrations in the Ser-Leu supplemented cells revealed rapid Ser-Leu accumulation, which remained constant over-time (Figure S13b). Neither serine nor leucine accumulated, at least during the duration of the Ser-Leu treatment, arguing that Ser-Leu was not degraded to its constituent amino acids (Figure S13cd).

Most conspicuously, Ser-Leu treatment led to a stark increase in the *de novo* production rate of the 3-phosphoglycerate (3-fold change), directly downstream of the Pgk1 activity, followed by increase in pyruvic acid production (Figure 7 and Figure S14). Interestingly, the excess of glycolytic 3-phosphoglycerate and pyruvate was directed away from the tricarboxylic acid cycle cycle, as *de novo* production rates of all of the measured tricarboxylic acid cycle intermediates, citric acid, succinic acid, fumaric acid and malic acid were decreased (Figure S15). Similarly, and possibly as a consequence, also *de novo* synthesis of tricarboxylic acid cycle-derived amino acids: methionine, saccharopine (intermediate in the metabolism of lysine), proline, arginine and aspartate was downregulated (Figures S16-S18). Moreover, Ser-Leu treatment led to i) up-regulation of *de novo* synthesis of 5'-GMP, and accumulation of 3'-AMP, and adenine all being intermediates of purine metabolism (Figure S19), ii) increased levels of two intermediates of sphingolipid metabolism, sphingosine and hydroxypalmitic acid (Figure S20), and iii) elevated *de novo* synthesis of cofactors $NADP^+$ and FAD^+ (Figure S21). Similarly, to Ser-Leu, also amino acid feeding resulted in number of metabolic changes. However, the observed effects were different; for instance, in contrast to the Ser-Leu, the amino-acid treatment did not affect *de novo* synthesis of the tricarboxylic acid cycle intermediates succinic acid, fumaric acid and malic acid.

Finally, and to complement our metabolic analysis we tested whether Ser-Leu supplementation affects yeast growth. For this purpose, starved yeast culture in stationary phase was supplemented with glucose together with either mock, 1 mM Ser-Leu or mixture of 1 mM serine and 1 mM leucine (Figure 8a). Yeast growth was monitored by measuring OD_{600nm} using an automatically recording incubator. Ser-Leu treatment affected yeast growth during early exponential phase and supplemented culture reached higher OD_{600nm} than mock. Ser-Leu treatment delayed diauxic shift for 30 minutes, therefore prolonged fermentation, and shortened respiration phase (Figure 8b). In comparison to treatment with dipeptide, supplementation with a mixture of serine and leucine affected yeast growth much later (4 h and 6 h upon treatment with dipeptide and amino acids, respectively) and did not delay the diauxic shift (Figure 8c).

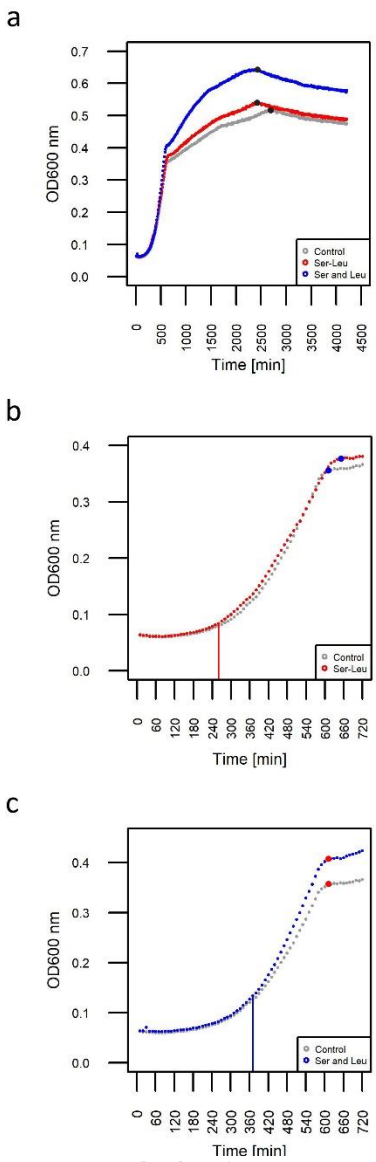


Figure 8: Analysis of YSBN2 growth response to Ser-Leu supplementation. a YSBN2 strain was supplemented with 1 mM Ser-Leu or mixture of 1 mM serine and 1 mM leucine. Growth was monitored by measuring the optical density at 600 nm wavelength for 75 hours using an automatically recording incubator. Data represents the means n = 3. Black dot represents end of respiration phase. b Cropped growth curve of YSBN2 supplemented with Ser-Leu (0 – 12 hours). Blue dot represents diauxic shift. Red, straight line indicates beginning of Ser-Leu treatment effect on yeast growth. c Cropped growth curve of YSBN2 supplemented with Ser and Leu (0 – 12 hours). Red dot represents diauxic shift. Blue, straight line indicates beginning of Ser and Leu treatment effect on yeast growth.

3.4 Discussion

Herein, we used PROMIS to chart a map of protein-small molecule interactions (PMIs) in the model yeast *Saccharomyces cerevisiae.* As a result, we report a unique data set resulting from an analysis of endogenous protein-metabolite and protein-protein complexes. Our most remarkable observation relates to the wealth of small molecules present in the protein complexes; this attests to the complexity of the protein–small molecule interactome and highlights an important but severely understudied role of small molecules as protein regulators. We report 225 previously predicted PMIs that could be validated using PROMIS. Considering that the STITCH database contains 87 true interactions for the same subset of metabolites and proteins, then a single PROMIS experiment was sufficient to nearly quadruple the number (from 87 to 312). We successfully queried the list of 225 validated interactions for binding events with a putative regulatory role, such as between Pnp1 and xanthine. However, *in vivo* significance of the xanthine inhibition of Pnp1 activity remains to be tested, xanthine binding to Pnp1 is an excellent example where querying a single PROMIS dataset is sufficient to retrieve regulatory interactions.

In addition to the previously predicted PMIs, the presented PROMIS data set can be mined for new binding events, assisting the discovery and functional characterisation of small molecule regulators. In line with an analogous PROMIS study in Arabidopsis proteogenic dipeptides stood out as a major group of protein-bound small molecules [22]. A role of dipeptides in the regulation of central metabolism has been discussed before. Increase of proteogenic dipeptides in

tumor-associated cells correlated with the glycolytic capacity of the tumor [48]. In comparison, treatment with the non-proteogenic dipeptide carnosine (β-alanyl-L-histidine) reduced the proliferative capacity of human gastric cancer cells by inhibiting glycolysis, mitochondrial oxidative phosphorylation and respiration [49]. Finally, the acidic dipeptide Tyr-Asp was found among small molecule ligands of a glycolytic enzyme, glyceraldehyde-3-P dehydrogenase [22]. Here, we could demonstrate that Ser-Leu affects glycolysis *via* direct binding and activation of Pgk1. Consistent with the *in vitro* results, Ser-Leu feeding led to a rapid accumulation of an important glycolytic intermediate 3-phosphoglycerate. 3-phosphoglycerate is eventually converted into pyruvate but can also be re-directed into serine biosynthesis. While serine is an entry point into one-carbon metabolism, pyruvate is utilized to produce energy *via* either the tricarboxylic acid cycle (respiration) or the ethanol production (fermentation). Reduced levels of the tricarboxylic acid cycle intermediates, and tricarboxylic acid cycle-derived amino-acids, measured in response to the Ser-Leu supplementation point to pyruvate being directed away from the respiration, most likely into fermentation. These data are in line with the measured growth effects. Ser-Leu treatment delayed the diauxic shift, which is indicative of Ser-Leu supporting fermentation over respiration. Moreover, and since Ser-Leu accumulation accompanies glucose depletion characteristic for the late logarithmic phase of growth, we propose that Ser-Leu, and possibly also other proteogenic dipeptides, reinforce sugar repression of the tricarboxylic acid cycle in yeast cells when the glucose levels fall low.

The role of Pgk1 in coordination of glycolysis and tricarboxylic acid cycle by increasing lactate production and suppressing mitochondrial pyruvate utilisation is well established in cancer cells [50-52]. Herein, and based on the similarities of yeast and cancer metabolism, we speculate that in addition to the posttranslational modifications of the mammalian Pgk1, such as phosphorylation and O-GlcNAcylation, that promote the switch from the tricarboxylic acid cycle, into lactate production, dipeptide binding may constitute an additional regulatory mechanism to promote the glycolytic capacity of cancer cells [48,53]. Notably, and in addition to being an enzyme, Pgk1 is also a protein kinase; known phosphorylation targets include pyruvate dehydrogenase kinase 1 (Pdhk1) and autophagy regulator Beclin1 [52,54]. Considering that Ser-Leu increases the Pgk1 affinity towards ATP, it will be interesting to test whether Ser-Leu binding, in addition to enzymatic, affects Pgk1 kinase activity.

Although, in the present study, we focused on the Ser-Leu regulation of Pgk1; it has to be noted that Ser-Leu protein interactome comprises numerous other enzymes involved in amino acid, purine, and NAD metabolism. Therefore, it is highly plausible that metabolic changes associated with Ser-Leu supplementation go beyond Pgk1 activation. For instance, Ser-Leu feeding inhibited valine production, despite the increased availability of pyruvate, which serves as a direct substrate for the synthesis of branched-chain amino acids. Obtained results indicate the presence of a regulatory interaction stopping excess of pyruvate from being directed towards the synthesis of valine. We hypothesize that such regulation

may be achieved by Ser-Leu inhibition of the regulatory subunit of the acetolactate synthase complex, Ilv6, which is among five high-confidence Ser-Leu protein targets. Based on the Ser-Leu co-elution with other branched chain-amino acid containing dipeptides, we also speculate that the function of Ser-Leu will be redundant with chemically similar dipeptides, and possibly even tripeptides, but as shown before for other dipeptides [55,56], different from the constituent amino acids, serine and leucine.

Finally, the regulatory role of dipeptides would become particularly important in conditions that promote protein degradation. We have recently shown that in response to abiotic stress, such as heat and dark, plants accumulate dipeptides in the autophagy-dependent manner [57]. Autophagy was also shown to account for the increase in dipeptides reported in the mammalian pro-tumorigenic cell lines [48]. Here, we could demonstrate that yeast accumulate dipeptides in response to glucose deprivation. However; it requires to be experimentally tested whether observed accumulation is autophagy dependent, glucose depletion was shown to trigger autophagy [58,59] and metabolic phenotype associated with Ser-Leu feeding such as accumulation of RNA degradation products, changes in lipid metabolism and cofactor production is reminiscent with the metabolic alterations downstream of autophagy [58,60-63].

In summary, the proteome and metabolome-wide map of the protein-protein and protein-metabolite complexes that we present here can be mined for regulatory small molecules, such as the here characterized proteogenic dipeptide Ser-Leu. Yeast growth strictly depends on the carbon availability; glucose being the primary carbon source [64]. The transition between growth on glucose to growth on ethanol is accompanied by acute metabolic rearrangement [65]. However, intensively studied, the underlying regulatory mechanisms are not entirely understood. Our work points to the involvement of proteogenic dipeptides in the control of yeast metabolism and diauxic shift, by direct regulation of enzyme activities and carbon flux. In a broader sense, presented data support proteogenic dipeptides' regulatory role at the nexus of protein degradation and central metabolism.

3.5 Methods

Yeast growth conditions, cell lysis, and extraction of native complexes for PROMIS.

The YSBN2 strain of *S. cerevisiae* was cultivated at 28 °C with moderate shaking until it reached the logarithmic phase (OD600 = 0.3–0.5) and used for preparation of soluble fraction containing endogenous complexes (Supplementary Methods).

Size exclusion chromatography.

2 mL of concentrated soluble fraction, corresponding to 40 mg of protein, was separated using a Sepax SRT SEC-300 21.2 × 300 mm column (Sepax Technologies, Inc., Delaware Technology Park,

separation range 1.2 mDa to 10 kDa) connected to an ÄKTA explorer 10 (GE Healthcare Life Science, Little Chalfont, UK) using a 7 mL/min flow rate, 4 °C. Equilibration of the column and separation were performed using 50 mM AmBIC pH 7.5, 150 mM NaCl, 1.5 mM $MgCl_2$, and 48 1-mL fractions were collected from the 39 mL to 86 mL elution volume. When compared with previous studies, the separation time decreased to less than 20 min [66]. The fractions were frozen by snap freezing in liquid nitrogen and subsequently lyophilised and stored at –80 °C for metabolite and protein extractions.

The chromatogram of the absorption at 280 nm indicates reproducible fractionation (R_{avg} = 0.98) of the native complexes present in the input samples (Figure S22). To correct for unspecific metabolite binding to the column matrix, a control experiment with a protein-free sample was performed. For this purpose, proteins were precipitated from the extract of native complexes using 80% acetone. An extract of total small molecules (bound and unbound) was then solubilised in a lysis buffer before fractionation on the SEC column. The mass features present in the SEC mobile phase (blank sample) were also quantified and filtered out as potential contaminants coming from chemicals.

Extraction of proteins and polar metabolites.

Proteins and metabolites from the lyophilised fractions were extracted using a methyl-tert-butyl ether (MTBE)/methanol/water solvent system, which separates molecules into pellets (proteins), organics (lipids), and an aqueous phase (primary and secondary metabolites) [67]. Molecules were extracted from each fraction by adding 1 mL of a homogenous mixture of -20 °C methanol:methyl-tert-butyl-ether:water (1:3:1), shaking for 10 min at 4 °C, incubating 10 min in an ice cooled ultrasonication bath and shaking again for 10 min at 4 °C. Next, 500 µL of UPLC grade methanol:water (1:3) was added to each fraction. The homogenates were vortexed and centrifuged for 5 min at 20,800 g, RT. Equal volumes of the polar fraction and protein pellet were dried in a centrifugal evaporator and stored at –80 °C until they were processed further. Qualitative and quantitative analysis of the fractionated proteins using the Bradford assay [68] and SDS-PAGE, respectively, showed that the majority of the proteins eluted in fractions corresponding to MW above 20 kDa (A6–C13, referred to as protein-containing fractions). Fractions C14–C15 contained low protein amounts with MWs below 20kDa. Therefore, fractions C14 to D9 were considered to contain mostly protein fragments and metabolites that were not bound to proteins.

LC-MS metabolomics.

After extraction, the dried aqueous phase was suspended in 100 µL of water and sonicated for 5 min using ultrasonication bath. Samples were centrifuged 10 min at 20,800 g, RT. Supernatant was transferred to UPLC glass vial. Polar metabolite extract was separated using a UPLC equipped with an HSS

T3 C18 reversed-phase column and mass spectra were acquired using an Exactive mass spectrometer in positive and negative ionization modes [67]. 3 µl of the sample was loaded onto the column for each ionization mode. To create the required gradient for metabolite measurement, mobile phase solutions were prepared as follows: buffer A (0.1% formic acid in H2O) and buffer B (0.1% formic acid in ACN). Metabolites were separated at 400 µl/min using the following gradient: 1 min 1% LC-MS mobile phase buffer B, 11 min linear gradient from 1% to 40% buffer B, 13 min linear gradient from 40% to 70% buffer B, then 15 min linear gradient from 70% to 99% buffer B, and hold a 99% buffer B concentration until 16 min. Starting from 17 min, use a linear gradient from 99% to 1% buffer B. Re-equilibrate the column for 3 min with 1% buffer B before measuring the next sample. Mass spectra were acquired using following settings: mass range from 100 to 1500 *m/z*, resolution set to 25,000, loading time restricted to 100 ms, AGC target set to $1e^6$, capillary voltage to 3kV with a sheath gas flow and auxiliary gas value of 60 and 20, respectively. The capillary temperature was set to 250 °C and skimmer voltage to 25V.

LC-MS/MS of proteins.

Proteins from each fraction were digested using LysC/Trypsin Mix (Promega Corp., Fitchburg, WI) according to the manufacturer's instructions. Digested proteins were desalted on self-made C18 Empore® extraction discs (3M, Maplewood, MN) STAGE tips [69]. Dried peptides were separated using C18 reversed-phase column connected to an ACQUITY UPLC M-Class system in a 120 min gradient (Supplementary Methods).

Data processing of LC-MS metabolite and protein data.

Data were processed using Expressionist Refiner MS 11.0 (Genedata AG, Basel, Switzerland) using settings described previously [66], with minor changes, and MaxQuant version 1.6.0.16 [70] and its built-in search engine, Andromeda [71]. Detailed settings and further data processing leading to determination of molecular complexes were described in Supplementary Methods.

Overexpression and purification of Pnp1 and Pgk1.

Pnp1 and Pgk1 overexpressing yeast strains were purchased from Dharmacon and are part of the yeast ORF collection [44]. Yeast cultivation and procedure of protein purification were described in Supplementary Methods.

Pnp1 enzymatic assay.

The method for Pnp1 enzymatic activity measurement was adapted from previous studies [34] (Supplementary Methods).

Affinity purification using Ser-Leu agarose beads.

Yeast cultivation and procedure of affinity purification were described in Supplementary Methods.

Thermal proteome profiling of the Ser-Leu-treated cell extracts.

Thermal proteome profiling of Ser-Leu-treated cell extracts was performed as described earlier [16] and analysed using a TPP package available on Bioconductor and NPARC [36] (Supplementary Methods).

Microscale thermophoresis.

Microscale thermophoresis measurements were performed using a Monolith NT.115 instrument (Nanotemper) (Supplementary Methods).

Pgk1 enzymatic assay.

Pgk1 activity was assayed using an optimised stopped assay, and the product was determined by an enzyme-cycling system, as described earlier, with minor modifications [45,46] (Supplementary Methods).

Dipeptide and amino acid accumulation during growth.

Yeast were cultivated at control conditions as described earlier [72] (Supplementary Methods).

Changes in growth and metabolism upon Ser-Leu supplementation (13C-isotope-labeling experiment)

Chemical treatment was applied by supplementing yeast culture with either mock, 100 µM Ser-Leu or a mixture of 100 µM serine and 100 µM leucine (Supplementary Methods).

Statistics and Reproducibility

Statistical analysis was performed using R [73]. One-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test or unpaired, two-tailed Student's t-test, was performed. The P-values <0.05 were considered significant and are represented as * $P < 0.05$, ** $P < 0.01$. For all statistical analysis data from at least three independent measurements was used. The exact number of replicates and detailed description of statistics performed are indicated in individual figure captions and methods.

Data Availability

The mass spectrometry proteomics data that support the findings of this study have been deposited in the ProteomeXchange Consortium via the PRIDE [74] partner repository with the dataset identifier PXD021530.

Code Availability

R code used for data processing and analysis was submitted to GitHub repository and deposited to Zenodo [75]. Code can be accessed at https://github.com/Marcin-Luzarowski/PROMIS.git or https://doi.org/10.5281/zenodo.4146637.

Acknowledgements

We thank Änne Michaelis and Fatıma Şen for excellent technical assistance, Krzysztof Bajdzienko, Dariusz Bieńkowski and Andreas Donath for help with establishing a user-friendly interface, and Prof. Dr. Zoran Nikoloski for a careful read of the manuscript and valuable comments.

Competing Interests

The authors declare no competing interests.

3.6 References – Publication #1

1 Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nature biotechnology 31, 357 (2013).

2 Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. Nature Reviews Microbiology 12, 327-340 (2014).

3 Escobar-Henriques, M. & Daignan-Fornier, B. Transcriptional regulation of the yeast gmp synthesis pathway by its end products. Journal of Biological Chemistry 276, 1523-1530 (2001).

4 Grove, A. Regulation of metabolic pathways by MarR family transcription factors. Computational and structural biotechnology journal 15, 366-371 (2017).

5 Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. Cell 141, 1117-1134 (2010).

6 Kudryashov, D. S. & Reisler, E. ATP and ADP actin states. Biopolymers 99, 245-256 (2013).

7 Hahn, S. & Young, E. T. Transcriptional regulation in Saccharomyces cerevisiae: transcription factor regulation and function, mechanisms of initiation, and roles of activators and coactivators. Genetics 189, 705-736 (2011).

8 Hackett, S. R. et al. Systems-level analysis of mechanisms regulating yeast metabolic flux. Science 354, aaf2786 (2016).

9 Machado, D., Herrgard, M. J. & Rocha, I. Modeling the Contribution of Allosteric Regulation for Flux Control in the Central Carbon Metabolism of E. coli. Frontiers in Bioengineering and Biotechnolgy 3, 154, doi:10.3389/fbioe.2015.00154 (2015).

10 Kochanowski, K. et al. Few regulatory metabolites coordinate expression of central metabolic genes in Escherichia coli. Molecular Systems Biology 13(1), doi:https://doi.org/10.15252/msb.20167402 (2017).

11 You, C. et al. Coordination of bacterial proteome with metabolism by cyclic AMP signalling. Nature 500, 301 (2013).

12 Kuhn, M. et al. STITCH 2: an interaction network database for small molecules and proteins. Nucleic acids research 38, D552-D556 (2010).

13 Szklarczyk, D. et al. STITCH 5: augmenting protein–chemical interaction networks with tissue and affinity data. Nucleic acids research, gkv1277 (2015).

14 Luzarowski, M. & Skirycz, A. Emerging strategies for the identification of protein-metabolite interactions. Journal of experimental botany, doi:10.1093/jxb/erz228 (2019).

15 Kosmacz, M. et al. Interaction of 2', 3'-cAMP with Rbp47b plays a role in stress granule formation. Plant Physiology 177, 411-421 (2018).

16 Savitski, M. M. et al. Tracking cancer drugs in living cells by thermal profiling of the proteome. Science 346, 1255784 (2014).

17 Lomenick, B. et al. Target identification using drug affinity responsive target stability (DARTS). Proceedings of the National Academy of Sciences 106, 21984-21989 (2009).

18 Piazza, I. et al. A map of protein-metabolite interactions reveals principles of chemical communication. Cell 172, 358-372 (2018).

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

20 Luzarowski, M. et al. Affinity purification with metabolomic and proteomic analysis unravels diverse roles of nucleoside diphosphate kinases. Journal of Experimental Botany 68, 3487-3499 (2017).

21 Haberkant, P. & Holthuis, J. C. Fat & fabulous: bifunctional lipids in the spotlight. Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids 1841, 1022-1030 (2014).

22 Veyel, D. et al. PROMIS, global analysis of PROtein-Metabolite Interactions using Size separation in Arabidopsis thaliana. Journal of Biological Chemistry 293, 12440-12453 (2018).

23 Sokolowska, E. M., Schlossarek, D., Luzarowski, M. & Skirycz, A. PROMIS: Global Analysis of PROtein-Metabolite Interactions. Current protocols in plant biology 4, e20101 (2019).

24 Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. Nature 425, 737-741 (2003).

25 Cherry, J. M. et al. Saccharomyces Genome Database: the genomics resource of budding yeast. Nucleic acids research 40, D700-D705 (2011).

26 Gorka, M. et al. Protein Complex Identification and quantitative complexome by CN-PAGE. Scientific reports 9, 1-14 (2019).

27 Havugimana, P. C. et al. A census of human soluble protein complexes. Cell 150, 1068-1081 (2012).

28 Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. Nature 402, C47 (1999).

29 Gauthier, S. et al. Co-regulation of yeast purine and phosphate pathways in response to adenylic nucleotide variations. Molecular microbiology 68, 1583-1594 (2008).

30 Saint-Marc, C. et al. Phenotypic consequences of purine nucleotide imbalance in Saccharomyces cerevisiae. Genetics, doi:10.1534/genetics.109.105858 (2009).

31 Iglesias-Gato, D., Martín-Marcos, P., Santos, M. A., Hinnebusch, A. G. & Tamame, M. Guanine nucleotide pool imbalance impairs multiple steps of protein synthesis and disrupts GCN4 translational control in Saccharomyces cerevisiae. Genetics 187, 105-122 (2011).

32 Breton, A. et al. Lethal accumulation of guanylic nucleotides in Saccharomyces cerevisiae HPT1 deregulated mutants. Genetics, doi:10.1534/genetics.107.083295 (2008).

33 Lecoq, K., Belloc, I., Desgranges, C., Konrad, M. & Daignan-Fornier, B. YLR209c encodes Saccharomyces cerevisiae purine nucleoside phosphorylase. Journal of bacteriology 183, 4910-4913 (2001).

34 Cattaneo, G. et al. Development, validation and application of a 96-well enzymatic assay based on LC-ESI-MS/MS quantification for the screening of selective inhibitors against Mycobacterium tuberculosis purine nucleoside phosphorylase. Analytica chimica acta 943, 89-97 (2016).

