TUDelft

Delft University of Technology

Assessing yeast proteome dynamics using high-resolution mass spectrometry

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Assessing yeast proteome dynamics using high-resolution mass spectrometry

MAXIME DEN RIDDER

Assessing yeast proteome dynamics using high-resolution mass spectrometry

Dissertation

For the purpose of obtaining the degree of doctor

at Delft University of Technology

by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen,

chair of the Board for Doctorates,

to be defended publicly on Friday 12 May 2023 at 12.30 o'clock

by

Maxime Jolijn DEN RIDDER

Master of Science in Life Science and Technology, Delft University of Technology, the Netherlands

Born in Zevenbergen, the Netherlands

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus	Chairperson
Prof. dr. ir. P.A.S. Daran-Lapujade	Delft University of Technology, promotor
Dr. M. Pabst	Delft University of Technology, copromotor

Independent members:

Prof. dr. S. Hubbard	University of Manchester, UK
Dr. J.A.A. Demmers	Erasmus University
Prof. dr. P. Branduardi	University of Milano-Bicocca, Italy
Prof. dr. C. Joo	Delft University of Technology
Prof. dr. B.M. Bakker	University of Groningen
Prof. dr. F. Hollmann	Delft University of Technology, reserve member

TUDelft

Contents

Samenvatting

Op massa spectrometrie-gebaseerde cellulaire proteomics heeft een prominente rol gespeeld in veel onderzoeksgebieden, waaronder biowetenschappen, biotechnologie en microbiële ecologie. Eiwitten zijn de cellulaire bouwstenen die de potentiële functie van genen rechtstreeks in stand houden door enzymatische katalyse, moleculaire signalering en fysieke interacties en daarom kan de functie niet rechtstreeks worden bepaald aan de hand van het genoom. Bovendien is het proteoom van een cel veel complexer dan het genoom, als gevolg van onder andere posttranslationele modificaties. Momenteel is massa spectrometrie de krachtigste techniek om de grote proteoomdiversiteit in cellen te onderzoeken, vaak toegepast in een shotgun (bottom-up) benadering. Toch worden geavanceerde, op massaspectrometrie-gebaseerde proteomics methoden, niet routinematig toegepast op microben. De celfabriek en het modelorganisme *Saccharomyces cerevisiae,* bijvoorbeeld, is een zeer goed gekarakteriseerd micro-organisme. Echter, veel onderzoeksvragen over de dynamiek van het proteoom onder verschillende groeiomstandigheden en de regulering van het complexe metabole netwerk via post-translationele modificaties, moeten nog worden beantwoord. In het bijzonder ontbreken momenteel kwantitatieve gegevens over proteoomdynamiek onder goed gedefinieerde omstandigheden, zoals in bioreactoren. Een beter begrip zal de ontwikkeling van geavanceerde industriële giststammen mogelijk maken en het gebruik van bakkersgist als modelorganisme verbeteren, bijvoorbeeld tijdens het analyseren van stofwisselingsziekten bij de mens. Daarom was het doel van dit proefschrift om een geoptimaliseerd monstervoorbereidingsprotocol op te stellen om de grootschalige kwantitatieve analyse van het gist proteoom mogelijk te maken onder zeer gecontroleerde omstandigheden. Daarnaast werd een nieuwe massaspectrometrische methode ontwikkeld om de mate van modificatie van metabole enzymen te kwantificeren, wat een beter begrip van hun rol in metabole regulatie mogelijk maakt.

In **Hoofdstuk 2** wordt het belang van post-translationele modificaties in proteomics-experimenten beschreven. Deze extra regulatielaag in het proteoom kan alleen worden gevisualiseerd door middel van gevoelige proteomics-analyses, aangezien deze modificaties niet in het genoom zijn gecodeerd. Over het algemeen blijft een groot aantal fragmentatiespectra in een typisch shotgun proteomics-experiment ongeïdentificeerd. Geschat wordt dat een groot deel van die niet-geïdentificeerde spectra afkomstig is van onverwachte post-translationele modificaties of natuurlijke peptidevarianten. Recente vorderingen in resolutie, nauwkeurigheid en kwantitatieve eigenschappen in massaspectrometrie, evenals ontwikkelingen in dataverwerking, maken nieuwe manieren mogelijk om eiwitmodificaties te onderzoeken. Waar de meeste gemodificeerde peptiden

gemakkelijk kunnen worden gedetecteerd door de veel-gebruikte methode, leidt het gebruik van database zoekalgortithmen vaak niet tot een positieve identificatie. Om de beperkingen van database-gelimiteerde methoden te overwinnen, zijn alternatieve algoritmen vastgesteld. Deze zoekhulpmiddelen voor open modificaties vereisen geen vaststelling van de modificatie vóór de analyse en kunnen daarom ook onverwachte modificaties identificeren in experimenten. In **Hoofdstuk 2** worden deze recente vorderingen in microbiële (gist) proteomics voor onbeperkte ontdekking van eiwitmodificatie besproken.

Bij gist proteomics wordt doorgaans een shotgun-benadering toegepast, waarbij eiwitten worden omgezet naar peptiden voorafgaand aan de LC-MS/MS-analyse. Monstervoorbereiding kan een van de belangrijkste bijdragen zijn aan datavariatie en slechte vergelijkbaarheid tussen proteomics-experimenten. Daarom wordt de invloed van een reeks monstervoorbereidingsprocedures op de proteomics-uitkomst beschreven in **Hoofdstuk 3**. Hier hebben we een systematische vergelijking van monstervoorbereidingsprotocollen uitgevoerd met behulp van een matrix van verschillende omstandigheden en reagentia die doorgaans worden toegepast op het lysaat van hele (gist) cellen, in bottom-up proteomics-experimenten. De protocollen werden geëvalueerd op verschillende criteria: de fractie van geïdentificeerde spectra, de proteoom- en aminozuursequentiedekking, de GO-term distributie en het aantal en de diversiteit van peptidemodificaties. De beste protocollen maakten de identificatie van ongeveer 65-70% van alle verkregen fragmentatiespectra mogelijk door een combinatie van database- en onbeperkte zoekmethoden voor modificaties te gebruiken. Een groot deel van de niet-geïdentificeerde spectra bestond uit een te lage kwaliteit om met vertrouwen te worden geïdentificeerd met behulp van aanvullende *de novo* sequencing. Interessant is dat een aantal onverwachte peptidemodificaties kunnen worden gekoppeld aan oplosmiddelen, additieven en andere veel gebruikte materialen. De verzamelde protocollen en grote sets onbewerkte massaspectrometrische data bieden een hulpmiddel om nieuwe protocollen te evalueren en te ontwerpen en de analyse van (natuurlijke) peptidemodificaties in gist te begeleiden.

Hoewel *S. cerevisiae* een van de meest gekarakteriseerde microben is, is er tot nu toe geen volledig begrip van de regulering van het metabolisme. Bovendien maakt de complexiteit ervan het moeilijk om biosynthetische routes voor de productie van chemicaliën te ontwikkelen en te optimaliseren. Recente studies hebben het potentieel van het gebruik van proteome toedeling aangetoond om voorspellingsmodellen voor metabolische processen te verbeteren, waarvoor zeer kwantitatieve en nauwkeurige proteoomgegevens nodig zijn. Daarom werden grootschalige kwantitatieve proteomics-experimenten van *S. cerevisiae* uitgevoerd om de reactie van *S. cerevisiae* op aerobe en anaerobe batchculturen te volgen in **Hoofdstuk 4**. Anaerobe omstandigheden lieten een aanzienlijk lagere eiwitdynamiek zien tijdens de overgang van proliferatie naar stationaire fase, wat te wijten is aan het gebrek aan een diauxische verschuiving in afwezigheid van zuurstof. Bovendien maakte de vergelijking met de giststam met minimale glycolyse, waarin alle overtollige minor iso-enzymen zijn

verwijderd, het mogelijk om de impact van genetische overtolligheid in gistglycolyse te onderzoeken. Hoewel er geen significante fysiologische reacties werden waargenomen bij het verwijderen van de kleine glycolytische paralogen, werd een aantal veranderingen op eiwitniveau waargenomen in de glycolyse van de gemuteerde stam, met name onder anaerobe omstandigheden.

Ten slotte, zijn er tot nu toe meer dan honderd verschillende soorten posttranslationele modificaties geïdentificeerd in alle domeinen van het leven. Desalniettemin is het merendeel van de modificaties zeldzaam en is hun biologische rol niet geïdentificeerd, terwijl deze de eiwitfunctie aanzienlijk kunnen beïnvloeden. Bovendien is crosstalk tussen modificaties aangetoond voor verschillende enzymen in gist, wat het belang benadrukt van het analyseren van het volledige modificatielandschap van een eiwit. Eiwitsequenties worden over het algemeen geanalyseerd via tryptische peptiden en daarom worden minder toegankelijke sequentiegebieden over het hoofd gezien. In **Hoofdstuk 5** wordt een proof-of-concept beschreven van een methode die het mogelijk maakt om de globale mate van modificatie te kwantificeren voor individuele enzymen van complexe cellysaten. Hier wordt de niet-gemodificeerde fractie van elk peptide gekwantificeerd waardoor alle soorten modificaties worden overwogen, ongeacht hun chemische aard. Kwantificering van de volledige eiwitsequentie wordt bereikt met behulp van een eiwitstandaard gegenereerd met celvrije synthese, in combinatie met een multi-proteasebenadering. Bovendien maakte het gebruik van TMT-labeling een high-throughput, gemultiplexte methode mogelijk. De aanpak wordt gedemonstreerd voor het glycolytische enzym Pyk1 in minimale glycolyse *S. cerevisiae*, waarvoor de globale eiwitmodificatieveranderingen werden gevolgd tijdens de overgang van proliferatie naar stationaire fase onder aërobe omstandigheden. Interessant genoeg vertoonden de meeste sequentiegebieden significante veranderingen op ten minste één punt tijdens de groeicurve. Verdere ontwikkeling van deze aanpak zou het mogelijk kunnen maken om het modificatie landschap van volledige metabole routes te volgen. Dergelijke experimenten maken het mogelijk om de rol van post-translationele modificaties in de regulatie van metabole routes in gist en daarbuiten te onderzoeken.

Summary

Mass spectrometry-based cellular proteomics has taken a prominent role in many fields of research, including life sciences, biotechnology and microbial ecology. Proteins are the cellular building blocks that directly maintain the potential function of genes through enzymatic catalysis, molecular signalling and physical interactions and therefore function cannot be directly determined from the genome. Furthermore, the proteome of a cell is far more complex than its genome, due to processes including post-translational modifications. Currently, mass spectrometry is the most powerful technique to investigate the large proteome diversity present in cells, commonly applied in a shotgun (bottom-up) approach. Still, advanced mass spectrometry-based proteomics methods are not routinely applied to microbes. The cell factory and model organism *Saccharomyces cerevisiae*, for example, is very well characterized, however, many research questions surrounding proteome dynamics under different growth conditions, and the regulation of its complex metabolic network via post-translational modifications, remain to be answered. In particular, quantitative proteome dynamics data under well-defined conditions (in bioreactors) are currently lacking. A more comprehensive understanding will enable development of advanced industrial strains and improve its use as model organism when *e.g.* analysing metabolic diseases in humans. Therefore, the aim of this thesis was to establish and apply optimized protocols to enable the large-scale quantitative analysis of proteome dynamics under highly controlled conditions. In addition, a novel mass spectrometric approach was established to quantify the degree of modification of metabolic enzymes, that allows for a better understanding of their role in metabolic regulation.

In **Chapter 2** the importance of post-translational modifications in proteomics experiments is described. This additional layer of regulation in the proteome can only be visualized through sensitive proteomics analyses, as these modifications are not encoded in the genome. Generally, in shotgun proteomics, many fragmentation spectra remain unidentified. A good fraction of those unidentified spectra likely originates from unexpected protein modifications or natural sequence variants. Recent advances in resolution, accuracy and quantitative properties in mass spectrometry, as well as developments in data processing enable new ways to explore protein modifications. Alternative algorithms (such as open modification search approaches) have been established over the past years, to overcome the limitations of database-restricted approaches. In **Chapter 2**, these recent methodological advancements enabling unrestricted protein modification searchers are reviewed and discussed.

In yeast proteomics, a shotgun approach is typically applied in which proteins are digested into peptides prior to LC-MS/MS analysis. Sample preparation protocols are a common source of biases in proteomics experiments. Therefore, the influence of a range of sample preparation procedures on the proteomics outcome is described in **Chapter 3**. Thereby, we performed a comparison of different sample preparation protocols which are commonly employed in whole cell lysate proteomics experiments. The outcomes (after applying the different protocols) were evaluated for several criteria: the fraction of identified spectra, the proteome and amino acid sequence coverage, the GO-term distribution and the number and diversity of peptide modifications. The best performing protocols allowed to identify approx. 65–70% of all fragmentation spectra. A large fraction of the unidentified spectra consisted of too low quality to be confidently identified using additional *de novo* sequencing. Interestingly, a number of unexpected peptide modifications could be linked to solvents, additives and other routine materials. The established protocols and large sets of mass spectrometric raw data provide a resource optimise yeast proteome studies and to validate (native) peptide modifications.

Even though *S. cerevisiae* is one of the most characterized microbes, a full understanding of the regulation of its metabolism has not been achieved thus far. Furthermore, its complexity makes it difficult to engineer and optimize biosynthetic routes for production of chemicals. Recent studies demonstrated the potential of proteome allocation data to improve prediction models for metabolic processes, that requires highly quantitative and accurate proteome data. Therefore, large-scale quantitative proteomics experiments were performed to monitor the proteome dynamics in aerobic and anaerobic batch cultures, in **Chapter 4**. Anaerobic conditions showed substantially lower protein dynamics during the transition from proliferation to stationary phase, accountable to the lack of diauxic shift in the absence of oxygen. Furthermore, comparison with the minimal glycolysis yeast strain, in which all redundant minor isoenzymes are removed, allowed to investigate the impact of genetic redundancy in yeast glycolysis. While no significant physiological responses were observed when deleting the minor glycolytic paralogs, a number of protein-level alterations was observed in glycolysis of the mutant strain in particular under anaerobic conditions.

Finally, more than hundred different types of post-translational modifications have been identified to date across all domains of life. Nevertheless, the majority of modifications is rare and their biological roles have not been identified, while these can significantly affect the protein properties and localisation. Furthermore, crosstalk between modifications has been demonstrated for various enzymes in yeast, highlighting the importance of analysing the full modification landscape on a protein. Protein sequences are commonly analysed via tryptic peptides and less accessible sequence regions are therefore overlooked. In **Chapter 5**, a proof-of-concept is described for an approach that enables to quantify the global degree of modification for individual enzymes from complex cell lysates. Here, the unmodified

fraction of each peptide is quantified, and therefore all types of modifications regardless their chemical nature are considered. Quantification of the full protein sequence is accomplished with the help of a protein standard generated with cell free synthesis, in combination with a multi-protease approach. Moreover, the use of TMT labelling allowed for a high-throughput, multiplexed method. The approach is demonstrated for the glycolytic enzyme Pyk1 in minimal glycolysis *S. cerevisiae*, for which the global protein modification changes were monitored during transition from proliferation to stationary phase under aerobic conditions. Interestingly, the majority of sequence regions showed significant changes at least at one point during the growth curve. Further development of this approach could allow to monitor complete metabolic pathways. Such experiments allow to explore the role of post-translational modifications in the regulation of metabolic pathways in yeast and beyond.

Chapter 1

General introduction

Over the past decades, mass spectrometry-based cellular proteomics emerged as key technology in many fields of research, such as life sciences, biotechnology and microbial ecology [1–4]. In contrast to genomics approaches, proteomics targets the final gene products, namely the proteins [5]. The proteins translate the catalytic and metabolic functions, enable cellular signalling and regulation, and have structural roles [3]. However, the proteome of a cell can be magnitudes of orders more complex than its genome, due to processes such as alternative RNA splicing, amino acid variation, post-translational -modification and -cleavage. Therefore, mass spectrometry is currently (and likely also in the foreseeable future) the most powerful approach to study the large proteome diversity present in cells. Interestingly, the advanced application of mass spectrometry-based proteomics to microbes and their complex ecosystems is still far from routine. This holds true for microbes such the model organism and cell factory yeast. For example, there are many open questions surrounding the proteome dynamics under different growth conditions, and the regulation of its complex metabolic network, or more specifically, the involvement of post-translational modifications. Providing a deeper understanding to these questions will support the development of advanced strains for industry, and enhance its use as model organism, *e.g.* when studying metabolic diseases such as cancer.

From the discovery of the electron to large-scale proteomics

Technically, mass spectrometers measure the mass-to-charge ratio of ions, which strictly involves ionisation of the peptide or protein, separation of the ions and finally measurement of the mass-to-charge ratio. Nevertheless, it took more than 75 years of development until mass spectrometers found its first large-scale application in life sciences. The origin of mass spectrometry is attributed to the physicist J.J. Thomson, towards the end of the 19th century. Thomson experimented with electric fields inside a cathode ray tube, which ultimately led to the discovery of the electron that brought him the Nobel Prize in Physics in 1906 [6, 7]. Subsequently, the mass of charged atoms, and several stable isotopes could be determined for the first time [8, 9]. Decades later, A. Nier developed a mass spectrometer to obtain pure uranium-235 by mass fractionation, which helped to demonstrate that this isotope was responsible for nuclear fission, rather than uranium-238 [10]. This enabled later to build the first atomic bomb in World War II. In the late 50s and early 60s, the first small organic molecules were measured. K. Biemann, for example, was one of the first who used mass spectrometry to validate the structure of natural products [11]. In the mid-70s, the birth of the nowadays widely employed tandem-mass spectrometer was attributed to D. Hunt [12]. Mass spectrometers could now also selectively fragment molecular ions to obtain structural information. Nevertheless, it was only in the 1980s that mass spectrometry found its first applications in biology. This was attributed to the newly developed soft ionisation techniques, namely electrospray ionisation and matrix assisted laser desorption ionisation. These

techniques brought another Nobel Prize to the field of mass spectrometry, namely for K. Tanaka and J. Fenn in the year 2002 [7].

However, the vast proteome diversity of cells required additional developments before mass spectrometry could finally coin the term 'cellular proteomics'. One of these developments was the coupling to high performance separation techniques. The second was the development of bioinformatics approaches that ultimately enabled the analysis of large-scale sequencing data, *e.g.* via database searching and validation procedures such as scoring and target/decoy approaches. Furthermore, other technical advancements improved sensitivity and resolving power of mass spectrometers. For example, high-resolution mass analysers such as the time-of-flight mass analyser, the Orbitrap mass analyser and later the integration of ion mobility techniques enabled many new insights into molecular features, such as covalent protein modifications [13–17].

Over the past decades, a range of dedicated proteomics approaches have been developed. Those allow now to answer a broad spectrum of proteome-related questions, such as large-scale protein discovery using shotgun (data-dependent) and data-independent acquisition, protein quantification using targeted proteomics and labelling techniques, discovery of protein modifications using enrichment strategies, study the protein turnover and substrate utilisation using stable isotope probing, analysis of metabolite/protein interactions using thermal proteome profiling, the discovery of novel enzymes using activity-based proteomics, determining protein/protein interactions using chemical cross linking experiments, protein conformation studies using hydrogen deuterium exchange mass spectrometry, and determination of protein complex stoichiometry via native mass spectrometry [2, 18]. Today, many of these approaches are routinely employed to perform system-wide analyses of cells, body fluids and tissues. However, currently, most proteomics experiments are performed on human cells and a few model organisms. The application to microbes and their complex ecosystems is increasing but still far from routine [19].

State-of-the-art in large-scale cellular proteomics

Liquid chromatography coupled to high-resolution tandem mass spectrometry is currently the most common instrumental setup when performing cellular proteomics studies. In bottom-up proteomics, the complex protein extract, as derived from cells or tissues, is proteolytically digested into peptides prior to chromatographic separation and mass spectrometric detection [6, 20–23]. For this, the protease Trypsin is considered as gold standard in proteomics, even though many other proteases (with alternative cleavage specificities) are available. Trypsin cleaves at the carboxyl side of the amino acids lysine and arginine (except when a proline follows), generating mid-sized (8–30 AAs long) peptides with basic residue at the C-terminus. This is highly suitable for chromatographic separation,

ionisation (protonation) and sequencing using fragmentation experiments [24]. Peptides are typically separated using reverse-phase liquid chromatography to avoid co-elution and competitive ionisation and to allow the detection of as many peptides as possible by employing extended chromatographic gradients [25, 26]. Additionally, the peptide mixtures can be fractionated to maximise identifications before or in conjunction with the online chromatographic separation, *e.g.* by using strong cation exchange chromatography [27, 28]. Electrospray ionisation is the most common ionisation technique in proteomics, in which the analytes are ionized (protonated) out of solution, making it compatible with liquid chromatography separation techniques [28]. Another well-known ionization technique is matrix-assisted laser desorption/ionization (MALDI), where the samples are co-crystallised with a liquid matrix, and ionized following laser pulses [29]. Due to incompatibility of MALDI with liquid chromatography, electrospray ionisation is preferred as it allows for high efficiency and sample throughput [21].

In a conventional shotgun (data-dependent acquisition) experiment, the mass-to-charge ratio of the peptide ions are measured at high resolution to obtain a full scan (MS1) spectrum (**Figure 1**). The most abundant peptide ions of every full scan are then selected for fragmentation via collision induced dissociation, higher-energy dissociation, or alternative approaches such as electron capture dissociation and electron transfer dissociation [21, 23, 30]. This needs to be achieved within the time peptides elute from the separation column which is usually at peak widths in the range of seconds. The spectrum of the resulting fragments (MS/MS or MS2 spectrum, **Figure 1**) is then subjected to *de novo* sequencing, spectral library searching or to the more common database searching approach, to identify the peptide sequence, possible modifications, and the parent protein [21].

Nowadays, high-resolution mass analysers such as the Orbitrap, time-of-flight mass analysers are the most commonly employed analysers in cellular proteomics [3, 31, 32]. On the other hand, the first mass analyser in tandem (or hybrid) mass spectrometers are often simple mass filers, such as the quadrupole mass analyser [3]. The quadruple-Orbitrap mass spectrometers emerged as one of the most popular instruments for cellular proteomics studies due to its advanced performance in regard to mass resolution, accuracy and fragment ion coverage [13, 33]. Nevertheless, time-of-flight analysers enable fast scan speeds of 100 microseconds per spectrum [31]. In combination with ion mobility as a third dimension of separation (*e.g.* trapped ion mobility, as employed in the timsTOF Pro instrument) [34–36], this provides a powerful alternative to the widely used quadruple-Orbitrap mass spectrometers. This has been recently demonstrated in a scan mode termed parallel accumulation serial fragmentation 'PASEF' where peptide ions that co-elute are accumulated in parallel at specific ion mobility spaces [37]. Those are then serially ejected from the ion mobility cell for further fragmentation and analysis. This process increases proteome

coverage without loss of sensitivity. Ion mobility devices have also been recently developed and flexibly implemented by several other mass spectrometer instrument developers [38].
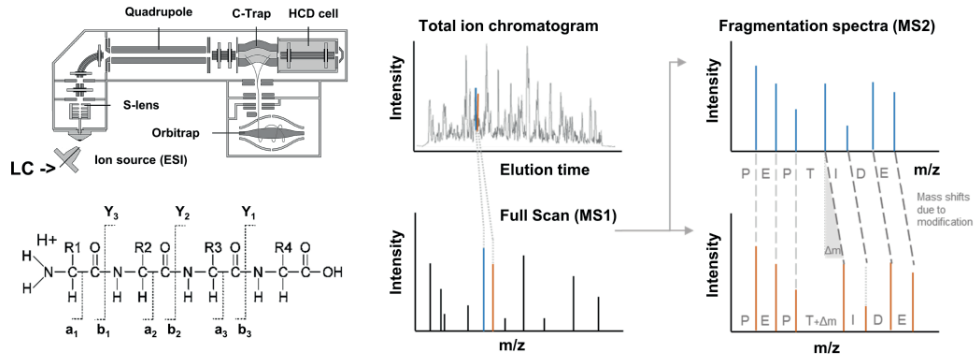
Figure 1. High-resolution tandem mass spectrometry for sequencing of complex peptide mixtures as used in cellular proteomics. The upper left shematic shows a quadrupole-Orbitrap mass spectrometer, which has been widely employed for cellular proteomics studies over the past decade. A liquid chromatographic separation system is usually directly coupled to the ion source. The chromatogram (upper middle graph) shows the measured total ion intensity obtained from the chromatographic separation and mass spectrometric detection of the proteolytic digest. The full scan mass spectrum (MS1) shows mass peaks derived from peptides that elute at a given time point during the chromatographic separation, for which the mass-to-charge (m/z) ratio and the ion intensity are measured. Modified peptides show an increased m/z value in relation to the chemical nature of the modifications. The fragmentation spectra (MS2) of the peptides are obtained in a consecutive fragmentation step (right graphs). This is performed in a dedicated collision cell. This is commonly achieved following isolation of the m/z value by a quadrupole mass filter and subsequent fragmentation by collision induced dissociation (CID, HCD) or electron transfer dissociation ETD. The blue and orange lines show the fragmentation spectra from a theoretical peptide and its modified counterpart. The sequencing spectrum finally represents the amino acid sequence of the peptide (as shown for the theoretical sequence "PEPTIDE" (right graphs). The modified peptide shows a mass shift on the Threonine (T) residue, which results in the peptide sequence PEPT(+Δm)IDE. Unknown modifications pose a challenge to automated data processing in cellular proteomics experiments.

Nevertheless, the final step of a proteomics experiment is the interpretation of the acquired sequencing spectra. Some 500K sequencing spectra can nowadays be achieved during a conventional two hour chromatographic separation [13, 33, 37, 39]. Most common algorithms used to investigate fragmentation spectra are based on database searching, where the experimental spectra are compared to *in silico*-generated spectra from predefined protein sequence databases [21, 39, 40]. Yet, different search engines have been developed over the past decades, which often employ very different scoring and filtering strategies. Therefore, the combined application of different search engines provides usually an increased number of peptide spectrum matches [41–48].

More recent developments improve the reproducibility of large-scale proteomics experiments by overcoming the stochastic sampling in shotgun experiments via data-independent acquisition methods (commonly abbreviated with 'DIA' or 'SWATH' MS) [49]. The ions of each window are then analysed by a high-resolution mass spectrometer. These strategies enable highly reproducible and rapid proteomics experiments, but usually also depend on advanced database or spectral library searching algorithm [50–53]. Additionally, hypothesis-driven studies allow to focus on a subset of the proteins. These targeted applications allow the application of focused data acquisition methods (*e.g.* parallel reaction monitoring (PRM), single reaction monitoring (SRM), or multiple reaction monitoring (MRM)) to perform a very sensitive quantification of hundreds of preselected proteins in single runs [54–57].

Apart from achieving a high proteome and amino acid sequence coverage in cellular proteomics studies, another challenge is the detection and quantification of covalent protein modifications and unexpected single nucleotide polymorphisms. Such events are commonly not captured in the reference sequence databases due to mass changes to the reference peptide and therefore require advanced experimental as well as bioinformatics approaches. Latest developments include sophisticated enrichment strategies that enable the sensitive detection and quantification of selected modifications, as well as open database searching approaches that allow the discovery of unexpected modifications in complex datasets [58–60].