35 Du, F., Navarro-Garcia, F., Xia, Z., Tasaki, T. & Varshavsky, A. Pairs of dipeptides synergistically activate the binding of substrate by ubiquitin ligase through dissociation of its autoinhibitory domain. Proceedings of the National Academy of Sciences 99, 14110-14115 (2002).

36 Childs, D. et al. Nonparametric Analysis of Thermal Proteome Profiles Reveals Novel Drug-binding Proteins. Molecular & Cellular Proteomics 18, 2506-2515 (2019).

37 Narayanan, A., Pullepu, D. & Kabir, M. A. The interactome of CCT complex–A computational analysis. Computational biology and chemistry 64, 396-402 (2016).

38 Pang, S. S. & Duggleby, R. G. Expression, purification, characterization, and reconstitution of the large and small subunits of yeast acetohydroxyacid synthase. Biochemistry 38, 5222-5231 (1999).

39 Liu, B., Sutton, A. & Sternglanz, R. A yeast polyamine acetyltransferase. Journal of Biological Chemistry 280, 16659-16664 (2005).

40 Inoue, Y., Kawai-Noma, S., Koike-Takeshita, A., Taguchi, H. & Yoshida, M. Yeast prion protein New1 can break Sup35 amyloid fibrils into fragments in an ATP-dependent manner. Genes to Cells 16, 545-556 (2011).

41 Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D. & Duhr, S. Protein-binding assays in biological liquids using microscale thermophoresis. Nature communications 1, ncomms1093 (2010).

42 Ito, K. et al. Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library. Nature communications 4, 2502 (2013).

43 Diether, M., Nikolaev, Y., Allain, F. H. & Sauer, U. Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli. Molecular Systems Biology 15 (2019).

44 Gelperin, D. M. et al. Biochemical and genetic analysis of the yeast proteome with a movable ORF collection. Genes & development 19, 2816-2826 (2005).

45 Gibon, Y., Vigeolas, H., Tiessen, A., Geigenberger, P. & Stitt, M. Sensitive and high throughput metabolite assays for inorganic pyrophosphate, ADPGlc, nucleotide phosphates, and glycolytic intermediates based on a novel enzymic cycling system. The Plant Journal 30, 221-235 (2002).

46 Steinhauser, M.-C. et al. Enzyme activity profiles during fruit development in tomato cultivars and Solanum pennellii. Plant Physiology 153, 80-98 (2010).

47 Buescher, J. M. et al. A roadmap for interpreting 13C metabolite labeling patterns from cells. Current opinion in biotechnology 34, 189-201 (2015).

48 Chaudhri, V. K. et al. Metabolic Alterations in lung cancer–associated fibroblasts correlated with increased glycolytic metabolism of the tumor. Molecular Cancer Research 11, 579-592 (2013).

49 Shen, Y. et al. Carnosine inhibits the proliferation of human gastric cancer SGC-7901 cells through both of the mitochondrial respiration and glycolysis pathways. PloS one 9, e104632 (2014).

50 Nie, H. et al. O-GlcNAcylation of PGK1 coordinates glycolysis and TCA cycle to promote tumor growth. Nature Communications 11, 36, doi:10.1038/s41467-019-13601-8 (2020).

51 Zhang, Y. et al. Macrophage-associated PGK1 phosphorylation promotes aerobic glycolysis and tumorigenesis. Molecular cell 71, 201-215. e207 (2018).

52 Li, X. et al. Mitochondria-translocated PGK1 functions as a protein kinase to coordinate glycolysis and the TCA cycle in tumorigenesis. Molecular cell 61, 705-719 (2016).

53 Naka, K. et al. Dipeptide species regulate p38MAPK–Smad3 signalling to maintain chronic myelogenous leukaemia stem cells. Nature communications 6, 8039 (2015).

54 Qian, X. et al. Phosphoglycerate kinase 1 phosphorylates Beclin1 to induce autophagy. Molecular cell 65, 917-931. e916 (2017).

55 Kanegawa, N., Suzuki, C. & Ohinata, K. Dipeptide Tyr-Leu (YL) exhibits anxiolytic-like activity after oral administration via activating serotonin 5-HT1A, dopamine D1 and GABAA receptors in mice. FEBS letters 584, 599-604 (2010).

56 Zhang, Z. et al. The novel dipeptide Tyr-Ala (TA) significantly enhances the lifespan and healthspan of Caenorhabditis elegans. Food & function 7, 1975-1984 (2016).

57 Thirumalaikumar, V. P., Wagner, M., Balazadeh, S. & Skirycz, A. Autophagy is responsible for the accumulation of proteogenic dipeptides in response to heat stress in Arabidopsis thaliana. The FEBS Journal, doi:https://doi.org/10.1111/febs.15336 (2020).

58 Galluzzi, L., Pietrocola, F., Levine, B. & Kroemer, G. Metabolic control of autophagy. Cell 159, 1263-1276 (2014).

59 Reggiori, F. & Klionsky, D. J. Autophagic processes in yeast: mechanism, machinery and regulation. Genetics 194, 341-361, doi:10.1534/genetics.112.149013 (2013).

60 Alexaki, A. et al. Autophagy regulates sphingolipid levels in the liver. Journal of lipid research 55, 2521-2531 (2014).

61 Harvald, E. B., Olsen, A. S. B. & Færgeman, N. J. Autophagy in the light of sphingolipid metabolism. Apoptosis 20, 658-670 (2015).

62 Huang, H. et al. Bulk RNA degradation by nitrogen starvation-induced autophagy in yeast. The EMBO journal 34, 154-168 (2015).

63 Liu, Y. et al. Autophagy-dependent ribosomal RNA degradation is essential for maintaining nucleotide homeostasis during C. elegans development. Elife 7, e36588, doi:10.7554/eLife.36588 (2018).

64 Gancedo, J. M. J. M. & reviews, m. b. Yeast carbon catabolite repression. 62, 334-361 (1998).

65 Broach, J. R. Nutritional control of growth and development in yeast. Genetics 192, 73-105 (2012).

66 Veyel, D. et al. System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. Scientific Reports 7, doi:10.1038/srep42387 (2017).

67 Giavalisco, P. et al. Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. The Plant Journal 68, 364-376 (2011).

68 Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry 72, 248-254 (1976).

69 Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Analytical chemistry 75, 663-670 (2003).

70 Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. Nature biotechnology 26, 1367-1372 (2008).

71 Cox, J. et al. Andromeda: a peptide search engine integrated into the MaxQuant environment. Journal of proteome research 10, 1794-1805 (2011).

72 Kanshin, E., Kubiniok, P., Thattikota, Y., D'Amours, D. & Thibault, P. Phosphoproteome dynamics of Saccharomyces cerevisiae under heat shock and cold stress. Molecular systems biology 11 (2015).

73 R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/ (2018).

74 Perez-Riverol, Y. et al. The PRIDE database and related tools and resources in 2019: improving support for quantification data. 47, D442-D450 (2019).

75 Luzarowski, M. Marcin-Luzarowski/PROMIS: First release of code allowing analysis of PMI using PROMIS. Zenodo, doi:10.5281/zenodo.4146638 (2020).

4. Publication #2 – PROMISed: a novel web-tool for CF-MS data analysis

PROMISed: A novel web-based tool to facilitate analysis and visualization of the molecular interaction networks from co-fractionation mass spectrometry (CF-MS) experiments

Authors:	**Dennis Schlossarek**, Marcin Luzarowski, Ewelina Sokołowska, Michał Górka, Lothar Willmitzer, Aleksandra Skirycz
Year of Publication:	2021
Journal:	Computational and Structural Biotechnology Journal, Volume 19, Pages 5117-5125
DOI:	https://doi.org/10.1016/j.csbj.2021.08.042
Personal Contribution:	I streamlined and expended the previously used data-analysis pipeline, and, from there, conceptualized, implemented and tested the software-tool. I benchmarked the software using an available data set. I wrote the manuscript together with my supervisor.

4.1 Abstract

Co-fractionation mass spectrometry (CF-MS)-based approaches enable cell-wide identification of complexes formed by molecules belonging to different biochemical families (proteins, nucleic acids and small molecules) present in the cellular lysate. CF-MS combines biochemical separation of molecular complexes with an untargeted mass-spectrometry-based proteomics and/or metabolomics analysis of the obtained fractions, and is used to delineate putative interactors. CF-MS data are a treasure trove for biological discovery. To facilitate analysis and visualization of original or publically available CF-MS datasets, we designed PROMISed, a user-friendly tool available online via https://myshiny.mpimp-golm.mpg.de/PDP1/ or as a repository via https://github.com/DennisSchlossarek/PROMISed. Specifically, starting with raw fractionation profiles, PROMISed (i) contains activities for data pre-processing and normalization, (ii) deconvolutes complex fractionation profiles into single, distinct peaks, (iii) identifies co-eluting protein–protein or protein–metabolite pairs using user-defined correlation methods, and (iv) performs co-fractionation network analysis. Provided with multiple CF-MS datasets generated for species grown to distinct developmental stages or under various environmental conditions, PROMISed allows to select for proteins and metabolites that are components of dynamic interactions and whose fractionation profiles are significantly shifted between the datasets. In this way, it also enables the identification of protein–protein and protein–metabolite pairs that always co-elute together as they form stable complexes. PROMISed enables users to (i) easily adjust parameters at each step of the analysis, (ii) download partial and final results, and (iii) select among different data-visualization options. PROMISed renders CF-MS data accessible to a broad scientific audience, allowing users with no computational or statistical background to look for novel protein–protein and protein–metabolite complexes for orthogonal experimental validation.

4.2 Background and Summary

Comprehensive identification of protein–protein interactions (PPIs) is crucial for understanding the intricate mechanics behind all biological processes. In the past, a plethora of PPIs have been unraveled by large-scale studies using affinity purification coupled with mass spectrometry (AP-MS) and yeast-two-hybrid (Y2H) screens. Recent progress in mass-spectrometry proteomics has seen an advent of a complementary high-throughput method named co-fractionation mass spectrometry (CF-MS). In contrast to AP-MS and Y2H, CF-MS obviates the need for a protein bait and enables proteome-wide characterization of PPIs in a single experiment. CF-MS is based on separation of protein complexes, utilizing different biochemical techniques such as size-exclusion chromatography (SEC) [1,2], ion exchange (IEX) chromatography [3], blue native gels [4] or density-gradient centrifugation [5], followed by mass-spectrometry-based proteomics analysis of the obtained fractions. Similarity between separation profiles—referred to as co-fractionation—is used to delineate putative interactors. We have recently extended the CF-MS workflow to the analysis of protein–metabolite interactions (PMIs), dubbing our approach PROMIS (PROtein–Metabolite Interactions using Size separation). PROMIS combines SEC-based separation of protein–protein and protein–metabolite complexes followed by both proteomics and metabolomics analysis of the collected fractions. PROMIS builds upon the observation that metabolites remain in protein complexes during mild cell lysis and biochemical fractionation. Analogously to PPIs [2,3], PMIs are delineated by correlating the fractionation profiles of a protein–metabolite pair [6–9]. Table 1 gives an overview of the research performed in the past decade using co-fractionation mass spectrometry to investigate either protein-protein or protein-metabolite interactions.

While a CF-MS protocol is not experimentally challenging, data analysis requires computational expertise that is not always present in an experimental group. Analysis of a CF-MS dataset entails normalization, the selection of single peaks from complex fractionation profiles, identification of co-fractionating molecule pairs, and finally charting of the interaction network. Here, we describe PROMISed (**PROMIS E**asy **D**ata analysis), a novel web tool designed to analyze, integrate, visualize and mine data obtained from CF-MS-based studies. Starting with datasets containing absolute or relative quantification of fractionated molecules (proteins, metabolites, nucleic acids), PROMISed allows users to i) perform pre-processing steps, including normalization, smoothing and replicate pooling, ii) split complex fractionation profiles into single peaks, iii) integrate data to identify co-fractionating molecules, e.g. protein–metabolite or protein–protein pairs, and iv) create and describe co-fractionation networks. Additionally, we implemented a statistical workflow which, given multiple datasets, looks for differential fractionation

profiles indicative of a novel interaction, e.g. associated with a particular developmental stage or environmental condition. It is based on determining significant differences in the Manhattan distances calculated between fractionation profiles within the replicates and between experimental conditions. The main challenge with CF-MS is to differentiate true complexes from those merely having similar fractionation properties. For instance, in a single PROMIS experiment, every metabolite co-fractionates with several hundred proteins, of which possibly only one is a true binder. PROMISed enables users to identify pairs of molecules that co-migrate across multiple datasets; and in that way narrows down the list of putative interactors. In summary, PROMISed allows users with no computational experience to mine their own or available CF-MS datasets for novel protein–protein and protein–metabolite interactions.

Table 1: Published datasets derived from co-fractionation-based methods. CN-PAGE: clear native PAGE, IEF: isoelectric focusing, IEX: ion-exchange chromatography, SEC: size-exclusion chromatography, SDG: sucrose density gradient. *1) the nine model species include: *Caenorhabditis elegans, Drosophila melanogaster, Mus musculus, Strongylocentrotus purpuratus, Homo sapiens, Xenopus laevis, Nematostella vectensis, Dictyostelium discoideum* and *Saccharomyces cerevisiae*; *2) plant species studied: *Arabidopsis thaliana, Brassica oleracea, Glycine max, Cannabis sativa, Solanum lycopersicum, Chenopodium quinoa, Zea mays, Oryza sativa ssp. japonica, Triticum aestivum, Cocos nucifera, Ceratopteris richardii, Selaginella moellendorf* and *Chlamydomonas reinhardtii.*

Method	Biological material	Identifications	Data available	Main focus	Reference
IEX, IEF, SDG	Human cell culture	3006 proteins, 622 putative complexes	Yes	A census of human soluble protein complexes	[3]
IEX	*E. coli*, *S. cerevisiae*		No	Target identification by chromatographic co-fractionation: monitoring of drug-protein interactions without immobilization or chemical derivatization	[21]
SEC	*A. thaliana*	713 cytosolic proteins	Yes	A proteomic strategy for global analysis of plant protein complexes	[22]
SEC	9 model species (*1)	13386 protein orthologues	No	Panorama of ancient metazoan macromolecular complexes	[2]
SEC	*A. thaliana* cell suspension culture	5000 proteins, 140 metabolites	Yes	PROMIS, global analysis of protein–metabolite interactions using size separation in *Arabidopsis thaliana*	[9]
SEC	HEK293	2127 proteins, 462 complexes	Yes	Complex-centric proteome profiling by SEC-SWATH-MS	[18]
SEC	*A. thaliana*	3889 and 5563 proteins in two	Yes	Global identification of protein complexes within the	[23]

		replicates, respectively		membrane proteome of Arabidopsis roots using a SEC-MS approach	
SEC, IEX, IEF	13 plant species (*2)	141910 unique proteins corresponding to 23896 orthogroups	Yes	A pan-plant protein complex map reveals deep conservation and novel assemblies	[19]
SEC	Human cell culture, mouse embryonic stem cells	1012 (human) and 1165 (mESCs) RNA-associated proteins	No	Systematic discovery of endogenous human ribonucleoprotein complexes	[17]
CN-PAGE	*A. thaliana*	2338 (end of day) and 2469 (end of night) proteins	Yes	Protein complex identification and quantitative complexome by CN-PAGE	[4]
SDG	*A. thaliana*	216 ribosomal proteins	Yes	Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes	[5]
IEX, SEC, SDG	*Synecho-cystis sp*. PCC 6803	2062 proteins, 291 multiprotein complexes	Yes	Global landscape of native protein complexes in *Synechocystis sp*. Pcc 6803	[24]
SEC	*S. cerevisiae* (log-phase)	3982 proteins, 74 metabolites	Yes	Global mapping of protein–metabolite interactions in *Saccharomyces cerevisiae* reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity	[6]
SEC	*Chaetomium thermos-philum*	3286 proteins, 257 metabolites	No	Coupling proteomics and metabolomics for the unsupervised identification of protein-metabolite interactions in *Chaetomium thermophilum*	[25]

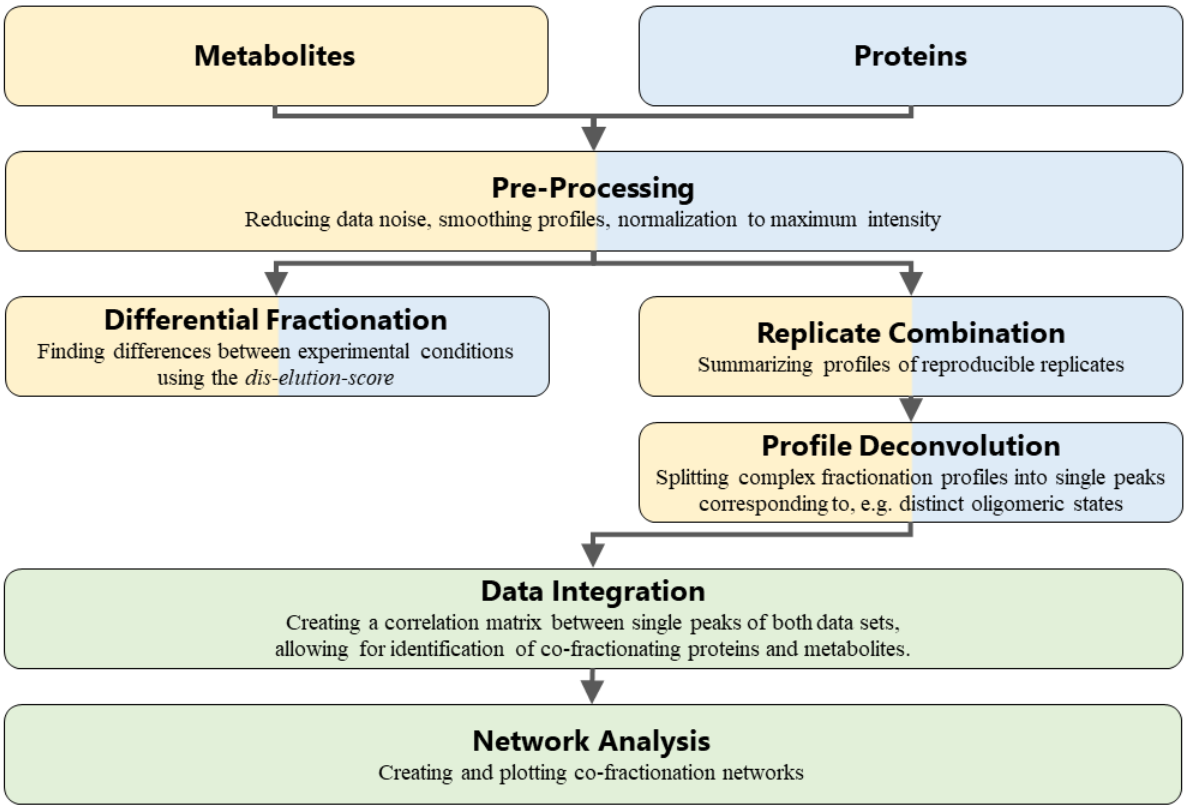


Figure 9: Schematic workflow for the PROMISed web tool. PROMISed is designed to process and integrate two different CF-MS datasets, e.g. protein and metabolite profiles obtained by SEC. Separate metabolite and protein datasets are uploaded, then processed in parallel using the same methods and settings chosen by the user. After deconvolution, the datasets are integrated into a correlation matrix, which can be mined for co-fractionating pairs, and from which co-fractionation networks can be drawn.

4.3 Software Descriptions and Methods

General design and implementation

PROMISed is accessible via a frontend web interface built using the shiny R package [10], as the backend of PROMISed is written in the R environment. The web tool aims to generate co-fractionation networks from raw fractionation profiles. The PROMISed user interface provides individual tabs to guide the user through the different steps of data analysis. Each step provides options to customize data analysis parameters and a plotting area displaying (intermediate) results, such as normalized profiles or protein–metabolite co-fractionation networks. For convenience, two example data files containing metabolite and protein fractionation profiles obtained upon cell lysate fractionation using SEC are available in order to demonstrate the functionality and requirements of PROMISed.

As shown in **Figure 1**, data analysis using PROMISed comprises i) a pre-processing step, reducing data noise and applying normalization and smoothing to ensure comparability between fractionation profiles obtained from independent CF-MS experiments, ii) statistical analysis based on the dis-elution score, identifying molecules whose fractionation profiles differ significantly between experimental conditions, iii) replicate pooling, creating a single fractionation profile for proteins or metabolites which replicates pass a chosen similarity threshold, iv) peak deconvolution, splitting complex fractionation profiles into single and distinct peaks for the downstream identification of co-fractionating molecules, v) data integration, creating a correlation matrix between the deconvoluted profiles which can be used to mine for co-fractionating partners or to vi) create co-fractionation networks.

PROMISed is available online via https://myshiny.mpimp-golm.mpg.de/PDP1/, located at the Max Planck Institute of Molecular Plant Physiology. Additionally, the source code as well as a docker image is available at https://github.com/DennisSchlossarek/PROMISed.

In the following we explain the main steps of the PROMISed workflow in detail, emphasizing the arguments adjustable by the user. A list of default settings is given in **Table 2**.

Input data requirements

The data files accepted by PROMISed are simple tables in a tab-delimited text format containing absolute or relative quantification of proteins and/or metabolites upon fractionation across collected fractions. Inexperienced users can easily generate the tables: the rows correspond to identified proteins or metabolites, each labeled with a unique name, and columns correspond to fractions obtained from different replicates and/or conditions. Column names should therefore include information about the name of the condition, the name and number of a replicate, and the number of a fraction, and should be constructed as:

NameOfCondition_NameOfReplicate_#Replicate_ #Fraction

Subsequent fractions of the same separation should be sorted in ascending order of fraction numbers. The tables may also contain columns with additional information, such as retention time and m/z for annotated metabolites or number of unique peptides used for protein identification. In such cases PROMISed asks the user to select a window of columns containing measured intensities in chromatographic fractions. Currently, additional information is ignored by PROMISed and will not be appended to the result tables. An example of CF-MS input data tables can be found in **Supplementary Tables 1 and 2** or in the demo data available in PROMISed.

Pre-processing

Prior to statistical analysis and peak deconvolution, different data pre-processing steps can be applied, if necessary. These include data-noise reduction, normalization, and smoothing of fractionation profiles. In case of complex separation using size-exclusion chromatography, fractionation profiles of proteins and metabolites span several consecutive fractions, with the peak width depending on experimental setup such as SEC column pore size. Therefore, measured metabolite or protein abundance in one fraction, but not directly neighboring fractions, can be considered as data noise, which might interfere with downstream processes. *Remove Single Peaks* provides the option to remove this data noise by replacing single-fraction peaks with zeroes. An additional way of reducing data noise is *Profile Smoothing*, which is achieved here by applying a local 2nd-degree polynomial regression fitting controlled by the *Span Value* parameter. To reduce data noise even further, fractions with low relative intensity can be set to 0. This might be necessary if measured intensities are close to the detection limit of the instruments used, e.g. sensitivity of mass spectrometer.

For data obtained from multiple, independently performed separations, normalization might be required to ensure comparability between fractionation profiles of different conditions or replicates. Normalization in PROMISed is performed separately for each fractionation profile, by normalizing every fraction to the fraction with maximum intensity within the fractionation profile. Normalization is required prior to statistical analysis of differentially eluting profiles, which uses Manhattan distances to compare peak intensities. All parameters and options included in the pre-processing step are listed below, and a schematic overview of their effects on fractionation profiles is shown in **Figure 2a**.

Remove Single Peaks: replaces single values surrounded by zeroes with zero, eliminating data-noise "peaks" that only span one fraction.

Normalize: fractionation profiles are normalized to their maximum intensity.

Minimum Relative Intensity: sets a threshold for the minimum relative intensity (after normalization). Values below this threshold are replaced by 0.

Profile Smoothing: reduces data noise by using a local 2nd-degree polynomial regression fitting, controlled by *Span Value* as an additional parameter. Default *Span Value* is 0.15 for profiles of length 40. **Sup. Figure 1** gives an overview of minimal *Span Values* for profiles of different length.

Span Value: controls the degree of smoothing by defining the number of neighboring fractions involved. Corresponds to proportion of total fractions and is given in a range of 0 to 1. Default is 0.15 for profiles

spanning 40 fractions. In such examples, six neighboring fractions will be used for smoothing. Profiles differing in the number of fractions may require optimization of Span Values.

Table 2: Default settings of the adjustable parameters in the PROMISed web tool.

Processing step	Parameter	Default
Pre-Processing	Remove Single Peaks	TRUE
	Normalize	TRUE
	Minimum Relative Intensity	0.10
	Profile Smoothing	TRUE
	Span Value	0.15
Replicate Combination	Correlation Method	Pearson
	Reproducibility Threshold	0.70
	Keep Single Replicates	FALSE
Deconvolution	Minimum Relative Intensity	0.20
	Minimum Intensity of Local Maxima	0.20
	Minimum Incline	0.80
Data Integration	Correlation Method	Pearson
	Correlation Threshold	0.70
Network Analysis	Filter Network	No Filter
	Node Colour	Cluster
	Layout	Force-directed
Differential Fractionation	p-Value Threshold	0.05

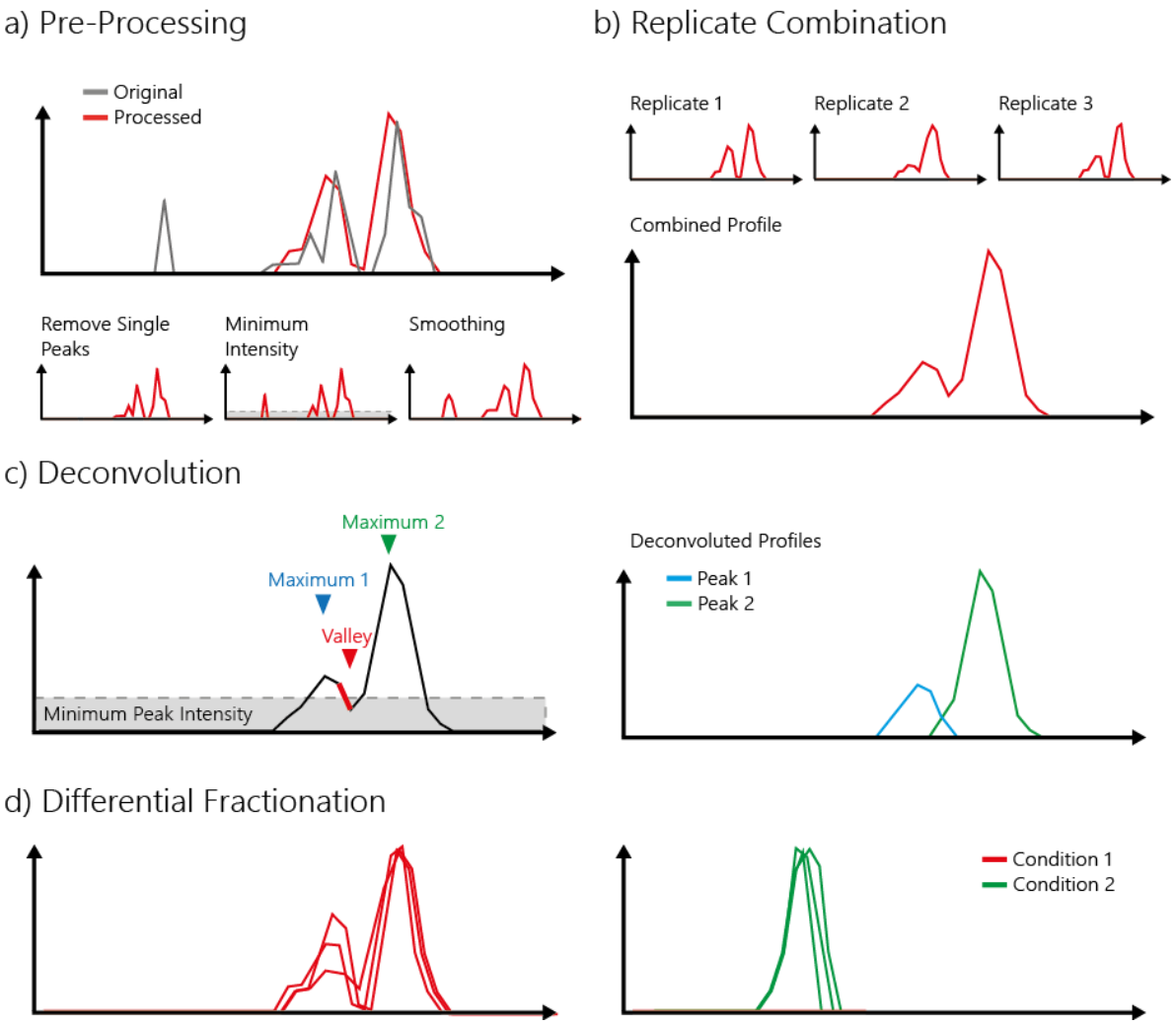


Figure 10: Effect of processing options in PROMISed on fractionation profiles. a) Top: Resulting profile (red) after applying all pre-processing steps compared to the original profile (gray). Bottom: Effect of individual pre-processing options on the shape of the fractionation profile. b) Fractionation profiles of three replicates and the combined profile. c) Left: Identification of local maxima and valleys in the combined profile. Both local maxima are greater than the *Minimum Peak Intensity*. The valley is considered a cutting point, since the incline to the previous fraction (marked in red) exceeds the *Minimum Incline* threshold. The resulting peaks are shown on the right side. d) Example of a protein or metabolite profile fractionating differentially between two conditions (left and right).

Replicate pooling

The pooling of replicates is achieved by summing up reproducible replicates of fractionation profiles (**Figure 2b**). Reproducibility is calculated using either one of three correlation methods. An adjustable threshold gives the user control over the strictness of replicate combination. In addition, for datasets comprising two replicates, a single profile is treated as reproducible when the other replicate comprises an empty profile. Two of our previous studies exploiting PROMIS for system-wide detection of

protein–metabolite complexes in *Arabidopsis thaliana* and *Saccharomyces cerevisiae* showed high reproducibility of fractionation (PCC > 0.9) between biological replicates [6,9].

Correlation Method: the user can choose between three methods of calculating correlation: *Pearson correlation, Kendall's tau* or *Spearman's rank correlation.*

Reproducibility Threshold: sets a threshold for the minimum correlation coefficient between replicates to sum up fractionation profiles.

Keep Single Replicates: allows to keep fractionation profiles of one replicate when the other replicates are uniformly 0. Used as default when uploaded data contains only one or two replicates.

Deconvolution

Since native proteins and metabolites can be involved in multiple protein–protein–metabolite complexes of different molecular weights, obtained CF-MS fractionation profiles are often complex and contain several local maxima. Profile deconvolution aims at splitting a complex fractionation profile into several profiles containing one peak each, corresponding to independent homomeric or heteromeric states of a protein or a protein partner of a metabolite. Profile deconvolution based on identifying local maxima was first used in studies aiming to identify protein–protein complexes using CN-PAGE [4] and later adapted for identification of protein–metabolite complexes using PROMIS [6]. In a nutshell, deconvolution is achieved by first identifying local maxima and subsequently finding the shape of the underlying fractionation peak (**Figure 2c**).

Local maxima are found using the turnpoint function (pastecs R package [11]) and filtered against the *Minimum Intensity of Local Maxima* parameter. Next, the shape of the fractionation peak is identified by "scanning" over the profile and making a decision at each fraction as follows: starting from a local maximum, each subsequent fraction is considered part of the underlying peak until its maximum normalized intensity is below *Minimum Relative Intensity* or until the fraction constitutes a major local valley, with an incline greater than *Minimum Incline* to one of the surrounding fractions. The fractions prior to the first local maximum, and fractions in-between the last fraction of an identified peak and the next maximum are assigned to the following true local maximum, as long as their intensity passes the *Minimum Relative Intensity* criteria. Profile deconvolution can be bypassed by checking the *No Deconvolution* box.

Minimum Relative Intensity: sets a threshold for the minimum relative. Values below this threshold are replaced by 0.

Minimum Intensity of Local Maxima: sets a threshold for the minimum relative intensity with which a local maximum is considered an independent peak.

Minimum Incline: minimum incline around a local minimum for cutting the peak into two.

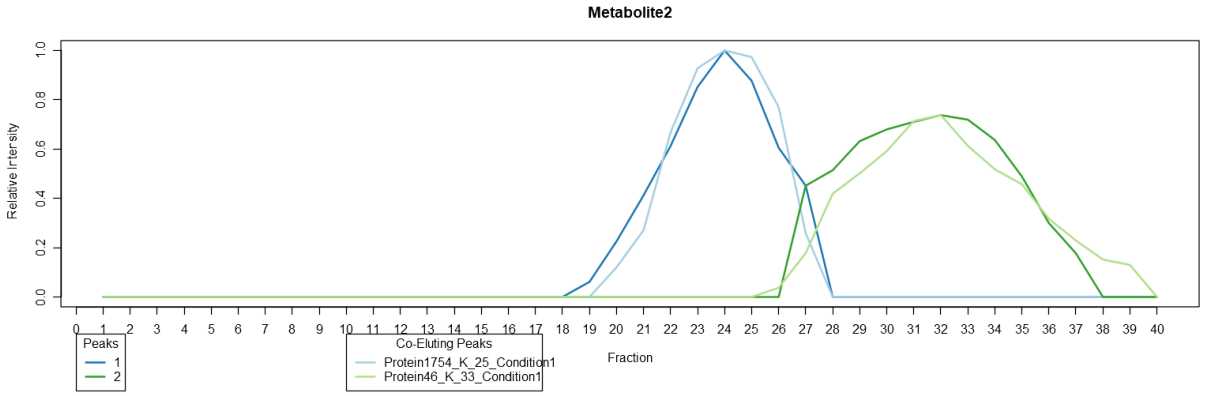
Data integration

The integration of the two datasets is accomplished by calculating a correlation matrix between the deconvoluted profiles of both datasets. The method of correlation can be either *Pearson correlation, Kendall's tau* or *Spearman's rank correlation*. In our hands and for the purpose of screening for protein–metabolite complexes in *S. cerevisiae* using SEC, Pearson correlation and Kendall's tau served equally well for retrieving known protein–metabolite assemblies [6]. The resulting correlation matrix is filtered using the *Correlation Threshold* and can be downloaded as a text file. Proteins and metabolites passing the threshold are further considered as co-fractionating. Previous studies showed that a threshold of 0.7 (default) gives a good agreement between sensitivity and specificity for detection of known molecular complexes [6,9].

PROMISed offers multiple ways of mining the obtained results in a targeted and untargeted manner (see "Network analysis"). Focusing on a pre-selected molecule, in the "Data Integration" tab users have the possibility to identify co-fractionating partners of a chosen entry across all provided datasets ("Intersections of Conditions") or under a defined condition (e.g. for single species, developmental stage or growth conditions). Additionally, the deconvoluted fractionation profiles of a target from each dataset can be directly compared in one plot. Lastly, the protein or metabolite peaks with the highest correlation to the target can be plotted together with the target, and a subset of the correlation matrix is displayed in the user interface (**Figure 3**).

Correlation Method: gives the user the option to choose between three methods of calculating correlation: *Pearson correlation, Kendall's tau* or *Spearman's rank correlation*.

Correlation Threshold: sets a threshold for the minimum correlation coefficient to consider metabolite and/or protein peaks as co-eluting.



Show 10 entries Search:

	Metabolite2_K_25_Condition1	Metabolite2_K_33_Condition1
Protein1754_K_25_Condition1	0.983	-0.209
Protein1238_K_24_Condition1	0.976	-0.211
Protein1825_K_26_Condition1	0.970	-0.218
Protein251_K_26_Condition1	0.968	-0.226
Protein694_K_25_Condition1	0.967	-0.224

Figure 11: Top scored co-fractionating protein peaks of a selected metabolite. Upper panel: Deconvoluted fractionation profile of a selected metabolite (Metabolite2) in dark, and the deconvoluted protein peaks with the best scores (here: Pearson correlation) in light colors. The first Metabolite2 peak (blue) co-elutes with Protein1754, the second Metabolite2 peak (green) with Protein33. Lower panel: Table depicting correlation scores, here Pearson correlation, between the two peaks originating from the Metabolite2 fractionation profile and all protein peaks. The table is sorted to show the highest 5 scores for the first Metabolite2 peak, revealing Protein1754 as the best co-eluting protein peak with a PCC of 0.98.

Network analysis

PROMISed constructs co-fractionation networks from the filtered correlation matrix using the igraph R package [12]. Networks are created as weighted correlation networks, depicting single protein or metabolite peaks as nodes, and the correlation coefficient between those peaks as edges. Networks can be restricted to selected entries instead of using the whole correlation matrix. To allow the user to recreate and analyze the obtained networks using different tools such as Cytoscape [13], the edgelists and nodelists can be readily downloaded.

The networks are displayed and made interactive using the visNetwork R package [14]. Nodes can be color-highlighted to depict either information about connectivity, calculated as k-coreness, or communities, identified using the Louvain method for community detection [15]. Additionally, the user can choose between three different options for the network's layout: network *components*, a *force-directed* layout, or *automatic* detection of the most appropriate layout. An example of network visualization is given in **Figure 4**.

Filter Network: determines whether a network should be created using all available data (*No Filter*) or only around a selected protein or metabolite (*Selection 1* and *2*).

Node Color: changes node colors to highlight either communities (*Cluster*) (**Figure 4a**) or node connectivity measured as k-Coreness (**Figure 4b**). Default is a *Uniform* light blue.

Layout: controls the layout of the depicted network with two options: *Force-directed* uses the network layout algorithm by Fruchterman and Reingold (1991), and *Circles* creates a circle of nodes for each network component.

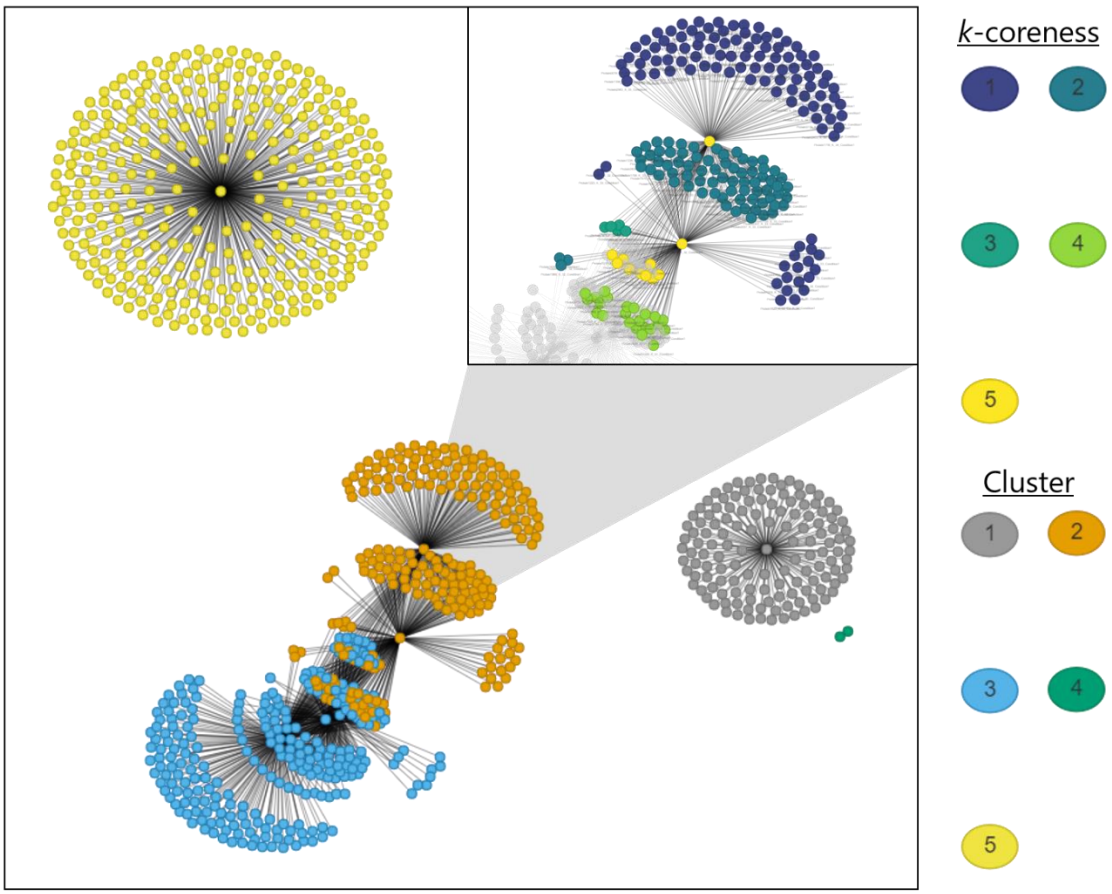


Figure 12: Visualization of co-fractionation networks. The network is depicted using the Force-directed layout and Cluster color option. The inset shows Cluster 2 colored according to its k-coreness.

Differential fractionation

Describing differences in fractionation profiles (as shown in **Figure 2d**) of the same protein or metabolite is a crucial, yet challenging step in identifying interactions responsive to experimental conditions. One attempt to meet this challenge was made by Mallam and colleagues, who developed a computational framework named DIF-FRAC score. The DIF-FRAC score compares a normalized Manhattan distance between fractionation profiles of treated and untreated proteins with those of known non-responsive proteins. This method proved to be powerful for the search of novel RNA-associated proteins in the presence and absence of RNA, where a true-negative list could be generated [17]. However, for most untargeted research questions this is not the case.

To circumvent this requirement, we developed the ***dis-elution score***, which relies on the statistical comparison of Manhattan distances between conditions with distances within conditions. In detail, we calculate the Manhattan distances of all combinations of replicates between conditions X and Y ($\|\vec{x} - \vec{y}\|_1$) and within the conditions ($\|\vec{x} - \vec{x}\|_1$ and $\|\vec{y} - \vec{y}\|_1$). The resulting vectors of Manhattan distances are then statistically compared using a one-way analysis of variance (ANOVA). A post-hoc Tukey test is then used to check whether $\|\vec{x} - \vec{y}\|_1$ is significantly larger than $\|\vec{x} - \vec{x}\|_1$ and $\|\vec{y} - \vec{y}\|_1$).

The "Differential Fractionation" tab allows the user to calculate the ***dis-elution score*** as a pairwise comparison between experimental conditions. The results can be downloaded in a table format, where FDR-corrected p-values of profiles passing the Tukey test are reported or viewed for each protein or metabolite in each pair of conditions depicted as a boxplot (see Metabolite 1–3 from example data). Please note that p-values shown on boxplots in the user interface are not FDR corrected and may differ from values in the results table.

p-Value Threshold: controls p-value threshold to filter dis-elution scores.

4.4 PROMISed captures predicted AHA2 interactors

To test the applicability of PROMISed to analyze previously published CF-MS experiments, we used the publicly available dataset from Gilbert&Schulze, 2018 **(Supplementary Tables 3 and 4)**. The authors first investigated membrane protein-protein complexes isolated from the *Arabidopsis thaliana* roots and subsequently focused on one selected protein: H+-ATPase AHA2. Comparison of the 174 proteins co-eluting with AHA2 with the list of 32 proteins identified in the AHA2 pulldown experiment revealed an overlap of 13 proteins. We downloaded original protein elution profiles available as a supplement dataset

to compare published analysis with the data analysis pipeline embedded in PROMIsed. The experiment comprises two biological replicas, which we analyzed separately, using the default settings, except for the Data Integration Correlation Threshold, which was set to 0.89 to replicate the author's high-confidence threshold. We then filtered the generated networks for AHA2 and obtained the network's edge-list. In other words, a list of co-fractionating proteins. In total, we identified 489 and 590 co-fractionating proteins in replicate one and replicate two, respectively, of which 161 were common. We then compared 161 identified interactors with the list of AHA2 protein partners reported in the original work (**Figure 5**). The overlap contained 64 of the previously predicted AHA2 protein interaction partners, of which 19 were also retrieved in the pulldown experiment, which is six more than in the original analysis. The discrepancy originates from an additional AHA2 elution peak resulting from the deconvolution step.

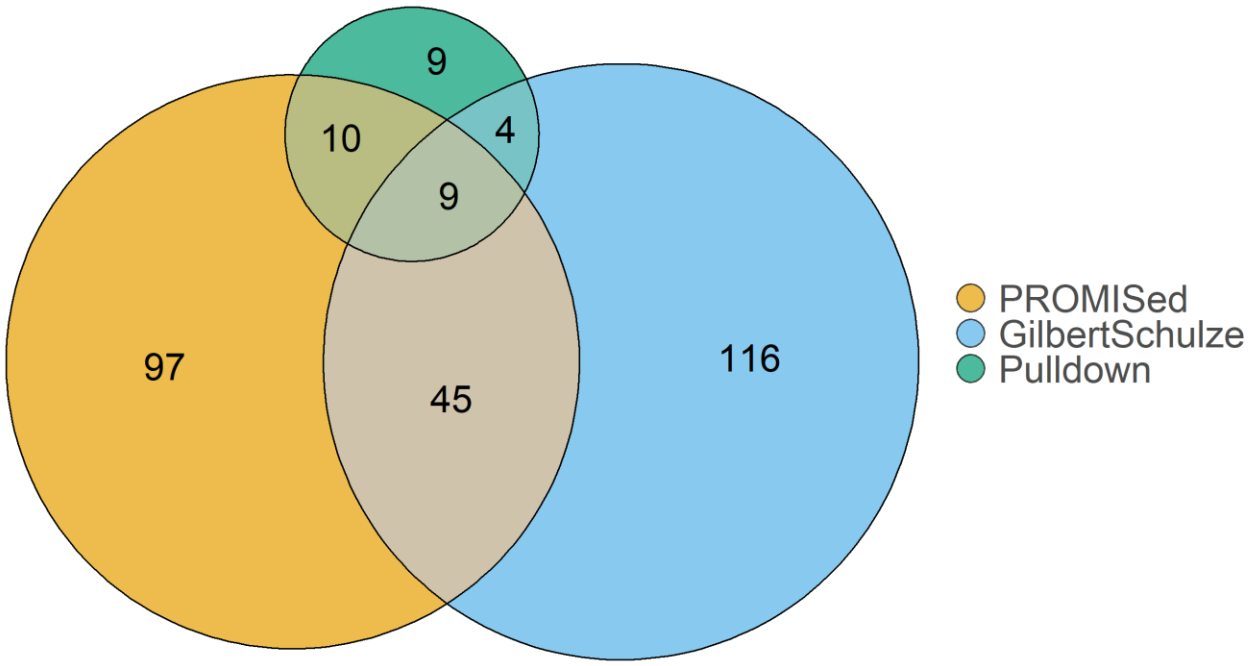


Figure 13: PROMISed can capture previously predicted protein interactors of AHA2. Comparison of the number of proteins co-fractionating with AHA2 in the data from Gilbert&Schulze, 2018. The data was analysed using PROMISed default settings with the Data Integration Correlation Threshold of 0.89, to reproduce the original work. In total, PROMISed captured 161 interaction partners shared in both replicates, compared to the 174 in the original work (Gilbert&Schulze). Additionally, PROMISed captured 19 proteins that were also identified in a pull-down experiment performed in the original work.

4.5 Novelty and Applications

Recent years have seen an advent of CF-MS-based approaches for the characterization of molecular complexes. Multiple datasets for both model and non-model organisms, spanning different developmental stages and environmental conditions have been published and are available to the scientific community. Moreover, interactive tools such as SECexplorer-cc (https://sec-explorer.shinyapps.io/hela_cellcycle/) [18], plant.Map (http://plants.proteincomplexes.org) [19] and YeastPMI (http://promis.mpimp-golm.mpg.de/yeastpmi) [6] provide an access to PPIs and PMIs for a handful of those published datasets. Here, we introduce a freely accessible and intuitive web tool designed for analysis, integration, visualization and mining of co-fractionation data, which does not require a computational or statistical background. PROMISed is not restricted by the identity of interactors (protein, metabolite etc.) or type of separation (SEC, IEX, native gel), and the only input it requires is raw fractionation profiles. PROMISed enables adjustment of multiple parameters at each analysis step, allowing tailoring of the process to a particular dataset or a biological question and making it more interactive and consequently more attractive to potential users. Moreover, at each data-processing step users can download and inspect the results file. When it comes to data mining, PROMISed can be used in several ways. In the simplest scenario, users can mine PROMISed-generated interaction networks (matrix) for the putative interactors of their protein or metabolite of interest. As already mentioned in the Introduction, a single CF-MS dataset may not be sufficient to differentiate true complexes from those merely having similar fractionation properties. A proven strategy to address this problem [19], implemented in PROMISed, is to look for repeating co-fractionation across multiple datasets, differing, for instance, in the separation method or source of the starting material. In addition to searching for repeating co-fractionation, another unique feature implemented in PROMISed allows users to look for proteins and metabolites that change their fractionation profile in multi-condition experiments. A difference in the fractionation profile can have different biological explanations, but it unequivocally attests to rewiring of the interactome. Overall, PROMISed is a unique tool designed for analysis, integration, visualization and mining of co-fractionation data, accessible to users without computational or statistical background.