Nevertheless, the qualitative information obtained from large-scale discovery experiments finds often only limited application. Many systems biology questions require additional quantitative information about the proteomic state of the cell. Unfortunately, peptide signal abundances cannot directly be translated into absolute peptide, or protein quantities. Albeit that over the past decades useful strategies have been investigated to obtain more quantitative parameters (*e.g.* spectral counting, PAI and emPAI indices [61]), without the application of additional synthetic peptide standards the proteome information remains semi-quantitative. Nevertheless, a relative quantification (*e.g.* fold change of the same protein between different conditions) can be achieved for several thousands of proteins at high confidence, using label free as well as chemical-labelling approaches. For example, stable isotope labelling by amino acid in cell culture (SILAC) is a frequently used labelling technique that enables relative quantification of different cultures by addition of selected stable isotope amino acids into the cell culture medium [62]. Other chemical-labelling methods include the isobaric tagging for relative and absolute quantification (iTRAQ) and the tandem mass tag (TMT) strategies (**Figure 2**) [63]. These allow to quantify thousands of proteins from multiple conditions, or time points, in single analysis runs. The intact masses of the chemical labels remain the same, while the mass normalizer and the mass reporter carry isotopes in different combinations to distinguish between samples. Peptides with the same sequence from different samples will elute from the separation column simulatenoulsy, and appear as a single peptide mass peak in the mass spectrum (**Figure 2**). Therefore, the number of peaks in the mass spectra does (in

theory) not increase when increasing the number of samples. Quantification is performed based on reporter ions in the MS2 spectrum, as the different mass tags only differ upon peptide fragmentation. The main advantage of TMT quantification is the ability to multiplex up to 16 samples [64], while for label-free quantification, samples are measured separately and thus significantly increase instrument time (**Figure 2**). Quantification reproducibility of the TMT approach does therefore not depend on performance variations LC-MS setup [65], whereas a label free method is cost-efficient compared to TMT quantification. Finally, label free quantification is often superior in terms of protein coverage/identifications [66] and TMT quantification (MS2 based) is prone to co-isolation which may lead to less accurate results [67]. To overcome this issue, MS3-based approaches allow to further isolte and fragment ions from MS2 spectra [68].
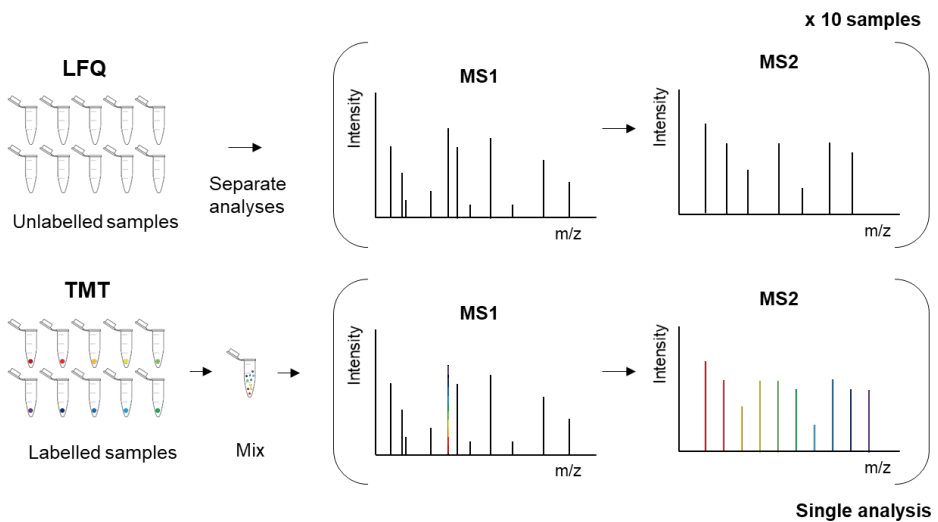
Figure 2. Protein quantification methods: label-free quantification (LFQ) and tandem mass tag (TMT). In LFQ, samples are prepared and analysed separately using LC-MS/MS. Quantification of the samples is performed for each separate analysis. TMT quantification enables the quantification of up to 16 samples simultaneously due to unique labelling of each sample and subsequent mixing of the samples. Peptides with the same amino acid sequence from different samples elute from the chromatographic separation system at the same time and show therefore a single peptide mass peak in the spectrum**.** Quantification is performed based on the reporter ions in the fragmentation (MS2) spectrum, as the different mass tags differ upon fragmentation.

Absolute quantification on the other hand, can be achieved by using ^{13}C or ^{15}N stable isotope labelled standards such as by employing 'AQUA peptides' [69], by using the 'QconCAT protein standards' [70], or the full-length stable isotope labelled proteins for absolute quantification 'FLEXIQuant' [71]. These are powerful, but also costly approaches, and therefore not applicable to the absolute quantification of a global microbial proteome.

The yeast *S. cerevisiae* as cell factory

Microbial fermentation has played an important role in human society since ancient times. Alcoholic fermented beverages have been produced with yeasts and consumed by people as early as Neolithic times [72, 73]. However, yeast was only visualized by the early microscopes of Antoni van Leeuwenhoek in the 17th century [74]. The connection to alcohol formation in the fermentation process was then demonstrated by Louis Pasteur two centuries later [75]. Nowadays, *S. cerevisiae* (baker's yeast) is the most well-known yeast and widely used for industrial and research applications. This is due to its genetic accessibility [76] and its GRAS (generally recognised as safe) status for the production of food and health-care products [77]. *S. cerevisiae* is also one of the most prominent cell factories because of its compatibility with high-density and large-scale fermentation, high tolerance against toxic inhibitors and final products [78] and its suitability as host for recombinant protein production [79].

Ultimately, *S. cerevisiae* became a popular metabolic engineering platform and a versatile industrial cell factory. As the world demands sustainable alternatives to petroleum-based products and fuels, microbial cell factories play an important role in the sustainable production of a broad range of products. Since yeast is already extensively used in beer and wine production, its use was extended towards bioethanol production [80], making it the most used biofuel to date. This alternative fuel is less toxic and more readily biodegradable than petroleum fuel such as higher octane number [81]. Additionally, many 'bio-based' chemicals have successfully been produced by *S. cerevisiae* [80], ranging from amino acids such as ornithine, which can be used as a dietary supplement [82], to pharmaceuticals products like opiods that are relevant for pain treatment [83]. However, many of the products can for now only be produced on laboratory scale in small amounts. To achieve cost-competitive bio-based processes, cell factories must convert raw materials into products of interest with high productivities and yields.

Therefore, extensive reengineering of the central carbon metabolism is often required, because not only the pathway of interest need optimisation, but also the supply of precursor metabolites and co-factors need to be ensured, while at the same time by-products should be eliminated [84]. Over the past decades, the yeast molecular biology toolbox expanded tremendously, most recently with the dynamic genome editing tool CRISPR-Cas which allows for the precise introduction of multiple genetic modifications in the yeast genome

simultaneously [85–87]. These tools could improve the production of (heterologous) bio-based products by industrial yeast strains. Nevertheless, multiple native genes need to be edited and heterologous multi-gene pathways to be implemented [88]. The central carbon metabolism remains essential to microbes, making simple elimination and replacement of biochemical reactions due to high complexity level of the genome of yeast often difficult. Therefore, metabolic pathway engineering is challenging, but essential for the production of bio-based chemicals and fuels.

The yeast *S. cerevisiae* as eukaryotic model organism

Yeast also became an important model organism for higher eukaryotes, in particular for human cells. *S. cerevisiae* shows a high degree of conservation for many key cellular processes when compared to human cells, such as for protein translocation and secretion, protein folding and chaperone functions. Furthermore, many key signal transduction processes are conserved between yeast and humans, indicating conserved elements for protein-protein interactions, regulation hierarchies and signal cross-talk [89]. Interestingly, 47% of the 414 essential yeast genes can be replaced one on one by their human orthologues [90]. Additionally, yeast can grow very fast (especially compared to mammalian cells), it shows simple nutritional requirements and harbours an accessible genome (approx. 6000 protein coding genes).

Therefore, yeast has become a popular single-cell model organism to study the carbohydrate metabolism of eukaryotes. For example, the Embden-Meyerhof-Parnas (EMP) pathway of glycolysis is similar in all eukaryotes and plays a central role in carbon metabolism. Glycolysis is involved in diseases such as cancer, via the 'Warburg effect' [91, 92]. Several features of fermenting yeast and human tumour cells (that show the Warburg effect) are shared. Cancer cells consume larger amounts of glucose compared to healthy cells. Most of this glucose is then converted into lactate, despite presence of oxygen, which would result in full oxidation to carbon dioxide generating more energy [92–94]. Yeast cells grow in a similar manner during the Crabtree effect, where ethanol is accumulated under aerobic conditions at high specific growth rates, as fermentation is preferred over respiration [95]. A recent study also highlights the beneficial use of yeast cells as a model organism for the human Warburg effect, because all parameters in the presence of yeast could be reconstituted. Furthermore, it was successfully demonstrated that oxidative phosphorylation repression is not necessary to boost cell growth [96]. Another example of *S. cerevisiae* as a model organism is in aging research. A remarkable number of aging pathways are conserved across human and yeast species, including genome instability and nutrient signalling [97, 98]. Therefore, baker's yeast remains an important and widely used model organism for higher eukaryotes.

Yeast growth on glucose

S. cerevisiae has the unique ability under yeast species to grow rapidly both under aerobic and anaerobic conditions. Growth of this yeast can be supported by a broad set of carbon sources, however the preferred carbon and energy sources are fermentable sugars, such as glucose. In glucose-limited, aerobic batch conditions, *S. cerevsisiae* encounters three phases throughout its growth curve; proliferation, diauxic shift and stationary phase. First, glucose is taken up in the yeast cell to support exponential growth and further metabolized through dissimilatory and assimilatory pathway. Glycolysis – the central pathway for sugar oxidation into pyruvate [99] – is coupled to fermentative production of ethanol under conditions of oxygen limitation and (or) sugar excess (Crabtree effect) (**Figure 3**) [100]. Nevertheless, *S. cerevisiae* can generate energy by fermenting glucose to ethanol to produce two ATP molecules per glucose or by respiration to obtain a significantly higher yield of ATP (16) on glucose molecule [101, 102]. Upon glucose exhaustion yeast cells undergo the so-called diauxic shift. Here, the previously generated ethanol is consumed by switching to gluconeogenesis using the glyoxylate shunt to provide its precursors and simultaneously increasing the respiration rate by upregulation of tricarboxylic acid (TCA) genes [103]. The main catalytic function of the TCA cycle is to provide reducing equivalents to the respiratory chain via oxidative decarboxylation of acetyl-CoA. Furthermore, the TCA cycle also supports biosynthetic metabolism and most intermediates are used by other metabolic reactions [104]. Oxidative phosphorylation ensures the generation of ATP through electron transport of the energy precursors from the TCA cycle. Finally, cell reach the stationary growth phase upon complete carbon source depletion and growth is arrested immediately. Carbon starved cells are characterized by thickened cell walls, increased resistance and accumulation of storage carbohydrates [105]. In this quiescence stage, extracellular sources for energy and carbon are absent and therefore the cells rely upon intracellular energy production. In the presence of oxygen, two types of storage polymers are available in yeast. The first type, storage carbohydrates trehalose and glycogen, are generated from glucose 6-phosphate during growth on glucose [106]. The second type of storage polymers are fatty acids, which are predominantly stored as di- and triacylglycerol esters. These fatty acids are catabolized through β-oxidation to produce acetyl-CoA, which subsequently enters the TCA cycle followed by energy generation through respiration. Interestingly, cellular chronological lifespan (CLS) is also modelled by stationary phase yeast cells, as they preserve viability and are able to grow upon encounter with nutrients. Understanding of stationary phase yeast cells has allowed for a better understanding of cellular mechanisms in aging, also applied to higher eukaryotes [107].

In the absence of oxygen, *S. cerevisiae* becomes auxotrophic for certain compounds, such as sterols and unsaturated fatty acids [108]. Consequently, anaerobic cells have a different lipid composition of the membrane (more saturated fatty acids, less sterol, ergosterol and squalene) compared to yeast cells in aerobic environments [109]. Furthermore, cells are not able to
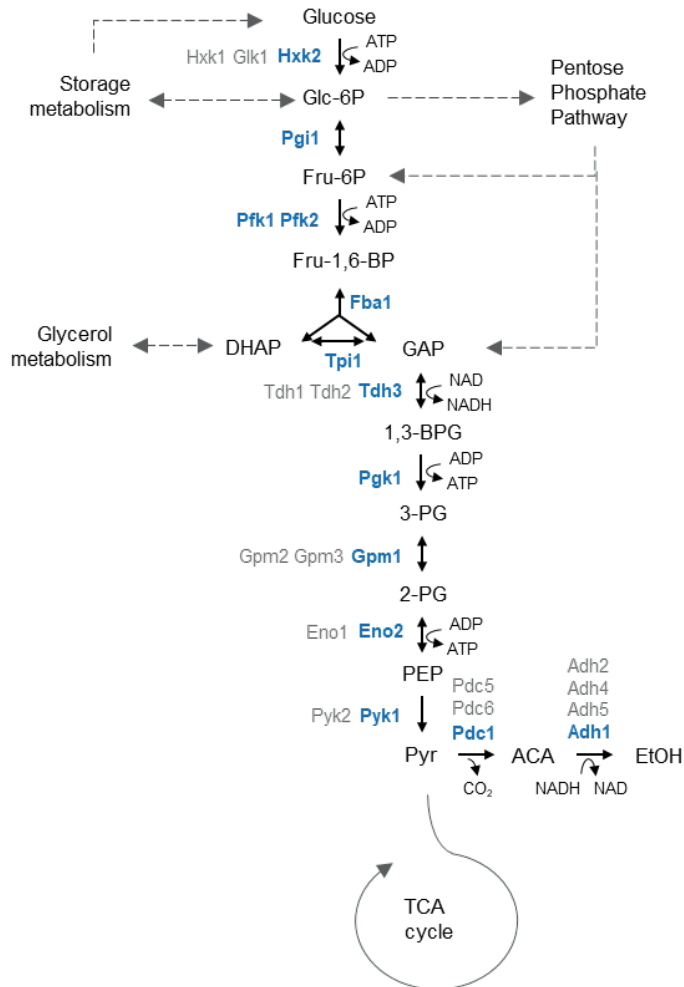
Figure 3. Schematic overview of *S. cerevisiae* glycolysis and associated pathways in the central carbon metabolism. The graphs outlines the major (blue) and minor (grey) *S. cerevisiae* isoenzymes. Only the major isoenzymes are retained in the minimal glycolysis strain.

respire using the TCA cycle under oxygen deprivation. To maintain a high growth rate, the *in vivo* activity of glycolysis is much higher in anaerobic cultures than in aerobic cultures [110]. Overall, alcoholic fermentation is a redox-neutral process. Nevertheless, growth is associated with anabolic processes that generally lead to a surplus of reducing compounds. In the presence of oxygen, these can be oxidised through the respiratory chain. On the other hand, glycerol is commonly produced to maintain redox balances during anaerobic growth [111]. Once glucose is depleted, the lack of oxygen ensures the entry directly into the

stationary phase. Here, the only known storage compounds, trehalose and glycogen, are utilized for energy generation. However, the conversion can only take place through alcoholic fermentation, thereby yielding 5–8 fold less ATP compared to respiratory dissimilation in aerobic conditions [106]. Little is known about the transition into or survival of budding yeast cells in stationary phase in anaerobic environments. However, yeast cells can pose as simplified model organism for cells in structured environments in multi-cellular organisms (*e.g.* tissues) that encounter micro-environments, including variations in oxygen concentrations, causing heterogeneity [112].

Layers of metabolic regulation in yeast

A major challenge in yeast growth is to understand how complex regulatory mechanisms control the metabolic pathways in response to changes in the environment. Generally, the metabolic flux depends on the capacities of the enzymes, which correlate with enzyme abundances, allosteric regulation but also with the presence of post-translational processes such as covalent protein modifications or irreversible cleavage of the protein backbone. Protein modifications are important post-translational regulators, which can alter protein (enzyme) activities and location spontaneously in a reversible manner. Post-translational modifications (PTMs) are not directly encoded in the genome and require therefore the analysis of the protein itself.

Since glycolysis is a pivotal pathway in cellular growth and division, the pathway is subjected to multiple layers of regulation. Even though glycolysis is one of the most studied pathways of the yeast metabolism, a profound understanding of its regulation is still not accomplished, partly by the genetic redundancy of the pathway. Genetic redundancy in eukaryotes such as *S. cerevisiae* is another level of regulatory complexity for experimental development of cell factories. Many genes have orthologues with apparently similar functions [113], but with poorly understood roles. In *S. cerevisiae*, eight of the 12 enzymatic reactions from glucose to ethanol are represented by multiple paralogous genes following whole genome duplication event(s), resulting in two or more isoenzymes for the same reaction [114]. In most cases, the major paralog of the isoenzymes has the largest contribution to the catalytic conversion. An exception is phosphofructokinase, as both paralogs support the activity equally as part of hetero-octameric enzyme [115]. The co-existence of more than one enzyme reaction in glycolysis has also not been fully understood and remains elusive even today. For some minor paralogs, expression depends on specific nutrient availability, while others exhibit a (moonlighting) function parallel to their glycolytic duty [116].

In general, the major paralogs are constitutively expressed at high levels, making up to 20% of the expressed proteins in solution in a yeast cell [117]. Furthermore, the high expression levels of glycolytic enzymes also allow for an overcapacity that result in rapid adaptation to changing environments [118], while expression levels are still prone to environmental cues

that can increase their expression levels up to five times [119, 120]. Overall, the minor glycolytic enzymes are transcribed at a much lower level compared to the major paralogs, although their expression is claimed to be condition-dependent. For example, *HXK1, GLK1, PYK2* and *ADH2* have a low expression in response to high glucose concentrations [121–124]. In addition, *PDC6*, minor paralog to *PDC1*, is strongly upregulated in sulphur-limited conditions, due to its lower sulphur amino acid content compared to its paralogs [125, 126]. In *S. cerevisiae*, for at least five glycolytic enzymes, secondary functions have been demonstrated. First, Hxk2 is not only involved in phosphorylating intracellular glucose, but also in the glucose repression process [127]. In these high glucose concentrations, approximately 15% of Hxk2 is localized to the nucleus where it interacts with the transcriptional repressor of the aforementioned repressible genes [128], including Hxk1 and Glk1. Fba1 is the second enzyme with an alternative function in yeast metabolism, while it lacks paralogs for its glycolytic activity. Fba1 is physically involved in the assembly of the subunits of vacuolar proton-translocating ATPases (V-ATPases), while Fba1 also plays an essential role in V-ATPase activity, that couple ATP hydrolysis to proton transport out of the cytosol into the vacuole [129, 130]. The enolase pair (Eno1 and Eno2) in glycolysis are both implied to be involved in vacuolar fusion [131], in addition to their enolase activity in glycolysis. Deletion of either enolase therefore also results in vacuole fragmentation in yeast cells [131]. Finally, Pyk1 and Pyk2 both catalyse the conversion of phosphoenolpyruvate to pyruvate, but only Pyk2 (minor paralog) allows for efficient growth on the three carbon-substrate dihydroxyacetone (DHA) [132]. Interestingly, in a study by Solis-Escalante *et al*., a minimal glycolysis (MG) yeast strain was constructed in which all the redundant minor isoenzymes were knocked-out (**Figure 3**) [120]. To this end, only 13 (out of 26) glycolytic enzymes were expressed in this yeast. The deletion of the glycolytic minor isoenzymes did not induce a phenotypic response under a range of tested growth conditions, confirming the suitability of such a strain for studying the regulation of metabolic enzymes. However, the underlying proteome level adjustments were not investigated.

Nevertheless, it has been demonstrated that the glycolytic flux in *S. cerevisiae* is predominantly determined by post-translation processes, such as allosteric regulation and protein modifications [133]. The glycolytic flux is thus partially determined by well-defined allosteric regulators. First, Hxk1 and Hxk2 are strongly inhibited by trehalose-6-phosphate, which acts as an intermediate in trehalose synthesis [134]. Secondly, Pfk1 and Pfk2 can be affected by multiple regulators, while the enzymes are influenced mainly by ATP inhibition, AMP, ADP and fructose-2,6,-biophosphate activation [135]. The final kinase in glycolysis is Pyk1 and this enzyme is regulated allosterically by fructose-1,6-bisphosphate, an glycolytic intermediate produced by phosphofructokinase, that acts as an feed-forward activator [124]. Inhibition of hexokinase and phosphofructokinase in combination with activation of pyruvate kinase allow for harmonization of the upper and lower part of glycolysis.

A slightly less characterized form of regulation is by PTMs on glycolytic enzymes. Many modifications, such as methylation [19], acetylation [20], succinylation [21], ubiquitination [22], nitration [23] and O-mannosylation [24] have been reported for glycolytic enzymes. Most proteome studies have focused on identifying phosphorylation in *S. cerevisiae* [136–140], however, only few can connect the modification to a biological function in the cell. For example, Hexokinase 2 is able to shuttle between the nucleus and cytoplasm due to a phosphorylation of the protein [141]. Moreover, it has been demonstrated recently that increased phosphorylation resulted in decreased enzyme activity and lower growth rates of *S. cerevisiae* under glucose-limited conditions, indicating a negative correlation between phosphorylation and activity of several glycolytic enzymes [142]. Yet, the functionality of many (un)identified modifications on glycolytic enzymes need to be determined.

State-of-the-art in yeast proteomics

Proteomics of microbes and their complex communities is still far from routine. For yeast, *S. cerevisiae* is by far the best investigated member considering proteomics studies [110, 143–153]. Yeast has been also subjected to several deep proteomic studies where 80–90% of the complete proteome could be identified [151, 152]. There were also numerous proteomic studies performed that focused on biological questions. For example, since yeast is able to grow under both aerobic and anaerobic conditions, several studies explored the influence of oxygen on the proteome. A study by de Groot *et al*. illustrated that most glycolytic enzymes were more abundant in the absence of oxygen compared to oxygen-rich conditions [110]. On the other hand, TCA cycle proteins were more abundant in aerobic conditions, as cells are not able to respire without oxygen. Helbig *et al*. discovered that the yeast respiratory chain complexes were still present in anaerobic conditions, although at reduced levels [145]. The genetic redundancy of enzymes in glycolysis was also addressed recently by Costenoble *et al*. This study showed that the minor isoenzymes Glk1 and Hxk1 redirect glucose to glycogen storage as opposed to glycolysis. In addition, the major ethanol-producing paralog Adh1 clustered with glycolytic enzymes, whereas the major ethanol-consuming paralog Adh2 clustered with gluconeogenesis proteins [148].

In batch cultures, the yeast encounters multiple environmental changes along the growth curve and the proteome needs to adapt accordingly. Especially under aerobic conditions, cells move from fermentation to respiration to growth arrest. Several studies therefore focused on the proteome adaptation during the diauxic shift in aerobic yeast cultures as cells reverse their metabolic flux from glycolysis to gluconeogenesis [146, 147]. According to a yeast proteomics study by Zampar *et al*., the diauxic shift is accomplished by the reduction of the glycolytic flux, production of storage compounds (before glucose depletion), which is followed by reversion of carbon flow through glycolysis and the glyoxylate cycle upon glucose exhaustion [146]. Finally, cells enter the stationary phase once the fermentation

1

products are consumed in the post-diauxic phase. Stationary phase yeast cells were analysed in comparison to exponentially growing cells, in which protein in log cells were enriched in various processes including biogenesis and RNA processing [153, 154]. In the absence of oxygen, however, cells enter the stationary phase directly after glucose exhaustion. Transcriptome changes of batch cultures were monitored throughout the complete (an)aerobic growth curve and revealed that the longevity and thermotolerance of yeast cells is strongly affected by oxygen availability [155]. The post-diauxic growth on ethanol was indicated to enable the cells with time and resource required for longevity and thermotolerance. Nevertheless, proteome analyses are required to confirm the changes on a transcriptome level during these conditions.

Recently, *in silico* models are generated to reproduce and predict the complex microbial metabolism to improve cell factories through metabolic engineering [156]. The prediction ability of these models is greatly enhanced with the use of resource allocation, including the cost of protein expression [157–161]. Mitochondrial proteome quantification and allocation, for example, revealed a trade-off between biosynthesis and energy generation during the diauxic shift [162]. In this study, the diauxic shift was suggested to represent the stage where major structural and functional reorganizations in mitochondrial metabolism occur. Another study demonstrated the influence of amino acid supplementation in aerobic and anaerobic cultivations. Here, the proteome reallocates from amino acid biosynthesis to ribosome to enable faster growth in rich media [160]. Proteome allocation was also applied to predict the Crabtree effect, explained by a reduced protein cost of generating ATP through fermentation than respiratory pathways [161]. Finally, proteome allocations changed linearly with the yeast specific growth rates. For example proteins involved in translation showed a remarkably linear correlation with the specific growth rate. On the other hand, glycolysis and chaperone proteins showed a linear decrease [163].

Nevertheless, a large number of spectra in (yeast) cellular proteomics experiments remain unidentified, likely resulting from unidentified modifications on peptides (**Figure 1**) [60]. In fact, the yeast proteome has been investigated for the presence of post-translational modifications. However, analysis of sub-stoichiometric modifications, with unknown chemical composition is very difficult to facilitate and are therefore often not explored at all. Interestingly, over the past decades, more than one hundred modifications have been detected (from different species) [164], where the fraction of modifications with a biological relevance is expected to be much lower. Therefore, the focus was strongly on the most common modifications such as phosphorylation, acetylation and methylation [139, 165–167]. However, the functionality of many modifications (sites) has not been validated to date. In addition, rare or difficult to analyse modifications may still have been left unnoticed so far. Most importantly, PTMs have commonly only been studied isolated, while many different modifications are known to modify one and the same enzyme simultaneously. Some

modifications have been also found competing for the same modification site or impacting on the modification stoichiometry of neighbouring modification sites through PTM cross-talk [168–171]. This either competitive or cooperative interplay has been already described as comparable to a large regulatory network [168, 170]. For example, cross-talk of phosphorylation and nearby U3-ligase dependent ubiquitinylation results in proteasomal degradation [169]. Another example is the potential interplay between histone methylation and adjacent regulatory phosphorylation sites [172]. PTM crosstalk is thought to be common in living cells, as more and more recent yeast studies demonstrate the co-occurrence of multiple PTMs at the same time [173–176]. Many technical as well as computational challenges however hamper the progress in this field and therefore there is still a large gap in understanding the diversity and regulatory impact of post-translational modifications, in particular for microbes such as yeast.

Still, a large number of peptide modifications are also introduced artificially during sample handling and preparation, underlining the need for mild preservative protocols, which is not common practice in the proteomics field. A significant fraction of unidentified spectra may therefore be a consequence of proteolytic cleavage or other sample preparation related artifacts. [177]. Therefore, an optimised sample preparation protocol is essential for achieving a deep proteome coverage and for studying (novel) protein modifications.

In summary, *S. cerevisiae* is one of the most extensively characterized model organisms, and has therefore also been subjected to many proteomics studies [110, 143–153]. However, the regulation of the complex yeast metabolism and the yeast proteome dynamics has not been fully explored to date. The proteome is a complex network in the cell that constantly adapts to meet the requirements of the cell. Therefore, to understand the regulatory networks in a cell in response to the environment, the proteome needs to be under close surveillance using the most advanced mass spectrometry-based methods. As the proteome is a highly dynamic entity in the cells in reaction to its environment, it is very important in proteomics experiments to closely control the conditions. However, many proteome features have been investigated in isolation (*e.g.* single protein modifications), or under poorly controlled growth conditions. Highly quantitative large-scale proteome data under tightly controlled culturing are currently lacking in the field that are required to assist predictive models of the metabolism.

Goal of the thesis and outline of the chapters

The single cell eukaryote *S. cerevisiae* is a highly important cell factory and model organism. Even though *S. cerevisiae* is one of the best characterized model organisms, a complete understanding of the metabolic network and its complex regulation has not been achieved to date. In particular quantitative proteome dynamics data established under highly controlled conditions are lacking. Therefore, this thesis aims to i) perform a comparative study on

cellular proteomic sample preparation protocols to enable standardised yeast proteomic studies, to ii) establish large-scale quantitative proteome dynamics data established under highly controlled conditions, and to iii) investigate novel mass spectrometric approaches to quantify the degree of protein modification of metabolic enzymes to better understand their role in metabolic regulation. The established data and methods aim to support metabolic engineering efforts in order to expand the scope of industrial applications, and to advance yeast in its role as model organism, *e.g.* when studying metabolic diseases.

Chapter 2 discusses recent advancements in data processing tools that support the discovery of post-translational modifications (in *S. cerevisiae*). Post-translational modifications represent an additional layer of regulation that is challenging to identify and to quantify in large-scale proteomics experiments. A large fraction of unidentified spectra is suggested to originate from unexpected post-translational modifications or natural amino acid sequence variants. Recent advances in mass spectrometry and data processing algorithm opened new ways to explore these important features. For example, so called "open modification" search approaches allow to simultaneously search for a large range of unexpected modifications in complex data.