4.6 Discussion and Conclusion

The use of CF-MS-based approaches has led to numerous and significant biological insights. For example, by separating the soluble protein fraction of human cell culture lines using ion-exchange chromatography, authors found as many as 13,993 interactions, corresponding to 622 putative protein complexes; more than half of them (364) were previously unannotated [3]. Using a workflow dubbed DIF-

FRAC, Mallam and colleagues identified 115 ribonucleoprotein complexes in human cell culture lines by comparing separation profiles of proteins obtained from lysates treated with RNAse to a non-treated control, inferring that 20% of human protein complexes contain an RNA component [17]. By expanding co-fractionation-based methods to protein–metabolite interactions, PROMIS has delineated putative protein interactors for over 140 metabolites in *A. thaliana* [9], and validated 225 previously predicted PMIs in yeast[6]. Following up on PROMIS results led to the identification of proteogenic dipeptides as novel regulators of metabolism [6,9].

While the number of methods using co-fractionation to study protein–protein and protein–metabolite interactions is growing, data-analysis pipelines and strategies greatly vary between research groups and are often highly specific for one experiment. For example, machine-learning approaches have been successfully used to create protein–protein interaction networks [2,3,19]. Additionally, a complex-centric experimental workflow using SEC-SWATH-MS has been developed [18], complemented by a machine-learning-based computational framework dubbed PCprophet [20]. However, machine-learning-based approaches require preselected data to train on, limiting transferability to other approaches and datasets.

In addition, neither approach has so far been applied to integrate the wide range of protein-binding metabolites. And finally, the complex-centric workflow, and especially any machine-learning approach, requires a certain level of bioinformatic skills, creating an entry barrier for many experimental researchers. We created a user-friendly application that does not require prior computational insight and that guides users through the processing steps, allows to adjust parameters of each activity to tailor various experimental setups, and helps to visualize co-fractionating pairs and interactome networks.

Code availability

PROMISed is available online via https://myshiny.mpimp-golm. mpg.de/PDP1/, located at the Max Planck Institute of Molecular Plant Physiology. Additionally, the source code as well as a docker image is available at https://github.com/DennisSchlossarek/PROMISed.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

Acknowledgments

We thank Dariusz Bieńkowski and Andreas Donath for help with hosting the software tool.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.csbj.2021.08.042.

4.7 References – Publication #2

1. Liu, X., Yang, W. Chu, Gao, Q. & Regnier, F. Toward chromatographic analysis of interacting protein networks. J. Chromatogr. A 1178, 24–32 (2008).
2. Wan, C. et al. Panorama of ancient metazoan macromolecular complexes. Nature 525, 339–344 (2015).
3. Havugimana, P. C. et al. A census of human soluble protein complexes. Cell 150, 1068–1081 (2012).
4. Gorka, M. et al. Protein Complex Identification and quantitative complexome by CN-PAGE. Sci. Rep. 9, 1–14 (2019).
5. Firmino, A. A. P. et al. Separation and paired proteome profiling of plant chloroplast and cytoplasmic ribosomes. Plants 9, 1–29 (2020).
6. Luzarowski, M. et al. Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. Commun. Biol. 4, 1–15 (2021).
7. Sokolowska, E. M., Schlossarek, D., Luzarowski, M. & Skirycz, A. PROMIS: Global Analysis of PROtein-Metabolite Interactions. Curr. Protoc. plant Biol. 4, e20101 (2019).
8. Veyel, D. et al. System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. Sci. Rep. 7, 1–8 (2017).
9. Veyel, D. et al. PROMIS, global analysis of PROtein-metabolite interactions using size separation in Arabidopsis thaliana. J. Biol. Chem. 293, 12440–12453 (2018).
10. Chang, W., Cheng, J., Allaire, J., Xie, Y. & McPherson, J. shiny: Web Application Framework for R. R package version 1.5.0. 21, 1–9 (2020).
11. Grosjean, P. & Ibanez, F. pastecs: Package for Analysis of Space-Time Ecological Series. R package version 1.3.21. (2018).
12. Csárdi, G. & Nepusz, T. The igraph software package for complex network research. InterJournal Complex Sy, 1695 (2006).
13. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research 13, 2498–504 (2003).
14. Almende, B. V., Thieurmel, B. & Robert, T. visNetwork: Network Visualization using 'vis.js' Library. R package version 2.0.9. (2019).
15. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. 2008, 1–12 (2008).
16. Fruchterman, T. M. J. & Reingold, E. M. Graph drawing by force-directed placement. Softw. Pract. Exp. 21, 1129–1164 (1991).
17. Mallam, A. L. et al. Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. Cell Rep. 29, 1351-1368.e5 (2019).
18. Heusel, M. et al. Complex-centric proteome profiling by SEC - SWATH - MS . Mol. Syst. Biol. 15, 1–22 (2019).

19. McWhite, C. D. et al. A pan-plant protein complex map reveals deep conservation and novel assemblies. Cell 181, P460-474.E14 (2020).

20. Fossati, A. et al. PCprophet: a framework for protein complex prediction and differential analysis using proteomic data. Nat. Methods (2021). doi:10.1038/s41592-021-01107-5

21. Chan, J. N. Y. et al. Target identification by chromatographic co-elution: Monitoring of drug-protein interactions without immobilization or chemical derivatization. Mol. Cell. Proteomics 11, (2012).

22. Aryal, U. K. et al. A Proteomic strategy for global analysis of plant protein complexes. Plant Cell 26, 3867–3882 (2014).

23. Gilbert, M. & Schulze, W. X. Global Identification of Protein Complexes within the Membrane Proteome of Arabidopsis Roots Using a SEC-MS Approach. J. Proteome Res. 18, 107–119 (2019).

24. Xu, C. et al. Global Landscape of Native Protein Complexes in Synechocystis sp. PCC 6803. Genomics. Proteomics Bioinformatics (2021). doi:10.1016/j.gpb.2020.06.020

25. Li, Y. et al. Coupling proteomics and metabolomics for the unsupervised identification of protein – metabolite interactions in Chaetomium thermophilum. PLoS One 1–13 (2021). doi:10.1371/journal.pone.0254429

5. Publication #3 – Rewiring of the Interactome during the Diauxic Shift

Rewiring of the protein-protein-metabolite interactome during the diauxic shift in *Saccharomyces cerevisiae*

Authors:	**Dennis Schlossarek**[1], Marcin Luzarowski[1], Ewelina M. Sokołowska, Venkatesh P. Thirumalaikumar, Lothar Willmitzer, Jennifer Ewald, Aleksandra Skirycz [1]: These authors contributed equally
Year of Publication:	In preparation
Journal:	-
DOI:	-
Personal Contribution:	I analyzed the data and provided data visualization. I, together with Marcin Luzarowski and my supervisor, wrote the manuscript. I assisted in the experiments.

5.1 Abstract

In budding yeast *Saccharomyces cerevisiae*, the switch from aerobic fermentation to respiratory growth is separated by a period of growth arrest, known as the diauxic shift, accompanied by a significant metabolic rewiring, including the derepression of gluconeogenesis and the establishment of mitochondrial respiration. Previous studies reported hundreds of proteins and tens of metabolites accumulating differentially across the diauxic shift transition. To assess the differences in the protein-protein (PPIs) and protein-metabolite interactions (PMIs) yeast samples harvested in the glucose-utilizing, fermentative phase, and early and late ethanol-utilizing, respiratory phases were analysed using isothermal shift assay (iTSA) and a co-fractionation mass-spectrometry approach, PROMIS. Whereas iTSA monitors changes in protein stability and is informative towards protein interaction status, PROMIS uses co-elution to delineate putative PPIs and PMIs. The resulting dataset comprises 1627 proteins and 247 metabolites, hundreds of proteins, and tens of metabolites characterized by differential thermal stability and/or fractionation profile, constituting a novel resource to be mined for the regulatory PPIs and PMIs. The examples discussed here include (i) dissociation of the core and regulatory particle of the proteasome in the late-ethanol phase, (ii) the differential binding of a cofactor pyridoxal phosphate to the enzymes of amino acid metabolism, (iii) and the putative, phase-specific interactions between proline containing dipeptides and enzymes of central carbon metabolism. We demonstrate that the reported changes translate into a significant rewiring of the protein-metabolite interactome

5.2 Introduction

Budding yeast S*accharomyces cerevisiae* grown on glucose undergoes two growth phases: glucose-utilizing and ethanol-utilizing, separated by a period of growth arrest, referred to as diauxic shift [1]. The depletion of both glucose and ethanol is associated with the cell cycle arrest and the onset of the stationary phase [2]. During the glucose-utilizing phase, yeast uses glucose to produce ATP and pyruvate through glycolysis. Pyruvate is further converted to ethanol, which accumulates in the medium until glucose is depleted. The presence of glucose suppresses oxidative phosphorylation (OXPHOS), gluconeogenesis, and the use of alternative carbon sources. The repression of respiration in the presence of oxygen is referred to as the Crabtree effect and is reminiscent of the Warburg effect described for cancer cells [3]. Decreasing glucose concentration triggers first gradual and, once glucose is depleted, abrupt changes to the metabolism characterized by de-repression of gluconeogenesis and establishment of mitochondrial respiration [1].

The transition from glucose to ethanol-based growth is orchestrated by the central energy-signaling pathways including the PKA (cAMP-dependent protein kinase), TORC1 (target of rapamycin), and Snf1 (sucrose non-fermenting) kinases [4,5]. Down-stream targets of PKA, TORC1, and Snf1 comprise multiple transcriptional regulators, which drive the massive transcriptional reprogramming reported for the diauxic shift [6]. Changes in gene expression are reflected at the protein and metabolite levels; hundreds of proteins and tens of metabolites displaying differential accumulation across glucose to ethanol transition [7]. For instance, abundance of glycolytic enzymes and metabolic intermediates decreases, whereas enzymes and metabolites of the glyoxylate cycle accumulate during the diauxic shift [1].

Biological entities such as proteins and metabolites rarely act independently but rather as a part of a larger complex. Resulting protein-protein (PP) and protein-metabolite (PM) interactions have diverse functional consequences from structural to regulatory with implications to all known cellular processes. The dramatic changes in the protein and metabolite abundance reported for the diauxic-shift likely translate into a significant rewiring of the PP and PM interactome. New interactions can be driven by the changes in the protein and metabolite concentrations and an alteration in the interactor status, such as, for instance, by a post-translational modification (PTM) of the protein partner. However, although the importance of protein interactions is undisputed, many interactions are still poorly characterized. This is especially true for protein-metabolite interactions.

While there are many approaches to study PPIs and PMIs; only a few allow a bird-eye view into the entirety of the interactomes across cellular transitions such as the diauxic shift. One such method is thermal proteome profiling (TPP) [8]; also referred to as cellular thermal shift assay CETSA [9]. Here, the interaction status of a protein is gauged from the difference in the temperature stability expressed as melting

temperature - the temperature at which 50 % of a protein is unfolded, compared between the different conditions. Initially developed to look for protein partners of small-molecule ligands [8], TPP/CETSA has been successfully applied to assess global changes in the protein interactomes [10,11]. The change in the temperature stability measured between two cellular states, whether it is a genetic, environmental or developmental perturbation [10–12] is indicative of a change in the protein interaction status; where the interacting partner can be e.g., a protein, metabolite or nucleic acid. Protein thermal stability can be also affected by the change in the PTMs status [13–15]. TPP/CETSA experiments are ideal to assess the global change to a protein interactome. Moreover, as proteins in a complex show coordinated changes in their melting behaviour when done across a large number of cell-states, TPP/CETSA experiments can be used to predict the composition of protein complexes [12].

A complementary approach that can be used to capture differential interactomes is co-fractionation mass spectrometry (CF-MS) (reviewed e.g. by [16]). CF-MS combines separation of complexes utilising different biochemical techniques such as size-exclusion chromatography (SEC) e.g. [17–19], ion exchange (IEX) chromatography e.g. [19,20], blue native gels e.g. [21] or density-gradient centrifugation e.g. [22] with mass-spectrometry (MS) analysis of the obtained fractions and uses co-elution to delineate interactors. First established for PPIs [18–23], CF-MS methods can also be used to resolve protein-metabolite [20,24–26] and protein-RNA interactions [27].

In the past we used a CF-MS approach, that we dubbed PROMIS to build a protein-metabolite interaction map of a budding yeast *Saccharomyces cerevisiae* during the glucose-utilising, logarithmic growth stage [25]. In doing so we reported hundreds of known and unknown small molecules separating together with proteins, attesting to the previously postulated [28,29] complexity of the protein-metabolite interactome, and the notion that many more small molecules than are known today interact with, and modulate the function of, their protein partners.

In the current study, we extended our analysis to explore the dynamics of PPIs and PPMs during the diauxic shift transition to understand how the dramatic changes in the protein and metabolite abundance reported for the diauxic-shift would translate into a rewiring of the PP and PM interactome. To this end, yeast harvested in the glucose-utilizing, fermentative phase, and early and late ethanol-utilizing, respiratory phases were subjected to isothermal shift assay (iTSA, a variation on the TPP/CETSA [30]) and PROMIS analysis. We could demonstrate that, the diauxic shift transition is associated with major changes in the PP and PM complexes. Moreover, our work attests to the suitability of PROMIS to capture changes in PMIs, as shown before for PP complexes e.g., [31].

5.3 Results and Discussion

Proteomics analysis of diauxic shift

To examine changes in the PM and PP interactome across the diauxic shift transition we used the YSBN2 strain of *Saccharomyces cerevisiae*. YSBN2 is a prototrophic strain closely related to the model strain S288c [32], the main difference being the presence of a resistance marker. New culture was started using a single colony and yeast were grown on standard YPD medium supplemented with glucose. OD600 was used to monitor growth and samples were taken after 6, 24 and 72 hours of cultivation. The 6 h time-point corresponds to the logarithmic, glucose - -utilizing phase, 24 h to post-diauxic, early ethanol-utilizing phase and 72 h to the late ethanol-utilizing phase (**Fig. 1A**).

To examine how well our study compares with other diauxic shift omics and multi-omics experiments we first measured protein abundances across the three sampled growth phases. Of the 1627 proteins present in our dataset, 697 proteins showed a significant (FDR < 0.01) fold-change (FC) greater than 2 between the glucose-utilizing phase and late ethanol-utilizing phase, 300 between glucose-utilizing and early ethanol-utilizing phase, and 288 between early and late-ethanol -utilizing phases, respectively (**Fig. 1B, Table S1-S2**). We performed a KEGG enrichment analysis on the 697 proteins differing between glucose and late ethanol phases (**Fig. 1C**). Among the 116 proteins we found accumulating in the late ethanol- in comparison to glucose-utilising phase, proteins associated with the TCA cycle and glyoxylate and dicarboxylate metabolism were significantly enriched. The list included the rate limiting enzymes of gluconeogenesis, fructose-1,6-bisphophatase (FBP1) and phosphoenolpyruvate carboxykinase (PCK1), TCA/glyoxylate cycle enzymes such as citrate synthase (CIT1 and CIT2), and trehalose phosphate synthases (TPS1 and TPS2). This is in line with previous studies showing that the major metabolic events associated with the diauxic shift include the reversal of flux from glycolysis to gluconeogenesis, accumulation of trehalose and glycogen, and activation of the TCA / glyoxylate cycle [1]. Conversely, in accordance with what was reported by [7], proteins associated with ribosome biogenesis and RNA polymerases were enriched among 581 down regulated proteins depleted in the late ethanol versus glucose-utilising phase. The list included the cytoplasmic GTPase RIA1 involved in the 60S ribosomal subunit biogenesis, ribosomal proteins (RPL29, RPS2), chaperones, and RNA polymerase subunits such as RPB3, RPB11 or TFG1. In summary, proteomics analysis of the three sampled yeast growth phases revealed known signatures of diauxic shift transition.

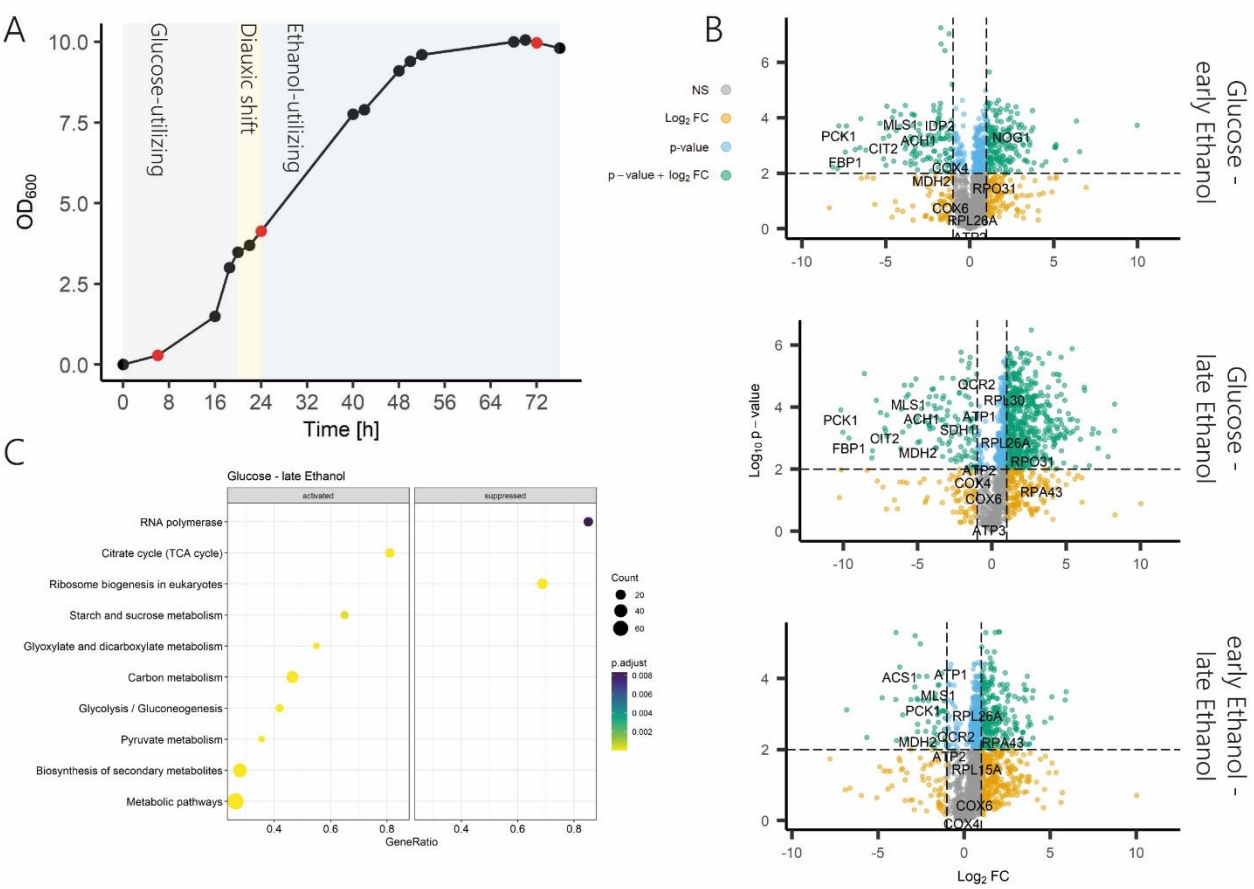


Figure 1: Changes in relative protein abundance between glucose- and ethanol-utilizing growth in *Saccharomyces cerevisiae*. A) *S. cerevisiae* strain YSBN2 growth curve in glucose containing complete medium. Samples were collected after 6, 24 and 72 hours corresponding to the glucose-utilizing phase and early and late ethanol-utilizing phases, respectively. B) Volcano plot visualization of changes in relative protein abundances between growth-phases tested using a two-sided t-test. The horizontal, dashed line indicates FDR = 0.01, vertical, dashed lines indicate a fold-change greater than 2. Proteins with significant (FDR < 0.01) fold changes of more than 2 are highlighted in green, and proteins involved in central carbon metabolism are labeled. C) KEGG enrichment analysis of proteins which significantly increase (activated) or decrease (suppressed) in abundance in the late ethanol-utilizing phase compared to the glucose-utilizing phase.

Thermal Proteome Profiling reveals major changes between fermentative and respiratory phase

We next wanted to understand whether these major changes in protein abundance across the diauxic shift are accompanied by global changes in the protein interactome. We used a simplified thermal proteome profiling protocol called isothermal shift assay (iTSA) [30]. Instead of a temperature gradient, protein stability is measured in a single temperature selected based on the average melting temperature of the proteome (**Fig. 2 A**).

To better represent the "meltome" we used three different "melting temperatures", 48, 52 and 56 °C. We additionally took samples at room temperature (RT) to normalize for the differences in the initial

protein abundance. Statistical analysis revealed hundreds of differential proteins, the differences being most pronounced at 56 °C (**Figure S1**). Interestingly, and again similarly to what was already reported [10,33], there was no correlation between protein abundance at room temperature and thermal stability at 56 °C (Pearson correlation coefficient: -0.026, $p = 0.12$) (**Fig 2 C**). Additionally, we could confirm a weak, but statistically significant negative correlation between a protein's chain length and thermal stability, as described before [34–36] varying from R = -0.064 in the early ethanol -utilizing phase to R = -0.24 in the glucose -utilizing phase (**Figure S2**).

We found that hundreds of proteins changed their stability across the three growth phases. Most differences (FDR < 0.05 and FC > 2 or < 0.5 at 48, 52 or 56 °C) were measured between the glucose and late ethanol phases (**Figure 2 B and D, Table S3**), with fewer differences measured between consecutive growth phases. 637 proteins (53.5 %), were significantly affected in their thermal stability between the glucose and late ethanol, in comparison to 131 proteins (11 % of all measured proteins) between the glucose and early ethanol, and only 17 (1.4 %) between the early and late ethanol phases. As would be expected 97 of the 131 proteins (74 %) differential between the glucose and early ethanol were also differential between the glucose and late ethanol phases. A single protein, heat shock protein 26 (HSP26), was differential in all comparisons.

A KEGG enrichment analysis on the 637 proteins changed between the glucose and late ethanol phases (**Fig2 E**) identified pathways involved in the biosynthesis of secondary metabolites (sce01110) and amino acids (sce01230), specifically valine, leucine and isoleucine biosynthesis (sce00290) and glycine, serine and threonine metabolism (sce00260), proteasome (sce03050), ribosome (sce03010) and carbon metabolism (sce01200).

The reported differences in the thermal stability likely reflect changes either in a protein interaction or its PTM status [10,13,14,37], or both, since PTMs are known to affect formation of PP and PM complexes [38]. Moreover, thermal stability can also serve as a proxy for enzyme activity, reflecting a change in the substrate occupancy [10]. In summary, proteome-wide analysis of changes in thermal stability across the transition from fermentative to respiratory metabolism reveals hundreds of proteins changing in stability, attesting to the significant changes in the protein interaction and/or PTMs status.