Chapter 3 describes the influence of a range of sample preparation procedures on the proteomics outcome. In yeast proteomics, a shotgun approach is typically applied in which proteins are digested into peptides prior to LC-MS/MS analysis. Sample preparation can be one of the most significant contributors to biases in proteomics experiments. Therefore, we performed a systematic comparison of sample preparation protocols commonly applied in whole cell lysate proteomics experiments. The results were evaluated for a range of key performance criteria. Finally, the best protocols allow to identify approx. 65–70% of all fragmentation spectra. Additional *de novo* sequencing suggests that the remainng spectra were largely of too low quality in order to provide spectrum matches. Interestingly, several unexpected peptide modifications could be linked to chemicals such as solvents and additives. The established protocols and mass spectrometric data provide a large resource for optimising and evaluating yeast proteomics studies.

Chapter 4 investigates with large-scale quantitative proteomics the proteome dynamics of *S. cerevisiae* in response to dynamic glucose concentrations in batch cultures under both aerobic and anaerobic conditions. Experiments were performed under highly-defined growth conditions and continuous measurements of off-gas, pH, temperature, dissolved oxygen and sampling without disturbing the controlled environment. Anaerobic conditions showed substantially lower protein dynamics during the transition from exponential to stationary phase, that could suggest that anaerobically growing cells lack the time and resources needed to adapt to the changing in environment, more particularly the depletion of carbon source. Additionally, comparison to the minimal glycolysis *S. cerevisiae* strain allowed to investigate

roles of the minor isoenzymes under these dynamic conditions. Interestingly, no significant physiological responses and only a few protein-level alterations were observed when deleting the redundant minor glycolytic paralogs.

Chapter 5 provides proof of concept for an approach that enables to quantify the global degree of modification for individual enzymes from complex cell lysates. To date, more than hundred different types of modifications have been observed. Modifications such as phosphorylation, acetylation or glycosylation, are very abundant and are known to translate important functions. However, studies on post-translational modifications usually consider only a single or a few modifications, while modifications have been demonstrated to interact. Furthermore, studies commonly only investigate sequence regions that are accessible via tryptic peptides. This approach quantifies the unmodified fraction of each peptide, and therefore considers all types of modifications, regardless their chemical nature. Quantification of the complete protein sequence is realised by the aid of a cell free synthesis produced protein standard, TMT labeling and a multi-protease approach. Finally, we exemplify this approach by monitoring the protein modification changes of the glycolytic enzyme Pyk1 from *S. cerevisiae*, during transition from proliferation to stationary phase under aerobic conditions. Further development and additional cross validation of the established data will allow to expand this approach to complete metabolic pathways.

Chapter 2

Shotgun proteomics: why thousands of unidentified signals matter

Maxime den Ridder, Pascale Daran-Lapujade and Martin Pabst

Abstract

Mass spectrometry-based proteomics has become a constitutional part of the multi-omics toolbox in yeast research, advancing fundamental knowledge of molecular processes and guiding decisions in strain and product developmental pipelines. Nevertheless, post-translational protein modifications (PTMs) continue to challenge the field of proteomics. PTMs are not directly encoded in the genome; therefore, they require a sensitive analysis of the proteome itself. In yeast, the relevance of post-translational regulators has already been established, such as for phosphorylation, which can directly affect the reaction rates of metabolic enzymes. Whereas the selective analysis of single modifications has become a broadly employed technique, the sensitive analysis of a comprehensive set of modifications still remains a challenge. At the same time, a large number of fragmentation spectra in a typical shot-gun proteomics experiment remain unidentified. It has been estimated that a good proportion of those unidentified spectra originates from unexpected modifications or natural peptide variants. In this review, recent advancements in microbial proteomics for unrestricted protein modification discovery are reviewed, and recent research integrating this additional layer of information to elucidate protein interaction and regulation in yeast is briefly discussed.

Introduction

Mass spectrometry (MS) is currently the most powerful technology for the identification, characterisation and quantification of complex mixtures of proteins [178–180]. The performance of mass spectrometers has steadily improved over recent decades, providing a near-complete yeast proteome coverage using short chromatographic separation times and minimum amounts of sample [151, 181]. The actual proteome is approximately 2 to 3 orders of magnitudes more complex than can be predicted from its genome [182], in which diversification processes such as post-translational modifications (PTMs) contribute substantially. PTMs are enzyme-mediated covalent modifications (or cleavage products) introduced following biosynthesis of the amino acid backbone, thereby expanding the range of possible protein isoforms (proteoforms) and functions without the need for changing the genetic code itself [183]. For example, the central carbon metabolism has been fine-tuned to exactly meet the requirements for building blocks and Gibbs free energy in conjunction with cell growth. Thus, when cells experience environmental changes, their metabolism immediately aims to adjust [133, 184, 185]. Apart from the relatively slow adjustment of enzyme abundances, cells utilise fast and dynamic routes such as allosteric regulation but also through the above-mentioned (reversible) protein side chain modifications of the enzyme itself [133, 186] (**Figure 1**). The number of different modification types reported across various species are in the range of hundreds [164, 187], but most have been only observed at low frequency and low stoichiometry, and a potential biological relevance remains elusive [188].
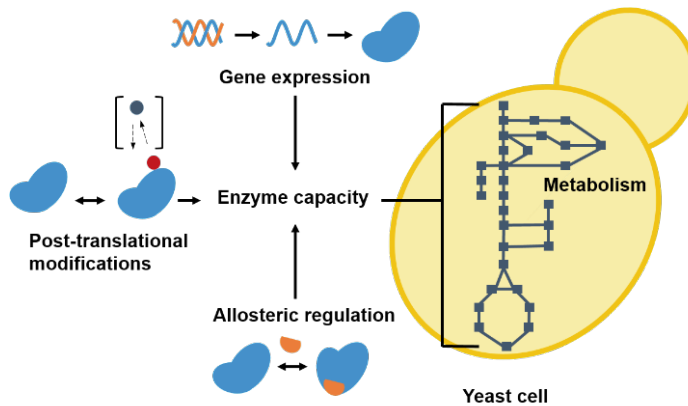
Figure 1. Overview of metabolic regulation mechanisms in a yeast cell. The metabolic flux in cell depends on the capacities of the metabolic enzymes, which depend on enzyme abundances, allosteric regulation but also on the occurrence of covalent modifications on the protein itself, termed post-translational modifications (PTMs).

PTMs observed for the yeast model *Saccharomyces cerevisiae* have recently been compiled in the database YAAM [189]. Here, statistics on 12 experimentally confirmed modification types (ubiquitination, phosphorylation, acetylation, lipidation, oxidation, succinylation, glycosylation, methylation, sumoylation, nitration, disulfide bond formation and N-terminal acetylation) are described that are experimentally confirmed by MS, point mutation or functional evidence (**Table 1**). According to this database, more than 70% of the complete yeast proteome has been observed post-translationally modified (albeit under different growth and experimental conditions), demonstrating the importance of these regulators for cellular processes [189]. Nevertheless, low frequent or difficult to analyse modifications, such as lipoylation [190, 191] are rarely addressed, of which some may still have been left unnoticed to date.

Table 1. Number of entries found for most abundant PTMs in yeast. Number of entries for each PTM described for *Saccharomyces cerevisiae* as collected in the YAAM database (http://yaam.ifc.unam.mx, November 2019). Each entry represents a modified residue described in literature, therefore a specific PTM site could be entered more than once [189]. The number of modified proteins in terms of the complete proteome of *S. cerevisiae* are also given.

Modification	# Mods	# Proteins
Acetylation	10052	1814
Disulfide	264	79
Glycosylation	1972	424
Lipidation	183	128
Methylation	287	143
Nitration	16	15
N-terminal Acetylation	762	687
Oxidation	875	605
Phosphorylation	87739	3955
Succinylation	1754	570
Sumoylation	138	48
Ubiquitination	14883	2355
Coverage		
Estimated total modified proteins.		4759
Modified proteins of the proteome (%)		70.82

Protein phosphorylation appears to be the most commonly observed PTM in yeast (**Table 1**), however, this may be heavily overestimated since many studies specifically focus on this type of modification. A similar trend was observed in the Uniprot database entries for the proteins of *S. cerevisiae*, in which also phosphorylation appears to be the most frequent type of modification in yeast [192]. In addition to phosphorylation, other common modifications

such as acetylation and methylation have been frequently described and were found involved in a variety of cellular processes. For example in yeast glycolysis, the shuttling back and forth of Hexokinase 2 between the nucleus and cytoplasm is regulated by phosphorylation [141]. Furthermore, studies over the past years have demonstrated that phosphorylation seems to affect many more processes, such as cell signalling [193], glycerol metabolism [194], regulation of nucleotide and amino acid biosynthesis [195], regulation of the outgrowth of autophagosomal membranes in autophagy [196] and DNA damage checkpoint signalling [197]. Moreover, lysine acetylation (and glutamine methylation) is an evolutionarily highly conserved modification, which regulates chromatin accessibility and therefore affects gene expression directly [198–200]. Recent studies on PTMs in yeast are summarized in **Supplementary Table 1**. For a more extensive reviews on post-translational modifications in yeast, in particular for phosphorylation, we refer readers to recent reviews from Oliveira and Sauer (2012), Tripodi *et al.* (2015) and Chen and Nielsen (2016) [167, 180, 201]. Nevertheless, the functionality of many modification sites and (unknown) modifications have not been investigated to date.

In this review, recent advancements in proteomics for unrestricted modification discovery are summarised, as PTM data analysis remains very challenging in identifying modifications. Moreover, we discuss the critical trade-off between maximum proteome coverage versus maximum sequence coverage in the context of a comprehensive functional characterisation of PTMs.

Advances in bioinformatics tools for unrestricted PTM discovery

Currently, the most frequently employed approach in discovery proteomics is referred to as shotgun proteomics (**Figure 2**). Thereby, (in a bottom-up approach) protein extracts are analysed following proteolytic digestion using liquid chromatography (LC) coupled to tandem mass spectrometry (MS/MS) [178]. After chromatographic separation of the proteolytic digest, peptides are analysed for (accurate) mass that further triggers automatic fragmentation to obtain peptide sequence information [23, 202–204]. The spectra are then matched against a predefined protein database, generally derived from public repositories or genome sequencing itself. In common, cellular shotgun proteomics experiments create enormous amounts of sequencing spectra. For example, as much as 80,000 MS/MS spectra could be obtained for an one hour yeast proteome experiment [151].

State-of-the-art proteomics workhorses, such as the quadrupole Orbitrap mass spectrometer [13], are capable of acquiring spectra at high speed (such as >20 Hz), high resolving power and high mass accuracy fragmentation spectra, supporting the identification of several
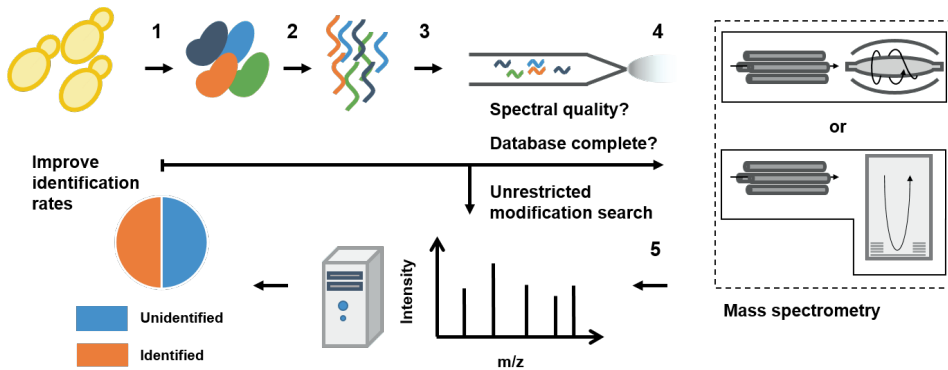
Figure 2. Protein identification by (a bottom-up) shot-gun proteomics experiment. The typical workflow consists of cell lysis, protein extraction from (yeast) cells and subsequent digestion into peptides using specific proteases (step 2). In step 3, the peptides are separated by liquid chromatography and further detected following electrospray ionization (ESI) by data dependent analysis, which automatically collects fragmentation spectra from (top intense) peptide signals, in step 4. In step 5, the fragmentation spectra (MS/MS) are commonly identified by database search approaches, matching fragmentation spectra to in-silico peptides derived from the target proteome database. Finally, the result of a typical shotgun proteomics experiment is depicted, in which roughly half of the spectra could not be assigned to a target protein sequence of the database. Similar identification rates have been reported in a survey reinvestigating hundreds of shot-gun proteomics experiments by Griss et al. (2016) . To improve the identification rate, several approaches can be taken. First, the quality of the spectra should be verified. In addition, modifications could be added to the database to ensure that the database is complete and modified peptides could be identified. Finally, an unrestricted search could be performed to identify (unexpectedly) modified peptides.

thousands of proteins in less than one hour of analysis time [33, 205]. Nevertheless, only a fraction of fragmentation spectra (on average less than half) is confidentially matched to the proteome database (**Figure 2**) [60]. It is estimated that a good portion of those unidentified, but high-quality spectra, may originate from unexpected (including not considered) modifications (or sequence variants) not present in the database [60, 206]. Where the majority of modified peptides may be readily detected by the employed method, their confident identification however, often fails. Peptides containing modifications could, for example, not only be shared by different proteins (protein inference), but several overlapping modified peptides could also share the same modified site [207].

Identification is commonly realised by a database matching algorithm, where the acquired fragmentation spectra are matched against *in silico* spectra from predefined protein sequence databases [21]. Modifications are thereby detected as mass deviations from the native peptide mass peak, and that, in the ideal case, can be further allocated to single amino acids (**Figure 3a**).
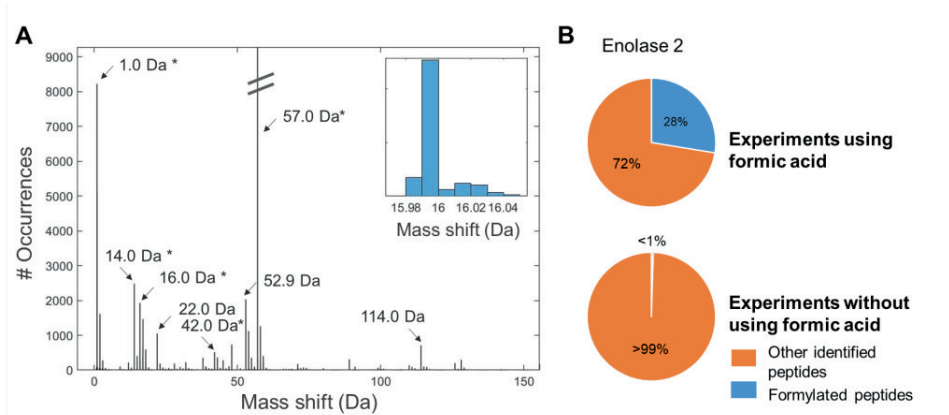
Figure 3. Peptide modification search. A. Unrestricted modification search as obtained by using the Byonic software tool (**Table 2**). The observed mass shifts (between 0 and 150 Da) on tryptic peptides from a yeast shot-gun proteomics experiment are represented in a histogram. Commonly, many mass additions derive from artificial modifications introduced during sample preparation, while others are of natural origin. However, the distinction between both types is a delicate step during data processing and evaluation. Common mass shifts found are indicated by an arrow: Deamidation (0.98 Da) ^{13}C isotope precursor selection (1.0 Da), methyl (14.0 Da), oxidation (15.98 Da), sodium adduct (22.0 Da), acetyl (42.0 Da), iron adduct (52.9 Da), carbamidomethyl (57.0 Da) and dicarbamidomethyl or ubiquitin (114.0 Da). Phosphorylation (80.0 Da) was not observed frequently in this search, even though it appears to be the most commonly observed PTM in yeast [189]. However, this modification type usually occurs at low stoichiometry and therefore enrichment methods are commonly applied, which was not performed here. (*) Indicated mass shifts can also represent amino acid substitutions, which require case-by-case investigations. Inset: Assignment of high-resolution mass shifts. Where high-resolution mass spectrometry can resolve 15.99 (oxidation) from a closely related addition at 16.02 Da. Other amino acid substitutions such as Ala->Ser show exactly the same composition and therefore mass shift. B. The degree of formylation of Enolase 2 (*S. cerevisiae*) is quantified following a common shotgun proteomics experimental set-up with or without the use of formic acid as a solvent for protein solubilisation. The use of formic acid during sample handling introduced a considerable number of formylated peptide artefacts, as shown for Enolase 2.

However, identification of PTMs by the simple addition of mass increments (modifications) is restricted. Firstly, it assumes prior knowledge of the modifications present in a sample. Secondly, consideration of multiple modifications leads to an exponential increase in search space, impacting on computational efforts and challenging common statistical parameters, resulting in increased false negative as well as false positive identifications [58, 208, 209]. To overcome limitations of database-restricted approaches, alternative algorithms have been established, with the most common tools and latest developments being listed in **Table 2**. These open modification search tools do not require specifying the modification before analysis and can therefore also identify unexpected modifications. Current bioinformatics tools use different strategies such as multi-round search, *de novo* sequencing, sequence-tagging or spectral library-based approaches. PeaksPTM, for example, is a multi-round

search tool incorporated into the PEAKS proteomics software solution [210], enabling efficient searches for all modifications listed in the UNIMOD database simultaneously [211, 212].

Following identification of the proteins present in a sample, modifications are searched using a 'one-PTM-per-peptide' limitation to avoid exponential growth of the search space. During this process, sample-common PTM types are identified, generating a finite set of relevant modifications used to search for peptides containing two or more PTMs [212]. Another multi-round tool—G-PTM-D—employs a similar strategy, but uses a mass-tolerant open search strategy in the first round [213]. A further advanced version was published by Solntsev *et al.* (2018) that increases the speed and accuracy. Alternatively, *de novo* sequencing derives the peptide sequence from a tandem mass spectrum without a protein database. This strategy is used by deNovoPTM to identify modified peptides. However, algorithms are computationally demanding, consider only a restricted number of modifications and require high-quality fragmentation spectra [215]. A more efficient *de novo* PTM identification workflow, termed Open-pNovo, has been published by Yang *et al.* (2017) [216].

The first developed unrestricted search tools included sequence-tag based approaches [203, 217, 218]. Here, only *de novo* sequence-tags (extracted peptide sequence fragments of 3 to 4 amino acids) are required to find possible peptide matches from a sequence database. The differences between the expected and observed mass of the match are then assumed to be mutations or modifications. Sequence-tag-based approaches were then also the baseline for faster, further developed tools such as MODa, a 'multi-blind' spectral alignment algorithm [209]. Many software tools further employ hybrid approaches to improve accuracy and speed. Open-pFind, for example, uses first a sequence-tag based approach followed by a restricted search in which modification types and protein sequence entries are set by semi-supervised machine learning [219]. Kong *et al.* (2017) established a novel fragment-ion indexing method implemented in a database search tool termed MSFragger [220]. This tool provided a substantial improvement in search speed and made unrestricted modification searches feasible for particularly large data sets. More recently, TagGraph was established by Devabhaktuni *et al.* (2019) which is an unrestricted *de novo* sequence-tag approach utilising a fast string-based search including a probabilistic validation model optimised for PTM assignments [221]. Another commercial software package that enables advanced modification searches is Byonic [222]. This tool includes an option for Modification Fine Control™ that is also a fully unrestricted search approach for unanticipated modifications, termed Wildcard Search™ (**Figure 3a**). An alternative peptide modification search strategy utilises spectral libraries. Thereby, identification of modified peptides are interpolated from the identification of unmodified reference spectra [208, 223–226]. This process results in improved accuracy and higher identification rates but is limited to peptides being present in the database. More recent developments here include the ANN-SoLo tool from Bittremieux *et al.* (2018) [227] and the SpecOMS tool developed by David *et al.* (2017) [228]. It should

be noted that most of the recently developed algorithms have not been applied in microbial or yeast proteomics to any significant degree, potentially due to difficulties in obtaining accurate FDR estimations and the risk of extensive false positive identifications.

Table 2. Recently developed open modification tools for PTM discovery. *Many tools utilise hybrid-type approaches. **Commercial platform(s).

Software tool	(Main) approach*	Fully unrestricted?	Download link	Reference
MSFragger	Error tolerant search	Yes	http://www.nesvilab.org/software	[220]
DeNovoPTM	*De novo* sequencing	No	http://www.mybiosoftware.com/denovoptm-ms-based-peptide-identification-software-tool.html	[215]
Open-pNovo	*De novo* sequencing	No	http://pfind.ict.ac.cn/software/pNovo/index.html	[216]
TagGraph	*De novo* sequencing / error tolerant search	Yes	http://sourceforge.net/projects/taggraph	[221]
ANN-Solo	Spectral library / error tolerant search	Yes	https://github.com/bittremieux/ANN-SoLo	[227]
SpecOMS	Spectral library / error tolerant search	Yes	https://github.com/matthieu-david/SpecOMS	[228]
G-PTM-D	Multi-round / error tolerant search	Yes	https://github.com/smith-chem-wisc/gptmd	[213]
MetaMorpheus	Multi-round search	No	https://github.com/smith-chem-wisc/MetaMorpheus	[229]
PeaksPTM**	Multi-round search	No	http://bioinfor.net/ptm	[212]
Byonic**	Various types, including sequence tagging / error tolerant search	Yes	www.proteinmetrics.com/products/byonic/	[222]
MODa	Sequence tagging / error tolerant search	Yes	https://omictools.com/moda-2-tool	[209]
Open-pFind	Sequence tagging / multi-round / error tolerant	Yes	http://pfind.ict.ac.cn/software/pFind3	[219]
PIPI	Sequence tagging / error tolerant search	Yes	http://bioinformatics.ust.hk/pipi.html	[230]

2

Sample preparation artefact or natural PTM?

Whenever a peptide modification is detected, analysts have to make a decision regarding whether the observed modification is a genuine proteoform variant or whether it is in fact an artefact introduced during the experiment. However, this is sometimes a very delicate process. Formylation, for example, is a natural histone modification [231], which however can also be introduced during sample preparation when using formic acid-containing buffers to increase the solubility of hydrophobic peptides and aggregates (**Figure 3b**) [232]. The same holds for carbamylation, which is frequently introduced when using buffers containing (high-molarity) urea [233], or unspecific alkylation reactions introduced by extensive iodoacetamide treatment, broadly used during sulfhydryl alkylation reactions [234, 235]. Furthermore, chemically labile amino acid residues may undergo oxidation, deamidation, pyroglutamate formation, dehydration or metal ion adduct formation [236–238]. Many of those chemically introduced modifications may however occur naturally in protein ageing processes and are therefore difficult to discriminate from sample preparation artefacts regardless of the method employed [239]. In addition to preserving the native state of a peptide during sample preparation, chromatographic and ionisation properties during the analysis process also need to be chosen thoughtfully, particularly for modifications such as phosphorylation, which undergo rapid enzymatic hydrolysis and in-source fragmentation leading to neutral loss [240]. Where otherwise comparable protocols are used for cell lysis, protein extraction and proteolytic digestion compared to conventional discovery proteomics experiments, the analysis and interpretation of post-translational modifications requires additional careful considerations [241]. Harsh conditions not only produce ambiguous identifications, but also induce mass spectrometric signal multiplications, reducing the discovery rates for both native and naturally modified peptides [233, 242, 243].

Maximise proteome- or sequence coverage?

The proteome-wide discovery of post-translational modifications is challenged by factors such as sub-stoichiometric occurrence, competitive ionisation (sensitivity), computational limitations and lack of effective validation strategies. Thus, when a proteome-wide analysis of a certain type of modification is performed, peptide fractionation techniques are employed to reduce the number of signals and therefore increase sensitivity. Suitable methods are based on any physicochemical properties such as size, charge or hydrophilicity but also on affinity [244, 245]. Alternatively, targeted mass pre-selected subset of modifications has also been performed, but typically only for quantification purposes rather than discovery [246–248]. For tools supporting quantitation, we refer to a review from Allmer (2012) [249]. However, over recent decades, it has become apparent that most proteins are modified by more than one modification at a time, and many modifications do not function in an isolated manner but seem instead to interact with modification sites from the same or other proteins, a process referred to as PTM cross-talk [168–171, 250]. This process, however, is changing the view

on how protein modifications are ideally investigated. For example, for the sake of improving sensitivity, an enrichment for a specific modification increases coverage for this particular modification tremendously, but it may abolish information essential for understanding a complete process.

Hence, to explore the functional aspects and interactions, the analysis should aim to maximise protein sequence coverage of related pathways. In bottom-up proteomics, trypsin is predominantly used for digestion of the proteome due to its high specificity and ease of use [24]. However, full sequence coverage is almost never achieved because digestion also generates peptides with sub-optimal length for MS detection [251]. To increase sequence coverage, multi-proteolytic digestion approaches have been proposed [252, 253]. Here, the proteome is subjected to digestion with multiple proteases in parallel, resulting in complementary parts of the protein sequence and thus higher sequence coverage. The use of trypsin in combination with LysC has therefore become common practice in the field of shotgun proteomics. Even though both proteases share lysine as a cleavage site, LysC/Trypsin digests were found most efficient to yield fully cleaved peptides [254]. Furthermore, the use of 4 alternative proteases (LysC, ArgC, AspN and GluC) in addition to trypsin led to nearly a 3-fold improvement of sequence coverage for proteins at low abundances in yeast [251].

In this context, prediction tools have been developed to support experiments designing a full protein sequence coverage. PTMselect is an example of such an open-source software tool, which simulates multi-enzyme digestion to tailor the optimal set of proteases for the discovery of global or targeted modification from any single or multiple proteins [255]. This approach allows sequence coverage to be achieved; however, it does not solve the sensitivity issues. On the other hand, sensitivity for labile or very large modifications, such as phosphorylation and glycosylation [256], could be increased using alternative fragmentation techniques such as electron transfer dissociation (ETD) [257–259].

Finally, using alternative bottom-up MS technologies to the commonly employed data-dependent acquisition (DDA), could lead to increased detection of modified peptides. PTMs are generally observed at low stoichiometry and therefore not selected for fragmentation using a DDA approach, in which only ions with highest intensity are chosen. To overcome this stochastic precursor ion selection, DIA methods could be employed [260]. Here, all precursor ions are systematically fragmented in predefined retention time and precursor ion mass to charge (m/z) range. However, proper data analysis software tools should be utilized to correctly identify, localise and quantify the modifications [261, 262]. Moreover, these tools are important for discrimination of co-isolated modified peptide isoforms resulting from large precursor isolation windows [261].

Outlook

Research on post-translational modifications has advanced the understanding of protein phosphorylation in metabolic flux control and the understanding of modification cross-talk in yeast [263, 264]. However, many of the latest developments for the analysis, discovery and quantification of larger sets of post-translational modifications still challenge the field. Because many proteins undergo more than one modification at a time, a comprehensive exploration will require an examination beyond the most commonly investigated modifications such as phosphorylation and acetylation, by a simultaneous increase in protein sequence coverage.

A recent study on fission yeast by Telekawa *et al.* (2018) demonstrated the comprehensive characterisation of a protein complex following affinity purification [265]. This work provided particularly high sequence coverage and gave insight on almost 40 modification sites of three different types of modifications within one complex. A similar study was performed by Šoštarić *et al.* (2019), who demonstrated the impact of acetylation and phosphorylation on subunit interaction in 3 large yeast complexes [266].

Considering that many modifications can influence binding affinities, modifications are often considered to be functionally associated [267]. A phosphoproteomics study in yeast illustrated that phosphorylated proteins engage in many more protein-protein interactions than their unmodified counterparts [268]. A better understanding of the impact of modifications on protein complex formation and on protein-protein or enzyme-substrate interactions may open effective intervention points and targets for engineering.

Acknowledgements

Supplementary information

Supplementary Excel tables can be found online: https://doi.org/10.1093/femsyr/foz088

Chapter 3

A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications

Maxime den Ridder, Ewout Knibbe, Wiebeke van den Brandeler, Pascale Daran-Lapujade and Martin Pabst

Abstract

The importance of obtaining comprehensive and accurate information from cellular proteomics experiments asks for a systematic investigation of sample preparation protocols. In particular when working with unicellular organisms with strong cell walls, such as found in the model organism and cell factory *S. cerevisiae*. Here, we performed a systematic comparison of sample preparation protocols using a matrix of different conditions commonly applied in whole cell lysate, bottom-up proteomics experiments. The different protocols were evaluated for their overall fraction of identified spectra, proteome and amino acid sequence coverage, GO-term distribution and number of peptide modifications, by employing a combination of database and unrestricted modification search approaches. Ultimately, the best protocols enabled the identification of approximately 65–70% of all acquired fragmentation spectra, where additional *de novo* sequencing suggests that unidentified spectra were largely of too low spectral quality to provide confident spectrum matches. Generally, a range of peptide modifications could be linked to solvents, additives as well as filter materials. Most importantly, the use of moderate incubation temperatures and times circumvented excessive formation of modification artefacts. The collected protocols and large sets of mass spectrometric raw data provide a resource to evaluate and design new protocols and guide the analysis of (native) peptide modifications.