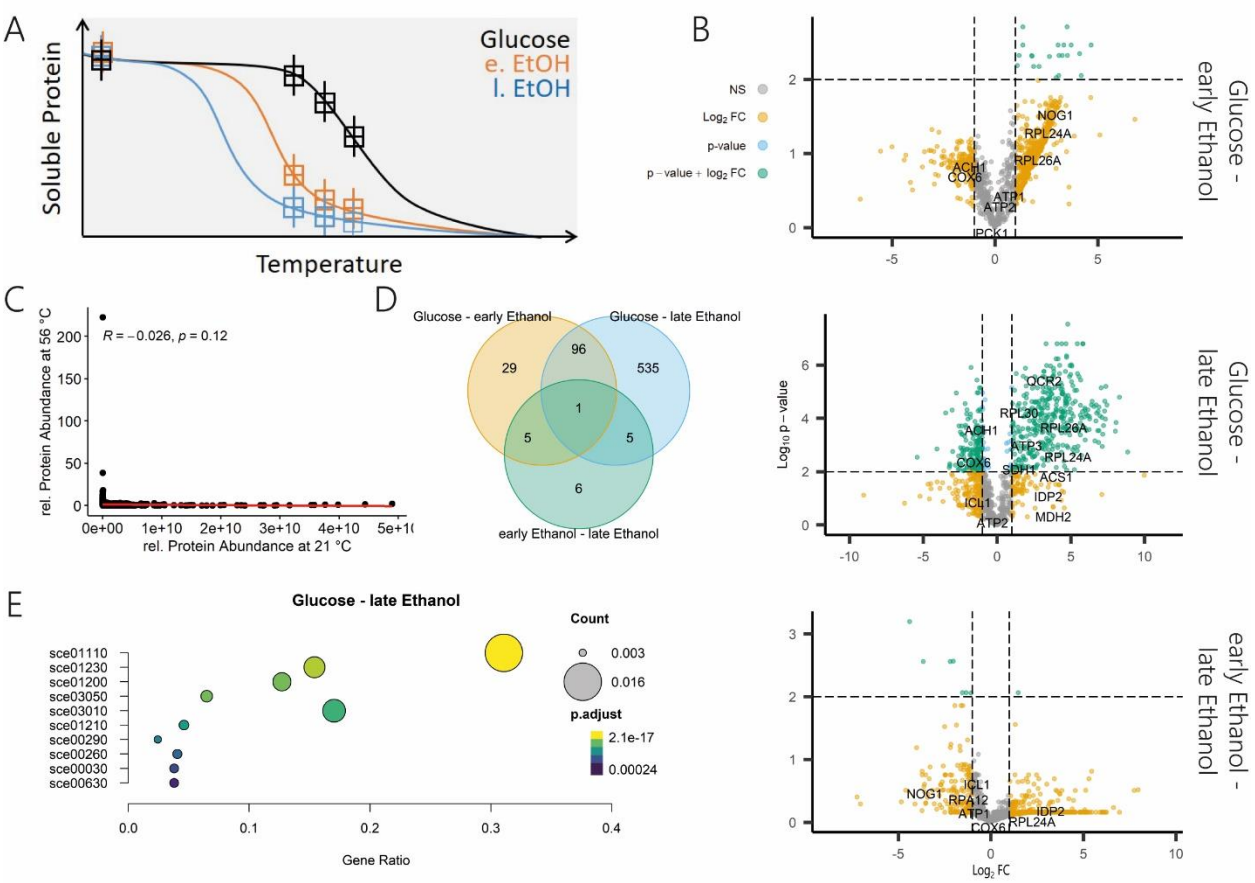


Figure 2: Growth phase dependent changes in protein thermal stability. A) Schematic differential thermal gradient curves of a protein sampled in three different growth phases. Protein thermal stability was measured after treatment with three elevated temperatures – 48 °C, 52 °C and 56 °C – as described in materials and methods. B) Volcano plots showing the differences in protein thermal stability at 56 °C between glucose- and early ethanol-utilizing phase (upper panel), glucose- and late ethanol-utilizing phase (middle panel) and early ethanol- and late ethanol-utilizing phase (lower panel), respectively, tested using a two-sided t-test. The horizontal, dashed line indicates FDR = 0.01, vertical, dashed lines indicate a fold-change greater than 2. Proteins with significant (FDR < 0.01) fold changes greater than 2 are highlighted in green, and proteins involved in central carbon metabolism are labeled. C) Relative protein abundance and protein thermal stability at 56 °C are not correlated (PCC = -0.26). D) Venn diagram showing the overlap of proteins with significant changes (FDR < 0.01, FC > 2) in thermal stability between the growth phases in all temperature treatments. E) KEGG enrichment of the 97 proteins with significant changes in thermal stability between glucose-utilizing and late-ethanol -utilizing phase. Only the ten most significant enrichments are shown. sce01110: Biosynthesis of secondary metabolites, sce02130: Biosynthesis of amino acids, sce03050: Proteasome, sce01200: Carbon metabolism, sce03010: Ribosome, sce01210: 2-Oxocarboxylic acid metabolism, sce00290: Valine, leucine and isoleucine biosynthesis, sce00260: Glycine, serine and threonine metabolism, sce00030: Pentose phosphate pathway, sce00630: Glyoxylate and dicarboxylate metabolism.

Diauxic shift is associated with major changes in the metabolite-protein interactome

Building on the results from iTSA, to further examine growth phase-dependent changes in the protein-protein (PP) and protein-metabolite (PM) complexes, we used a CF-MS approach developed in the group dubbed PROMIS for PRotein Metabolite Interactions using Size Separation [24,25,39,40]. PROMIS relies

on the size-based separation of molecular complexes present in native cell lysate: metabolites bound to protein complexes separate into earlier-eluting high molecular weight fractions, whereas unbound small molecules separate in late-eluting low molecular weight fractions [24,39].

The yeast cultures harvested at the different growth phases were used to prepare native cell lysates. PP and PM complexes were separated using size exclusion chromatography (SEC), yielding 60 fractions. Forty of the collected fractions are protein-containing and span a size range between 5 mDa to 20 kDa, as examined using commercial reference proteins of known size. Fractions were extracted and relative abundances of proteins and metabolites were measured using mass spectrometry-based untargeted proteomics and metabolomics (**Fig 3A**). We obtained a dataset containing 2812 proteins and 275 metabolites, which we annotated using an in-house library of authentic reference compounds (**Table S4 and S8**). Elution profiles were normalised, deconvoluted and correlated using PROMISed, a novel web-based tool to facilitate analysis and visualization of the molecular interaction networks from co-fractionation mass spectrometry (CF-MS) experiments [41].

First, to understand the relation between changes in metabolite abundance and PMIs, we calculated the total amount of each metabolite by summing the relative intensity across the 60 chromatographic fractions, whereas the 40 protein containing fractions represent the protein bound pool, and the 20 protein-free fractions represent the free, unbound pool. Comparison of the total amounts revealed major differences between the growth phases, specifically, comparison of the late ethanol and glucose phases revealed 72 increased and 12 decreased metabolites (FDR < 0.05 and FC > 2 or < 0.5) (**Figure 3 C, Figure S5, Table S5**). The majority of accumulating metabolites were proteogenic dipeptides, in line with our previous study where we reported that proteinogenic dipeptides start to accumulate prior to the diauxic shift transition and remain high in ethanol grown yeast [25]. An interesting exception is a subgroup of proline containing dipeptides such as Pro-Glu, which does not change or even decreases upon glucose depletion. Also decreasing in abundance are nucleotides 2'-AMP, 5'-AMP, 5'-GMP, 5'-GDP, 5'-UMP, 5'-UDP. The exception being, 5'-ADP, which significantly increased in abundance in the early ethanol-utilizing phase by 2.7 fold, and then drastically decreased again in the late ethanol-utilizing phase by 13.8 fold.

Changes in metabolite concentrations can drive non-covalent protein-metabolite interactions; the concentration at which half of the protein is occupied is defined as binding affinity. Binding affinity also depends on a protein's status, such as presence or absence of PTMs or oligomerization state. To understand whether metabolite binding affinities globally change during the diauxic shift we calculated the ratio between the bound and free pool of each metabolite across the three growth phases (**Figure 3 E, Figure S6, Table S5**). Whereas the median ratio for glucose and early ethanol-phases is approximately 0.019 and 0.026, respectively, it significantly increases in the late-ethanol phase to 0.147 ($p < 0.001$). Among the 48

metabolites, which have a significantly higher bound to free ratio (FDR < 0.05 and FC > 2 or < 0.5, **Figure 3 D**) in the late ethanol versus glucose phase are mainly dipeptides and nucleotides. Additionally we find the niacin precursor kynurenine and the amino acid leucine. This increase in retention of the identified metabolites in the protein containing fractions is highly intriguing in terms of both mechanism and function. Metabolite binding is known to enhance protein stability and in that way has proteoprotective properties [42], which may be especially important under conditions associated with the accumulation of damaged proteins and protein aggregates as encountered e.g. during ageing [43].

Next we mined PROMIS datasets for proteins and metabolites that differ in their elution profiles across the three growth phases. Since the fractionation profile of a metabolite is dependent on its protein interaction partners, a change in PMIs between the growth phases would be reflected in the metabolite´s profile. We therefore used a statistical workflow for the pairwise comparison of fractionation profiles, dubbed dis-elution-score [41], as well as the presence or absence of a metabolite in a given growth phase, to identify metabolites whose fractionation profiles differ across the examined growth phases (**Figure 3B, Table S7, Table S9**). Our analysis identified 77 metabolites that differed in their elution profile in at least one comparison. The list comprised dipeptides, amino acids, nucleotides, co-factors, and metabolic intermediates such as kynurenine and methionine sulfoxide. This differential elution can be driven by multiple factors, such as a change in a metabolite and protein concentration, or a change in a protein oligomerization, interaction or PTMs status.

In summary, our analysis identified tens of metabolites and hundreds of proteins that differ in their elution profile across the diauxic shift transition attesting to the significant changes to the PP and PM interactomes.

Diauxic Shift is accompanied by changes in protein interaction status

To identify differentially fractionating proteins, we applied a simplified approach, in which, for each protein, we calculated the manhattan distances between the growth phases and compared it to the median of all manhattan distances obtained this way. A protein with a distance of at least 1.5 times the median distance was considered differential. We found that, with 755, the highest number of differential fractionating proteins were found between the glucose -utilizing and the late-ethanol -utilizing phase, compared to the 557 between glucose and early ethanol and 669 between the ethanol -utilizing phases. Since the numbers of differential proteins identified with the simplified dis-elution-score (see Materials and Methods) was similar (421, 427, 445, respectively), we attribute the large differences to the lower number of identified proteins in the late ethanol -utilizing phase. Taken together, this indicates that the protein

oligomerization states changes dramatically in the course of the diauxic shift, hinting towards global changes in the protein-interaction landscape.

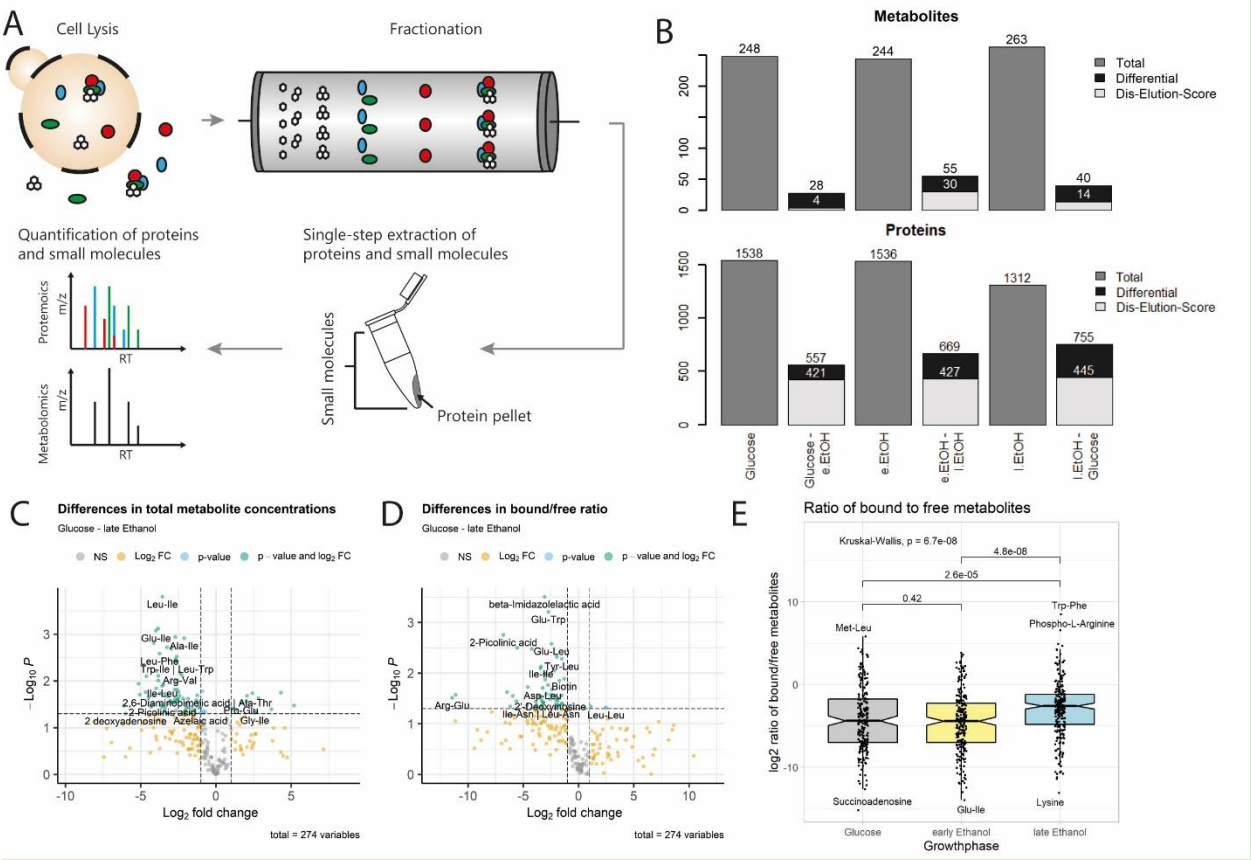


Figure 3: Growth phase dependent differential fractionation of protein-metabolite complexes. A) Schematic workflow of PROMIS: Endogenous protein – small molecule complexes were extracted and fractionated using size exclusion chromatography. Proteins and small molecules were extracted from SEC-fractions using an organic-solvent based method, which denatures proteins (resulting in formation of protein pellets) and releases small-molecules from binding pockets. Proteins and small molecules were separately analyzed using LC-MS. Metabolites were analyzed in biological triplicates, proteins were measured in one replicate. B) Number of differentially fractionating small molecules (upper panel) and proteins (lower panel) between the growth phases. Differentially fractionating small molecules were identified using the dis-elution score (light grey, Schlossarek et al, 2021) and differences in presence/absence (dark grey). For proteins, a simplified approach was used comparing the Manhattan-distance of a protein to the median Manhattan-distance of all proteins between two growth phases (see Materials and Methods). C) Volcano plot showing the differences in total metabolite abundance between glucose-utilizing and late ethanol-utilizing phases. Total abundances were estimated as the sum of the metabolite fractionation profile. D) Changes in metabolite interaction status between glucose-utilizing and late ethanol-utilizing phases as measured as the ratio between metabolite abundances in protein bound and protein free fractions. E) Boxplot comparing the $\log_2$-transformed bound/free ratio of metabolites in each of the three growth phases. In the late ethanol-utilizing phase, metabolites show a higher fraction of bound metabolites (14.7 %) compared to the glucose-utilizing phase (1.9 %) and early ethanol-utilizing phase (2.6 %, $p < 0.001$).

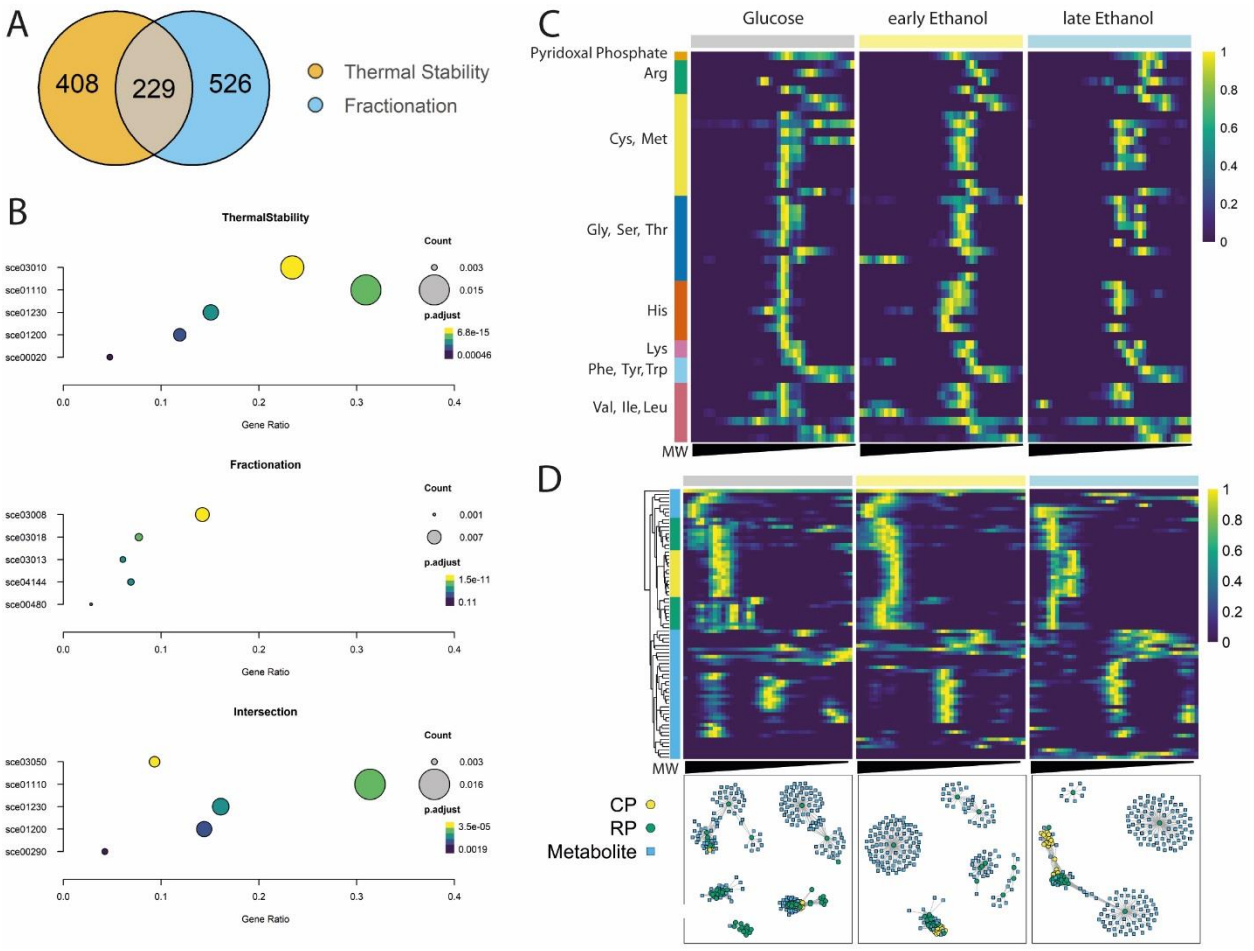


Figure 4: Migration pattern of core and regulatory particles of the 26S proteasome. A) Venn diagram showing the overlap of proteins with significant changes in thermal stability and differential fractionation. B) Top five most significant enriched KEGG pathways of the three venn diagram sections. KEGG identifiers: sce03010: Ribosome, sce01110: Biosynthesis of secondary metabolites, sce01230: Biosynthesis of amino acids, sce01200: Carbon metabolism, sce00020: Citrate cycle (TCA cycle), sce03008: Ribosome biogenesis in eukaryotes, sce03018: RNA degradation, sce03013: Nucleoplasmic transport, sce04144: Endocytosis, sce00480: Glutathione metabolism, sce03050: Proteasome, sce00290: Valine, leucine and isoleucine biosynthesis. C) Normalized fractionation profiles of pyridoxal phosphate (top row) and proteins involved in various amino acid metabolic pathways obtained for the glucose-utilizing phase and ethanol -utilizing phases. Theoretical molecular weight was calculated using reference proteins of known mass. Proteins involved in more than one pathway are included multiple times. D) Disassembly of the proteasome in the late ethanol -utilizing phase: Heatmap showing fractionation profiles of 13 proteasomal core particle (CP) proteins and 19 regulatory particle (RP) proteins, as well as 44 co-fractionating metabolites across the growth phases. Relative protein intensities were normalized to the maximum intensity of the fractionation profile. Theoretical molecular weight, ranging from 20 kDa to 5 mDa, was calculated using reference proteins of known mass. Correlation networks of subunits of the proteasomal core particle (yellow), regulatory particle (green) and co-fractionating metabolites (blue). Only correlations of 0.7 or higher are displayed.

Comparison of Thermal Stability and Differential Fractionation

We next compared the proteins showing differential thermal stability and fractionation, focusing on the comparison between the glucose and late ethanol-utilizing phase. 229 proteins were characterized by the differential fractionation pattern and altered thermal stability, 408 proteins were only affected in their thermal stability and 526 proteins in their elution profile (**Figure 4A, Table S13**). The 229 proteins found in the overlap we interpret as part of dynamic protein complexes, as exemplified by the varying assembly of the proteasome (see below). The 408 proteins with differential thermal stability only were interpreted as undergoing changes in their PTM status or interactions with metabolites, or both, that do not change their oligomeric status. These were significantly enriched for proteins associated with the ribosome, secondary metabolism and biosynthesis of amino acids (**Figure 4 B**), including glycine, serine and threonine metabolism, such as CYS3 (see below, **Table S13**).

Whereas the change in the melting stability but not the fractionation pattern can be explained by the change in a ligand concentration or a protein PTMs status (discussed above) the opposite is harder to interpret. Presented here includes 526 proteins which show differential fractionation, despite being not affected in their thermal stability (**Figure 4 A**). This group is significantly enriched for proteins involved in ribosome biogenesis, RNA degradation, nucleocytoplasmic transport and endocytosis (**Figure 4 B**), and enzymes of the central carbon metabolism. There are many factors that affect thermal stability, and therefore susceptibility to thermal proteome profiling [34]. As discussed in the previous section protein length and thermal stability show a generally weak, but statistically highly significant negative correlation [34–36]. We therefore hypothesised that proteins with differential fractionation, but similar thermal stability are, on average, smaller than other proteins identified in both experiments. However, the opposite is the case: We found that, with a median length of 468.5 aa, proteins with differential fractionation are significantly larger than proteins which are only affected in their thermal stability (median length = 332.5 aa, p = 1.4e12) or the median of all proteins shared between both experiments (median length = 423 aa, p = 0.0016, **Figure S5**), and also larger than the reported median length of yeast proteins of 379 aa [44]. A much simpler explanation is that PROMIS is better suited to select even a relatively minor change of a protein oligomeric state. A single protein can exist in multiple oligomeric states, reflected by multiple maxima in the PROMIS elution profile. The distribution of a protein between the different maxima will vary and in our analysis we accept all the maxima above the 10 % of the main maxima. In contrast, because TPP/CETSA/iTSA gives average stability of all the oligomeric states by default it will mainly reflect the stability of a dominant oligomeric state and hence may miss interactions. Therefore, TPP and CF-MS are complementary approaches and can be used in combination to unravel dynamics in protein-protein interactions that would be missed using one approach alone.

Regulation of amino acid metabolism by dynamic PMIs

Among the 408 proteins affected by thermal stability only, we found the enzyme cystathionine gamma-lyase 3 (CYS3), which catalyses the formation of cysteine from cystathionine [45–47]. We queried our data set for metabolites interacting with CYS3, and found pyridoxal phosphate (PLP), the known cofactor of CYS3 [48]. **Figure 4C** displays a heatmap of fractionation profiles of PLP and proteins involved in amino acid biosynthesis that require PLP as a cofactor. PLP shows a statistically significant differential fractionation pattern between the glucose -utilizing phase and early ethanol -utilizing phase (dis-elution-score (DES): 9.2e-04) and between the ethanol -utilizing phases (DES: 5.29e-07), which is not the case for CYS3 and other proteins involved in amino acid metabolism (**Figure 4C**). In the glucose-utilizing phase, CYS3 and PLP cofractionate (pearson correlation = 0.922) with a peak at around 172 kDa, the size of enzymatically active CYS3 tetramer [48]. In the late ethanol -utilizing phase, CYS3 shows an additional, smaller peak, corresponding to the size of the dimer at ~ 85 kDa. The PLP elution maxima shifts to even smaller sized fractions, showing no co-fractionation with CYS3 (pearson = 0.52 for the larger peak, 0.11 for the smaller peak). This loss of interaction is accompanied by an observed decrease in thermal stability of CYS3 in the late ethanol -utilizing phase.