Introduction

Despite recent advancements in mass spectrometric instrumentation, a large fraction of fragmentation spectra (MS/MS) from bottom-up shotgun proteomics experiments usually remain unidentified [60, 206]. Amongst the many possible reasons, a decreased identification rate can result from the presence of non-peptidic contaminants (particularly surfactants), increased spectral complexity due to co-fragmentation of multiple precursor ions, low-quality spectra due to poor and incomplete peptide fragmentation, unexpected peptide sequence variants or co/post-translational modifications that are not covered by the database, as well as deficiencies of the database search scoring schemes used to match the spectra. The latter led to the development of iterative approaches or to the combined use of orthogonal database search algorithms [45, 46, 269–271]. However, a significant fraction of unassigned MS/MS spectra may also be a consequence of incomplete or nonspecific proteolytic cleavage or result from unintended peptide modifications introduced during sample preparation when using highly reactive chemicals. In fact, sample preparation can be one of the most significant contributors to data variation and poor comparability between proteomics experiments [177]. Hence, selecting the most suitable sample preparation protocol is essential for enabling deep proteome coverage and for confidently identifying and quantifying (novel) peptide modifications. Moreover, the (native) *in vivo* state of the proteins should be preserved as good as possible. Therefore, sample preparation protocols usually include quenching of the cellular metabolism subsequently after sampling. This is particularly important when investigating reversible post-translational modifications, that may be rapidly cleaved by the many hydrolytic enzymes present in every cell [240]. However, different quenching strategies, such as using ethanol or trichloroacetic acid (TCA) [242, 272], have been shown to differently impact the outcome of proteomics experiments [272]. Following protein extraction, proteins are solubilised and commonly denatured (*e.g.*, with Urea). Disulphide bonds of proteins are subsequently reduced by *e.g.*, Dithiothreitol (DTT) or Tris(2-carboxyethyl)phosphine (TCEP) and alkylated using reagents such as iodoacetamide (IAA) or acrylamide (AA) to avoid reoxidation of the sulfhydryl groups [235, 273]. Finally, the proteins are proteolytically cleaved (typically with trypsin) to create protein fragments that are sufficiently small for efficient measurement, but that retain sufficient unique sequence information for the subsequent protein identification step. Chemicals, or their employed concentrations, are often not compatible with the following steps in the protocol and, therefore, require removal by methods such as protein precipitation *e.g.*, with acetone or Trichloroacetic acid (TCA), or filter-aided approaches [274–280]. The type and pH of the buffer used during the final proteolytic digestion may ultimately also impact the amino acid cleavage site specificity [281]. Lastly, non-peptidic compounds and non-volatile salts are removed before analysis by the aid of solid-phase extraction (SPE), using different types of reverse-phase or mixed-mode resins [282].

However, combinations of different sample preparation procedures are expected to differently bias the proteomic analysis outcome. The selected protocol, therefore, may influence the proteome fraction or native modifications that can be detected and quantified significantly [283]. In addition, when investigating the biological significance of peptide modifications, the observed modification needs to be traced back to its co/post-translational or sample processing origin. However, without appropriate controls, this is a highly challenging procedure. For example, formylation can occur as a natural histone modification [231], but may also be introduced during sample preparation when using formic acid-containing buffers, such as those used to increase the solubility of hydrophobic peptides [232, 284]. Similarly for carbamylation, this modification has been associated with severe renal and cardiovascular disorders; however, it may also originate from high molarity Urea-containing buffers used during protein denaturation [285]. Alkylation of cysteine residues has become standard practice in proteomics experiments, which however due to over- or off-target alkylation reactions frequently introduces unintended peptide modifications. For example double alkylation artefacts, which for the case of IAA (114.04 Da) mimics the C-terminal glycine residue of ubiquitin [234, 235, 273, 286]. Nevertheless, recent studies suggest that disulphide reduction and cysteine alkylation may not be essential to achieve a good proteome coverage. This would therefore eliminate the exposure of the peptides to highly reactive alkylating reagents [287]. Furthermore, chemically labile amino acid residues may undergo sample preparation induced oxidation, deamidation, pyroglutamate formation, dehydration or excessive metal ion adduct formation [236–238]. Many of those modifications (or adducts) may, however, also occur as consequence of the natural protein ageing processes within the cell and are, therefore, difficult to discriminate from sample preparation artefacts [239].

A systematic evaluation – using a matrix-like approach, which leaves out one chemical at a time – and which investigates the impact on i) the % of identifications, ii) proteome coverage, iii) GO-term distribution and iv) post-isolation modifications, has not been performed for a unicellular organism to date. We therefore constructed a matrix of conditions, which include steps frequently employed in whole cell lysate proteomics and applied it to the preparation of well-controlled chemostat grown yeast cells. The outcomes of the proteomics experiments were compared for their obtained proteome and amino acid sequence coverage and for their GO-term profiles. Moreover, we performed an unrestricted modification search using TagGraph, to identify reagent-induced modifications. Finally, we investigated the quality of the unidentified spectra using *de novo* peptide sequencing. In summary, this study provides a systematic evaluation of sample preparation protocols for bottom-up proteomics experiments of the increasingly studied model organism and cell factory yeast. Thereby, we demonstrate the large impact of the different sample preparation procedures on proteome coverage and overall identification rates.

Ultimately, the performed study and publicly available large proteomics datasets, provide a valuable resource to select for the most suitable sample preparation elements for different experiments, and support the analysis of (native) peptide modifications.

Materials and Methods

Yeast strain, growth media and storage. In this study, we used the minimal glycolysis (MG) yeast strain IMX372 (*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1::KlURA3, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4*) [120], which under the selected growth conditions shows no phenotypic alterations compared to the parent CEN.PK lineage [288]. Shake flask and chemostat cultures were grown in synthetic medium (SM) containing 5.0 $g·L^{-1}$ $(NH4)_2SO_4$, 3.0 $g·L^{-1}$ KH_2PO_4, 0.5 $g·L^{-1}$ $MgSO_4·7H_2O$ and 1 $mL·L^{-1}$ trace elements in demineralized water. The medium was heat sterilized (120˚C) and supplemented with 1 $mL·L^{-1}$ filter sterilized vitamin solution [289]. For shake flask cultures 20 $g·L^{-1}$ heat sterilized (110˚C) glucose (SMG) was added. In chemostat cultures, 7.5 $g·L^{-1}$ glucose was added, and the medium was supplemented with 0.2 $g·L^{-1}$ antifoam Pluronic PE 6100 (BASF, Ludwigshafen, Germany). Frozen stocks of *S. cerevisiae* cultures were prepared by the addition of glycerol (30% v/v) in 1 mL aliquots for storing at -80˚C.

Yeast chemostat cultures and sampling. Aerobic shake flask cultures were grown at 30°C in an Innova incubator shaker (New Brunswick™ Scientific, Edison, NJ, USA) at 200 rpm using 500 mL round-bottom shake flasks containing 100 mL medium. Duplicate aerobic chemostat cultures were performed in 2 L laboratory fermenters (Applikon, Schiedam, The Netherlands) with a 1 L working volume in duplicate. SM-medium was used and maintained at pH 5 by the automatic addition of 2 M KOH. Mixing of the medium was performed with stirring at 800 rpm. Gas inflow was filter-sterilized and compressed air (Linde Gas, Schiedam, The Netherlands) was sparged to the bottom of the bioreactor at a rate of 500 $mL·min^{-1}$. Dissolved oxygen levels were measured with Clark electrodes (Mettler Toledo, Greinfensee, Switzerland). The temperature of the fermenters was maintained at 30°C. The reactors were inoculated with exponentially growing shake flask cultures of *S. cerevisiae* strain IMX372 to obtain an initial optical density (OD_{660}) of approximately 0.4. Following the batch phase, the medium pump was switched on to obtain a constant dilution rate of 0.10 h^{-1}. Chemostat cultures were assumed to be in steady state when, after five volume changes, the culture dry weight, oxygen uptake rate and CO_2 production rate varied less than 5% over at least 2 volume changes. The dilution rate and carbon recovery were determined after each experiment.

Analytical methods. OD_{660} measurements to monitor growth were performed on a JENWAY 7200 spectrophotometer (Cole-Parmer, Stone, UK). The biomass dry weight was

determined in duplicate by extracting 10 mL of broth and filtrating it over a filter with 0.45 µm Ø pores while a vacuum was applied to the filter. The filters were washed twice with 10 mL of demineralized water. Prior to use, the filters were dried in the oven for at least 24 hours at 70°C. After filtration, the filters were dried in a microwave oven at 360 W for 20 min leaving only dry biomass. For extracellular metabolite determinations, broth samples were centrifuged for 5 min at 13,000 g and the supernatant was collected for subjection to an Aminex HPX-87H ion exchange column (Agilent, Santa Clara). The HPLC was operated at 60°C and 5 mM of H_2SO_4 was used as mobile phase at a rate of 0.6 mL·min^{-1}. Off-gas concentrations of CO_2 and O_2 were measured using an NGA 2000 analyzer. Proteome samples (1 mL, at approx. 3.6 g·L^{-1} dry weight) were taken from steady state cultures. The samples were collected in multifold in trichloroacetic acid (TCA) (Merck Sigma, Cat. No. T0699) with a final concentration of 10% or in five volumes of ice-cold methanol (MeOH) (Thermo Fisher, Cat. No. 15654570). Samples were centrifuged at 4000 *g* for 5 min at 4°C. Cell pellets were frozen at -80°C [272].

Proteomics sample preparation protocols (an extended version of all sample preparation protocols is provided in the SI document). Yeast cell culture pellets were resuspended in lysis buffer composed of 100 mM triethylammonium bicarbonate (TEAB) (Merck Sigma, Cat. No. T7408) containing 0.1%, 1% sodium dodecyl sulphate (SDS) (Merck Sigma, Cat. No. L4522) or 8 M Urea (Merck Sigma, Cat. No. U5378) and phosphatase/protease inhibitors. Yeast cells were lysed by glass bead beating using a Mini-Beadbeater-16 (Biospec Products, USA) and thus shaken 10 times for 1 minute with a bead beater alternated with 1 minute rest on ice. For in-solution methods, proteins were reduced by addition of 5 mM DTT (Merck Sigma, Cat. No. 43815) or 5 mM TCEP (Merck Sigma, Cat. No. C4706) and incubated for 1 hour or 30 min at 37°C or 56°C, respectively. Subsequently, the proteins were alkylated for 30 min or 1 hour at room temperature in the dark by addition of 15 mM iodoacetamide (Merck Sigma, I1149) or 50 mM acrylamide (AA) (Merck Sigma, Cat. No. A9099), respectively. Protein precipitation was performed by addition of four volumes of ice-cold acetone (-20°C) (Merck Sigma, Cat. No. 650501) or TCA to a final concentration of 20% and proceeded for 1 hour at -20°C or 30 min at 4°C, respectively. The proteins were washed twice with acetone and subsequently solubilized using 100 mM ammonium bicarbonate (ABC) (Merck Sigma, Cat. No. 09830). Alternatively, for filter-aided sample preparation (FASP), proteins were loaded to a filter (Merck-Millipore, Microcon 10 kDa, Cat. No. MRCPRT010) after bead beating and reduced by addition of DTT and alkylated with iodoacetamide, as described earlier. After alkylation, proteins were washed four times with TEAB and ABC buffers. For all protocols, proteolytic digestion was performed by Trypsin (Promega, Cat. No. V5111), 1:100 enzyme to protein ratio (v/v) and incubated at 37°C overnight. For filter-aided sample preparation protocols, peptides were eluted from the filters after digestion using ABC and 5% acetonitrile (ACN) (Thermo Fisher, Cat. No. 10489553) / 0.1% formic acid (FA) (Thermo Fisher, Cat. No. 10596814) buffers

consecutively. For all protocols, solid phase extraction was performed with an Oasis HLB 96-well µElution plate (Waters, Milford, USA, Cat. No. 186001828BA). Peptide fractions were eluted using MeOH buffer containing trifluoroacetic acid (TFA) (Merck Sigma, Cat. No. 302031), FA or ABC. Eluates were dried using a SpeedVac vacuum concentrator. Dried peptides were resuspended in 3% ACN / 0.01% TFA prior to MS-analysis to give an approximate concentration of 500 ng per µL.

Shotgun proteomic analysis. For each protocol an aliquot corresponding to approx. 750 ng protein digest was analysed using a one-dimensional shotgun proteomics approach [290]. Each sample was analysed in two technical replicates. Briefly, the samples were analysed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL·min^{-1} over a linear gradient from 5% to 30% solvent B over 90 min, then from 30% to 60% over 25 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 120 min. Solvent A was H_2O containing 0.1% FA, and solvent B consisted of 80% ACN in H_2O and 0.1% FA. Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70,000 resolution in full MS mode with a maximum ion injection time (IT) of 100 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 17,500 resolution with AGC target of 2E5 and IT of 75 ms, 2.0 m/z isolation width and normalized collision energy (NCE) of 28.

Mass spectrometric raw data processing.
De novo **sequence analysis.** *De novo* sequencing was performed using the algorithm available via PEAKS Studio X+ (Bioinformatics Solutions Inc., Waterloo, Canada) [210], allowing 10 ppm parent ion and 0.5 Da fragment ion mass error, and oxidation as variable modification, where the resulting *de novo* sequences were exported to 'de novo peptide.csv' files for further unrestricted modification search using TagGraph, as described below. **Taxonomic profiling.** Taxonomic purity assessment using the same *de novo* peptide sequences was performed as described recently by H.B.C. Kleikamp *et al.* (2021) [291].

Database searching. Database searching against the proteome database from *S. cerevisiae* (Uniprot, strain ATCC 204508 / S288C, Tax ID: 559292, June 2020, excluding 13 glycolytic isoenzymes) was performed using PEAKS Studio X+, allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, carbamidomethyl or acrylamide as fixed (or none), and methionine oxidation and N/Q deamidation as variable modifications. To control false-positive peptide identifications, a uniform 1% false discovery rate (FDR)

was applied to peptide spectrum matches (PSM), and subsequently the protein identifications required ≥2 unique peptides. Results from the PEAKS DB search were exported to 'proteins.csv' and 'DB search psm.csv' files, containing the identified proteins and identified DB search peptide-spectrum matches, respectively.

Unrestricted modification search. TagGraph [221] was used to perform unrestricted global peptide modification search using the (mzML-formatted) mass spectrometric raw data and the *de novo* sequences obtained from PEAKS Studio, using the yeast proteome database plus the CRAPome contaminant sequences [292]. The analysis was performed allowing for 10 ppm precursor mass tolerance, cysteine carbamidomethylation or acrylamidation as static modifications, and methionine oxidation as a differential modification, as described by Devabhaktuni *et al.* (2019) [221]. TagGraph.1.8 was installed on a Windows desktop Docker container, and the processing of multiple files was automated via a PowerShell script. An FDR of 1%, 10 ppm parent ion and a maximum absolute deviation of 0.1 Da between experimental and database modification mass were applied to the analysis results and these were exported to '.txt' files for further analysis. Interconversion of mass spectrometric raw data. Conversion of the mass spectrometric raw data was performed using peak picking 'vendor' into '.mzML' and '.mzXML' using the msConvertGUI tool (ProteoWizard) [293].

Identification of scans originating from glycopeptides. First, the fragmentation spectra were searched for the presence of glycan-typical HexNAc oxonium fragment ions (204.087, $[M+H^+]^+$) using functions from the Matlab 'sugar miner' script as described recently [294]. Those scans indicate glycopeptides–which remain unidentified by the chosen database, or open modification search parameters. Scans with strong oxonium ion signals were summarized in an '.xlsx' table.

Data processing and visualization. Overall number of modified peptides and types of modifications. The results of the unrestricted modification search using TagGraph were used to determine the overall volume as well as the types of modifications found after applying the different protocols. First, the mass shifts (=deviations from the unmodified peptide mass) were collected from the ".txt" files and binned using the Matlab 'histcounts' function, at a bin width of 0.01 Da. This procedure was done for the combined dataset as well as for each protocol separately. Mass shifts with more than 5 occurrences per averaged conditions were tested for significant changes across all protocols using Matlabs 'anova1' function. Mass shifts with significant changes ($p<0.01$) across all conditions (2 biological and 2 technical replicates) were visualized by Euclidean distance clustering using the Matlab 'clustergram' function, standardizing along the rows (mass shifts) of data, clustering along the columns of data, and then clustering along the rows of row-clustered data. Default color variation has been used which shows for values between -3 and 3, where values above and below show the same maximum color tone.

Spectral Average Local Confidence (ALC) score histograms. After the scan numbers of each protocol were allocated to different categories, the Average Local Confidence (ALC) score distribution for identified and unidentified MS/MS scans was determined to evaluate the quality of the unidentified spectra. The ALC score was extracted from the *de novo* sequences '.csv' files and were therefore only determined for the spectra that were *de novo* sequenced. The ALC scores of the identified and unidentified scans were binned separately using the 'histcounts' function in Matlab. The resulting distributions were plotted as a bar graph and exported as table.

Proteome and amino acid sequence coverage. The proteome coverage was determined using the PEAKS-DB search results 'proteins.csv'. For this, the number of proteins per protocol was calculated from proteins with >2 unique peptides per protein. The percentage of observed proteins for each protocol was subsequently calculated by normalizing to the number of total proteins in the *S. cerevisiae* (Uniprot) protein sequence database. The average protein sequence coverage per protocol was further extracted from the PEAKS 'proteins.csv' output files. The sequence coverage (%) was moreover used to determine the amino acid coverage, in which the sequence coverage of each identified protein was summed and related to the total number of protein sequences present in the sequence database. The average for the proteome and amino acid coverage was calculated for each biological replicate and the deviations from the average for each protocol were plotted as bar graphs. A two-tailed unpaired Student's t-test was performed to determine if the relative coverage change between protocols was statistically significant.

Ontology analysis. The 'proteins.csv' files obtained through PEAKS-DB were used to determine the differences in the cellular component distribution. Python 3.8 [295] was used to programmatically link the Uniprot accession numbers of the identified proteins (with 2 unique peptides) to the Gene Ontology (GO) [296] terms using Retrieve/ID Mapping function on Uniprot [211]. The GO library was imported using the 'goatools' module in Python to retrieve the cellular component terms in the GO hierarchy. Functions of the proteins (in absolute numbers) were summarized in pie charts based on their cellular component GO terms. A protein could be allocated to multiple cellular components.

Overall % of identified fragmentation spectra. The overall number of identified fragmentation (MS/MS) spectra for each protocol was determined by combining the outputs obtained from TagGraph, PEAKS database search and the sugar oxonium ion search. First, scan numbers that did not result into amino acid sequence candidates by PEAKS *de novo* sequencing were allocated to the 'No ALC' category. Next, identifications were extracted from the TagGraph files. Thereby, identified scans were allocated to two categories, 'unmodified peptides' and 'modified peptides '. The scan numbers of the peptides identified

with the second search engine were extracted from the PEAKS-DB search peptide-spectrum matches files 'DB search psm.csv'. The scan numbers were compared with the scans of the *de novo* sequences and TagGraph results. Identifications that were only found with PEAKS database search, were allocated to the 'second search engine' category. The category "sugars" represents the scans that were identified as potential glycopeptides. If scans were present in multiple categories, including TagGraph, the scans were allocated to TagGraph identifications (modified or unmodified peptides). If scans were identified containing sugar fragments and by PEAKS database search, the scans were allocated to PEAKS-DB search identifications. Furthermore, 'No ALC'-allocated scans that could be identified containing sugar fragments or via PEAKS database search, were allocated to one of the respective categories. Finally, the MS/MS scans that could not be identified by any of the before mentioned categories, were allocated to the 'unidentified scans' category. As final check, the summed scans of all established categories required to be equal to the sum of MS/MS scans in the raw mzXML files. The distribution of the scan identifications amongst different categories was visualized using stacked bar graphs, in which the number of scans in each category was normalized against the total number of MS/MS scans in percentages, for each protocol.

Data availability. Mass spectrometric raw data, protein sequence database, and search files have been deposited at ProteomeXchange server and are publicly available under the project code PXD026806.

Results and Discussion

Comparison of yeast whole lysate sample preparation protocols
A systematic comparison of a matrix of different sample preparation procedures for bottom-up yeast proteomics experiments was performed to investigate the impact on spectral identification rates and quality, achieved proteome coverage, amino acid sequence coverage, GO-terms distribution and reagent-induced peptide modifications (**Table 1**, **Figure 1**). Duplicate aerobic chemostat cultures of the IMX372 *S. cerevisiae* strain [120] were cultured in glucose-limiting conditions (Supplementary Table S1), which provided highly reproducible yeast cell biomass. Typical yeast shotgun proteomics experimental workflows are complex and consist of multiple steps, including sample collection, cell lysis, extraction and denaturation of proteins, (filter-aided) protein precipitation, proteolytic digestion, peptide purification (and/or enrichment) and finally, LC-MS/MS analysis (**Table 1**) [151, 279]. To this end, proteome samples were taken during steady-state conditions from both biological replicates, in which the cellular metabolism was quenched using ice-cold TCA or methanol to preserve the proteome and associated post-translational modifications. Cell lysis was performed with SDS- or Urea-containing buffers employing bead-beating in all cases. Crude cell lysates were then treated with an in-solution or FASP approach, in which proteins

were reduced with TCEP or DTT and alkylated with iodoacetamide, acrylamide or alternately, were left untreated. For in-solution digestion methods, proteins were precipitated with acetone or TCA. After overnight digestion with trypsin, solid-phase extraction (SPE) was performed to remove contaminants by using Oasis HLB cartridges, which make use of a co-polymer of divinylbenzene and vinyl pyrrolidinone that shows an enhanced retention of polar peptides (Waters). Elution from the cartridges was achieved with a variety of buffers to assess the influence of different additives (**Table 1**). Finally, the samples of the different protocols were analysed using a (short) one-dimensional gradient with duplicate injections.

Table 1. Matrix of investigated sample preparation protocols (1–14). Abbreviations: SPE, solid phase extraction; TCA, trichloroacetic acid; MeOH, methanol; SDS, sodium dodecyl sulphate; DTT, Dithiothreitol; TCEP, Tris(2-carboxyethyl)phosphine; IAA, iodoacetamide; AA, acrylamide; TFA, trifluoroacetic acid; FA, formic acid; ABC, ammonium bicarbonate. The detailed protocols are provided in the S.I. materials.

3

#	Sampling method	Lysis buffer	Reducing agent	Alkylation reagent	Protein purification	SPE buffers
1	TCA	SDS	DTT	IAA	Acetone	TFA
2	TCA	SDS	DTT	IAA	Acetone	FA + ABC
3	TCA	SDS	DTT	IAA	Acetone	ABC
4	TCA	SDS	DTT	IAA	Acetone	MeOH
5	TCA	SDS	DTT	IAA	Acetone	FA
6	MeOH	SDS	DTT	IAA	Acetone	TFA + ABC
7	TCA	SDS	DTT	IAA	Acetone	TFA + ABC
8	TCA	SDS	DTT	IAA	TCA	TFA + ABC
9	TCA	SDS	DTT	AA	Acetone	TFA + ABC
10	TCA	SDS	DTT	-	Acetone	TFA + ABC
11	TCA	SDS	TCEP	IAA	Acetone	TFA + ABC
12	TCA	UREA	DTT	IAA	Acetone	TFA + ABC
13	TCA	UREA	DTT	IAA	FASP	TFA + ABC
14	TCA	SDS	DTT	IAA	FASP	TFA + ABC

On average, 35420±5909 (biological replicate 1) and 37261±8271 (biological replicate 2) MS/MS scans were obtained across all protocols (**Figure 2**, **Supplemental Figure S1**). Furthermore, the analysis of the protocols resulted in 23631±5237 and 23477±6590 peptide spectrum matches (PSMs) (**Figure 2a**), resulting in 66% and 61% MS/MS spectrum identifications on average (**Figure 2b**). However, protocols 8 and 14 resulted consistently in a considerably lower number of MS/MS spectra for both biological replicates, albeit that a comparable amount of proteolytic digest (750 ng) was analysed.

For protocol 8, protein precipitation was performed with TCA/acetone instead of acetone solely. Difficulties in re-dissolving the protein pellet can arise for the TCA/acetone precipitation approach and may ultimately impact the overall recovery. It therefore becomes necessary to use stronger buffers, larger volumes and additional mechanical disruption of the pellet, to solubilise the protein [275, 278, 297]. In this study, protein pellets were dissolved in ammonium bicarbonate buffers to allow subsequent trypsin digestion, which did not prove difficult for acetone-precipitated samples. However, the TCA/acetone procedure led to a partially insoluble pellet, and a considerably lower number of proteins were subsequently identified, compared to when only using acetone-precipitation (20% vs. 29% proteome coverage for protocol 8 = TCA/acetone, vs protocol 7 = acetone; **Figure 2c**). Similar results were observed in other studies, which was attributed to the increased protein denaturation caused by this approach [275, 298–300].

Samples treated according to protocol 14 were subjected to FASP using an SDS containing buffer. SDS is known to interfere with binding and elution during the reverse-phase separation of the peptides and severely suppresses ionisation by electrospray ionisation [301–303]. Multiple rounds of centrifugation were applied to ensure SDS removal, however, residual SDS may have still impacted the LC-MS/MS analysis, resulting in a lower total number of MS/MS scans (13232 MS/MS spectra on average for protocol 14, vs. 36340 MS/MS spectra on average across all protocols). However, an earlier study buffer exchanged SDS with Urea following solubilisation of the proteins, allowing for high protein identification rates, because SDS was likely successfully withdrawn from the sample [279]. Hence, FASP sample preparation in combination with SDS-containing buffers is not recommended unless the complete removal of SDS can be ensured.

Figure 1. Whole cell lysate sample preparation study for yeast and established data processing pipeline. A. The employed study performed yeast cultivation in aerobic chemostats, sample collection and quenching, cell lysis by bead beating, protein extraction from cells and subsequent reduction and alkylation of the proteins. Proteins are then concentrated using protein precipitation or a filter-aided approach. Enriched proteins are subsequently digested using Trypsin, where the obtained peptides are subjected to solid phase extraction (SPE) to remove contaminants prior to shotgun proteomic analysis. B. The obtained MS/MS spectra were de novo sequenced using PEAKS and subsequently identified through database (PEAKS-DB) and unrestricted modification searches (TagGraph). A data analysis pipeline was established to determine the % of identified spectra, spectral quality, proteome and amino acid sequence coverage. Peptides identifications obtained by TagGraph were furthermore used to determine the modification profiles for every protocol. Finally, proteome were annotated with GO-terms to investigate the distribution according to 'cellular components'.

Overall proteome coverage is strongly affected by sample preparation procedures

MS/MS spectra were identified using database searching employing common statistical filtering criteria (see materials and methods) and requiring at least 2 unique peptides per protein identification. The proteome coverage was calculated based on the proteins identified per protocol as a percentage of the total number of proteins in yeast (known ORFs). An average proteome coverage of approximately 26% and 27% was achieved amongst the different methods for biological replicates 1 and 2, respectively (**Figure 2c**). Small differences in the average depth of the proteome coverage between both biological replicates were observed, most likely because the experiments were conducted at different time periods and, therefore, the instrumental performance may have been slightly different. Three protocols (6, 8 and 14) differed substantially from the other procedures. As discussed above, for protocols 8 and 14, LC-MS/MS analysis was likely compromised by incompatibility with the used reagents.

Protocol 6, on the other hand, resulted in an average number of 39152 MS/MS spectra, which is slightly higher than the average of 36341 MS/MS spectra across all protocols (**Figure 2b**). Since the total ion count was very comparable, its suggestive that a similar amount of proteolytic digest was injected to the LC-MS system. This was the only sample that was subjected to quenching with methanol as opposed to TCA during sampling. Methanol quenching (applied to yeast) might, therefore, significantly impact the number of identifiable proteins. TCA as a quenching solution has been also proven useful in quantitative phosphoproteomics measurements recently [272]. Methanol, on the other hand, is routinely employed to rapidly arrest the cellular metabolism when performing metabolomics studies [304]. Methanol has been also used to co-extract metabolites and proteins from yeast, only recently [305]. In the present study, the methanol quenching bias was observed consistently for all biological and technical replicates. Protein aggregation due to exposure to methanol followed by poor resolubilisation is likely the cause for the low proteome coverage, however, the exact mechanism remains to be explored. Nonetheless, the highest proteome coverage (31% and 32%, for biological replicates 1 and 2, respectively) was obtained with sample 13, in which a filter-aided approach was used in combination with Urea buffer. Similar results have been found in previous studies, in which FASP outperformed in-solution approaches in terms of the number of protein identifications [306, 307].
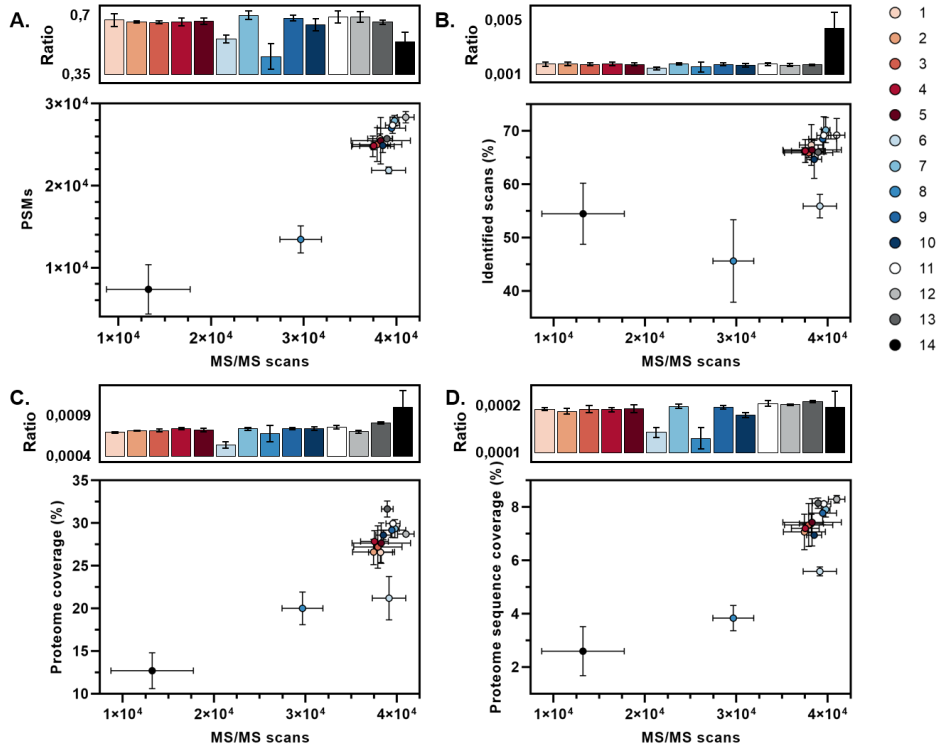
Figure 2. Achieved % of identified fragmentation spectra, proteome and sequence coverage for the different sample preparation protocols. The number of Peptide Spectrum Matches (PSMs) (A); identified scans (%) (B); proteome coverage (%) (C) and proteome sequence coverage (D), were plotted against the number of MS/MS scans obtained per protocol. Each coloured circle represents the average of averages of two biological replicates with each 2 technical replicates (2x2), while the standard deviation is represented by the error bars. In addition, the bars depicted above each plot show the ratios of the acquired PSMs, identified scans and proteome (sequence) coverage vs. MS/MS scans for each protocol. The ratios are averages of two biological replicates with each 2 technical replicates (2x2), while the standard deviation is shown as error bars. The proteome coverage was calculated based on the identified proteins per protocol as a percentage of the total number of proteins in yeast (known ORFs). In addition, the proteome sequence coverage was calculated based on the identified amino acids in the proteomics experiments, as a percentage of the total proteome amino acid sequence (sequence coverage).

The proteome coverage does not seem to have been significantly impacted by the type of detergent used during lysis (SDS or Urea) for samples 7 and 12, respectively. Typically, SDS is used as an ionic and denaturing surfactant that can disrupt cell membranes and cause protein denaturation by disrupting protein–protein interactions. Urea, on the other hand, is a chaotrope that can bind to proteins, thereby causing protein unfolding [308]. Following protein extraction, protein reduction was performed by either TCEP or DTT. The type of reducing agent used during sample processing was also not a key determinant in the outcome of the proteome analysis. This was very similarly observed for yeast in another study [273].

In this study, proteins were subsequently alkylated with iodoacetamide, acrylamide or alkylation was left out completely (samples 7, 9 and 10, respectively). Expectedly, and in line with a recent publication [287], the absence of an alkylation step dramatically reduced the detection of cysteine-containing peptides, although the depth of the proteome coverage remained nearly unchanged. However, the number of MS/MS scans and obtained peptide spectrum matches, as well as sequence coverage, was steadily lower for the non-alkylated sample (protocol 10), when compared to the alkylated samples (protocols 7 and 9).

Furthermore, no significant changes in the number of identified proteins could be observed for the different solvents used for solid-phase extraction (protocols 1–5, 7, **Figure 2** and Supplemental **Table S2** and **S3**). Though, the number of MS/MS scans, identified proteins and proteome (sequence) coverage was consistently higher for protocol 7 (**Figure 2**, **Supplemental Table S2 and S3**), in which a combination of basic and acidic MeOH buffers were used for elution, thereby maximising the recovery of peptides across a large pH range. Moreover, TFA (protocol 7) appeared to be a better choice for peptide elution compared to formic acid (FA) (protocol 2 and 5) because the proteome (sequence) coverage was repeatedly higher when using TFA compared to FA (**Supplemental Table S2 and S3**).

A gene ontology (GO) analysis [296] was performed to investigate whether the different protocols bias towards specific cellular components, such as cytosolic soluble proteins, or oppositely, towards hydrophobic membrane proteins. Therefore, GO terms were assigned to proteins based on their 'cellular component'. Thereby, the distribution of the cellular localisation of the observed proteins was comparable across the different protocols, which is exemplified in **Figure 3** for protocol 13 with the most represented GO terms (biological replicate 2, technical replicate 1, **Supplemental Table S4**). Detected proteins were predominantly assigned to intracellular organelle functional GO categories, consisting of *e.g.* cytosolic, mitochondrial and ribosomal proteins, which could be explained by their high expression levels [309, 310]. Irrespective of the protocol, proteins were extracted using bead beating, which is a relatively harsh and likely reproducible approach. Nevertheless, a recent study demonstrated that extraction methods can significantly contribute to the variability of proteomics experiments [177].
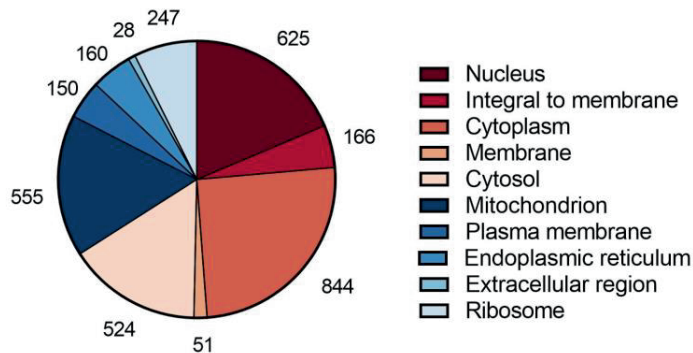
Figure 3. Distribution of Gene Ontology (GO) terms for identified proteins. Based on the classifications of GO annotation, the identified proteins (with at least 2 unique peptides) were categorized into cellular components and displayed in pie chart format, exemplified here for protocol 13 (biological replicate 2, technical replicate 1) with absolute protein numbers.

Commonly, proteins are already reported with a few peptides, *e.g.* 2 unique peptides. While acceptable for qualitative and quantitative analysis, further characterisation of individual proteoforms as well as analysis of post-translational modifications demand a higher sequence coverage. Therefore, we quantified the fraction of the total proteome amino acid sequence space covered for every protocol. An average proteome sequence coverage of 6.7% and 7.0% was observed across all protocols for biological replicates 1 and 2, respectively (**Figure 2d**). Hence, for the employed (relatively short) one-dimensional gradient—despite the good proteome coverage—a rather small fraction of the amino acid sequence space was actually observed in our experiments. In the present study, an average protein sequence of 26% and 25% was obtained amongst the identified proteins (**Supplemental Figure S2**). This value is proportional to the average amino acid sequence coverage of 28% obtained by Herbert *et al*. (2014), who obtained through extensive experiments a near-complete proteome coverage [151]. Nevertheless, the proteome coverage strongly depends on the protein amino acid sequences, and therefore, on the cleavage specificity of the employed protease(s) [311].

One of the challenges to improve proteome sequence coverage arises from signal suppression during electrospray ionisation (particularly for short one-dimensional gradients), where peptides with higher basicity tend to ionise preferably, and lower abundant peptides or highly hydrophobic or acidic peptides may remain undetected. Longer LC separations, additional peptide pre-fractionation, as well as 2-dimensional gradients significantly increases the number of identifications [26, 152, 244, 245]. Nevertheless, material requirements, sample processing and analysis time will increase proportionally, when using multi-dimensional

chromatography or off-line peptide fractionation. Finally, multi-protease digestion has been demonstrated to boost not only the protein but also the proteome sequence coverage [251, 253].

As observed for the proteome coverage, protocols 6, 8 and 14 resulted in a much lower proteome sequence coverage (**Figure 2d**, **Supplemental Table S2 and S3**). Overall, while lysis buffer, reducing agent or SPE elution buffers seemingly did not impact proteome coverage, the alkylation step was critical for obtaining a high amino acid sequence coverage.

The observed variability in peptide modifications
By using an unrestricted modification search (TagGraph) 221751 and 181496 mass shifts were observed across all used protocols, which derived from 991759 and 1043314 MS/MS spectra for biological replicates 1 and 2, respectively (**Figure 4**). A highly comparable mass shift profile was observed for both technical and biological replicates, confirming the reproducibility of the chemostat cultivation [184], the sample preparation protocols and the shotgun analysis. To discriminate between peptide modifications from biological (co/post-translational) and sample processing origin, we searched for alterations (or mass shifts) that were predominantly (or exclusively) present when using certain sample preparation protocols. Hence, we searched for mass shifts that showed a significant change in frequency ($p < 0.01$) across all conditions and replicates, of which the averaged occurrences across all protocols are shown in **Figure 5** (**Supplementary Table S5**). Albeit we employed combinations of well-established sample preparation steps, we observed a distinct number of condition-specific modifications, of which some had not been linked to sample preparation artefacts before.

For example, two different approaches to quench the cellular metabolism were employed, using either TCA or methanol. A mass shift of -89.04 Da was prominent in TCA-arrested yeast cells, with the notable exception of protocol 14 (SDS/FASP protocol). Furthermore, we used reagents for cell lysis and protein denaturation such as SDS and Urea. Interestingly, we did not observe any mass shifts related to the use of those reagents. In aqueous solutions, Urea dissociates upon heating over time. One of its degradation products is iso-cyanate, which can react with the N-termini of proteins/peptides and at the side-chain amino groups of lysine and arginine residues, to mimic *in vivo* carbamylation (+43 Da). This reaction is enhanced by high temperatures; however, Urea solutions in our study were only heated to 37°C for 1 hour, whereas 56°C is commonly used for protein reduction with DTT or TCEP [233]. Furthermore, the use of ammonium-containing buffers (and the removal of Urea before overnight digestion) seemed to minimise protein carbamylation when using Urea solutions [312], which was also observed in our study. Therefore, the relatively mild procedure used by our protocols seemingly avoided carbamylation reactions taking place.
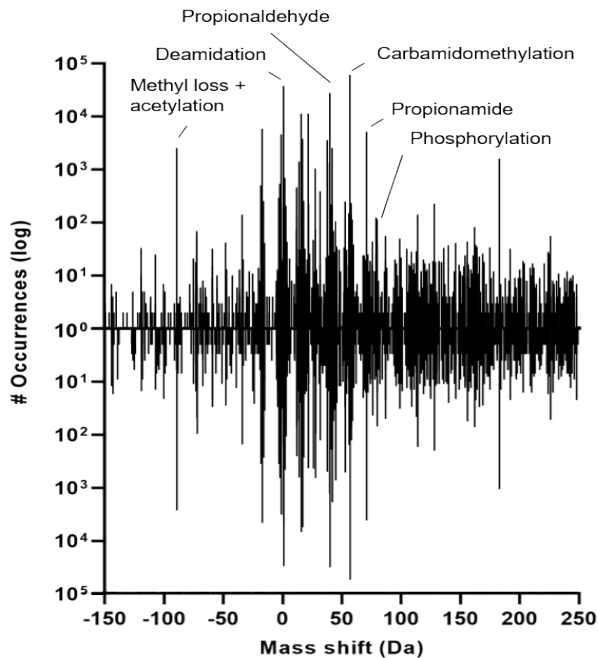
Figure 4. Sum of observed mass shifts in the different protocols using an unrestricted modification search approach (TagGraph). The number of occurrences corresponds to the number of peptide spectrum matches containing the mass deviation (log scale) for biological replicate 1 (upper histogram) and 2 (lower histogram). The total of both biological replicates show a highly comparable mass-shift profile.

No significant mass shifts were observed between samples reduced with TCEP or DTT. This is in line with an earlier study demonstrating that both reducing agents resulted in the same number of protein identifications [235, 273]. Nevertheless, acrylamide appeared there to be more efficient in combination with DTT, whereas iodoacetamide performed better with TCEP [235]. As expected, alkylation reagent-related mass shifts were the most abundant artefacts (*e.g.*, +57.01 for IAA with 61582 and 53462 occurrences). However, the frequently reported methionine alkylation and the subsequent neutral loss of -48 Da from the molecular peptide ion $[M-48+H^+]^+$, which is the loss of the side-chain from oxidised methionine [235], was rarely detected. Moreover, double alkylation with the characteristic mass delta of +114 Da was compared to counts for single alkylation (+57 Da) for all samples (on average) below <0.20%. That mass addition was also observed in the non-alkylated sample, suggesting that this mass shift resulted rather from ubiquitinylation ('GlyGly') and not from excessive alkylation. The modification analysis further indicated that the applied alkylation procedure was sufficiently mild (in particular when removing excess reagent before overnight

digestion) with only little off-target, and hardly any excessive, alkylation taking place. On the other hand, all protocols that used IAA showed additional +39.98 Da mass additions, which presumably derive from an N-terminal S-carbamoylmethyl- cysteine cyclisation product [234, 313, 314]. Acrylamide showed only some additional -73.05 Da mass shifts at low frequency, the mechanism of which, however, remains to be investigated. Interestingly, the protocol (10) employing reduction, but no alkylation resulted in an additional series of (albeit low frequent) mass shifts (such as -59.05, 79.9, 94.9, 99.0), which likely originate from reactions of the free sulfhydryl group with compounds (naturally) being present in cellular extracts.

An unexpected observation was the very abundant mass addition of +40.03 Da (labelled in Unimod as 'propionaldehyde' [164]). This mass shift was present in all protocols that employed protein precipitation with acetone but was strongly reduced for protocols that used TCA with acetone wash, and which was absent in protocols using FASP. This mass shift was in a recent study related to acetone artefacts, which in fact can impact some 5% of the proteome (peptides with glycine as the second N-terminal amino acid) [277]. The 40 Da mass shift, however, can also be misinterpreted for an amino acid substitution from glycine to proline. Acetone modification involves aldimine formation between the ketone and the amino groups of a peptide, which is regarded as being acid labile and, therefore, rather rearranges to a more stable form (imidazolidinone), resulting in the observed +40.03 Da mass addition [277]. Interestingly, we observed some few mass additions of +41.03 Da, which may indicate the presence of the proposed aldimine intermediate. Other artefacts that presumably also derive from acetone, such as +98 Da and to a lesser extent +84 and +112 Da [315], were only observed at trace levels or were not present at all. However, it is suggested that those artefacts seemed to be more pronounced at elevated pH and, therefore, our protocols might have prevented their formation. Even though acetone may increase peptide complexity substantially, this procedure is still commonly used in sample preparations, likely because it efficiently removes organic compounds such as inhibitors or compromising detergents, such as SDS [297].

The +28 Da mass adduct was nearly exclusively found for protocols in which formic acid was used during solid-phase extraction, confirming that yeast (at the given growth conditions), showed hardly any double methylation events. Previous studies provided similar findings, thereby highlighting that the proteome analysis outcome can be strongly affected by the use of this reagent [232, 284]. A recent work demonstrated that these unwanted modifications can be largely avoided by processing the samples at low temperatures [284].
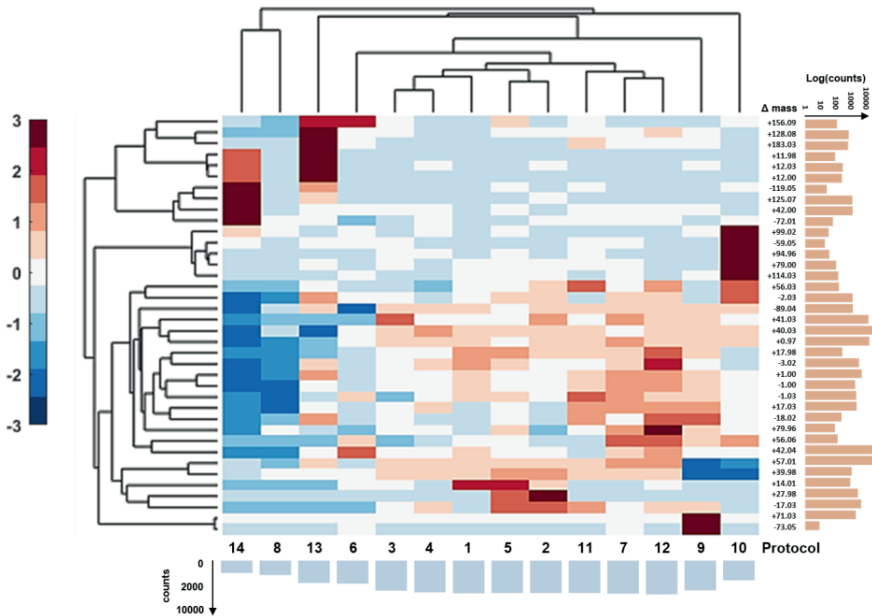
Figure 5. Clustered heat map of most prominent modifications identified by TagGraph across the different sample preparation protocols. Dendrogram and standardized clustered heat map of the relative change (%) of mass shifts observed across the sample preparation protocols. The mass shifts found for each protocol originating from each technical and biological replicate (2x2), were averaged per sample and normalized to the number of obtained MS/MS spectra. The bars below the heat map show the total counts of modifications observed for each sample and the bars on the y-axis show the total counts of the respective mass shift (modification) across the investigated samples.

Another interesting observation in regard to solid-phase extraction was that the number of ammonium adductions (+17 Da) was hardly impacted by the nature of the buffers used to elute the peptides from the solid-phase extraction cartridges. Those adducts were supposedly carried over from the ammonium bicarbonate buffers, used for proteolytic digestion.

For the samples prepared with the alternative filter-aided protocols, a number of mass shifts were observed that were detected at a much lesser frequency in samples prepared by in-solution digestion (*e.g.*, one protocol: +41.03, 42.04 or both protocols: -89.05, 11.0 and -2.0). On the other hand (albeit at low frequency), a series of mass modifications such as -119.0, -72.0, 11.98, 12.0, 42.0, 125.07, 128.07, 156.09 and 183.03 were increased or exclusively observed, when using filter-aided protocols.

Common naturally occurring modifications were also found in the top mass shifts that show a significant change in frequency across all protocols. Methylation (+14.01), for example, was detected in a higher frequency for protocol 1 and 5, representing protocols in which only acidic buffers (TFA and FA, respectively) were used during SPE. Another common post-translational modification is acetylation (+42.01), which appeared much more frequent in protocol 14 (FASP, SDS). However, the mass shifts of these naturally occurring modifications can also be a result of various amino acid substitutions [164].

Considerable variations in the fraction of overall identified spectra observed for the different sample preparation protocols

The number of identified MS/MS spectra for each sample was determined by combining identifications from TagGraph, PEAKS database search and the sugar oxonium ion fragment search. First, the spectra were *de novo* sequenced, where a fraction of all spectra (2.6 % and 3.8%, on average, for replicates 1 and 2, respectively) did not provide any *de novo* spectra, which were further allocated to the 'unsequenced' category ('no ALC score', **Figure 6, Supplemental Table S3**). The *de novo* sequences were then used to perform an unrestricted peptide (modification) search using TagGraph, which resulted in the identification of 28% and 26% of unmodified and 17% and 14% of modified peptides, on average, for replicates 1 and 2, respectively (at a 1% peptide FDR). Moreover, PEAKS was used as an additional/orthogonal search engine to maximise identifications of unmodified peptides. Thereby, the number of confidently identified peptides considerably increased to 20% and 22%, on average, for replicates 1 and 2, respectively. Because different search engines employ individual approaches and scoring matrices, a combination of multiple search engines typically achieves significantly more matches [41–48]. The PEAKS database (PEAKS-DB) is a hybrid approach that combines elements of *de novo* sequencing (short sequence tag extraction) and database searching [316]. TagGraph, contrarily, is an unrestricted *de novo* sequence-tag approach employing fast string-based searches followed by a probabilistic validation model optimised for the identification of modified peptides [221]. Furthermore, spectra were screened for known carbohydrate fragments, which indicate a glycopeptide spectrum that would likely remain unidentified by any of the above-employed approaches and, therefore, were allocated to the category 'sugars' (0.8% and 0.6%, on average, for replicates 1 and 2, respectively). Finally, the remainder of the *de novo* sequences that could not be allocated to one of the aforementioned categories were categorised as 'unidentified'. Here, approximately 31 and 34% remained unidentified for replicates 1 and 2, respectively.
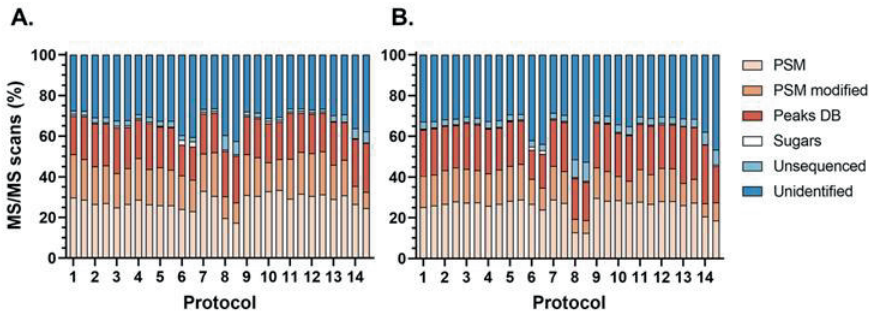
Figure 6. Overall spectral coverage obtained for the different sample preparation protocols. Allocation of total number of MS/MS spectra of each sample (1–14) into different categories: PSMs (TagGraph), PSMs modified (TagGraph), identified spectra by second search engine (PEAKS-DB), spectra containing oxonium fragments, 'unsequenced' and unidentified spectra for biological replicate 1 (A) and 2 (B).

As mentioned above, small differences in the identification rates between the two biological replicates were observed, presumably due to operational performance differences of the mass spectrometer at the different time periods. Still, a very similar pattern in allocation of the sequencing spectra is observed between the replicates (**Figure 6**). Expectedly, protocols 6 (methanol quenching), 8 (TCA precipitation) and 14 (FASP with SDS buffer) resulted in the lowest number of spectral identifications due to possible protein aggregation and following losses of protein during precipitation and LC-MS/MS incompatibility of reagents, respectively. Furthermore, the number of additional identifications with PEAKS-DB was considerably lower than for the other protocols when the samples were quenched with methanol. Conversely, a higher number of spectra from potential glycopeptides was observed using this protocol. This observation might be explained by the slightly increased fraction of plasma membrane proteins (Supplementary Table 4), when using this protocol (number 6). The other protocols resulted in highly similar profiles and identification rates, with few exceptions. A considerably lower number of modified peptides was found for protocols 10 (no alkylation) and 13 (FASP, Urea). As speculated in the previous paragraph, this presumably results from the loss of cysteine-containing peptides, and acetone induced modifications for protocols 10 and 13, respectively. Finally, sample 13 (FASP with Urea buffer) resulted in one of the highest identification rates with a relatively low number of modified peptides. Nevertheless, a large fraction of all spectra remained unidentified.

Investigating 'the nature of the unidentified fraction'

As also observed in other studies [60, 206], approximately one third of the MS/MS spectra remained unidentified in the current study across all protocols. Recurrently, protocols 6, 8 and 14 led to a significantly higher fraction of unidentified spectra, on average, 44%, 54% and 45%, respectively. Other protocols obtained similar results, ranging from 30–35% unidentified scans. Protocol 7 (in-solution, IAA) resulted in the highest number of identified spectra (70%), showcasing the key role of the type of protocol used on the identification rate. The inability to identify spectra can have various causes, such as impurities (or contaminations) in the sample, excessive co-fragmentation, poor quality spectra, unexpected modifications or peptide sequences not covered by the database (*e.g.*, sequence variants and microbial impurities). First, the taxonomic purity of the samples was confirmed by a recently published *de novo* metaproteomic profiling approach [291], in order to exclude the presence of sample carry-over from foregoing analyses and to exclude the presence of bacterial contaminants. This was performed with protocol 13 for both biological replicates, for which no impurities could be identified (data not shown), confirming that no carry-over took place and that the yeast cultures were of highest purity.

A high quality of the fragmentation spectra is a prerequisite for confident peptide spectrum matching. Therefore, we assessed the quality by *de novo* sequencing and the associated average local confidence score (ALC), as determined by PEAKS [210]. In general, peptides with higher ALC scores indicate a more complete peptide fragment ion coverage and, therefore, are expected to have a higher chance of correct identification. As example, the ALC scores of the (un)identified spectra of biological replicate 1 processed with protocol 13 is visualised in **Figure 7** (**Supplemental Table S6**). The ALC distribution is, as expected, very different for the fraction of identified and unidentified spectra. The ALC scores of the unidentified spectra spread predominantly around low numbers, whereasall identified spectra showed very high ALC scores). This suggests that reduced quality (peptide fragment ion coverage) is the main reason for the lack of identification. Nevertheless, a small fraction of the unidentified spectra showed high ALC scores. Most of the *de novo* sequences were close to sequences in the yeast protein sequence database and may therefore represent sequence variants (*e.g.*, single nucleotide polymorphisms SNPs, or isoform- and allele-specific variants of proteins) or additional peptide modifications, which were not covered by the database or could not even be identified by the open modification search approach. Furthermore, peptides containing unexpected semi-tryptic or nonspecific cleavages may also add to this fraction [281].
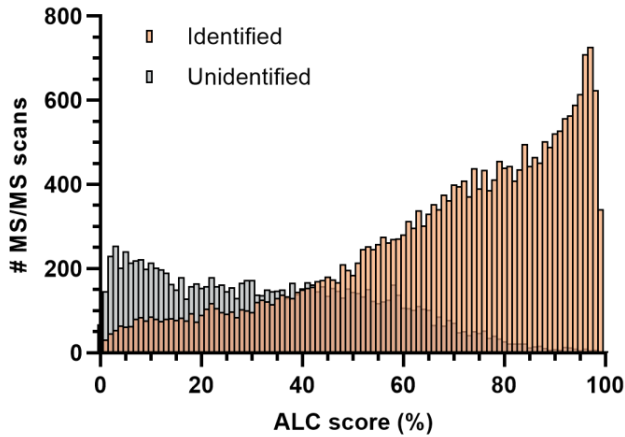
Figure 7. Assessing the spectral quality of the acquired fragmentation spectra. The Average Local Confidence (ALC) score distribution of identified (orange) and unidentified (grey) spectra, as shown for protocol 13 (biological replicate 1). In general, peptides with higher ALC score are more likely to provide a confident match during database search (because of the better fragment ion coverage of the peptide backbone, and the possibly 'cleaner' spectrum). The unidentified fraction from protocol 13 (grey) shows an ALC score distribution at predominantly lower numbers, indicating that the unidentified spectra largely derive from very low-quality spectra, insufficient to provide a confident peptide sequence match. The fraction of higher ALC scoring spectra presumably derive from sequence variants or unexpected modifications not present in the database. Taxonomic impurities or sample carry over can be excluded.