An important driver of protein-metabolite interactions is the concentration of a protein and its ligand. However, both PLP and CYS3 show no significant differences in abundance across the course of the diauxic shift. Another factor regulating PMIs are PTMs, and it has been shown that protein thermal stability can be affected by PTMs [13], especially of the phosphosites that affect protein structure [14,37]. In human, the cystathionine gamma-lyase, CTH, is phosphorylated by PKG1-β. Yeast CYS3 and human CTH share a 51% protein sequence identity [48], and the region around CTH Ser377 and CYS3 Ser372 shows a conserved sequence. Moreover, NetPhos 3.1 [49,50] predicts a phosphosite at Ser372 of CYS3. Therefore, we hypothesized that the interaction between CYS3 and PLP might be regulated by a dynamic phosphorylation of CYS3. However, in a recent study tracking the phosphorylation states of yeast proteins during the diauxic shift, CYS3 did not show phosphorylation at Ser372, but at Ser40. Contrary to our hypothesis, the phosphorylation state at Ser40 showed no changes in the course of the diauxic shift [51]. Therefore, the observed "loss of interaction" with PLP can not be explained by a dynamic phosphorylation status. However, phosphorylation is not the only PTM, and the dynamic regulation of CYS3-PLP interaction could be mediated by e.g. acetylation or methylation, presumably at the PLP binding Lys203, or at distant sites, conferring a conformational change to the enzyme.

Reorganization of the Proteasome across the Diauxic Shift

As mentioned above, the 229 proteins found in the overlap of iTSA and PROMIS contain multiple proteasomal proteins (**Figure 4 B**), specifically 13 of the 14 core-particle (CP), and 18 of the 21 regulatory-particle (RP) subunits. To learn about the re-arrangement of the proteasome across the diauxic shift transition we used PROMISed [41] to create interaction networks restricted to proteasomal subunits and cofractionating metabolites (PCC > 0.7) (**Figure 4 D**). We used the Louvain method [52] to detect communities within the networks, and obtained seven, six and five clusters, for the glucose, early ethanol and late ethanol phase, respectively (**Figure 4 D, Tables S14-S18)**. We interpret these clusters as stable or transient sub-assemblies of proteasomal subunits and co-eluting metabolites. The clusters can be divided into major clusters containing multiple subunits that correspond to large complexes and minor clusters composed of few or even single subunits that correspond to small complexes or monomeric proteins.

In the glucose-utilizing phase there are two major clusters, one containing the 13 CP and 16 RP subunits, and 23 metabolites, and the second consisting of 17 RP subunits and 14 metabolites. The remaining five minor clusters contain up to 5 RP subunits and dozens of metabolites. In the early ethanol-utilizing phase, we again identified two major clusters, one containing the 13 CP and 3 RP subunits, and 4 metabolites, and the second consisting of 17 RP subunits and 11 metabolites. The remaining four minor clusters contain between one and four RP subunits. Finally, in the late ethanol-utilizing phase, two major clusters were found, one is made up of 19 CP, a single RP subunit, and five metabolites. The other cluster consists of 19 RP and 5 CP subunits and 10 metabolites. Three minor clusters contain a single RP subunit and several metabolites.

There are three observations that can be made from the network analysis. First, the number and size of the minor clusters decrease in the ethanol-utilizing phases. Second, the interaction between the two sub complexes, CP and RP, is changing. In the glucose-utilizing phase, the majority of CP and RP subunits are clustered together. In contrast, both sub complexes form largely separate clusters in the late ethanol-utilizing phase. The early ethanol-utilizing phase represents an intermediate behaviour, where RP and CP form distinctive clusters, which are tightly connected, based on their co-fractionation profiles. Third, minor clusters contain more metabolites (up to 89) than major-protein clusters (up to 23).

Our findings are in line with Bajorek and colleagues, who reported the disassembly of the yeast 26S proteasome (CP and RP) into the 19S (RP) and 20S (CP) sub-complexes in the stationary phase [53]. Moreover, in Arabidopsis, proteins involved in the RP and CP, respectively, possess different extremes of thermostability, in that the CP is highly stable, while the RP is a thermo-labile complex, hinting towards a higher degree of conformational flexibility of the RP [33]. In human cells under hypoxia, the CP can act as a stand-alone 20S proteasome independently of the RP, and shows distinct features compared to the human

holoenzyme, as it is able to rapidly degrade unstructured proteins, generates longer peptides resulting from a distinct cleavage pattern and degrades conjugated ubiquitin instead of releasing it from the substrate [54]. Against this background we speculate that in the late ethanol-utilising phase CP functions as stand-alone 20S proteasome degrading unstructured, aggregated proteins that accumulate in the ageing yeast cells. An additional mechanism to the discussed above global increase in the metabolite retention in the protein complexes.

Our data further suggest that the remodelling of the proteasome is accompanied by changing metabolite binding. Intriguingly, among the 23 metabolites that we found co-eluting, in the glucose-utilizing phase, 22 are proteinogenic dipeptides. Dipeptides were also found co-eluting with the clusters corresponding to the 19S and 20S proteasome (see network description above), however they were outnumbered by other compounds, especially nucleotides. These results extend previous work, where 3 RP and 4 CP subunits were found among the 86 high confidence interactors of the dipeptide Ser-Leu in the glucose-utilizing phase [25]. A functional connection between dipeptides and proteasome would be highly interesting. This would demonstrate a novel regulatory loop where protein clearance is regulated by products of protein degradation, that is specific dipeptides (also see below).

Dipeptides and central carbon metabolism

Proteogenic dipeptides are a recently discovered class of metabolites with mostly unknown functions. In our previous work in yeast, we showed that the dipeptide Ser-Leu interacts with and activates the glycolytic enzyme phosphoglycerate kinase (Pgk1) by increasing the enzyme's affinity to ATP [25]. A different dipeptide, Tyr-Asp, was previously shown to interact with the glycolytic enzyme glyceraldehyde 3-phosphate dehydrogenase (*Ath*GAPC) in the model plant Arabidopsis [24]. This interaction promotes tolerance to oxidative stress by redirecting carbon flux into the pentose phosphate (PPP) pathway, and increasing NADPH levels [55]. Dipeptides originate from protein degradation, and, although they are always present in the cell, they accumulate in conditions associated with high rates of protein clearance, such as in response to stress [25,56–61]. We, for instance, found that in plants, dipeptide levels increase in response to heat and dark stresses in an autophagy dependent manner [56], and that specific dipeptides display diurnal oscillation in response to the change in a plant carbon status, downstream of TOR signalling [62]. In comparison, yeast accumulate dipeptides in response to glucose depletion *prior* to the diauxic shift transition [25].

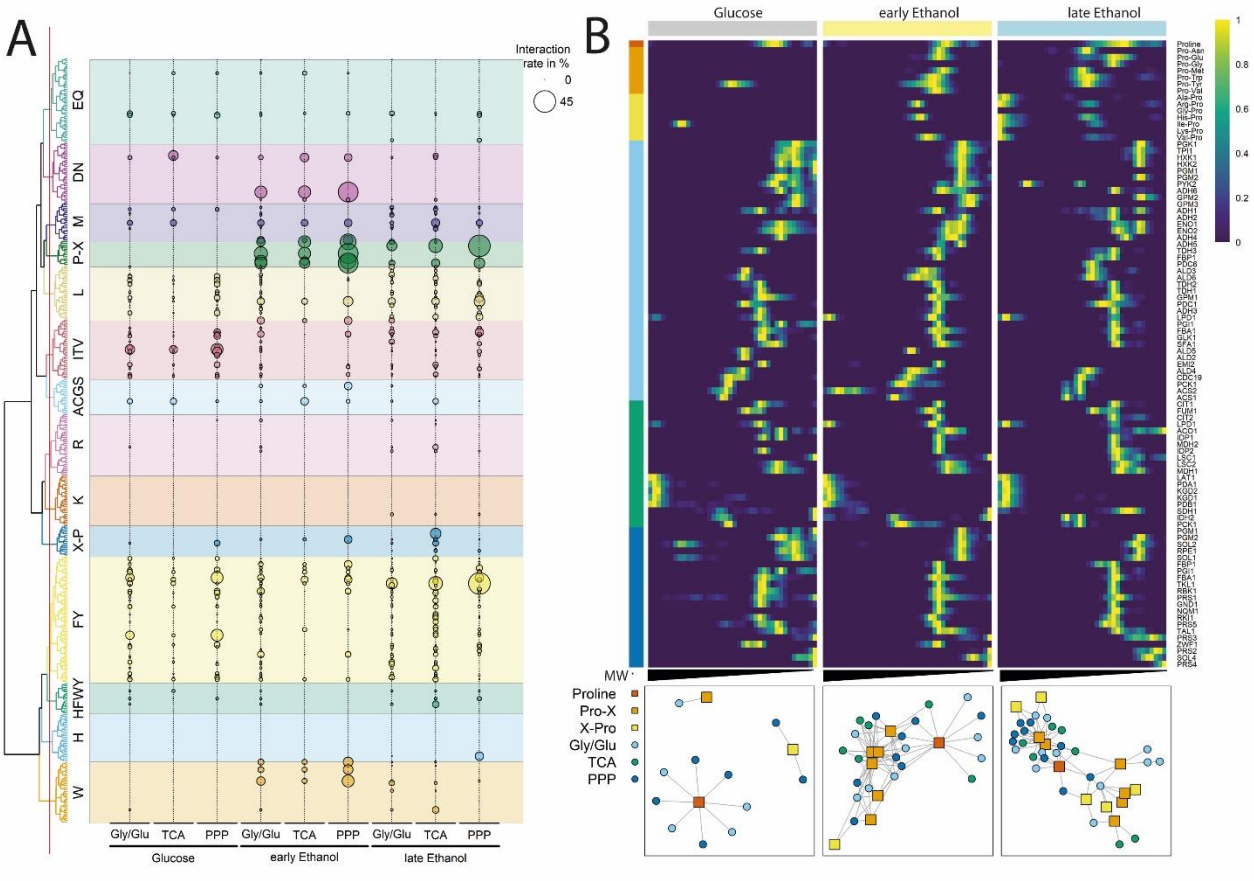


Figure 5: Changes in interaction rates between dipeptides and the central carbon metabolism during the diauxic shift. A) Interaction rates were calculated as the percentage of observed interactions in all possible interactions between the groups and are shown as size-scaled circle. Circle color corresponds to the dipeptide-group. Proteins were grouped into Glycolysis/Gluconeogenesis (**Gly/Glu**), tricarboxylic acid cycle (**TCA**) and pentose phosphate pathway (**PPP**). Dipeptides were clustered based on their chemical structure similarity, resulting in 14 groups: **EQ:** Glu- or Gln-containing, **DN:** Asp- or Asn-containing, **M:** Met-containing, **P-X:** N-terminal Pro-containing, **L:** Leu-containing, **ITV:** Ile-, Thr-, or Val-containing, **ACGS:** Ala-, Cys-, Gly- or Ser-containing, **R:** Arg-containing, **K:** Lys-containing, **X-P:** C-terminal Pro-containing, **FY:** Phe- or Tyr-containing, **HFWY:** N- and C-terminal His-, Phe-, Trp- or Tyr-containing, **H:** His-containing, **W:** Trp-containing. Original data can be found in supplementary table S19. B) Proline containing dipeptides co-fractionate with proteins of the central carbon metabolism in the ethanol -utilizing phase. Top panel: Heatmap showing the normalized fractionation pattern of proline, N-terminal Pro-containing (Pro-X), C-terminal Pro-containing (X-Pro) dipeptides, and proteins involved in Gly/Glu, TCA and PPP. Theoretical molecular weight, ranging from 20 kDa to 5 mDa, was calculated using reference proteins of known mass. Lower Panel: Correlation network of proline and proline containing dipeptides and proteins involved in central carbon metabolism. Only correlations of 0.7 or higher are displayed.

Given the major metabolic rewiring associated with the diauxic shift and the demonstrated role of dipeptides in regulating central carbon metabolism, we queried the current datasets for changing dipeptide-enzyme interactions. Among the 274 annotated protein-bound metabolites 145 (53 %) are proteogenic dipeptides. To get a better understanding of the protein-dipeptide interaction network, we first grouped the 400 proteogenic dipeptides using the ChemmineR package [63], resulting in 14 dipeptide groups (**Figure 5**

A). Four groups are composed of aromatic dipeptides, and 10 of non-aromatic dipeptides. We also grouped the enzymes of the central carbon metabolism into three categories based on the KEGG annotations: glycolysis/gluconeogenesis (Gly/Glu), tricarboxylic acid cycle (TCA) and the pentose phosphate pathway (PPP). We then calculated the interaction rate between dipeptides and enzymes of the different pathways as the percentage of observed interactions in the total of possible interactions (**Figure 5 A, Table S19+S20**).

The group of dipeptides that stood out was the N'-terminal proline containing dipeptides (P-X) (**Figure 5 A-B**). These were almost entirely absent in the protein-containing fractions in the glucose-utilizing phase with only one putative interaction. This is in stark contrast to both ethanol-utilizing phases. The number of co-fractionating proteins and resulting interactions increased to 17 and 69, in the early ethanol-utilizing phase, and to 26 proteins and 48 interactions in the late ethanol-utilizing phase. This increase in interactions cannot be attributed to changes in P-X dipeptide concentration, since the levels of P-X dipeptides do not change or even decrease upon glucose depletion. Enzymes co-fractionating with P-X include glyceraldehyde-3-phosphate dehydrogenase (TDH1, TDH2 and TDH3), fructose 1,6-bisphosphate aldolase (FBA1), phosphoglycerate mutase (GPM1), 3-phosphoglycerate kinase (PGK1), and citrate synthases (CIT1 and CIT2) (**Table S21+S22**). Interestingly, a similar list of putative interaction partners of proline-containing dipeptides has been reported in *Arabidopsis thaliana* [62], pointing to the putative conserved role of proline-containing dipeptides in the regulation of central carbon metabolism.

5.4 Conclusion

Here, we combined two approaches, iTSA and PROMIS, to investigate proteome- and metabolome-wide changes in the PPI and PMI landscape during the transition from fermentative to respiratory growth in *Saccharomyces cerevisiae*. *On par* with previous studies e.g., [7] we reported hundreds of proteins and tens of metabolites accumulating differentially across the diauxic shift transition. Most importantly, the reported changes translate into a significant rewiring of the protein-metabolite interactome; the most pronounced difference measured between glucose- and late-ethanol utilizing phases. Intriguingly, global analysis of the protein-bound to free metabolite ratios revealed that small-molecules are preferentially retained in the protein complexes during late-ethanol phase. We speculate it may be related to the metabolite-driven proteoprotection [42], particularly that entry into the stationary growth phase is associated with accumulation of aggregated proteins [64].

There is only a 20 % overlap of the differential proteins characterized by the change in both the thermal stability and elution profile. Thus, combining both methods is better to obtain a comprehensive view of the differential interactome. Whereas iTSA reveals a major change in a protein interaction or PTM status, PROMIS can capture the different protein oligomerization states and also provide a glimpse into the

nature of the interacting partners including small-molecules. We demonstrate that CF-MS, and specifically PROMIS, is suitable to study the dynamics of protein-metabolite complexes across multiple cell states. The biggest challenge for any CF-MS method is to distinguish between the true and coincidental co-elution, and hence to identify true interactors. We expect that analogously to what was done for the PPIs, we will be able to significantly improve the identification of true protein-metabolite complexes by combining multiple datasets and chromatographic methods, increasing the number of collected fractions, and by introducing machine learning approaches building on the known interactions, reviewed by [65].

The present dataset attests to the dynamic rewiring of metabolite-protein-protein complexes accompanying the switch from fermentative to respiratory growth in yeast. It provides a valuable resource for unravelling the role of PPI and PMI in adjusting yeast metabolism to the changes in glucose availability. We present three examples of the differential PP and PM interactions that we think constitute novel regulatory mechanisms governing the diauxic shift. (i) Pyridoxal phosphate (PLP) is a known cofactor of enzymes involved in transamination [66,67]. Amino acid homeostasis relies on three principal processes: amino acid uptake, *de novo* synthesis and recycling. During starvation (stationary phase) yeast mostly recycles amino acids through proteolysis [68]. Here, we report a loss of co-elution and hence a change in binding, between amino acid biosynthetic enzymes, exemplified by CYS3, and PLP in the late-ethanol utilizing phase. We speculate that the "loss of PLP binding" may be due to a change in the PTMs status, other than phosphorylation, which would lead to the inhibition of the enzymatic activity, contributing to the downregulation of *de novo* amino acid synthesis in the late-ethanol utilizing phase. (ii) Next, we report a gradual disassembly of the 26S proteasome into the 20S (CP) and 19S (RP) sub-complexes associated with the switch between fermentative and respiratory growth. Our observations are in line with previous literature findings that showed a disassembly of the proteasome into stable complexes, accompanied by a reduction in proteolytic activity [53]. Nevertheless, the CP is able to degrade its substrates independently of the RP in various eukaryotic models [54]. Against this background we speculate that in the late ethanol-utilising phase CP functions as stand-alone 20S proteasome degrading unstructured, aggregated proteins that accumulate in the ageing yeast cells. (iii) Lastly, and on par with our previous report from plants [62], we demonstrate a co-elution, and hence a putative interaction, between the proline containing, Pro-X, dipeptides, and various enzymes of the central carbon metabolism specifically in the ethanol-utilizing phases. Intriguingly the appearance of Pro-X dipeptides in the protein complexes is independent of the cellular concentrations pointing to the binding being dependent on e.g. on a change in a protein(s) PTMs status. And however, the identity of the Pro-X protein targets needs to be independently validated; it solidifies the shown before regulatory roles of dipeptides [25,55,60,69], beyond simply intermediates of the protein degradation.

In summary, the present dataset (i) attests to the dynamic rewiring of metabolite-protein-protein complexes accompanying the switch from fermentative to respiratory growth in yeast. It provides a valuable resource for unravelling the role of PPI and PMI in adjusting yeast metabolism to the changes in glucose availability. Notably, whereas the switch between fermentative to respiratory growth is also of interest to cancer metabolism research, the entry into the stationary phase shares molecular signatures with chronological ageing in other organisms including humans. (ii) Moreover, we demonstrate that CF-MS, and specifically PROMIS, is suitable to study the dynamics of protein-metabolite complexes across multiple cell states. Finally, as discussed before [25,55] the reported small molecules, represent just a small subset of all the metabolic features measured in the protein containing fractions, that we could annotate. In the future and by concentrating on the chemical identification of the "unknown" metabolic features showing differential elution we aim to identify novel small-molecule regulators of the diauxic shift transition.

5.5 Methods

5.5.1 Experimental Procedures

Yeast growth and cell lysis

Experiments were performed using the YSBN2 strain of *Saccharomyces cerevisiae* cultivated in YPD-medium at 30 °C with moderate shaking (120-140 RPM) using Innova Shakers. Cultures were inoculated using a single colony grown on a YPD plate. Cells were collected by centrifugation (4,000 g, 4 °C, 20 min) after 6, 24 and 72 hours of cultivation, corresponding to growth in the glucose-utilizing, early ethanol-utilizing and late ethanol-utilizing phase, respectively. Cell pellets were washed with AmBIC buffer (50 mM ammonium bicarbonate, 150 mM NaCl, 1.5 mM $MgCl_2$), transferred to 50 mL tubes and centrifuged again (4,000 g, 4 °C, 20 min). The pellets were then snap frozen in liquid nitrogen and stored at -80 °C until further use. Collected cells were mixed with an ice-cold lysis buffer (50 mM AmBIC, 150 mM NaCl, 1.5 mM $MgCl_2$, 5 mM DTT, 1 mM PMSF, 1x cOmplete EDTA-free Protease Inhibitor Cocktail (MERCK, 11873580001), 0.1 mM Na3VO4, and 1 mM NaF) and frozen (-20 °C) silica-zirconia beads (Biospec, 11079105z). Yeasts were homogenised by bead beating 10 × 30 s at 20 Hz using a Retsch Mixer Mill MM 400, and cooled in an ice-water bath for 1 min in-between bead beating. Cell debris and silica-zirconia beads were sedimented by a 10 min centrifugation at 4000 g and 4 °C. After centrifugation, the supernatant was used for either the Isothermal Shift Assay or PROMIS.

Isothermal shift assay

The protocol was adapted from [30]. The supernatant (see above) was transferred to Eppendorf tube and centrifuged for 10 min at 21,000 g and 4 °C. The protein concentration of the supernatant was determined using the Bradford assay. Equivalent of 200 µg of proteins (5 replicates for each of the 3 growth phases) was transferred to PCR tubes and heated to 48 °C, 52 °C or 56 °C for 3 min and then incubated for 3 min at room temperature. During this time samples were moved to 1.5 mL tubes and centrifuged for 20 min at 21,000 g and 4 °C. The supernatant was then transferred to fresh Eppendorf tubes and proteins were precipitated overnight at -20 °C using 80% Acetone (MS-grade). Next day, proteins were pelleted by centrifugation for 20 min at 21,000 g and 4 °C. Supernatant was removed and proteins were air-dried. Finally, proteins were digested and desalted as described in section LC-MS/MS of proteins. Dried peptides were suspended in 60 µL MS loading buffer (2% ACN, 0.2 % TFA), separated and quantified using LC-MS as described in section LC-MS/MS of proteins.

Separation of endogenous complexes using size exclusion chromatography

The protocol was adapted from [25]. The supernatant (see above) was transferred to ultracentrifuge tubes and centrifuged for one hour at 35,000 RPM (max. 148,862 g, avg. 116,140 g) at 4 °C to obtain the soluble fraction containing endogenous complexes. The soluble fraction was loaded onto previously pre-rinsed (15 mL wash buffer: 50 mM AmBIC, 150 mM NaCl, 1.5 mM MgCl2, 20 min, 4,000 g, 4 °C) Amicon Ultra-15 centrifugal filter units (10 kDa MWCO) and centrifuged for 20 min at 4,000 g, 4 °C.

Soluble fraction, corresponding to 40 mg of protein, was loaded on Sepax SRT SEC-300 21.2 × 300 mm column (Sepax Technologies, Inc., Delaware Technology Park, separation range 1.2 mDa to 10 kDa) connected to an ÄKTA explorer 10 (GE Healthcare Life Science, Little Chalfont, UK) and separated at 7 mL/min flow rate, 4 °C. 50 mM AmBIC pH 7.5, 150 mM NaCl, 1.5 mM MgCl2 was used to equilibrate the column and separate molecular complexes. Forty 1-mL fractions were collected from the 39 mL to 78 mL elution volume. The fractions were frozen by snap freezing in liquid nitrogen and subsequently lyophilized and stored at –80 °C for metabolite and protein extractions.

Extraction of proteins and polar metabolites

The extraction protocol was adapted and modified from [40]. Proteins and metabolites were extracted from the lyophilised fractions using a methyl-tert-butyl ether (MTBE)/methanol/water solvent system. Equal volumes of the polar fraction and protein pellet were dried in a centrifugal evaporator and stored at –80 °C until they were processed further.

LC-MS/MS of proteins

Fractionated proteins were quantified using the Bradford assay. Protein pellets from 40 fractions corresponding to 39 to 78 mL elution volume were suspended in 30 µL urea buffer (6 M urea, 2 M thiourea in 40 mM ammonium bicarbonate). 20 µg of proteins from each fraction were reduced, alkylated and enzymatically digested using LysC/Trypsin Mix (Promega Corp., Fitchburg, WI) according to the manufacturer's instructions. Self-made C18 Empore® extraction discs (3M, Maplewood, MN) STAGE tips were used for protein desalting [70]. Proteins were concentrated using the centrifugal evaporator to approximately 4 µL and stored at -80 °C until measured. Dried peptides were suspended in 60 µL MS loading buffer (2% ACN, 0.2 % TFA), and 3 µL (equivalent to 0.8–1.0 µg of peptides) were separated using C18 reversed-phase column connected to an ACQUITY UPLC M-Class system in a 120 min gradient. The gradient started from 3.2% and increased to 7.2% ACN in 20 min, to 24.8% ACN over 70 min and to 35.2% ACN over 30 min, followed by a 5 min washout with 76% ACN. The Thermo Q Exactive HF operated with a data-dependent method as follows: MS full scans were performed in FTMS with resolution set to 120,000, from 300.0 to 1600.0 m/z, a maximum fill time of 50 ms, and an AGC target value of 3e6 ions. A maximum of 12 data-dependent MS2 scans was performed in the ion trap set to an AGC target of 1e5 ions with a maximal injection time of 100 ms. Precursor ion fragmentation was achieved with collision-induced fragmentation with a normalised collision energy of 27 and isolation width of 1.2 m/z. Charge states of 1 and ≥7 were excluded.