Conclusion

Our systematic study on the unicellular organism yeast demonstrates the strong impact of several elements within sample preparation protocols, on various aspects of the proteomics analysis outcome. The total number of obtained peptide spectrum matches ranged from approx. $1x10^4$ to $3x10^4$, the proteome coverage ranged from below 15% to approx. 30%, and the overall matched spectra varied in the range from approx. 50 to 70%. The depth of the proteome coverage was heavily affected by the sample quenching protocol, where cells arrested by TCA resulted in a higher number of protein identifications compared to methanol-arrested cells. Furthermore, filter-aided procedures, when combined with Urea, outperformed other protocols in terms of the number of protein identifications.

The use of unrestricted modification search moreover enabled to examine for reagent-induced peptide modifications. The overall frequency of reagent-specific modifications was for some protocols significantly lower (particularly for protocols 8 and 14), albeit that also

correlated with the lower number of spectra and decreased total ion current (TIC) observed for those protocols. Our analysis confirmed previously identified sample processing induced modifications, but also revealed unexpected modifications, such as those related to acetone, alkylating agents as well as filter materials.

Approximately 70% of the overall acquired MS/MS spectra could be identified when using filter-aided as well as the best performing in-solution protocols. The unidentified spectra were predominantly of low-quality, lacking sufficient peptide fragment coverage for a confident identification. A maximum proteome coverage and reduced number of reagent-induced modifications were obtained for the filter-aided approach, even in the presence of Urea buffer (when using moderate incubation temperatures). However, this approach appeared to be more laborious and requires the availability of suitable filters and centrifuges. A comparable performance was obtained with the best performing in-solution protocols, such as 7 and 11. More specialised protocols, such as S-trap [317] and ultrasonic FASP digestion [318] were not evaluated, but are expected to enable a similarly good proteome coverage at reduced processing times.

The use of chemostats in this study ensured high experimental reproducibility, however, the results obtained here could be further applied to different types of cultivation (*e.g.* different growth conditions). Across all protocols, protein extraction was performed mechanically with bead beating, which proved a reproducible approach. More importantly, the type of sample collection proved crucial to obtain a high proteome coverage and should therefore be chosen carefully (also when applying to different conditions). Ultimately, we systematically evaluated a matrix of sample preparation protocols for the unicellular model eukaryote yeast and provide a large resource of protocols and associated mass spectrometric raw data. These enable the selection of suitable sample preparation elements, the design of more specialised procedures and support the evaluation and analysis of (native) peptide modifications in yeast.

Supplementary Information

Supplementary Excel tables can be found online: https://doi.org/10.1016/j.jprot.2022.104576
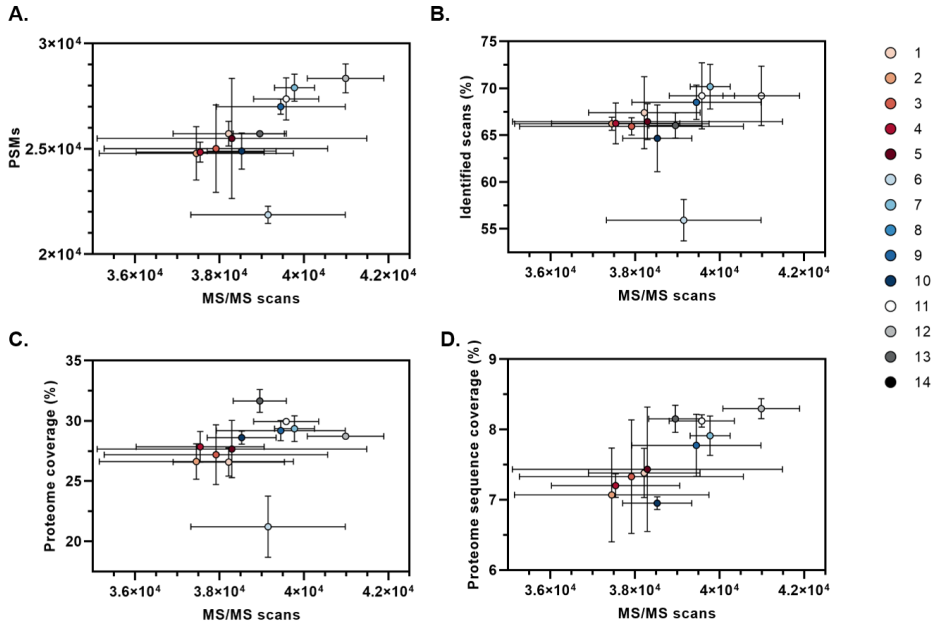
Figure S1. Achieved sequence identification rates, proteome and sequence coverage for the different sample preparation protocols (zoomed version of figure 2). The number of Peptide Spectrum Matches (PSMs) (A.), identified scans (%) (B.), proteome coverage (%) (C.) proteome sequence coverage (D.) were plotted against the number of MS/MS scans obtained per protocol. Each colored circle represents the average of four technical replicates obtained from two biological replicates, while the standard deviation is represented by the error bars. The proteome coverage was calculated based on the proteins identified per protocol as a percentage of the total number of proteins in yeast (known ORFs). In addition, the proteome sequence coverage was calculated based on the identified amino acids in the proteomics experiments, as a percentage of the total proteome amino acid sequence (sequence coverage).
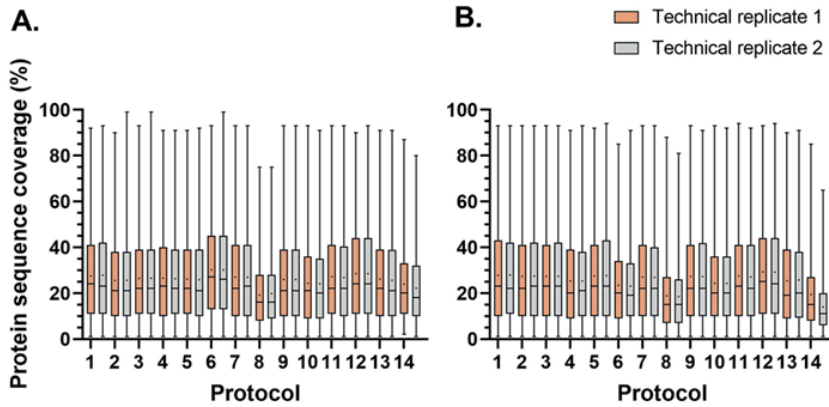
Figure S2. : Distribution of protein sequence coverage obtained after applying different sample preparation protocols (1–14). MS/MS spectra were identified using conventional database matching with PEAKS-DB in which at least 2 unique peptides had to be found for a reliable protein identification. The distribution of the sequence coverage of these identified proteins is shown here for each sample for both technical replicates (orange and gray). The average sequence coverage (·) and median (-) are shown in the boxplots. The average sequence coverage across all samples was 26 % and 25 %, for biological replicate 1 (A.) and 2 (B.), respectively.

Table S1. Growth characteristics of IMX372 (MG strain) in glucose-limited aerobic chemostat cultures at a dilution rate of 0.10 h⁻¹.

Bioreactor	Replicate 1	Replicate 2
Biomass yield ($g_x \, g_{sucrose}^{-1}$)	0.46	0.44
$Q_{glucose}$ (mmol $g_x^{-1} \, h^{-1}$)	-1.211	-1.264
Q_{co2} (mmol $g_x^{-1} \, h^{-1}$)	3.15	3.24
Q_{o2} (mmol $g_x^{-1} \, h^{-1}$)	-3.10	-3.22
Respiratory quotient	1.02	1.00
Carbon recovery (%)	100.01	97.09
Actual dilution rate (h^{-1})	0.101	0.101

Chapter 4

Proteome dynamics during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast

Maxime den Ridder, Wiebeke van den Brandeler, Meryem Altiner, Pascale Daran-Lapujade and Martin Pabst

Abstract

The yeast *Saccharomyces cerevisiae* is a widely used eukaryotic model organism and promising cell factory in industry. However, despite decades of research on yeast metabolism, its regulation is not fully understood and its complexity still represents a major challenge for engineering and optimizing biosynthetic routes for production of chemicals. Recent studies showed the potential of resource and proteome allocation data to enhance prediction models for metabolic processes. However the availability of comprehensive and accurate proteome allocation data is still limited. To address this knowledge gap, this study reports a proteome dynamics study that comprehensively covers the transition from exponential to stationary phase for both aerobically and anaerobically grown yeast cells. The combination of highly controlled reactor experiments, biological triplicates and standardised sample preparation protocols ensures the reproducibility and accuracy of this unique dataset. The CEN.PK lineage was chosen for its popularity and relevance for both fundamental and applied research. Together with the prototrophic, standard haploid strain CEN.PK113-7D, an engineered strain with genetic minimization of the glycolytic pathway was investigated, leading to the quantitative assessment of 54 proteomes and over 1700 proteins. Anaerobic cultures showed remarkably less proteome-level changes than aerobic cultures during the transition from exponential to stationary phase, accountable to the lack of diauxic shift in the absence of oxygen. These results support the notion that anaerobically growing cells lack time and resources to adapt to changes in the environment. The performed proteome dynamics study constitutes an important step towards better understanding the complex allocation of proteome upon glucose exhaustion and the impact of oxygen on this allocation. This new proteome dataset offers a valuable resource *e.g.* for the development of resource allocation model and metabolic engineering efforts.

Introduction

The yeast *Saccharomyces cerevisiae* is a widely used eukaryotic model organism and cell factory that represents a promising alternative to the fossil fuel-based production of chemicals. However, economic competitiveness is still a major hurdle for such bio-based processes. A shift towards cell factories therefore requires the development of improved strains with high product yield and productivity. However, constructing such cell factories involves extensive genetic engineering to rewire native genomes, which have been optimised over millions of years of evolution for cell growth and survival. Nevertheless, intensive research and development over the past decades have led to successful developments where yeast processes were brought to industrial scale, for example for the production of the drug precursor artemisinic acid [80, 319–321]. In parallel, *in silico* approaches to reproduce and predict the microbial metabolism have been developed to assist metabolic engineering efforts [156]. The complexity of yeast metabolism, however, limits the predictive power of these models. A promising approach to improve such models is to consider resource allocation, and more particularly, the cost of protein expression [157–161]. A prerequisite for this approach, however, is the availability of comprehensive and accurate proteome dynamics data established under tightly controlled conditions. Unfortunately, such data are commonly not available and are difficult to obtain.

S. cerevisiae displays a remarkable metabolic flexibility, as it tunes its metabolism between full respiratory sugar dissimilation and alcoholic fermentation, with different degrees of respiro-fermentative metabolism as a function of environmental cues, substrate and oxygen supply. The well-known Crabtree effect results in partial repression of respiration and therefore in respiro-fermentative growth in the presence of excess sugar (*e.g.* glucose, galactose) even in the presence of oxygen [100]. Conversely, gluconeogenic substrates as ethanol or acetate lead to strict respiratory metabolism in aerobic settings. In the absence of oxygen, *S. cerevisiae* will fully ferment carbon sources. However, respiratory and fermentative substrate dissimilation have a large impact on ATP yield, as full respiration of 1 mol of glucose results in 16 moles of ATP, while fermentation of the same amount of glucose only yields 2 moles of ATP [322]. The metabolic mode therefore strongly affects cellular resources, in particular their optimum allocation for growth and survival. To obtain a better insight on how *S. cerevisiae* responds to changes in substrate and oxygen supplies, we monitored its proteome employing tightly controlled bioreactors. Albeit various yeast proteomics studies have been performed over the past decade [110, 147, 149, 162, 310, 323–326], the dynamic proteome responses to substrate availability during all growth phases of yeast cultures (exponential, diauxic and stationary growth) were monitored here for the first time under both, aerobic and anaerobic conditions using standardized sample preparation protocols [327]. Interestingly, only little is known to date about the proteome dynamics under anaerobic conditions, in particular during the transition from exponential to stationary phase.

Considering eukaryotes such as *S. cerevisiae*, genetic redundancy is another level of complexity for *in silico* design and experimental development of *e.g.* cell factories. Many genes, more particularly in metabolism, have orthologues with similar functions [113], often with a poorly understood physiological role. With minimal genomes in mind, several studies have explored the requirement for these redundant genes and implemented top-down approaches to reduce genetic redundancy [120, 328, 329]. Such minimized genomes have the potential to facilitate the complete redesign and construction of entirely synthetic yeast genomes. Moreover, by eliminating isoenzymes with different regulatory and kinetic properties, genetic minimization of key metabolic pathways can facilitate the formulation and validation of mathematical models. Solis-Escalante *et al.* constructed a yeast strain in which the 26 genes encoding enzymes of the Embden-Meyerhof-Parnas pathway of glycolysis, main pathway for sugar utilization, were minimalized to a set of 13 genes in the Minimal Glycolysis (MG) strain (**Figure 1a**) [120]. While this genetically reduced strain appeared physiologically comparable to its parent strain (with the full set of glycolytic genes) the underlying proteome dynamics and potential protein level adjustments were not investigated. In this study, the engineered MG strain and its parental *S. cerevisiae* were investigated side by side, during transition from exponential to stationary phase, in the presence or absence of oxygen. The temporal proteome dynamics across all growth phases were monitored from triplicate bioreactor cultures. Quantitative shotgun proteomics experiments were performed using 10-plex tandem mass tag (TMT) isobaric labelling. The use of tightly-controlled reactor experiments in combination with robust sample preparation protocols allowed for highly accurate quantitative data. Those therefore constitute highly valuable resources for *in silico* approaches, *e.g.* to assist metabolic engineering efforts. Furthermore, the established proteome data expand the current understanding of protein dynamics in yeast during carbon-limited growth–under both aerobic and anaerobic conditions–switching from proliferation to stationary phase. The comparison to the minimal glycolysis mutant moreover measures the impact of the loss of the minor glycolytic isoenzymes on the global proteome.

Materials and Methods

Yeast strains and media. The MG yeast strain IMX372 (*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1::KlURA3, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4*) and CEN.PK113-7D (*MATa MAL2-8C SUC2*) used in this study share the CEN.PK genetic background [120, 288]. Shake flask and batch cultures were grown in synthetic medium (SM) containing 5.0 g/L $(NH_4)SO_4$, 3.0 g/L KH_2PO_4, 0.5 g/L $MgSO_4 \cdot 7H_2O$ and 1 mL/L trace elements in demineralized water, set at pH 6. The medium was heat sterilized (120°C) and supplemented with 1 mL/L filter sterilized vitamin solution and 20 g/L heat sterilized (110 °C) glucose (SMG) [289]. The bioreactor medium was supplemented with 0.2 g/L antifoam Emulsion C (Sigma, St. Louise, USA) or with 0.2

g/L antifoam Pluronic PE 6100 (BASF, Ludwigshafen, Germany) for anaerobic and aerobic cultures, respectively. In case of anaerobic cultivations, the medium was also supplied with anaerobic growth factors, 10 mg/L ergosterol (Sigma-Aldrich, St. Louis, MO) and 420 mg/L Tween 80 (polyethylene glycol sorbate monooleate, Merck, Darmstadt, Germany) dissolved in ethanol. Frozen stocks of *S. cerevisiae* cultures were prepared by the addition of glycerol (30% v/v) in 1 mL aliquots for storage at -80 °C.

Bioreactor cultures. Aerobic shake flask cultures were grown at 30°C in a Innova incubator shaker (New Brunswick™ Scientific, Edison, NJ, USA) at 200 rpm using 500 mL round-bottom shake flasks containing 100 mL medium. Triplicate aerobic batch cultures of control and MG yeast were performed in 2 L laboratory fermenters (Applikon, Schiedam, The Netherlands) with a 1.2 L working volume under aerobic and anaerobic conditions. SM-medium was used and maintained at pH 5 by the automatic addition of 2 M KOH. Mixing of the medium was performed with stirring at 800 rpm. Gas inflow was filter sterilized and compressed air (Linde Gas, Schiedam, The Netherlands) or nitrogen (<10 ppm oxygen, Linde Gas) was sparged via the bottom of the bioreactor at a rate of 500 mL/min, for aerobic and anaerobic cultures, respectively. Dissolved oxygen levels were measured with Clark electrodes (Mettler Toledo, Greifensee, Switzerland). The temperature of the fermenters was maintained at 30°C. The reactors were inoculated with exponentially growing shake flask cultures of *S. cerevisiae* strain IMX372 and CEN.PK113-7D to obtain an initial optical density (OD_{660}) of approximately 0.2. Sampling for HLPC and OD_{660} measurements was done every 90 minutes. Proteome samples were taken at 6, 9, 12, 16.5, 27 and at 7.5, 10.5, 13.5, 16.5 hours in aerobic and anaerobic conditions, respectively.

Biomass, metabolites and gas measurements. To monitor growth, OD_{660} measurements were performed on a JENWAY 7200 spectrophotometer (Cole-Parmer, Stone, UK). The biomass dry weight was determined in duplicate as described earlier [289]. For extracellular metabolite determinations, broth samples were centrifuged for 5 min at 13,000 g and the supernatant was collected for analysis with an Aminex HPX-87H ion exchange column (Biorad, Hercules, CA, USA). The HPLC was operated at 60°C and 5 mM of H_2SO_4 was used as mobile phase at a rate of 0.6 mL/min. Off-gas concentrations of CO_2 and O_2 were measured using an NGA 2000 analyzer. Proteome samples (~3-5 mg dry weight) were taken from batch cultures. The samples were collected in multifold in trichloroacetic acid (TCA) (Merck Sigma, Cat. No. T0699) with a final concentration of 10%. Samples were centrifuged at 4000 g for 5 min at 4°C. Cell pellets were frozen at -80°C [272].

Yeast cell lysis, protein extraction and proteolytic digestion. Cell pellets of the aerobic and anaerobic cultures were resuspended in lysis buffer composed of 100 mM Triethylammonium bicarbonate (TEAB) containing 1% SDS and phosphatase/protease inhibitors. Yeast cells were lysed by glass bead milling by 10 cycles of 1 minute shaking

alternated with 1 min rest on ice. Proteins were reduced by addition of 5 mM DTT and incubation for 1 hour at 37°C. Subsequently, the proteins were alkylated for 60 min at room temperature in the dark by addition of 50 mM acrylamide. Protein precipitation was performed by addition of four volumes of ice-cold acetone (-20°C), followed by 1 hour freezing at -20°C. The proteins were solubilized using 100 mM ammonium bicarbonate. Proteolytic digestion was performed by Trypsin (Promega, Madison, WI), 1:100 enzyme to protein ratio, and incubated at 37°C overnight. Solid phase extraction was performed with an Oasis HLB 96-well μElution plate (Waters, Milford, USA) to desalt the mixture. Eluates were dried using a SpeedVac vacuum concentrator at 50°C and frozen at -80°C.

Quantitative temporal proteome analysis. Desalted peptides were reconstituted in 100 mM TEAB and TMT10-plex reagents (Thermo) were added from stocks dissolved in 100% anhydrous acetonitrile (ACN). Peptides were mixed with labels in a 1:8 ratio (12.5 μg to 100 μg) and incubated for 1 hour at 25°C and 400 rpm and the labelling reaction was stopped by addition of 5% hydroxylamine to a final concentration of 0.4%. Labelled peptides were then mixed in at approx. equal quantities. Two bridging samples were included in each TMT10-plex experiment to improve comparability between different experiments. The bridging sample was a mixture of the three biological replicates of MG yeast under aerobic conditions in the mid-stationary phase. Peptide solutions were diluted with water to obtain a final concentration of acetonitrile (ACN) lower than 5%. Solid phase extraction was performed to desalt the final peptide mixture. Desalted peptides were subsequently frozen at -80°C for 1 hour and dried by vacuum centrifugation. Peptides were finally resuspended in 3% ACN/0.01% TFA prior to MS-analysis to give an approximate concentration of 500 ng per μL. Samples were labelled as indicated in **SI table 2**.

Shotgun proteomic analysis. An aliquot corresponding to approximately 1 μg protein digest was analysed using an one dimensional shot-gun proteomics approach [290]. Briefly, the samples were analysed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 μm x 150 mm, 2 μm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL/min over a linear gradient from 5% to 25% solvent B over 180 min, then from 25% to 55% over 60 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 240 min. Solvent A was H_2O containing 0.1% formic acid (FA), and solvent B consisted of 80% ACN in H_2O and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70 K resolution in full MS mode with a maximum ion injection time (IT) of 75 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 35 K

resolution with AGC target of 1E5 and IT of 100 ms, 1.2 m/z isolation width and normalized collision energy (NCE) of 32.

Processing of mass spectrometric raw data. Data were analysed against the proteome database from *Saccharomyces cerevisiae* (Uniprot, strain ATCC 204508 / S288C, Tax ID: 559292, July 2020) using PEAKS Studio X (Bioinformatics Solutions Inc., Waterloo, Canada) [210], allowing for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, acrylamide and TMT10 label as fixed and methionine oxidation and N/Q deamidation as variable modifications. Peptide spectrum matches were filtered against 1% false discovery rates (FDR) and identifications with ≥2 unique peptides. Changes in protein abundances between different time points using the TMT quantification option provided by the PEAKSQ software tool (Bioinformatics Solutions Inc., Canada). Auto normalization was used for quantitative analysis of the proteins, in which the global ratio was calculated from the total intensity of all labels in all quantifiable peptides. Quantitative analysis was performed using protein identifications containing at least 2 unique peptides, which peptide identifications were filtered against 1% FDR. The significance method for evaluating the observed abundance changes was set to ANOVA and the significance score was expressed as the $-10 \times \log 10(p)$, where p is the significance testing p-value. The p-value represents the likelihood that the observed change is caused by random chance. Results from PEAKSQ were exported to 'proteins.csv', containing the quantified proteins.

Pathway analysis, functional enrichment, and data visualisation. Briefly, the exported 'proteins.csv' files from PEAKSQ, listing the quantified proteins for each experiment, were directly imported into the Python environment. Normalization between data was performed using a bridging sample. A function was further established that links Uniprot accession numbers and yeast genes (as obtained from https://www.uniprot.org/docs/yeast.txt, and which subsequently was used to annotate identified proteins from the experiments with correct gene names. The biological triplicates per condition (aerobic and anaerobic) and strain (control and MG) were treated separately. Furthermore, each biological replicate consisted of two additional technical replicates. To analyse the technical and biological replicates, clustermaps were made using a self-built Python function based on the clustermap function from the Seaborn package in Python [330], using the Euclidean distances metric and the average linkage method. Only proteins detected in all three biological replicates were used for the cluster analysis. The fold change of each protein in a specific condition was calculated relative to the bridging sample. The average fold changes of the technical replicates were subsequently used to determine the standard deviations of the biological replicates. The averages of the biological replicates were determined to obtain the four sub-datasets i) WT aerobic, ii) MG aerobic iii) MG anaerobic and iv) WT anaerobic. All graphs ultimately show the analyses of these biological-replicate averages and their corresponding standard deviations.

To study how protein abundances changed in individual cellular pathways, the obtained proteomics data were analysed using the KEGG (Kyoto Encyclopaedia of Genes and Genome) pathway database [331]. All the up-to-date KEGG pathways were retrieved with the constructed 'KEGG_tool.py' code. Here, the Bio.KEGG.REST module from the Biopython package in Python was used [332]. Thereby, the functions 'kegg_list' was used to list all pathways for *S. cerevisiae*, and 'kegg_get' to retrieve gene names that are assigned to a specific pathway. Since many pathways have an extensive list of members, the pathways in the central carbon metabolism (CCM) were reduced to the most important genes in order to enable meaningful visualisation in graphs. Using the above mentioned clustermap function, the protein fold changes of the CCM were plotted on a heatmap for each of the experiments, without any clustering. For better visualisation of the trends, the data were normalised to the mid-exponential (ME) phase. The same function was moreover used to display the average absolute intensity of every protein throughout the whole growth curve, using log10 and absolute scale, respectively. The significance of a difference in biological-replicate-average fold changes between two datasets was assessed by performing a two-sided two-sample unpaired t-test (also known as Welch's t-test), using the 'ttest_ind_from_stats' function from the 'SciPy.stats' module in Python [333]. Global proteome changes between two experiments or phases were visualised in volcano plots, where the -10log10(p) is plotted against the log2(fold change) between the two conditions. These plots were generated using the 'gene_exp.volcano' a modified version of the GeneExpression.volcano function from the 'Bioinfokit.visuz' module in Python [334]. This function enabled the division of the fold changes between two experiments into i) insignificant changes, ii) statistically significant changes (but not necessarily biologically significant), iii) statistically and likely biologically significant changes. For this study the statistical significance threshold was generally set to $p < 0.05$. The (presumed) biological significance threshold was set to a log2 fold change threshold of +/- 0.32 (indicating a 1.25 absolute fold change).

A functional enrichment analysis using the STRING database was performed in order to determine whether specific GO-terms or KEGG-pathways are enriched under a particular condition [335]. For this, Python was used to programmatically accesses the STRING database via an API. This created a dictionary containing the more and less abundant proteins, the species identifier (4932 for *S. cerevisiae*), the functional categories that should be assessed, the FDR threshold (<0.05 in this study), and an optional set of 'background genes' with as alternative background the whole species proteome. The function 'backgroundgene_2_string' retrieves the protein-specific string identifiers for the back-ground genes/proteins, which in this case were all proteins detected across the experiments. Estimation of the average protein content for the aerobic and anaerobic growth conditions using emPAI and PAI indices was performed according to Yasushi Ishihama *et al.*, 2005 [61]. Circle graphs were made using the 'surf' function in Matlab, where circle areas represent the obtained emPAI values.

Data availability. Mass spectrometric raw data have been deposited to the ProteomeXchange Consortium [336] via the PRIDE [337] partner repository and are publicly available under the project code PXD031412.