LC-MS metabolomics

After extraction, the dried aqueous phase was measured using ultra-performance liquid chromatography coupled to a Q-Exactive mass spectrometer (Thermo Fisher Scientific) in positive and negative ionisation modes, as described earlier [71].

5.5.2 Data Analysis

Data pre-processing: LC-MS metabolite data

Expressionist Refiner MS 12.0 (Genedata AG, Basel, Switzerland) was used for processing the LC-MS data with the following settings. Repetition was used to reduce the volume of data and to speed up processing. All types of data except Primary MS Centroid Data were removed using Data Sweep. Chemical Noise Subtraction activity was used to remove artefacts caused by chemical contamination. Snapshot of chromatogram was saved for further processing. Further processing of chromatogram snapshot were

performed as follows: chromatogram alignment (RT search interval 0.5 min), peak detection (minimum peak size 0.03 min, gap/peak ratio 50%, smoothing window 5 points, centre computation by intensity-weighted method with intensity threshold at 70%, boundary determination using inflection points), isotope clustering (RT tolerance at 0.02 min, m/z tolerance 5 ppm, allowed charges 1–4), filtering for a single peak not assigned to an isotope cluster, charge and adduct grouping (RT tolerance 0.02 min, m/z tolerance 5 ppm). A more detailed description of the software usage and possible settings was published before [40]. In-house library of authentic reference compounds was used to identify molecular features allowing 0.005 Da mass deviation and dynamic retention time deviation (maximum 0.2 min). Processing of fractionated samples resulted in annotation of 282 small molecules across three growth phases.

Protein data analysis was restricted to the 1627 proteins commonly identified across all experiments.

Analysis of protein abundance

Protein intensities were measured in three growth phases in five replicates. Protein intensities were normalized to the median intensity across all samples as $Int_{norm} = Int_{protein} * (Median_{sample}/Median_{global})$. To identify differences in protein abundance between growth phases in a pairwise manner, we calculated the fold-changes and performed a two-sided student's t-test.

Analysis of thermal stability

After treatment with 3 temperatures (48 °C, 52 °C and 56 °C) and subsequent removal of denatured proteins by centrifugation, the abundance of soluble proteins was measured in five replicates. Protein intensities were normalized to the median intensity within each temperature group, and scaled to the protein abundance at room temperature. Proteins with significant (FDR < 0.01) fold changes between growth phases in the same temperature group were considered as having a differential thermal stability. At 56 °C, 21 proteins showed a differential thermal stability between glucose-utilizing phase and early ethanol-utilizing phase, 531 between glucose-utilizing phase and late ethanol-utilizing phase and 8 between early- and late-ethanol -utilizing phases, respectively.

Analysis of differential protein elution profiles

For each protein, the Manhattan-distance (MD) was calculated between processed elution profiles of all three growth phases in a pairwise manner. From this, the median MD was determined, and proteins with a MD greater than 1.5 of the median were considered as differentially eluting.

Protein data integration

To further narrow down the list of proteins of interest, we integrated the different datasets using the VennDiagram package [72]. We then performed KEGG-enrichment analysis for proteins present in the different sections of the Venn diagram using the clusterProfiler package [73].

Volcano plots

Volcano plots are generated using the EnhancendVolcano R package [74]. The x-axis shows the $\log_2$ fold-change (FC), the y-axis the $-\log_{10}$ p-value. The horizontal, dotted line indicates the p-value threshold (0.01), the vertical, dotted lines the thresholds of fold-changes (-2 and 2). Points correspond to proteins and are coloured as follows: Grey: FC between -2 and 2, $p > 0.01$, Green: FC outside of thresholds, $p > 0.01$, Blue: FC between thresholds, $p < 0.01$, Red: FC outside of thresholds, $p < 0.01$. Dipeptide clustering based on chemical structure similarity.

Dipeptide clustering

Smiles codes of dipeptides were obtained from https://pubchem.ncbi.nlm.nih.gov/ on 12.10.2021. Pairwise distances between dipeptides were calculated on atom pair libraries using cmp.cluster() of the ChemmineR package [63]. Clustering was performed using Ward´s minimum variance method [75]. The resulting dendrogram was cut at height 1.6, resulting in 14 dipeptide groups named after the predominant amino acids.

Calculation of interaction rates

Fractionation profiles of dipeptides and proteins involved in central carbon metabolism were analyzed with PROMISed [41], using default settings, and correlation tables were obtained. PROMISed splits fractionation profiles into distinct, single peaks, and calculates Pearson correlation between all obtained metabolite and protein peaks. We filtered the obtained correlation table for co-fractionating pairs using a threshold of 0.7. We then grouped all proteins based on their involvement in Glycolysis/Gluconeogenesis,

tricarboxylic acid cycle (TCA cycle) and the pentose phosphate pathway (PPP). The interaction rate between dipeptides and these groups was calculated as the percentage of observed interactions in all possible interactions as

$$InteractionRate = n_{Interactions}/(n_{DipeptidePeaks} * n_{ProteinPeaks}) * 100$$

where $n_{Interactions}$ is the number of observed co-fractionations (PCC > 0.7), $n_{DipeptidePeaks}$ the number of peaks originating from the dipeptide and $n_{ProteinPeaks}$ the number of peaks of proteins in that group.

Acknowledgements

We thank Änne Michaelis and Aleksandra Brzezińska for excellent technical assistance.

Competing interests

Authors declare no competing interests.

5.6 References – Publication #3

1. Zampar, G. G. *et al.* Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. *Mol. Syst. Biol.* **9**, (2013).
2. Werner-Wahsburne, M., Braun, E., Johnston, G. C. & Singer, R. A. Stationary Phase in the Yeast Saccharomyces cerevisiae. **135**, 97–109 (1996).
3. Hammad, N., Rosas-Lemus, M., Uribe-Carvajal, S., Rigoulet, M. & Devin, A. The Crabtree and Warburg effects: Do metabolite-induced regulations participate in their induction? *Biochim. Biophys. Acta - Bioenerg.* **1857**, 1139–1146 (2016).
4. Galdieri, L., Mehrotra, S., Yu, S. & Vancura, A. Transcriptional regulation in yeast during diauxic shift and stationary phase. *Omi. A J. Integr. Biol.* **14**, 629–638 (2010).
5. Broach, J. R. Nutritional control of growth and development in yeast. *Genetics* **192**, 73–105 (2012).
6. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Chemtracts* **12**, 148–152 (1999).
7. Murphy, J. P., Stepanova, E., Everley, R. A., Paulo, J. A. & Gygi, S. P. Comprehensive temporal protein dynamics during the diauxic shift in Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **14**, 2454–2465 (2015).
8. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science (80-.).* **346**, (2014).
9. Molina, D. M., Jafari, R., Ignatushchenko, M. & Seki, T. Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay. *Science (80-.).* **341**, 84–88 (2013).
10. Becher, I. *et al.* Pervasive Protein Thermal Stability Variation during the Cell Cycle. *Cell* **173**, 1495-1507.e18 (2018).
11. Dai, L. *et al.* Modulation of Protein-Interaction States through the Cell Cycle. *Cell* **173**, 1481-1494.e13 (2018).
12. Mateus, A. *et al.* The functional proteome landscape of Escherichia coli. *Nature* **588**, 473–478 (2020).
13. Huang, J. X. *et al.* High throughput discovery of functional protein modifications by Hotspot Thermal Profiling. *Nat. Methods* **16**, 894–901 (2019).
14. Potel, C. M. *et al.* Impact of phosphorylation on thermal stability of proteins. *Nat. Methods* **18**, 757–759 (2021).
15. Sun, W. *et al.* Monitoring structural modulation of redox-sensitive proteins in cells with MS-CETSA. *Redox Biol.* **24**, 101168 (2019).
16. Salas, D., Stacey, R. G., Akinlaja, M. & Foster, L. J. Next-generation interactomics: Considerations for the use of co-elution to measure protein interaction networks. *Mol. Cell. Proteomics* **19**, 1–10 (2020).
17. Aryal, U. K. *et al.* A Proteomic strategy for global analysis of plant protein complexes. *Plant Cell* **26**, 3867–3882 (2014).
18. Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).
19. Hu, L. Z. M. *et al.* EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* **16**, 737–742 (2019).
20. Chan, J. N. Y. *et al.* Target identification by chromatographic co-elution: Monitoring of drug-protein interactions without immobilization or chemical derivatization. *Mol. Cell. Proteomics* **11**, (2012).
21. Gorka, M. *et al.* Protein Complex Identification and quantitative complexome by CN-PAGE. *Sci. Rep.* **9**, 1–14 (2019).

22. Geladaki, A. *et al.* Combining LOPIT with differential ultracentrifugation for high-resolution spatial proteomics. *Nat. Commun.* **10**, 1–15 (2019).

23. McWhite, C. D. *et al.* Co-fractionation/mass spectrometry to identify protein complexes. *STAR Protoc.* **2**, 100370 (2021).

24. Veyel, D. *et al.* PROMIS, global analysis of PROtein-metabolite interactions using size separation in Arabidopsis thaliana. *J. Biol. Chem.* **293**, 12440–12453 (2018).

25. Luzarowski, M. *et al.* Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. *Commun. Biol.* **4**, 1–15 (2021).

26. Li, Y. *et al.* Coupling proteomics and metabolomics for the unsupervised identification of protein–metabolite interactions in Chaetomium thermophilum. *PLoS One* **16**, 1–13 (2021).

27. Mallam, A. L. *et al.* Systematic Discovery of Endogenous Human Ribonucleoprotein Complexes. *Cell Rep.* **29**, 1351-1368.e5 (2019).

28. Lindsley, J. E. & Rutter, J. Whence cometh the allosterome? *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10533–10535 (2006).

29. Li, X. & Snyder, M. Metabolites as global regulators: A new view of protein regulation. *BioEssays* **33**, 485–489 (2011).

30. Ball, K. A. *et al.* An isothermal shift assay for proteome scale drug-target identification. *Commun. Biol.* **3**, 1–10 (2020).

31. Heusel, M. *et al.* A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133-155.e6 (2020).

32. Canelas, A. B. *et al.* Integrated multilaboratory systems biology reveals differences in protein metabolism between two reference yeast strains. *Nat. Commun.* **1**, (2010).

33. Volkening, J. D., Stecker, K. E. & Sussman, M. R. Proteome-wide analysis of protein thermal stability in the model higher plant arabidopsis thaliana. *Mol. Cell. Proteomics* **18**, 308–319 (2019).

34. Jarzab, A. *et al.* Meltome atlas—thermal proteome stability across the tree of life. *Nat. Methods* **17**, 495–503 (2020).

35. Leuenberger, P. *et al.* Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science (80-.).* **355**, (2017).

36. Mateus, A. *et al.* Thermal proteome profiling in bacteria: probing protein state in vivo . *Mol. Syst. Biol.* **14**, 1–15 (2018).

37. Smith, I. R. *et al.* Erratum to: High throughput discovery of functional protein modifications by Hotspot Thermal Profiling (Nature Methods, (2019), 16, 9, (894-901), 10.1038/s41592-019-0499-3). *Nat. Methods* **18**, 760–762 (2021).

38. Duan, G. & Walther, D. The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. *PLoS Comput. Biol.* **11**, 1–23 (2015).

39. Veyel, D. *et al.* System-wide detection of protein-small molecule complexes suggests extensive metabolite regulation in plants. *Sci. Rep.* **7**, 1–8 (2017).

40. Sokolowska, E. M., Schlossarek, D., Luzarowski, M. & Skirycz, A. PROMIS : Global Analysis of PROtein-Metabolite Interactions. *Curr. Protoc. Plant Biol.* **4**, 1–18 (2019).

41. Schlossarek, D. *et al.* PROMISed: A novel web-based tool to facilitate analysis and visualization of the molecular interaction networks from co-fractionation mass spectrometry (CF-MS) experiments. *Comput. Struct. Biotechnol. J.* **19**, 5117–5125 (2021).

42. Verma, K. *et al.* Distinct metabolic states of a cell guide alternate fates of mutational buffering through altered proteostasis. *Nat. Commun.* **11**, (2020).

43. Sampaio-Marques, B. & Ludovico, P. Linking cellular proteostasis to yeast longevity. *FEMS Yeast Res.* **18**, 1–11 (2018).

44. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33**, 3390–3400 (2005).

45. Ono, B. -I, Naito, K., Shirahige, Y. -I & Yamamoto, M. Regulation of cystathionine γ-lyase in Saccharomyces cerevisiae. *Yeast* **7**, 843–848 (1991).

46. Ono, B. *et al.* Cloning and characterization of the ALG3 gene of Saccharomyces cerevisiae. *Glycobiology* **6**, 439–444 (1996).

47. Cherest, H., Thomas, D. & Surdin-Kerjan, Y. Cysteine biosynthesis in Saccharomyces cerevisiae occurs through the transsulfuration pathway which has been built up by enzyme recruitment. *J. Bacteriol.* **175**, 5366–5374 (1993).

48. Messerchmidt, A. *et al.* Determinants of enzymatic specificity in the cys-met-metabolism PLP-dependent enzymes family: Crystal structure of cystathionine γ-lyase from yeast and intrafamiliar structure comparison. *Biol. Chem.* **384**, 373–386 (2003).

49. Blom, N., Gammeltoft, S. & Brunak, S. Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351–1362 (1999).

50. Blom, N., Sicheritz-Pontén, T., Gupta, R., Gammeltoft, S. & Brunak, S. Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* **4**, 1633–1649 (2004).

51. Gassaway, B. M., Paulo, J. A. & Gygi, S. P. Categorization of Phosphorylation Site Behavior during the Diauxic Shift in Saccharomyces cerevisiae. *J. Proteome Res.* **20**, 2487–2496 (2021).

52. Blondel, V. D., Guillaume, J. L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, 1–12 (2008).

53. Bajorek, M., Finley, D. & Glickman, M. H. Proteasome Disassembly and Downregulation Is Correalted with Viabolity during Stationary Phase. *Curr. Biol.* **13**, 1140–1144 (2003).

54. Sahu, I. *et al.* The 20S as a stand-alone proteasome in cells can degrade the ubiquitin tag. *Nat. Commun.* **12**, (2021).

55. Moreno, J. C. *et al.* Tyr-Asp inhibition of glyceraldehyde 3-phosphate dehydrogenase affects plant redox metabolism. *EMBO J.* **40**, 1–16 (2021).

56. Thirumalaikumar, V. P., Wagner, M., Balazadeh, S. & Skirycz, A. Autophagy is responsible for the accumulation of proteogenic dipeptides in response to heat stress in Arabidopsis thaliana. *FEBS J.* **288**, 281–292 (2021).

57. Yang, L. *et al.* Deciphering drought-induced metabolic responses and regulation in developing maize kernels. *Plant Biotechnol. J.* **16**, 1616–1628 (2018).

58. Strehmel, N. *et al.* Stress-related mitogen-activated protein kinases stimulate the accumulation of small molecules and proteins in Arabidopsis thaliana root exudates. *Front. Plant Sci.* **8**, 1–13 (2017).

59. Williams, A. *et al.* Metabolomic shifts associated with heat stress in coral holobionts. *Sci. Adv.* **7**, (2021).

60. Naka, K. *et al.* Dipeptide species regulate p38MAPK-Smad3 signalling to maintain chronic myelogenous leukaemia stem cells. *Nat. Commun.* **6**, (2015).

61. Chaudhri, V. K. *et al.* Metabolic alterations in lung cancer-associated fibroblasts correlated with increased glycolytic metabolism of the tumor. *Mol Cancer Res* **23**, 1–7 (2014).

62. Calderan-Rodrigues, M. J. *et al.* Proteogenic Dipeptides Are Characterized by Diel Fluctuations and Target of Rapamycin Complex-Signaling Dependency in the Model Plant Arabidopsis thaliana. *Front. Plant Sci.* **12**, 1–15 (2021).

63. Coa, E. Y., Horan, K., Backman, T. & Girke, T. ChemmineR. v. 3.14. 2022 (2022). doi:10.18129/B9.bioc.ChemmineR

64. O'Connell, J. D. *et al.* A proteomic survey of widespread protein aggregation in yeast. *Mol. Biosyst.* **10**, 851–861 (2014).

65. Wagner, M., Zhang, B., Tauffenberger, A., Schroeder, F. C. & Skirycz, A. Experimental methods for dissecting the terraincognita of protein-metabolite interactomes. *Curr. Opin. Syst. Biol.* **28**, 100403 (2021).

66. Toney, M. D. Reaction specificity in pyridoxal phosphate enzymes. *Arch. Biochem. Biophys.* **433**, 279–287 (2005).

67. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1407–1418 (2011).

68. Alvers, A. L. *et al.* Autophagy and amino acid homeostasis are required for chronological longevity in Saccharomyces cerevisiae. *Aging Cell* **8**, 353–369 (2009).

69. Xia, Z. *et al.* Substrate-binding sites of UBR1, the ubiquitin ligase of the N-end rule pathway. *J. Biol. Chem.* **283**, 24011–24028 (2008).

70. Rappsilber, J., Ishihama, Y. & Mann, M. Stop And Go Extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).

71. Giavalisco, P. *et al.* Elemental formula annotation of polar and lipophilic metabolites using 13C, 15N and 34S isotope labelling, in combination with high-resolution mass spectrometry. *Plant J.* **68**, 364–376 (2011).

72. Chen, H. VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. (2018).

73. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. *Omi. A J. Integr. Biol.* **16**, 284–287 (2012).

74. Blighe, K., Rana, S. & Lewi, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. R package version 1.8.0. (2022).

75. Ward, J. H. Hierarchical Grouping to Optimize an Objective Function. *J. Am. Stat. Assoc.* **58**, 236 (1963).

6. General Discussion

Parts of this chapter will be published in a review by Schlossarek and Skirycz (2021)

In the presented work, we have focused on studying dynamic protein-metabolite-complexes (PMC) between different metabolic states. To do so, we used a co-fractionation mass-spectrometry (CF-MS) based method previously developed in our group, dubbed PROtein-Metabolite Interactions using Size separation (PROMIS) [1]. PROMIS combines the chromatographic separation of native PMCs using size exclusion chromatography with untargeted proteomic and metabolomic analysis of the resulting fractions. The method was originally developed to capture protein-metabolite-interactions (PMI) in the model plant *Arabidopsis thaliana* [1]. We therefore first adapted and applied PROMIS to the budding yeast *Saccharomyces cerevisiae*, a fast growing and well-established model organism [2], and found 74 metabolites co-fractionating with 3982 proteins (chapter 3). Form this, we provided experimental evidence for 225 previously predicted interactions, as well as novel interactions, for example between xanthine and purine nucleoside phosphorylase (Pnp1). Using orthogonal approaches, we could further show that xanthine has a regulatory effect on PnP1 *in vitro*, reducing its enzymatic activity, proving that PROMIS is suitable to study PMIs in the model organism *S. cerevisiae*.

We then asked if we could use the established method to capture changes in the protein-protein and protein-metabolite interactome between different cellular and metabolic conditions, namely across the diauxic shift. The diauxic shift comprises the switch from the glucose-utilizing, fermentative growth to the ethanol utilizing, respiratory growth in yeast, and is accompanied by drastic changes in the proteome [3] and metabolome [4], which we could partially recapitulate in our dataset (chapter 5). We used PROMIS in combination with an orthogonal approach, the isothermal shift assay (ItSA [5]) to study changes in the protein-protein and protein-metabolite interactions, and were able to find major differences in the fractionation profiles of proteins as well as protein-bound metabolites. Interestingly, in the late ethanol-utilizing phase, the identified metabolites were retained in the protein containing fractions to a greater degree than in the glucose-utilizing phase, hinting towards a more extensive PMI network during respiratory growth. Moreover, we found growth-phase dependent differences in protein-metabolite interactions for many biosynthetic pathways. For example, in the ethanol utilizing phase, the interaction between cystathionine gamma-lyase (CYS3), involved in amino acid biosynthesis, and its cofactor pyridoxal phosphate (PLP) is disrupted, presumably through a dynamic post-translational modification of the enzyme. We also found growth-phase dependent differences in the co-fractionation of specific groups of proteogenic dipeptides and enzymes involved in central carbon metabolism (see below). Moreover, by integrating the fractionation behavior and thermal stability (TS), accessed by ItSA, we found that the proteasome dissembles into its 19S and 20S subcomplexes in the late ethanol-utilizing phase, each co-fractionating with

a distinct set of metabolites. This is in line with previous studies that described a disassembly of the proteasome in nutrient limited conditions [6], and showed that the 20S core particle acts as an ubiquitin independent protease on its own [7]. Lastly, there is only a 20 % overlap between proteins differentially affected in their TS or fractionation. We therefore argue that proteins affected by both, fractionation and TS, belong to larger multi-protein complexes that noticeably alter their complex composition between glucose- and ethanol consumption, as is the case for the proteasome. Proteins affected only in TS might show changes in either their post-translational modification status [8–10] or interaction with a metabolite partner [11], or both, that do not change their oligomeric state. Proteins affected by their fractionation behavior only are interpreted to have changed oligomeric states between the growth phases. Because ItSA gives an average stability of all the oligomeric states, by default, it will mainly reflect the stability of the dominant oligomere and therefore may miss interactions that would be captured by CF-MS.

Taken together, these findings show that i) various metabolic pathways and cellular processes, such as amino acid biosynthesis and proteolysis, are differentially regulated throughout the course of the diauxic shift and that ii) PROMIS is able to capture those dynamics in protein-metabolite complexes and that iii) CF-MS and thermal stability based methods can be used as complementary approaches.

Tracking down Dipeptide Specificity

In our two studies investigating PMIs in yeast, we found that 36 out of 74 and 145 out of 274 annotated metabolites, respectively, were proteogenic dipeptides (chapters 3 & 5). In various organisms, individual dipeptides have been shown to act e.g. as a neurotransmitter [12], anxiolytic compound [13,14] or enhancers of lifespan and stress tolerance [15,16]. Previously, we found the dipeptide Tyr-Asp to co-fractionate with GAPDH [1]. Further investigation showed that Tyr-Asp inhibits the enzyme activity of GAPDH, and supplementation of Tyr-Asp increases tolerance to oxidative stress in plants by redirecting glycolytic flux towards the pentose phosphate pathway and NADPH production [16].

In this work, we first investigated true interaction partners of one representative dipeptide found in yeast, namely Ser-Leu. Based on our previous knowledge from the Tyr-Asp-GAPDH interaction [16], we focused on proteins involved in central carbon metabolism, and found phosphoglycerate kinase (PGK1) as a co-fractionating protein of Ser-Leu. Using microscale thermophoresis, we validated that PGK1 specifically binds Ser-Leu, but not serine, in line with previous ligand detected NMR studies [17]. Next, we showed that Ser-Leu increases PGK1 activity *in vitro*, and Ser-Leu supplementation diverges glycolytic flux from the TCA cycle towards purine and sphingolipid metabolism (chapter 3). This metabolic rewiring

is accompanied by a prolonged fermentative growth and delayed diauxic shift, hinting towards a physiological role of this dipeptide.

The results discussed above demonstrate that two dipeptides, Tyr-Asp and Ser-Leu, are involved in the regulation of central carbon metabolism (CCM) in plants and yeast, respectively. However, even if we only consider the 20 canonical and unmodified proteogenic amino acids, dipeptides comprise 400 chemically diverse molecules, with mostly unknown biological roles. We therefore aimed to get an impression of the specificities of dipeptides, especially regarding their tendencies to co-fractionate with proteins involved in CMM. To this end, we clustered the 400 dipeptides based on their chemical similarity, resulting in 14 groups. We then calculated the interaction rates as the percentage of observed interactions in all possible interactions between CCM-proteins and dipeptides for all three growth phases. This analysis revealed that the 145 dipeptides annotated in the dataset showed partially group- and growth phase specific co-fractionation with CCM-proteins. For instance, phenylalanine or tyrosine containing dipeptides (FY) showed a consistent co-fractionation across all three sampled time points (chapter 5, Figure 5 A). In contrast, dipeptides containing proline in their N-terminal position (P-X) show no putative interactions with CCM-proteins in the glucose-utilizing phase, but co-fractionate in the ethanol-utilizing phase. Additionally, in the late ethanol-utilizing phase, we found a stark difference of dipeptides containing a proline in either the N-terminal (P-X) or C-terminal (X-P) position: P-X dipeptides predominantly showed an elution maximum in the mid-sized fractions, while the X-P dipeptides were eluting in the high-molecular fractions, indicating interactions with protein-complexes of different sizes. Although, as discussed above, we cannot declare biologically relevant *bona-fide* interactions from CF-MS experiments alone, this analysis suggests that the binding of dipeptides to their protein partners is specific, and this specificity is driven by their chemical structure.