Results

Capturing proteome dynamics of laboratory control and minimal glycolysis yeast in aerobic and anaerobic batch bioreactor cultures

To optimize data reproducibility and reliability, the batch cultures were performed in bioreactors in which mixing, aeration and pH were tightly controlled. To further increase biological significance, independent triplicate cultures were performed for the two investigated strains. The laboratory, prototrophic control strain *S. cerevisiae* CEN.PK113-7D, a popular lineage for biotechnology for which several omics datasets are already available, was chosen, as well as the 'minimal glycolysis' variant ('MG', IMX372) lacking glycolytic and fermentation minor isoenzymes (**Figure 1a**). The batch cultures were sampled during all growth phases, ranging from proliferation to growth arrest in stationary phase (**Figure 1b**). The presence or absence of oxygen strongly affects yeast physiology and results in differences in metabolism and growth phases. During growth on glucose, aerobic cultures both respire and ferment, producing ethanol and other fermentation products. This growth on glucose is followed by a diauxic growth phase during which the fermentation products are fully respired until stationary phase. Conversely, in the absence of oxygen, *S. cerevisiae* fully ferments glucose and does not respire. Dissimilation of fermentation products requires oxygen, anaerobic cultures therefore directly switch from exponential growth on glucose to stationary phase, without diauxic phase. These physiological differences were reflected in the growth and metabolites profiles (**SI Table 1** and **Figure 1b**). Aerobic proteome dynamics were monitored in time with sampling at 6, 9, 12, 16.5 and 27 hours of growth, corresponding to mid-exponential, late exponential, early-diauxic, mid-diauxic and stationary growth phases, respectively (**Figure 1b**). To align sampling time points to physiology, anaerobic cultures were sampled at 7.5, 10.5, 13.5 and 16.5 hours of growth, corresponding to mid- and late exponential, early stationary and stationary growth phase, respectively (**Figure 1b**). After cell lysis and trypsin digestion, peptide samples of three biological replicates per condition were labelled using TMT10-plex reagents, mixed equally and subjected to a 4-hour gradient shotgun proteomics experiment (**SI Table 2**). On average, 1175 and 1106 proteins were quantified in the control yeast under aerobic and anaerobic conditions respectively, with at least two unique peptides and 1% FDR. Similarly, 1131 and 1127 proteins were quantified confidently on average for the aerobic and anaerobic cultures of the MG strain. In total, across all TMT experiments, 1734 proteins were quantified (**SI Table 3**), which is close to 40% of the total proteome considering the expression of ~4500 proteins at any time [338]. Moreover, the absolute protein amount was estimated based on the number of sequenced peptides per protein [61]. The total protein content was estimated based on the protein identifications and
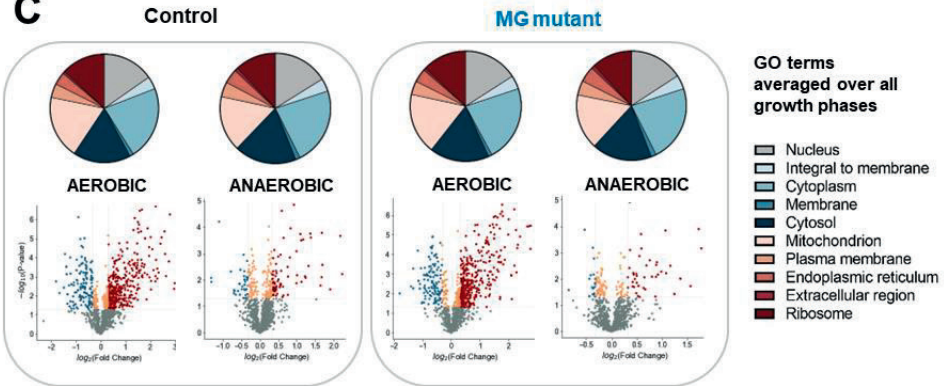
A
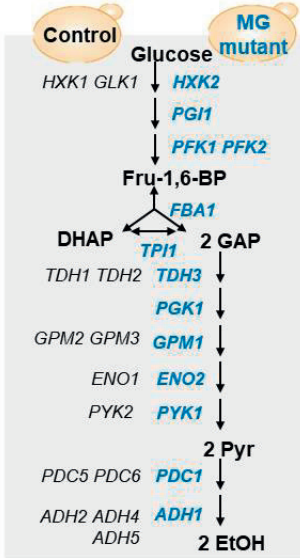
B
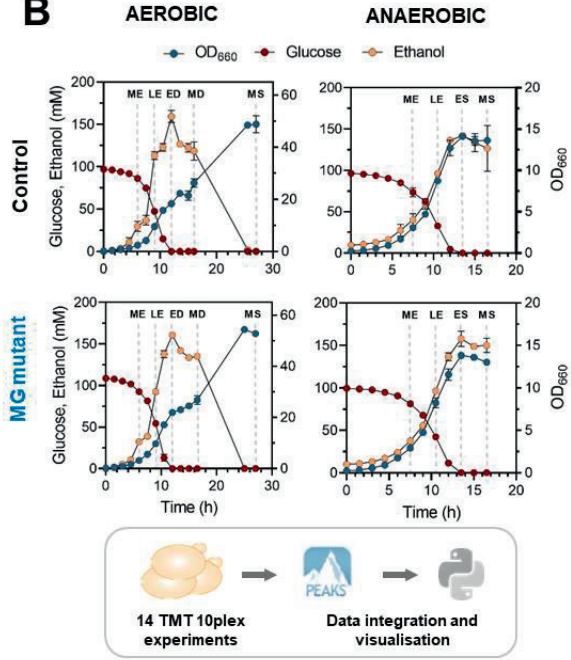
C

Global proteome fold change of mid-stationary normalized to mid-exponential growth phase

● Significantly more abundant ● Below significance threshold ● Significantly less abundant ● Significant

Figure 1. Comprehensive and temporal proteomic profiling of the transition from proliferation to stationary phase in laboratory control and minimal glycolysis yeast under aerobic and anaerobic conditions. A. Schematic overview of glycolysis in control (CEN.PK113-7D, black) and MG yeast (IMX372, blue). The blue proteins are retained in MG yeast. *Adh3 is a mitochondrial protein. B. Yeast growth in aerobic and anaerobic cultures. Glucose (red), ethanol (yellow) and OD_{660} concentrations were measured during the different growth phases of aerobic and anaerobic batch cultures of control yeast and MG strain. The values shown are from three biological replicate cultures, standard deviations indicated by the error bars. The dotted grey lines indicate time points at which samples were taken for proteome analysis. Proteome samples were therefore taken from each biological replicate in the aerobic cultures after 6, 9, 12, 16.5 and 27 hours of growth, in the mid-exponential ('ME'), late-exponential ('LE'), early-diauxic ('ED'), mid-diauxic ('MD') and (mid-) stationary ('MS') growth phase, respectively. Furthermore, proteome samples were taken of the anaerobic cultures after 7.5, 10.5, 13.5 and 16.5 hours of growth, in the ME, LE, early-stationary ('ES') and MS growth phase, respectively. Proteome samples were subjected to quantitative shotgun proteomics experiments, using 10-plex TMT isobaric labelling and a one-dimensional, 4-hour chromatographic separation. Database searching and quantitative analysis was performed using PEAKS X and by a tailor-made Python data processing pipeline. C. Annotation of yeast protein functions using Gene Ontology (GO) terms. Based on the classifications of GO annotation, the overall functions of the identified yeast proteins (with at least 2 unique peptides present) were categorized into cellular component, and displayed in pie chart format with absolute protein numbers (average of three biological replicates). The global proteome changes between the Mid-Exponential and Mid-Stationary phase under aerobic and anaerobic conditions in control and MG strain were visualised in volcano plots. The fold changes were normalized to the aerobic and anaerobic Mid-Exponential phases. The $\log_2$ of the abundance fold change between the two conditions was plotted against the significance ($-\log^{10}p$), using a p-value threshold of <0.05 and a fold change threshold of >1.25 (which corresponds to a $\log_2$ fold change threshold +/- 0.32). Significant protein fold changes of Mid-Stationary proteins were coloured red if higher, blue if lower, or peach if similar to their mid-exponential equivalents. The exact numbers can be found in Table 1.

further extrapolated to the proteins generally expressed in the cell, indicating that more than 99% of the protein content was quantified in our study (**SI Figure 1, SI Table 4**). The detected proteins were predominantly assigned to intracellular organelles functional GO categories, consisting of *e.g.* cytosolic, mitochondrial and ribosomal proteins (**Figure 1c, SI Table 5**), which could be explained by their high expression levels [309, 310]. A similar GO-term assignment was found for both strains and conditions.

The crude protein quantification profiles were further analysed using a Python data processing pipeline to enable a tailored visualisation and interpretation of the large-scale data. To this end, the data from the 14 separate TMT experiments were compared for their temporal and conditional protein abundance changes of control (CEN.PK113-7D) and MG yeast strains under (an)aerobic conditions. Data are represented as fold-change for each protein in a specific condition relative to a bridging sample. To improve comparability between

experiments, the bridging sample used in all TMT experiments was a mixture of the three biological replicates of the MG yeast aerobic stationary phase (**SI Table 2**).

To assess experimental reproducibility, global proteome data were compared with cluster analysis based on Euclidean distances using three biological replicates per strain and condition and two technical replicates per time point within each biological replicate. Heat map visualization showed clustering of replicates of the different growth phases per strain and condition, confirming the reproducibility of the reactor experiments and proteome analyses (**SI Figure 2**). The average protein abundance of three biological replicates per condition and strain was therefore used in following analyses.

Oxygen availability strongly affects global proteome dynamics across the growth curve
To explore the impact of oxygen availability on the yeast proteome, we first focused on the growth phases with the most marked change in the global proteome between aerobic and anaerobic cultures for the control strain CEN.PK113-7D, *i.e.* stationary and mid-exponential phases. While a similar number of proteins were quantified in the presence and absence of oxygen, the number of differently expressed proteins between stationary and mid-exponential phases varied significantly between both conditions (p-value <0.05 and fold change of +/- 1.25). For aerobic cultures, 364 proteins were significantly more abundant under carbon starvation (stationary phase), while this only accounted for 78 proteins under anaerobiosis (**Figure 1c, Table 1**). Furthermore, a significantly lower abundance was observed in the stationary phase for 174 proteins compared to exponential growth in the presence of oxygen, as to 42 proteins only in the absence of oxygen (**Figure 1c, Table 1**).

Deprived of usable carbon source, stationary phase yeast cells arrest growth, thereby entering a state of decreased metabolism and biosynthesis, and overall lower transcription and translation rates [339, 340]. Accordingly, ribosomal proteins have been shown to be expressed at lower levels in stationary phase [154, 341]. In good agreement with physiology, the abundance of proteins involved in processes associated with protein synthesis and cellular growth showed decreased abundance in the transition between exponential and stationary growth phase under both aerobic and anaerobic conditions, *e.g.* 'gene expression', 'ribosome assembly' and 'cellular macromolecule biosynthetic process' (**SI Table 6 and 7**).

In the presence of oxygen, yeast cells transition from respiro-fermentation on glucose to full respiration using ethanol as primary carbon source. This increase in respiratory activity was nicely reflected in the proteome as proteins more abundant in stationary phase than exponential phase were typically associated with mitochondrial respiration in aerobic conditions including processes as 'generation of precursor metabolites and energy', 'mitochondrion organization' and 'transmembrane transport' (**SI Table 6**). Understandably, in non-respiring, anaerobic cultures this response was not observed. In these cultures, mostly

proteins involved in carbohydrate catabolic and disaccharide metabolic processes were more abundant in stationary than in mid-exponential phase, potentially to ensure survival in growth-arrested cells. Interestingly, proteins in cellular components involving 'cell periphery' and the 'plasma membrane' were also found to be more abundant (**SI Table 7**).

Table 1. Number of proteins with significant changes during the transition to subsequent growth phases under aerobic and anaerobic conditions in CEN.PK113-7D. The total number of proteins indicates the number of proteins that were detected in at least two biological replicates in both phases that were compared. Proteins were normalized to the former growth phase. A p-value threshold of 0.05, and a fold change threshold of 1.25 (which corresponds to a log2 fold change of +/- 0.32) was used to select proteins that are represented in the table.

CEN.PK113-7D growth transition	# Proteins quantified	# More abundant	# Less abundant
Anaerobic			
Late exponential / Mid-Exponential	1092	1	22
Early stationary / Late Exponential	1092	52	3
Stationary / Early Stationary	1092	0	0
Stationary / Mid-Exponential	1092	78	42
Aerobic			
Late exponential / Mid-Exponential	998	12	21
Early Diauxic / Late exponential	998	67	4
Mid-Diauxic / Early Diauxic	1168	125	24
Stationary / Mid-Diauxic	1168	90	34
Stationary / Mid exponential	1168	364	174

Comparing proteome data across growth phases revealed that the diauxic shift is the event with the strongest impact on proteome rearrangement, with 24 proteins with lower and 125 proteins with higher abundance between the beginning of the diauxic growth phase and the mid-diauxic phase (**Table 1, SI Table 6**). The diauxic shift was characterized by an increased abundance in proteins involved in aerobic respiration, fatty acid metabolism, and the generation of precursor metabolites and energy, in line with the switch from respiro-fermentative to fully respiratory metabolism. Conversely, the set of proteins with decreased abundance during the diauxic shift was enriched for proteins involved in protein synthesis in the cytosol. This result is also in line with the decreased growth rate, and thereby protein synthesis rate, of yeast cells grown on ethanol media as compared to glucose [342].

Under anaerobiosis, most proteome changes occurred in the transition between exponential and stationary growth (55 proteins, *i.e.* 46% of all detected changes in abundance throughout

all phases). Notably, prolonged cultivation during the stationary phase under anaerobiosis did not further alter the proteome (**Table 1, SI Table 7**).

Impact of oxygen on proteome rearrangements in central carbon metabolism across growth phases

Central carbon metabolism (CCM) consists of key pathways required for the conversion of carbon sources into the 12 building blocks for the synthesis of cellular components, and encompasses ca. 150 transport proteins and enzymes [329]. The flow of carbon and electrons via the CCM therefore responds to carbon source nature and abundance. As oxygen availability dictates how much ATP molecules can be produced from the carbon source, CCM also responds to oxygen availability. The proteins involved in CCM are therefore expected to be considerably affected by glucose and oxygen availability. In our study, 101 out of 142 CCM proteins (**SI Table 8**) were successfully quantified in at least one of the main four conditions (**Figure 2a**). In the presence of oxygen, 71 proteins had significantly different abundance over the time-course, while in the absence of oxygen this only resulted in 18 significant protein abundance changes over the growth curve.

S. cerevisiae harbours a set of 17 proteins able to transport hexoses, known as Hxt proteins. Their expression is primarily dictated by hexose, and mostly glucose, abundance. Being membrane-bound and with low abundance, Hxt proteins are typically difficult to find in proteome studies, and their high level of homology makes their identification challenging. Nevertheless, six Hxt proteins were quantified in the present dataset: Hxt2, 3, 4, 5, 6 and 7. Four of these Hxts could be quantified in all samples, irrespective of strain and oxygen supply. Hxt6 and Hxt7 share high protein sequence similarity (>99%) and were therefore considered as a single protein group. These were the most abundant Hxts and were consistently more abundant upon glucose exhaustion in all tested conditions (**Figure 2a**), in good agreement with their high affinity for glucose [343]. Hxt3 and Hxt4, also identified in all conditions, responded differently in the presence and absence of oxygen. Abundance of Hxt4, high affinity transporter, was logically increased upon glucose exhaustion, but was then decreased when reaching stationary phase in aerobic cultures, while it remained high under anaerobiosis. Hxt3, low affinity transporter expressed at high and low glucose concentration [344], decreased in abundance across the growth curve, markedly in aerobic stationary phase. Hxt5 was only detected in the control strain in the presence of oxygen, but its abundance was in line with its induction by non-fermentable carbon sources and decreasing growth rates [343]. Hxt2, only detected in anaerobic cultures of the control strain, was also responding as expected for a high affinity transporter, with increased abundance upon glucose exhaustion.

Among the 26 glycolytic and fermentation proteins, 13 major isoforms are constitutively expressed with high abundance, while the remaining are minor isoforms with lower

abundance and condition-dependent expression [119, 120]. This notion was well reflected in the present comprehensive dataset, in which 23 of these proteins could be quantified in both conditions of the control strain. All major isoenzymes were found, their abundance remaining stable across growth phases under anaerobiosis, but was slightly decreased in stationary phase under aerobiosis (**Figure 2a**). Most minor isoenzymes were detected, although not always in all conditions. Tdh1, minor glyceraldehyde-3P dehydrogenase expressed with and without oxygen, was systematically more abundant upon glucose exhaustion. Similarly, Glk1 and Hxk1, glucose-repressed isoenzymes of the predominant hexokinase 2, were also more abundant upon glucose exhaustion. Adh2, alcohol dehydrogenase repressed by glucose and induced by ethanol [345], was only detected in the presence of oxygen and its abundance was strongly increased in mid-diauxic phase. In all conditions, the estimated glycolytic protein content (**Figure 2b** and **c**) was superior to the remainder of the CCM proteins.

The TCA cycle, a set of 22 proteins located in the mitochondrion, is particularly active during respiration. The majority of the TCA cycle proteins were detected and their abundance was generally strongly increased during the diauxic shift, but unchanged during anaerobic cultures (**Figure 2a**). A small group of proteins were unaffected by progression through the growth phase under anaerobiosis, among which Pda1, Pdb1 and Lat1, three of the five subunits of the pyruvate dehydrogenase complex, and Aco2 and Idp1. Interestingly, the temporal abundance of Frd1 and Osm1, proteins important for *e.g.* protein folding in anaerobic conditions [346], was also unaffected in the presence and absence of oxygen. Many CCM metabolites cross the mitochondrial membrane via transporter proteins, and increased respiratory activity is expected to increase the flux of metabolites between cytosol and mitochondria. Accordingly, the temporal profile of the seven quantified mitochondrial transporters increased after glucose exhaustion aerobically, but not anaerobically, with a marked increase for Odc1 and Sfc1, carboxylic acids antiporters. Despite the increase in respiratory activity during the diauxic shift and phase, reflected in the TCA cycle proteins, the abundance of anaplerotic proteins Pyc1, Pyc2 and Mae1 remained unchanged. Conversely, the abundance of the gluconeogenic proteins Fbp1 and Pck1 and glyoxylate cycle proteins MLs1 and Icl1, required for ethanol utilization, was strongly increased upon glucose exhaustion aerobically, but were logically very stable anaerobically. Growth on non-fermentable carbons sources requires a complex metabolic rearrangement to supply cytosolic and mitochondrial acetyl-CoA. Proteins involved in acetate and acetyl-CoA metabolism were accordingly overall increased in expression during the diauxic phase, particularly Acs1 and Ach1, but were not visibly affected by glucose exhaustion under anaerobiosis.

4
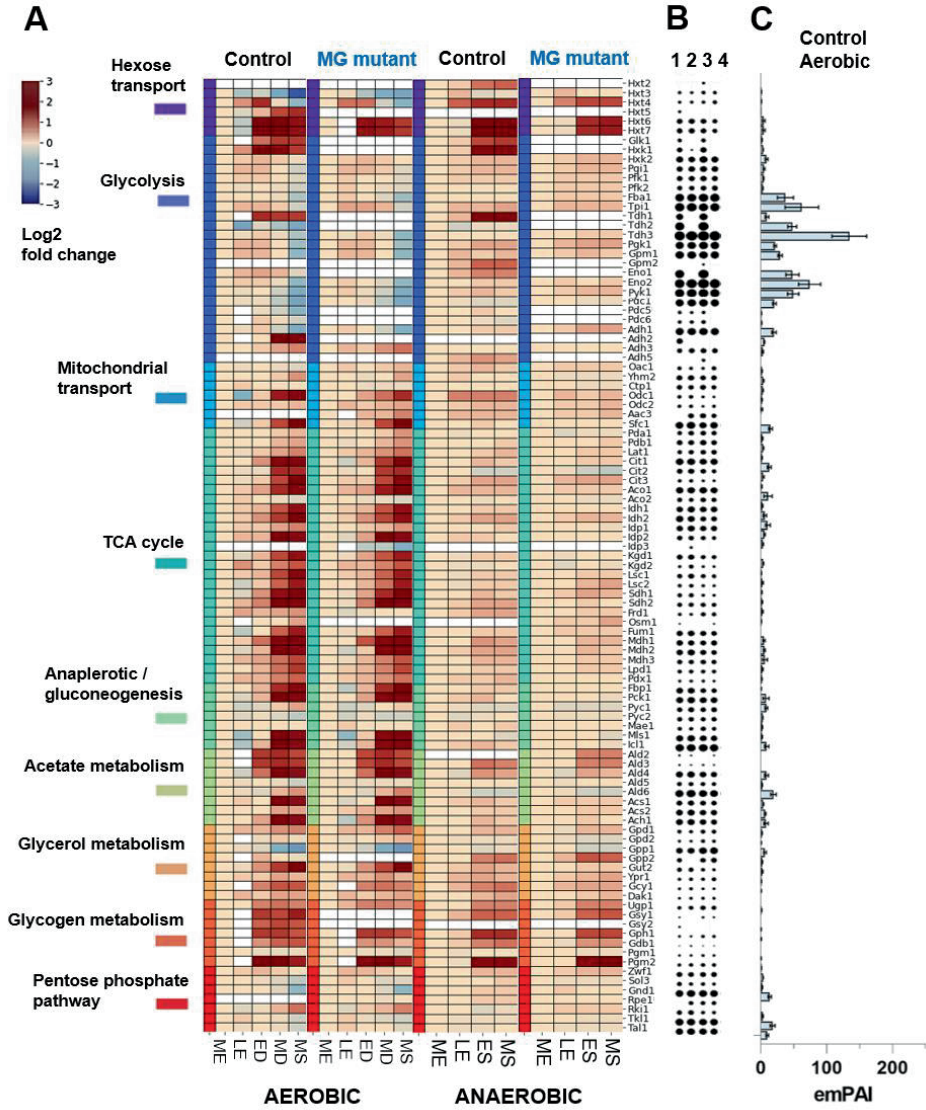
Figure 2. Yeast control and MG strain central carbon metabolism (CCM) protein abundance levels, under aerobic and anaerobic conditions. A. The heat map shows the temporal log2 fold changes of the enzymes of the CCM of control (CEN.PK113-7D) and MG (IM372) yeast for the mid-exponential (ME), late-exponential (LE), early-diauxic (ED), mid-diauxic (MD), and mid-stationary (MS) phases, compared to the mid-exponential (ME) phase for each condition separately. The proteins belonging to specific pathways of the CCM are each highlighted with different colours. White gaps in the figure indicate that the protein was not detected, or that it has been deleted in case of the MG mutant strain, in any of the biological replicates of the respective growth phase. No filtering for significance or fold-change thresholds was performed and all enzymes that were detected were included in the shown heatmap. B. The circle graph (1: control aerobic, 2: MG aerobic, 3: control anaerobic and 4: MG anaerobic) and the bar (C.) graph show the protein abundance expressed as the exponentially modified protein abundance index (emPAI) for the proteins of the CCM. The emPAI estimates the absolute protein amount in proteomics by the number of sequenced peptides per protein. The values are calculated from at least one biological replicate, standard deviations are indicated by the error bars in the bar graph. The circle areas directly correlate to the obtained emPAI values of the individual proteins for each condition. C. The absolute emPAI values of control strain CEN.PK113-7D under aerobic conditions, similar values were found for the other conditions.

4

Redox metabolism, key for cell survival, is balanced according to oxygen availability. While respiring cells can oxidize the NADH produced during glucose assimilation via oxidative phosphorylation, two-steps glycerol formation from dihydroxyacetone phosphate (DHAP) is the major electron sink for yeast when oxygen is absent. Abundance of the four paralogs, Gpd1 and Gpd2, and Gpp1 and Gpp2 was not strongly affected across the different phases and growth conditions (**SI Figure 3a**), in agreement their reported respective transcriptional regulation and (in)activation by post-translational modification (phosphorylation) [347]. The only marked changes were the decreased abundance of Gpp1 in the presence of oxygen, and Gpp2 increased abundance upon glucose exhaustion in the absence of oxygen. In aerobic conditions, NADH is oxidized by external or internal NADH dehydrogenases which shuttle the electrons into the mitochondrial electron transfer chain. Contrarily, anaerobic yeast cultures reoxidize the excess NADH formed in biosynthesis via glycerol production [348]. Finally, Gut2 marked increased abundance during the diauxic shift under aerobiosis can be explained both by its role in glycerol utilization or in maintaining the redox balance (**SI Figure 3a**).

Oxygen-dependent dynamics in other pathways
Respiration is an important mechanism for energy conservation in the presence of the electron acceptor oxygen. After the diauxic shift, respiration becomes the main ATP source for the cells. Accordingly, the abundance of the ATP synthases and cytochrome oxidases in the oxidative phosphorylation pathway increased significantly after glucose exhaustion (**SI Figure 6**). For anaerobic cultures, however, the abundance of these proteins remained constant over the entire growth curve. Respiring cells are prone to generation of reactive

oxygen species (ROS), for instance by the production of superoxide during electron transfer during oxidative phosphorylation, which can induce the expression of stress tolerance genes [349]. Accordingly, Sod1 and Sod2, enzymes that detoxify superoxide and produce hydrogen peroxide [350], were significantly more abundant under aerobic conditions, compared to anaerobic conditions (**SI Figure 6**). Moreover, their abundance increased by at least two-fold towards the stationary phase compared to log growth in the presence of oxygen, while the protein abundance remained constant under anaerobic growth. A similar protein profile was found for the peroxisomal and mitochondrial catalase A (Cta1) that detoxifies hydrogen peroxide. Nevertheless, the abundance of Ccs1, copper ion chaperone to Sod1, was relatively constant and similar between both conditions in the control strain (**SI Figure 6**).

Heme synthesis, encompassing a set of 8 Hem proteins, also depends on oxygen availability [351]. In this study, 7 Hem proteins could be quantified. Hem2, 3, 12 and 15 protein abundance were either similar for aerobic and anaerobic conditions, or higher aerobically (**Figure 3a**). However, Hem13, Hem13 and Hem14 were only quantified or higher abundant under anaerobic conditions, although Hem14 only in one biological replicate with a few peptides. On the other hand, Hem13 was confidently quantified with more than 10 peptides for each biological anaerobic replicate, while it was not found aerobically. Surprisingly, both Hem13 and Hem14 require oxygen for enzymatic activity.

Finally, sterol synthesis also requires oxygen and ergosterol is therefore supplied as anaerobic growth factor during anaerobic yeast cultures [108]. Here, 17 Erg proteins were found in either aerobic and/or anaerobic conditions (**Figure 3b**). Interestingly, Erg proteins that need oxygen, such as Erg1, Erg3, Erg11 and Erg25-28 were either solely present anaerobically or more abundant in anaerobic cultures compared to aerobic cultures.

Survival in stationary phase
A previous study has shown a strong effect of oxygen availability and the presence of transition through diauxic phase on yeast cells robustness during stationary phase [352]. This work proposed that oxygen availability had a positive effect on adenylate energy charge, longevity, stress response and thermotolerance during stationary phase. However, this study was based on changes in transcript levels, without confirmation on what happened at protein level. In particular little is known about the proteome dynamics under anaerobic conditions.

To fill this knowledge gap, the proteome differences between stationary phase in anaerobic and aerobic conditions were studied. In the presence of oxygen, 249 proteins were more abundant and 125 less abundant than in the absence of oxygen (p-value < 0.05 and fold change of ± 1.25, **SI Table 9, SI Figure 4**). In stationary phase, aerobic cells still rely on respiration and proteins involved in respiration were accordingly more abundant under aerobic conditions. Conversely, in the absence of oxygen, yeast cells enter the stationary

phase rather abruptly after glucose depletion. Proteins associated with processes as 'biosynthesis', 'glycolysis' and 'cytoplasmic translation' were more abundant in anaerobic stationary cells compared to aerobic conditions. Several ribosomal proteins were also more abundant, supposedly due to the lack of time and resources to adjust to the altering conditions. Furthermore, proteins involved in storage metabolism, in particular glycogen metabolism, were enriched anaerobically.

To ensure survival during stationary phase, yeast cells accumulate storage carbohydrates in sugar-rich conditions, that can be used as carbon and energy source during famine. Anaerobic cultures are entirely dependent on glycogen and trehalose as energy storage components. Conversely, the presence of oxygen enables yeast cells to metabolize other available nutrients such as lipids and amino acids [102, 106, 353]. Several proteins involved in glycogen metabolism were more abundant in anaerobic compared to aerobic cultures (**SI Figure 3b**). Glucose-6P is the starting point for glycogen metabolism and is converted into Glucose-1P by Pgm1 or Pgm2. Both proteins were detected under aerobic and anaerobic conditions, although Pgm2 was detected with higher coverage and confidence in each biological replicate as it is the major isoform. Pgm2 abundance increased strongly (approx. 2.8-fold) after glucose depletion under anaerobic conditions. Unfortunately, Pgm2 was only detected in aerobic conditions after glucose exhaustion, although it was less abundant compared to anaerobic conditions. Glucose-1P is subsequently converted to UDP-glucose by Ugp1. This enzyme was consistently more abundant in the absence of oxygen over the entire growth curve. Nevertheless, the protein profile was similar under both conditions, as the abundance of Ugp1 increased considerably after glucose was depleted. In the following step, glycogen synthesis is initiated by Glg1 and Glg2, which were not found in our analysis. Subsequently, glycogen is generated by Gsy1/2. Only Gsy1 was detected in our experiments with sufficient confidence in the anaerobic cultures. Glycogen is a polymer that can be branched by Glc3, however, this enzyme was not quantified in our study. Glycogen can also be utilised by Gph1 to form glucose-1P again or by Gdb1 for conversion to glucose. Here, only Gph1 showed a profile comparable to Pgm2.