To further investigate dipeptide binding specificities, we plan to perform large-scale nano differential scanning fluorimetry (nanoDSF) experiments, screening recombinant proteins against a library of 400 dipeptides, 20 amino acids as well as a subset of cyclo-dipeptides. In nanoDSF experiments, binding between a protein and small molecule is determined as a shift in the melting temperature (T_m) of the protein determined by tyrosine and tryptophan fluorescence. In a preliminary experiment, performed without replication, we considered dipeptides or amino acids affecting the melting temperature by more than 1.5 K as either stabilizing or destabilizing, and therefore putatively interacting. Figure 1 depicts preliminary results for two proteins involved in central carbon metabolism, namely glyceraldehyde-3-phosphate-dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK1). GAPDH was destabilized by a wide range of dipeptides, excluding dipeptides containing proline in the N-terminal position (P-X), as well as methionine- (M) or tryptophan-containing (W and HFWY) dipeptide groups. In contrast to the P-X group,

dipeptides containing proline in the C-terminal position (X-P) had a destabilizing effect, and Pro-Pro showed the largest effect with a ΔT_m of -6.43 K. For PGK1, tyrosine containing dipeptides (Y) had the most prominent effect on thermal stability, with an average shift in melting temperature of -2.5 K, and a standard deviation of 3.78 K, which is due to the large ΔT_m of Thr-Tyr (-23.8 K). Although clustered in the same group (FY), phenylalanine-containing (F) dipeptides showed no effect on thermal stability (mean ΔT_m = -0.42 ± 0.62 K). Besides tyrosine-containing dipeptides, we observed only a few dipeptides across the different groups which affected the thermal stability, e.g. Met-Ile (M, ΔT_m = -8.92 K), Val-Lys (K, ΔT_m = - 8.31 K) or Gln-His (H, ΔT_m = -5.42 K). Although these are only preliminary results, they show clearly that dipeptides can have drastic effects on the thermal stability of proteins, and that these effects are protein- and dipeptide specific, as exemplified by X-P and P-X dipeptides, and in line with the results discussed in chapter 5. Moreover, the case of F- and Y-dipeptides elucidates how specific this interactions can be, as these dipeptides are indistinguishable by our clustering method, differing only in one OH-group. However, the experiments were conducted with only one replicate and need to be repeated in order for us to make definite conclusions. Moreover, it would be interesting to test other enzymes not involved in central carbon metabolism, as well as from different organisms to elucidate binding specificities.

In the future, we also plan to investigate the specificity of dipeptide accumulation in different model organisms. Although the production of dipeptides in the cell is still enigmatic, it has been suggested that one source is the cleavage of proteins by peptidases following autophagy [18]. If this is the case, the *in-silico* digestion of the whole proteome into random peptides of length two should give us a "base-line" of dipeptide distribution, which we can further refine by considering accurate protein-levels as well as protein-turnover rates. At least for *S. cerevisiae,* these proteomics parameters have been addressed in the past [3,19,20]. By comparing experimentally observed dipeptide levels with the theoretical patterns, we can gain insight into the specificity of dipeptide accumulation. Furthermore, combining this analysis with genetic approaches would allow us to study processes affecting dipeptide levels *in vivo*, including, hypothetically, the generation of dipeptides by peptidases and the depletion by dipeptide hydrolases, extracellular transport, and conversion into other chemical compounds.

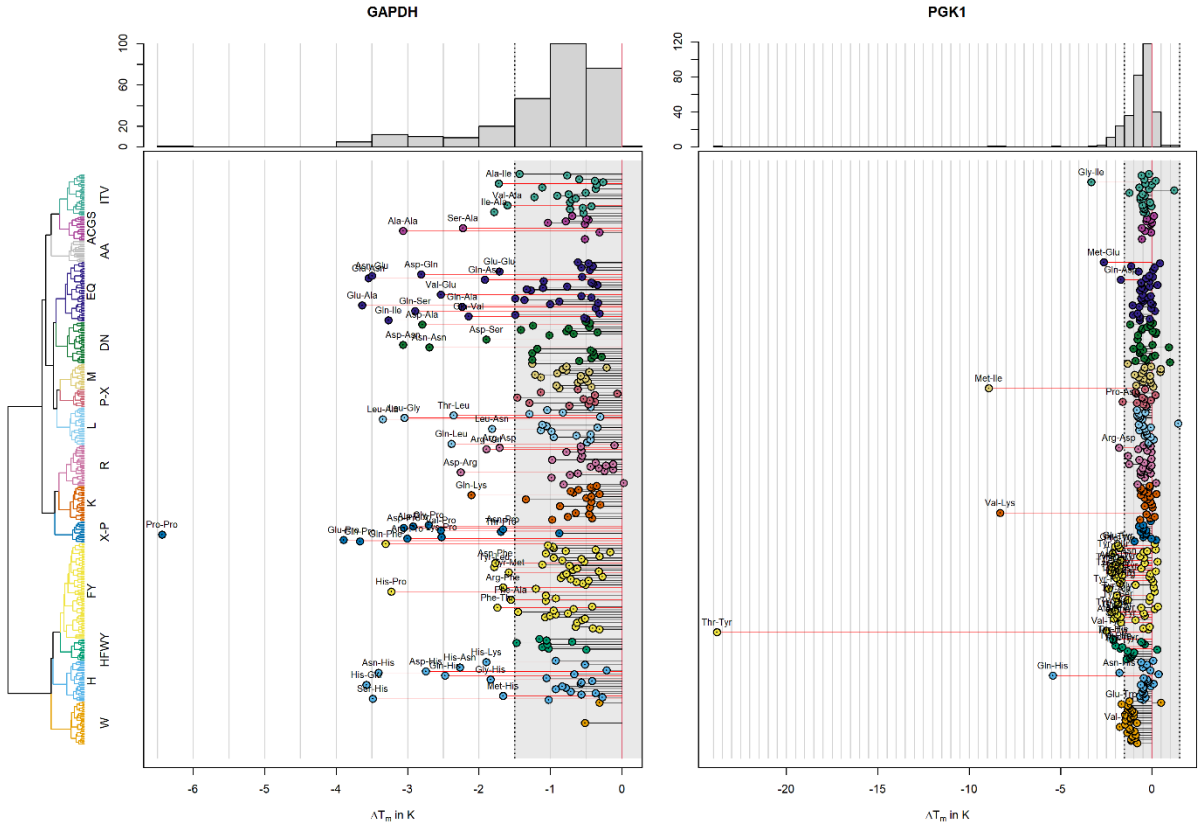


Figure 1: Shifts in protein thermal stability are dipeptide specific. The shift in melting temperature (ΔT_m in K) was calculated for every dipeptide in relation to the average T_m of the untreated protein (n = 1 for treatment, n = 7 for control). Grey area marks tempearute shift below the set threshold of 1.5 K. Dipeptides are clustered based on their chemical similarity as described in chapter 5, with addition of single amino acids. ITV: Ile-, Thr-, or Val-containing, ACGS: Ala-, Cys-, Gly- or Ser-containing, AA: Single, non-aromatic amino acids, EQ: Glu- or Gln-containing, DN: Asp- or Asn-containing, M: Met-containing, P-X: N-terminal Pro-containing, L: Leu-containing, R: Arg-containing, K: Lys-containing, X-P: C-terminal Pro-containing, FY: Phe- or Tyr-containing, HFWY: N- and C-terminal His-, Phe-, Trp- or Tyr-containing, H: His-containing, W: Trp-containing. Experiments were conducted using a Prometheus NT.48 from Nano Temper. Protein concentrations were adjusted to give optimal signals at 30 % excitation. GAPDH was in the presence of its cofactor, NAD. Dipeptide concentrations were 0.25 mg/ml.

Towards Reproducible Co-Fractionation Mass Spectrometry Experiments

Since it was first conceptualized by Liu and colleagues in 2008 [21], co-fractionation mass spectrometry has gained more and more popularity. However, although it has been successfully applied to a wide range of organisms [22–26] under various biological conditions [27] (see chapter 1), the method still has not reached maturity, and the field did not yet find a consensus about best-practices regarding data analysis and the experimental procedure itself [28,29]. For example, little agreement has been found about how many fractions should be collected to reach a sufficient chromatographic resolution, with some researches

collecting forty fractions [1,24,30] and others combining hundreds of fractions from different separation techniques [22,23,31], or which distance measure to use to address fractionation profile similarity. Consequently, it is hard to compare different datasets and make statements about e.g. putative interactions of a protein in different species, if those species were not investigated in the same study.

To alleviate this problem, two independent meta-analysis studies have come up with a set of concurrent guidelines for both, the experimental and data analysis aspects of CF-MS: Pang and colleagues evaluated novel and published yeast datasets to create a best-practice workflow [28]; and Skinnider and colleagues reanalyzed 206 published datasets, creating a coherent resource which they used for benchmarking data analysis strategies [29]. Here, we will give a brief summary of the postulated guidelines. Both studies agree that as few as 40 collected fractions yield a sufficient resolution for near-maximum recall of complexes in a cost-effective way. To further increase resolution, additional fractions should be collected either from replicate experiments, or from an orthogonal separation technique. However, the two studies do not agree on one single-best separation technique, as Skinnider favours SEC and BN-PAGE [29] whereas Pang reports highest resolution for IEX, as evidenced by higher average peak widths in the SEC dataset compared to IEX [28]. Therefore, the choice of separation technique should be made based on the specific experimental question, as e.g. BN-PAGE can be used for the separation of membrane-associated proteins, which is not possible with IEX to date. The use of stable isotope labeling using amino acids (SILAC) is not recommended, as it did not significantly increase the accuracy, but decreased the number of detected proteins [29]. Moreover, for label free approaches, IBAQ is favored over spectral counting in terms of accuracy and proteome coverage, whereas the MaxLFQ algorithm implemented in MaxQuant [32] is inferior to both. In terms of fractionation profile pre-processing, treating missing values as zeroes increased accuracy, while normalization of the profiles is not necessary [29].

The ultimate goal in CF-MS is to obtain a measure of profile similarity between the measured proteins (and metabolites), from which complexes can be inferred. Pearson correlation and Euclidian distance are both widespread metrics of profile similarity. Both studies agree that, while Pearson correlation, as well as Kendall and Spearman, are well-suited metrics, Euclidean distance alone performed no better than random chance at quantifying similarity between fractionation profiles [28,29]. However, when using machine learning (ML) approaches, the Pearson correlation and Euclidian distance show synergistic effects [29]. The best performing metrics – mutual information, distance correlation, cosine distance and weighted cross-correlation – all showed comparable recovery of known protein complexes [29]. For using machine learning, Skinnider describes two distinct regimes: For the low-data regime, simple linear classifiers perform best, but are sensitive to the input data, whereas in the data-rich regime, non-linear classifiers, such as random forest, perform best and are insensitive to input data [29]. Moreover, in general,

the accuracy is higher when features are calculated separately for each replicate instead of combining replicates beforehand. Lastly, predicting protein complexes from CF-MS experiments alone is sufficient, and the integration of additional genomic data, such as co-expression or high confidence PPIs, does not contribute to predicting novel complexes [28].

In this work, we presented PROMISed, a web-based tool for the analysis of CF-MS datasets based on profile deconvolution and correlation of the resulting fractionation peaks. PROMISed gives the user control over a set of important data-analysis parameters, and is in good agreement with the guidelines discussed above for "conventional" CF-MS data analysis. For example, PROMISed allows the user to choose between three correlation methods – Pearson, Kendall's, and Spearman – all of which showed to be well suited to address profile similarity. Moreover, PROMISed employs a range of preprocessing steps, such as normalization and deconvolution, which are necessary for the underlying workflow of PROMISed. Whereas not essential for data analysis according to the guidelines, these steps also do not diminish accuracy of resulting profile similarities. Lastly, besides PROMISed, multiple data-analysis pipelines employing machine learning, such as SLIMP [33], have been presented. With EPIC [34] and PRINCE [35,36], two ML workflows are available as software tools or packages. ML based approaches can be powerful to predict known and putative protein-complexes in well established model organisms, but can fall short when analyzing data of less known organisms. The big strength of PROMISed is the identification of co-fractionating molecules, irrespective of proteins or metabolites, solely relying on their fractionation profiles, without prior knowledge. For the future, we anticipate that the guidelines formulated by Pang et al and Skinnider et al will have a great impact on how CF-MS experiments are performed and analyzed, and hope that ML based data analysis tools will be made broadly accessible with experimental biologists in mind.

Outlook

With the methodological advances made to investigate protein-protein and protein-metabolite interactions in the past decade, using e.g. CF-MS and TPP/CETSA, we are now able to routinely chart large scale interaction networks of various organisms across biological conditions. However, irrespective of a potentially large fraction of false positive interactions, those networks remain highly incomplete for mainly two reasons: First, while the recent methods are well suited to identify interactions for soluble proteins, charting interactions within the membrane associated proteome is still a challenging task. Second, in contrast to the proteome, the characterization of the metabolome remains highly insufficient even in modern model organisms, which greatly constrains any attempt to deduce PMI networks. We expect that these technical limitations will be addressed in the near future, e.g. by employing more sophisticated separation techniques, applying protein-sequencing methods [37–39] to quantify the proteome, or by advances in

characterizing the *terra-incognita* of the metabolome [40]. Until these necessary advances make their way into the laboratories, there are, however, computational approaches to improve network coverage available today. For one, putative protein-metabolite interactions can be inferred solely from the protein sequences using deep learning [41,42], or missing interactions in existing protein-metabolite networks can be complemented using biclique extension based prediction [43].

Finally, the ultimate goal of interactomics should be to decipher the biological roles of the postulated interactions, a task that can not be fulfilled by one research group alone. Therefore, we expect that the incorporation of interactions derived from CF-MS studies into publicly available databases, such as STRING [44], STITCH [45] or CORUM [46], will pave the way for their functional characterizations, as those databases are used by computational and experimental biologists alike.

References – General Discussion

1. Veyel, D. *et al.* PROMIS, global analysis of PROtein-metabolite interactions using size separation in Arabidopsis thaliana. *J. Biol. Chem.* **293**, 12440–12453 (2018).
2. Nielsen, J. Yeast Systems Biology: Model Organism and Cell Factory. *Biotechnol. J.* **14**, (2019).
3. Murphy, J. P., Stepanova, E., Everley, R. A., Paulo, J. A. & Gygi, S. P. Comprehensive temporal protein dynamics during the diauxic shift in Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **14**, 2454–2465 (2015).
4. Zampar, G. G. *et al.* Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. *Mol. Syst. Biol.* **9**, (2013).
5. Ball, K. A. *et al.* An isothermal shift assay for proteome scale drug-target identification. *Commun. Biol.* **3**, 1–10 (2020).
6. Bajorek, M., Finley, D. & Glickman, M. H. Proteasome Disassembly and Downregulation Is Correalted with Viabolity during Stationary Phase. *Curr. Biol.* **13**, 1140–1144 (2003).
7. Sahu, I. *et al.* The 20S as a stand-alone proteasome in cells can degrade the ubiquitin tag. *Nat. Commun.* **12**, (2021).
8. Huang, J. X. *et al.* High throughput discovery of functional protein modifications by Hotspot Thermal Profiling. *Nat. Methods* **16**, 894–901 (2019).
9. Smith, I. R. *et al.* Erratum to: High throughput discovery of functional protein modifications by Hotspot Thermal Profiling (Nature Methods, (2019), 16, 9, (894-901), 10.1038/s41592-019-0499-3). *Nat. Methods* **18**, 760–762 (2021).
10. Potel, C. M. *et al.* Impact of phosphorylation on thermal stability of proteins. *Nat. Methods* **18**, 757–759 (2021).
11. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science (80-.).* **346**, (2014).
12. TAKAGI, H., SHIOMI, H., UEDA, H. & AMANO, H. MORPHINE-LIKE ANALGESIA BY A NEW DIPEPTIDE, L-TYROSYL-L-ARGININE (KYOTORPHIN) AND ITS ANALOGUE. *Eur. J. Pharmacol.* **55**, 109–111 (1979).

13. Mizushige, T. *et al.* Aromatic amino acid-leucine dipeptides exhibit anxiolytic-like activity in young mice. *Neurosci. Lett.* **543**, 126–129 (2013).

14. Kanegawa, N., Suzuki, C. & Ohinata, K. Dipeptide Tyr-Leu (YL) exhibits anxiolytic-like activity after oral administration via activating serotonin 5-HT1A, dopamine D1 and GABAA receptors in mice. *FEBS Lett.* **584**, 599–604 (2010).

15. Zhang, Z. *et al.* The novel dipeptide Tyr-Ala (TA) significantly enhances the lifespan and healthspan of: Caenorhabditis elegans. *Food Funct.* **7**, 1975–1984 (2016).

16. Moreno, J. C. *et al.* Tyr-Asp inhibition of glyceraldehyde 3-phosphate dehydrogenase affects plant redox metabolism. *EMBO J.* **40**, 1–16 (2021).

17. Diether, M., Nikolaev, Y., Allain, F. H. & Sauer, U. Systematic mapping of protein-metabolite interactions in central metabolism of Escherichia coli . *Mol. Syst. Biol.* **15**, 1–16 (2019).

18. Thirumalaikumar, V. P., Wagner, M., Balazadeh, S. & Skirycz, A. Autophagy is responsible for the accumulation of proteogenic dipeptides in response to heat stress in Arabidopsis thaliana. *FEBS J.* **288**, 281–292 (2021).

19. Christiano, R., Nagaraj, N., Fröhlich, F. & Walther, T. C. Global Proteome Turnover Analyses of the Yeasts S.cerevisiae and S.pombe. *Cell Rep.* **9**, 1959–1965 (2014).

20. Martin-Perez, M. & Villén, J. Determinants and Regulation of Protein Turnover in Yeast. *Cell Syst.* **5**, 283-294.e5 (2017).

21. Liu, X., Yang, W. chu, Gao, Q. & Regnier, F. Toward chromatographic analysis of interacting protein networks. *J. Chromatogr. A* **1178**, 24–32 (2008).

22. Wan, C. *et al.* Panorama of ancient metazoan macromolecular complexes. *Nature* **525**, 339–344 (2015).

23. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).

24. Luzarowski, M. *et al.* Global mapping of protein–metabolite interactions in Saccharomyces cerevisiae reveals that Ser-Leu dipeptide regulates phosphoglycerate kinase activity. *Commun. Biol.* **4**, 1–15 (2021).

25. McWhite, C. D. *et al.* A pan-plant protein complex map reveals deep conservation and novel assemblies. *Cell* **181**, P460-474.E14 (2020).

26. Xu, C. *et al.* Global Landscape of Native Protein Complexes in Synechocystis sp. PCC 6803. *Genomics. Proteomics Bioinformatics* (2021) doi:10.1016/j.gpb.2020.06.020.

27. Heusel, M. *et al.* A Global Screen for Assembly State Changes of the Mitotic Proteome by SEC-SWATH-MS. *Cell Syst.* **10**, 133-155.e6 (2020).

28. Pang, C. N. I. *et al.* Analytical Guidelines for co-fractionation Mass Spectrometry Obtained through Global Profiling of Gold Standard Saccharomyces cerevisiae Protein Complexes. *Mol. Cell. Proteomics* **19**, 1876–1895 (2020).

29. Skinnider, M. A. & Foster, L. J. Meta-analysis defines principles for the design and analysis of co-fractionation mass spectrometry experiments. *Nat. Methods* **18**, 806–815 (2021).

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

31. Shatsky, M. *et al.* Quantitative tagless copurification: A method to validate and identify protein-protein interactions. *Mol. Cell. Proteomics* **15**, 2186–2202 (2016).

32. Tyanova, S., Temu, T. & Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319 (2016).

33. Zühlke, B. M. *et al.* Authors/Affiliations. *bioRxiv* (2021) doi:https://doi.org/10.1101/2021.06.16.448636.

34. Hu, L. Z. M. *et al.* EPIC: software toolkit for elution profile-based inference of protein complexes. *Nat. Methods* **16**, 737–742 (2019).

35. Stacey, R. G., Skinnider, M. A., Scott, N. E. & Foster, L. J. A rapid and accurate approach for prediction of interactomes from co-elution data (PrInCE). *BMC Bioinformatics* **18**, 1–14 (2017).

36. Skinnider, M. A., Cai, C., Stacey, R. G. & Foster, L. J. PrInCE: An R/Bioconductor package for protein-protein interaction network inference from co-fractionation mass spectrometry data. *Bioinformatics* **37**, 2775–2777 (2021).

37. Restrepo-Pérez, L., Joo, C. & Dekker, C. Paving the way to single-molecule protein sequencing. *Nat. Nanotechnol.* **13**, 786–796 (2018).

38. Alfaro, J. A. *et al.* The emerging landscape of single-molecule protein sequencing technologies. *Nat. Methods* **18**, 604–617 (2021).

39. Brady, M. M. & Meyer, A. S. Cataloguing the proteome: Current developments in single-molecule protein sequencing. *Biophys. Rev.* **3**, (2022).

40. Wagner, M., Zhang, B., Tauffenberger, A., Schroeder, F. C. & Skirycz, A. Experimental methods for dissecting the terraincognita of protein-metabolite interactomes. *Curr. Opin. Syst. Biol.* **28**, 100403 (2021).

41. Sun, T., Zhou, B., Lai, L. & Pei, J. Sequence-based prediction of protein protein interaction using a deep-learning algorithm. *BMC Bioinformatics* **18**, 1–8 (2017).

42. Wang, L., Wang, H. F., Liu, S. R., Yan, X. & Song, K. J. Predicting Protein-Protein Interactions from Matrix-Based Protein Sequence Using Convolution Neural Network and Feature-Selective Rotation Forest. *Sci. Rep.* **9**, 1–12 (2019).

43. Thieme, S. & Walther, D. Biclique extension as an effective approach to identify missing links in metabolic compound–protein interaction networks. *Bioinforma. Adv.* **2**, 1–15 (2022).

44. Szklarczyk, D. *et al.* The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).

45. Szklarczyk, D. *et al.* STITCH 5: Augmenting protein-chemical interaction networks with tissue and affinity data. *Nucleic Acids Res.* **44**, D380–D384 (2016).

46. Giurgiu, M. *et al.* CORUM: The comprehensive resource of mammalian protein complexes - 2019. *Nucleic Acids Res.* **47**, D559–D563 (2019).

Acknowledgments

I would like to thank Aleksandra Skirycz and Lothar Willmitzer for their excellent supervision and their encouragement to develop my own scientific interests. I would also like to thank the members of my PhD advisory committee, namely Camila Caldana, Friedrich Kragler, Zoran Nikoloski and Alisdair Fernie for their scientific discussions and valuable inputs. I am also grateful to Alisdair Fernie for welcoming me in his group seminars and activities when our office became increasingly empty. A very special thanks to Marcin Luzarowski for introducing me to data analysis and programming, sharing his data and the great co-supervision, and Ewelina Sokolowska, for measuring hundreds of proteomic samples, the great scientific and non-scientific discussions and critical trouble-shooting. I am also very grateful to Änne Michaelis for LC-MS measurements, and Andreas Donath and the IT-Team for hosting the Shiny-apps and their great service. Many thanks also to Dariusz Bieńkowski for the help with Shiny and Docker, great discussions and hobby time. I am also very grateful to Jennifer Ewald and Deniz Irvali for hosting me in Tübingen and their help and advice on yeast biology, together with the Yeast Club. I would also like to thank Ina Talke for her support, the IMPRS for funding and all IMPRS students for their critical input and discussion. Many thanks also to all former and current members of or group and friends at the institute, especially Marcin and Ula of House Storchenhof, Ewka, Anne, Ayesha, Monika, Juan, Romina, Venka, Mateusz, Ewa, Nastia, Mona, Damoun, Weronika, Änne and Uta, and our former students, especially Dominika, whose work unfortunately didn't make it into this thesis. I am also very grateful for the assistance and encouragement I got from AG Hiltbrunner to apply for this PhD position in the first place.

I would like to thank my friends, first and foremost Sascha for his support and welcome distractions, the "Pen&Paper chicas" Lena, Caro, Alex and Leo, and the "DSA Internationals" Sissy, Max, Markus and Konstantin, as well as Richard for the pandemic online music festivals.

Finally, I would like to thank my family. I would have never come this far without the continuous support from my parents and my brothers Tim and Sebastian.