4
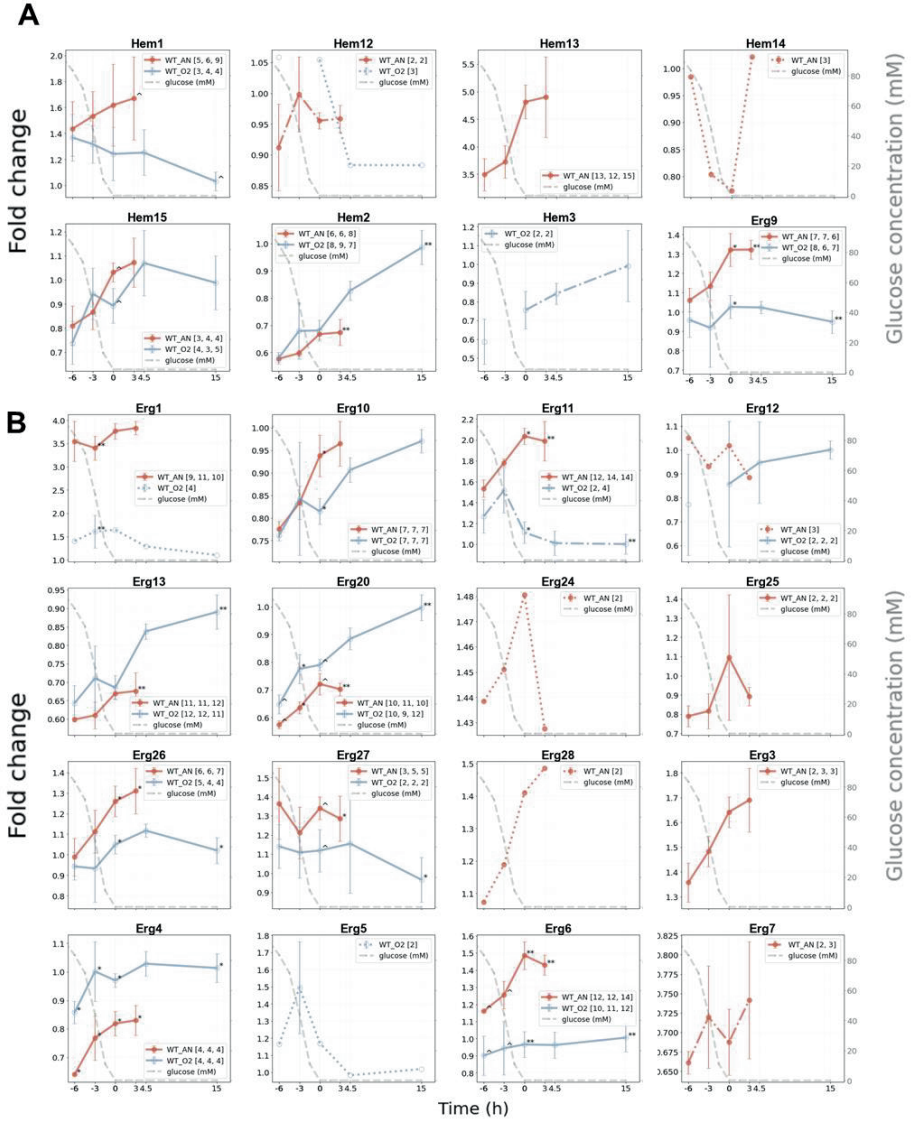
Figure 3. Protein profiles of CEN.PK113-7D in non-respiratory oxygen-dependent pathways during aerobic and anaerobic growth. The biological-replicate-averaged fold-change values were plotted against the time relative to glucose depletion in hours, for proteins involved in heme (A) and sterol (B) synthesis. The different colours of the line graphs represent: orange the control strain under anaerobic conditions (WT_AN), light blue: the control strain under aerobic conditions (WT_O2). The error bars show the standard deviation of the mean of the three biological replicates. The grey dashed line represents the glucose concentration in time in millimolar (mM), shown on the secondary y-axis. The number of quantified peptides per biological replicate are given in brackets. Asterisks (*) and circumflexes (ˆ) indicate the significance between the aerobic and anaerobic experiments. P-values are indicated as follows: < 0.001 (***), < 0.01 (**), < 0.05 (*), and < 0.1 (ˆ).

Trehalose is a second storage metabolite in yeast, which is synthesized from UDP-glucose. It is converted into trehalose-6P by Tps1, and subsequently into trehalose by Tps2. Trehalose is utilised by Nth1, Nth2 and Ath1 and converted into glucose again. In our study, the enzymes leading up to trehalose have similar protein profiles under aerobic and anaerobic conditions, in which the protein abundance increased significantly after glucose depletion (**SI Figure 3b**). Trehalose utilisation, however, was more difficult to capture as only Nth1 was quantified with a low number of unique peptides under aerobic and anaerobic conditions. Furthermore, in the absence of oxygen, yeast cells cannot catabolize fatty acid by beta-oxidation as energy reserve during carbon starvation. Proteins such as Fox2, Pox1 Cat2 and Crc1, showed relatively constant expression levels over the anaerobic growth curve, while abundance increased drastically after glucose depletion in the presence of oxygen (**SI Figure 6**).

Finally, aerobic stationary phase cells are known to acquire increased robustness and stress tolerance during the transition into stationary phase. However, previous studies indicate that this does not apply to anaerobic cultures to the same extent [352]. Stress proteins include a range of heat shock proteins (Hsp) with various functions. The fold changes and levels of Hsp proteins were comparable in the presence or absence of oxygen in exponential growth phase (**SI Figure 5**) but were more abundant in both aerobic and anaerobic conditions towards the end of the growth curve in the MS phase. This increase was markedly larger in the aerobically cultured cells, resulting in a significantly lower anaerobic fold change of the proteins in the stationary phase.

Proteome-level alterations following genetic minimization of the glycolytic pathway
The MG strain genetic reduction consisted in the removal of the 13 minor enzymes involved in glycolysis and fermentation, only leaving the 13 major isoenzymes. The present dataset shows that most minor isoenzymes were expressed and quantifiable in all samples from aerobic and anaerobic batch cultivations, with the exception of Gpm2 and Adh5 only detected in anaerobic cultures, Adh2 only in aerobic cultures, and Adh4, Gpm3 and Pyk2 that were

not detected at all (**Figure 4**). Tdh2, Eno1 and Adh2 were abundant in batch cultures and deletion of the minor isoenzymes might therefore affect yeast physiology (**Figure 5**). The MG strain was previously well characterized by physiological and transcriptome analysis in the presence of oxygen, revealing that the physiology and transcriptome of the MG strains was nearly identical to that of the control strain with a full set of glycolytic and fermentation
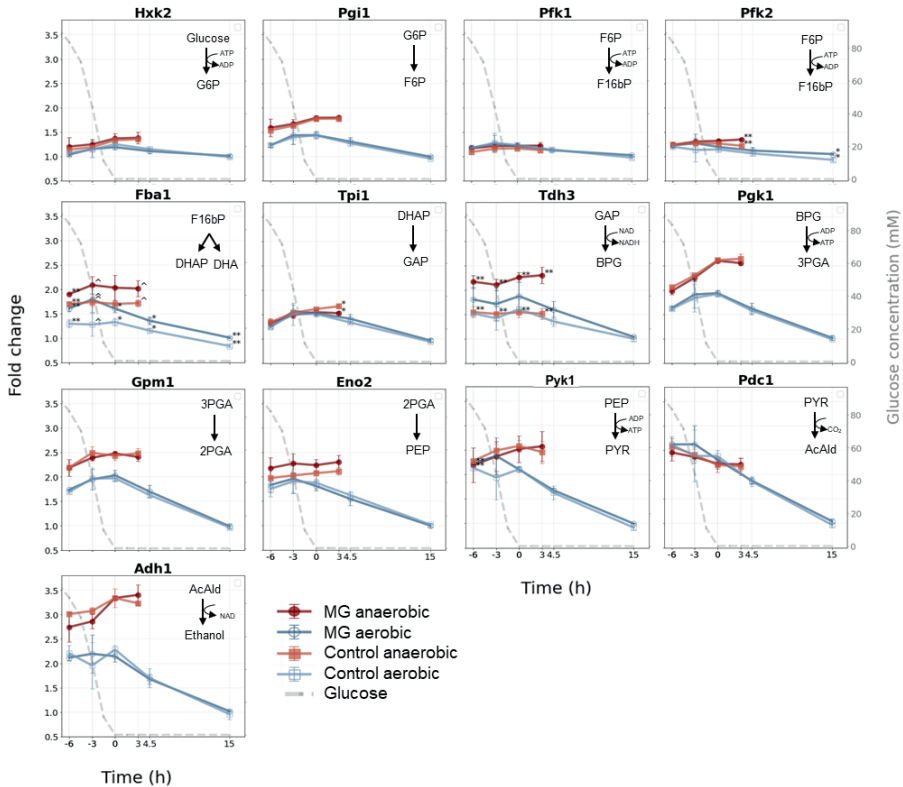
Figure 4. Comparison of major glycolytic enzyme abundances under aerobic and anaerobic growth between control yeast and minimal glycolysis (mutant) strain. The biological-replicate-averaged fold-change values were plotted against the time relative to glucose depletion in hours. The different colours of the line graphs represent: red for the MG (IMX372) strain under anaerobic conditions, dark blue the MG strain under aerobic conditions, orange the control (CEN.PK113-7D) strain under anaerobic conditions, light blue: the control strain under aerobic conditions. The error bars show the standard deviation of the mean of the three biological replicates. The grey dashed line represents the glucose concentration in time in millimolar (mM), shown on the secondary y-axis. Asterisks (*) and circumflexes (ˆ) indicate the significant changes between the control and MG strain. P-values are indicated as follows: < 0.001 (***), < 0.01 (**), < 0.05 (*), and < 0.1 (ˆ).

genes. The proteome of the MG strain was however not explored, and little is known about the response of the MG strain to anaerobiosis. Under anaerobiosis, the fluxes through glycolysis and the fermentation pathway are substantially higher than in aerobiosis [133], the deletion of all minor isoenzymes might therefore have a different cellular impact in the presence and absence of oxygen. As previously observed, the physiology of the MG and control strain in aerobic and anaerobic cultures was near-identical (**SI Table 1**) [120].

Table 2. Number of proteins with significant changes between CEN.PK113-7D and minimal glycolysis (IMX372) yeast under aerobic and anaerobic conditions. The MG protein abundances were normalized to the protein levels of control yeast (CEN.PK113-7D) of the same growth phase under the same condition. A p-value threshold of 0.05 was used to select for significant proteins and a fold change threshold of 1.25 (log2 fold change of +/- 0.32) was used to select for up or down-regulated proteins. The total number of proteins indicates the number of proteins that were detected in at least two biological replicates in both phases that were compared.

MG mutant vs. Control yeast	# Proteins quantified	# More abundant	# Less abundant
Anaerobic			
Mid exponential	978	10	13
Late exponential	978	10	8
Early stationary	978	8	18
Stationary	978	7	6
Aerobic			
Mid-Exponential	1019	18	10
Late-Exponential	950	33	13
Early-Diauxic	1019	15	9
Mid-Diauxic	1019	3	3
Stationary	1019	3	2

The difference in abundance of the glycolytic and fermentation proteins between the MG and the parental control strain was assessed by comparing the expression levels across the entire growth curve using a two-sided two-sample t-test. Remarkably, all major isoenzymes displayed identical time profiles between MG and control strain, very stable under anaerobiosis, and decreasing after mid-diauxic phase under aerobiosis (**Figure 4**). Also, with the exception of Fba1 and Tdh3, the abundance of glycolytic and fermentation proteins was well conserved between the MG and control strains. Under anaerobic conditions, Tdh3 abundance was consistently higher for all growth phases by 40–50% in the MG strain as compared to the control strain (p-value <0.01), but not aerobically. The estimated protein amount of Tdh2 in the control strain is approximately 1/3 of the estimated protein amount of Tdh3, both under aerobic and anaerobic conditions (**Figure 5**). Tdh1 levels were also markedly increased after glucose depletion in the control strain (**Figure 2a**). The loss of these relatively abundant minor isoenzymes and subsequent overall reduction of glyceraldehyde-
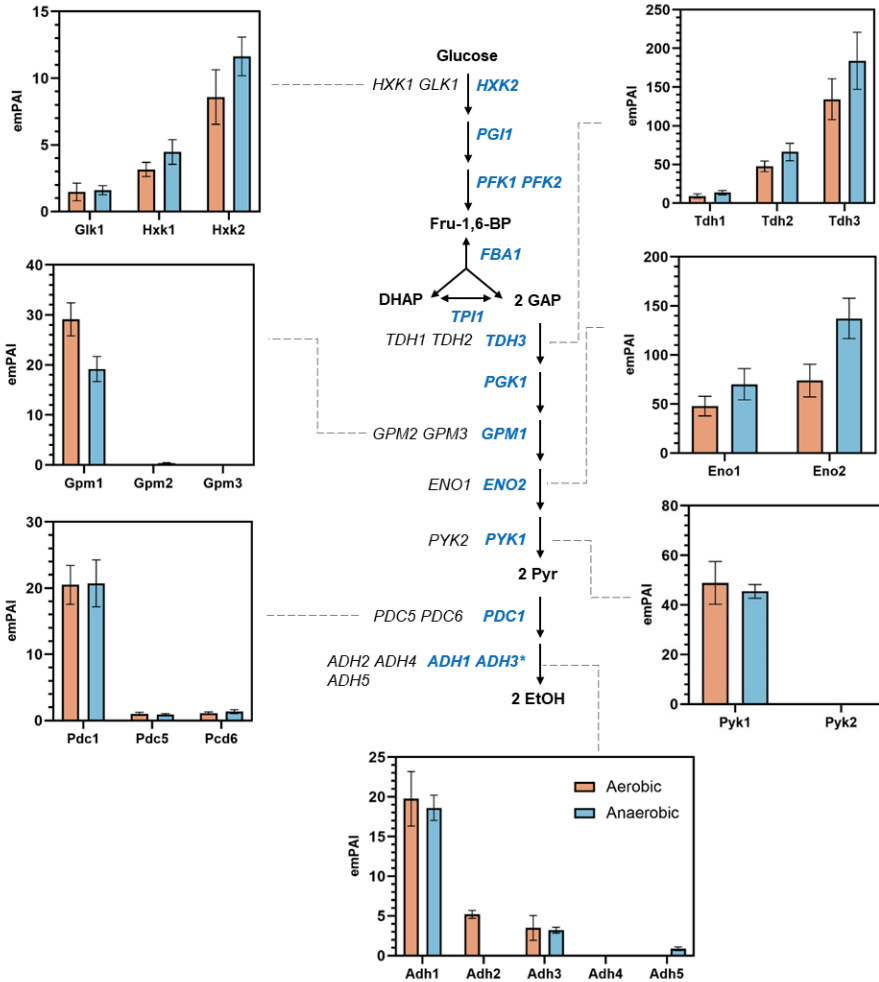
Figure 5. Estimated abundances of the glycolytic isoenzymes under aerobic and anaerobic conditions in CEN.PK113-7D. The Exponentially Modified Protein Abundance Index (emPAI) for the isoenzymes detected in glycolysis under aerobic (red) and anaerobic (blue) conditions, which estimates the absolute protein amount in proteomics by the number of sequenced peptides per protein. The values are calculated from at least one biological replicate culture, standard deviations indicated by the error bars. *Adh3 is a mitochondrial protein.

3P activity in the MG strain might have caused a cross-regulation and increased abundance of Tdh3. Interestingly, while the minor isoenzyme Eno1 was also abundant in the control

strain (**Figure 5**), its deletion had no visible effect on Eno2 level in the MG strain (**Figure 4**). Fba1 abundance was slightly (~20–40%) but significantly higher in the MG than in the control strain, both in the presence and absence of oxygen. Fba1 does not have isoenzymes, this difference in abundance could therefore not be attributed to cross-regulation. Fba1 is an abundant protein in yeast, operating far from saturation [133], and the flux through glycolysis is not increased in the MG as compared to the control strain. This increase in Fba1 abundance is therefore not likely explained by a need for higher aldolase capacity in the MG strain.

Overall, the enzyme level adjustments within the glycolytic and fermentation pathways following the deletion of the minor isoenzymes were remarkably small, but slightly more pronounced under anaerobic conditions. In line with this mild phenotype, the proteome was not visibly affected by the deletion of the minor isoenzymes (statistical significance cut-off of 0.05 (5%), fold change threshold of +/-1.25. **Table 2, SI Table 10 and 11**). Consequently, no enriched or depleted GO terms or KEGG pathways, could be identified. The metabolic comparability between both strains was therefore underlined by their similar CCM enzyme abundance and global proteome profiles (**Figure 2a**).

Discussion

Yeast proteome dynamics have been studied for aerobic and anaerobic conditions over the past decades [110, 144, 324, 326]. However, the transition from exponential into stationary phase under anaerobic conditions has not been investigated to date. This provides the most comprehensive study to date of the *S. cerevisiae* CEN.PK113-7D proteome response to oxygen and nutrient availability using tightly controlled batch bioreactor cultures. Moreover, biological triplicates and optimised sample preparation protocols ensure the accuracy and reproducibility of the present dataset. Additionally, this study investigates the impact of genetic minimization of the glycolytic pathway using the recently established 'minimal glycolysis' strain [23]. The present dataset therefore enables the quantitative analysis of 54 individual proteomes, thereby covering approx. 40% of all yeast proteins and *ca.* 99% of the complete protein biomass in yeast. The high reproducibility in the dataset was also highlighted by the strong similarities between the proteomes of the MG and control yeast. This comprehensive and accurate dataset therefore provides an ideal resource for applied and fundamental studies in yeast, and more particularly for *in silico* proteome allocation studies.

The most remarkable result from this study is the substantially smaller proteome response of yeast cells grown under anaerobiosis as compared to aerobiosis. In both conditions yeast cells have to tune their metabolism to the transition from glucose excess to exhaustion, leading to a shift from exponential to stationary growth. These drastic changes however triggered a far milder response than the aerobic transition to and out of diauxic phase, which represented 58% of the measured changes in protein abundance. During the diauxic shift, cells rewire the

proteome for respiratory growth on ethanol as main carbon source. This transition leads to a multitude of physiological and morphological changes, *e.g.* smaller cell size, increased mitochondrial volume, decreased growth rate, increased respiration rate and therefore increased ROS production, induction of gluconeogenesis and glyoxylate cycle, and large changes in fluxes in central carbon metabolism. These changes were well mirrored by the modifications in proteome allocation. For instance an increase in mitochondrial protein abundance was observed, while the glycolytic proteins were downregulated [143, 147, 149, 323]. Transition from diauxic phase to stationary phase led to the further modification of the abundance of ca. 120 proteins. Conversely, as little as 55 proteins showed altered abundance after glucose depletion under anaerobiosis, and prolonged cultivation in stationary phase did not further alter the proteome. Using the exact same strain and experimental set-up, Bisschops *et al*. (2015) [155] showed a stronger and faster decrease in viability upon glucose exhaustion for anaerobic than aerobic cultures. Based on physiological and transcriptome data, the authors attributed this lack of robustness to cells inability to adapt to glucose exhaustion in the absence of oxygen. Conversely, in the presence of oxygen the diauxic shift gives the time and resources to transition from fast growth to growth arrest. The present proteomics study supports this view in different ways. The small protein response during the transition from sugar excess to depletion suggests that the anaerobic cells do not have the means, or proper regulatory network to adjust to the new conditions. Furthermore, aerobic cultures acquire robustness and stress tolerance during the transition into stationary phase thanks to the expression of a set of 'stress-response' genes [354], such as heat shock proteins (Hsp) that increase tolerance to high temperatures. Both aerobic and anaerobic cultures showed a similar 'stress signalling' as shown by the increase in Hsp proteins towards stationary phase, however this increase was far less pronounced in anaerobic cultures. Abundance of Hsp was therefore substantially lower in the absence of oxygen, in line with the lower transcription of Hsps and lower thermo-tolerance of anaerobic stationary phase cultures as compared to aerobic cultures observed by Bisschops *et al.* [155]. Considering that, for practical and financial reasons, industrial-scale processes favour anaerobic environment, the present results bring valuable information for the construction of metabolism predictive models [156, 160, 163, 355–359].

In the presence of oxygen, yeast cells switch to respiration once glucose is depleted. Accordingly, the abundance of respiration-related proteins increased upon glucose exhaustion in aerobic cultures, while the protein profiles in anaerobic cultures remained constant. Several other non-respiratory pathways in *S. cerevisiae* are oxygen-dependent, such as fatty acid beta-oxidation, heme- and sterol synthesis. Expectedly, most detected proteins in these pathways were aerobically more abundant or contained similar abundance profiles to anaerobic cultures. Nevertheless, oxygen-dependent protein Hem13 involved in heme synthesis was only confidently quantified under anaerobic conditions and the lack of detection in aerobic conditions suggests that it is far less abundant. Transcription of *HEM13*

is repressed by oxygen and heme itself [360, 361], and therefore it lacks repression in the absence of oxygen and thus increasing its abundance. Similar protein profiles under anaerobic conditions were previously found for Hem1, Hem14 and Hem15 [326]. Several oxygen-dependent Erg proteins were also more abundant or solely detected under anaerobiosis. Here, transcription of various Erg proteins is regulated by oxygen and the absence leads to increased expression of Erg proteins [362] [363].

Glycolysis and alcoholic fermentation are well-studied pathways that play an important role in sugar conversion in the biotechnology industry. Earlier studies have shown that major glycolytic isoenzymes are abundant proteins which expression remains relatively stable irrespective of the growth environment, although these measurement often rely on transcript data, or enzyme assays that cannot distinguish between isoenzymes [120]. The present proteome dataset confirms that the abundance of most major glycolytic isoenzymes decreased during the diauxic shift and further during stationary phase in the presence of oxygen. Conversely, their abundance was unaffected by the transition to stationary phase under anaerobic environment. Aerobic cultures in stationary phase therefore display substantially lower glycolytic enzymes levels than anaerobic cultures, for instance a 2.5- to 3.5-times lower abundance of Pgk1, Gpm1, Pdc1 and Adh1. While this difference in abundance is not expected to affect survival in stationary phase, in which the glycolytic flux is extremely low or absent, it will influence the ability of stationary phase cells to reach fast growth when transitioned to sugar-rich medium. When exposed to anaerobic sugar excess, cells grown aerobically to stationary phase have to allocate resources to increase the abundance of glycolytic enzymes and reach fast growth, while cells pre-cultured anaerobically do not. This aspect should be considered during the start-up phase of anaerobic industrial fermentations and their modelling.

The expression of minor isoenzymes is condition-dependent and several are reported to have distinct functions, especially during changes in carbon source availability. The present dataset shows that the presence of oxygen only visibly affected the abundance of Eno1, Tdh1 and Hxk1, as the abundance was significantly higher after glucose depletion under anaerobic conditions compared to aerobic conditions (**Figure 2**). While some minor isoenzymes have substantial abundance in yeast cells (**Figure 5**, *e.g.* Eno1, Tdh2, Hxk1), their removal did not trigger substantial changes in abundance of the major isoenzymes. It is however notable that Tdh3, glyceraldehyde dehydrogenase major isoenzyme, showed a ca. 1.5–fold increase in abundance in the MG strain compared to the control strain under anaerobic conditions (**Figure 4**). The same trend was observed aerobically, albeit much less pronounced. As most glycolytic enzymes, glyceraldehyde dehydrogenase operates at overcapacity, meaning that the enzyme capacity largely exceeds the flux catalysed *in vivo* [364]. Therefore, Tdh3 increased abundance in the MG strain does most likely not result from the need to compensate for Tdh1 and Tdh2 deletion to maintain the glycolytic flux. The glyceraldehyde

dehydrogenase isoenzymes do not have any well-described moonlighting function. However, next to their cytosolic localization, they are also found in the cell wall in which they might play a yet uncovered role. The composition and structure of *S. cerevisiae* cell wall is affected by oxygen and several cell wall proteins are specifically enriched under anaerobiosis (*e.g.* cell wall mannoprotein of the Srp1p/Tip1p family) [365], which might explain the observed cross-regulation in the MG strain. Fba1 was also mildly but significantly upregulated in the MG mutant both aerobically (1.2 to 1.4–fold change) and anaerobically (1.15–fold change). As Fba1 does not have isoenzymes and is solely responsible for the glycolytic flux, its change in abundance is difficult to explain. Fba1 is also involved in vacuolar function as subunit of the vacuolar V-ATPase [130]. However, as no or little differences were observed in other components of the V-ATPase between MG and the control strain, the molecular mechanism leading to Fba1 slightly higher abundance in MG remains unclear. Many factors can alter the functionality of proteins, including post-translational modifications, protein localisation, or interactions with other proteins or biomolecules [167, 366]. A recent study suggested that phosphorylation regulates the activity of most glycolytic enzymes [158]. However, the stable abundance of glycolytic proteins between the MG and the control strain was well reflected in the stability of *in vitro* enzyme activity [120], suggesting the lack of differences in post-transcriptional regulation between the MG and control strains. Overall, remarkably few proteome changes were observed as consequence of glycolysis genetic reduction. This similarity between MG and control strain further support the role that MG can play as simplified model organism in proteome allocation studies, and to explore the role of post-translational modifications in the regulation of glycolysis.

The complete proteome dynamics and abundance data for the batch reactor cultured CEN.PK113-7D strain and the related 'minimal glycolysis' mutant strain for aerobic and anaerobic growth, can be found in the **SI Table 12**. Mass spectrometric raw data and unprocessed search files are publicly available via the PRIDE repository under the project code PXD031412.

Acknowledgments

Supplementary Information

Supplementary Excel tables and SI Figure 6 can be found online:
https://doi.org/10.1101/2022.09.23.509138

SI Table 1. Biomass-specific substrate consumption and product formation rates of MG (IMX372) and control (CEN.PK113-7D) strain during aerobic and anaerobic exponential growth. The average and standard deviations were calculated for three biological replicates per strain. The biomass specific consumption or formation rates were calculated for the substrate glucose (q_s), ethanol (q_{EtOH}), glycerol(q_{Glyc}), acetate (q_{Ace}), carbon dioxide (q_{CO2}), and oxygen (q_{O2}). Significant differences ($p<0.05$) between control and MG yeast were calculated with a two-sided two-sample unpaired t-test and are highlighted with an asterisk. #The qCO2 of IMX372 under anaerobic conditions is likely underestimated due to technical complications in the CO2 off gas analysis equipment.

	Aerobic				Anaerobic			
	CEN.PK113-7D		IMX372		CEN.PK113-7D		IMX372	
	Average	Stdev	Average	Stdev	Average	Stdev	Average	Stdev
μ^{max} (h^{-1})	0.413	0.012	0.396	0.002	0.366	0.003	0.355	0.012
q_s (mmol·gDW^{-1}·h^{-1})	-18.890	1.029	-18.585	0.580	-23.264	1.792	-20.677	0.659
q_{EtOH} (mmol·gDW^{-1}·h^{-1})	30.166	1.999	27.063	0.446	31.814	3.049	30.999	1.677
q_{Glyc} (mmol·gDW^{-1}·h^{-1})	1.736	0.074	1.518	0.147	4.607	0.395	4.297	0.198
q_{Ace} (mmol·gDW^{-1}·h^{-1})	0.494	0.082	0.493	0.080	0.560	0.092	0.540	0.045
q_{CO2} (mmol·gDW^{-1}·h^{-1})	33.024	2.056	28.175	0.912	32.037	1.213	27.053*#	1.262
q_{O2} (mmol·gDW^{-1}·h^{-1})	9.479	0.533	7.978*	0.149				
RQ	3.517	0.173	3.574	0.215				
$Y^{biomass/glucose}$ (gDW·gglucose^{-1})	-0.022	0.001	-0.021	0.001	-0.016	0.001	-0.017	0.000
$Y^{ethanol/glucose}$ (mol·mol^{-1})	1.598	0.111	1.458	0.025	1.379	0.032	1.503	0.129
$Y^{glycerol/glucose}$ (mol·mol^{-1})	0.092	0.009	0.082	0.009	0.199	0.027	0.208	0.004
$Y^{acetate/glucose}$ (mol·mol^{-1})	0.026	0.003	0.026	0.004	0.024	0.004	0.026	0.001
Carbon recovery	1.000	0.000	1.000	0.000	1.000	0.000	1.000	0.000

4
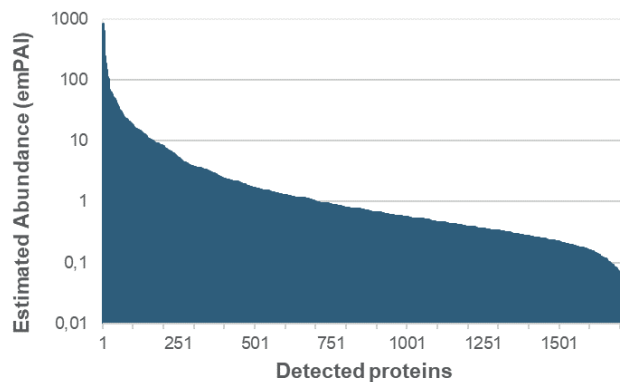
SI Figure 1. Estimated absolute abundance of detected proteins across all experiments. The absolute abundance was estimate using the Exponentially Modified Protein Abundance Index (emPAI) for all identified proteins across all experiments. The absolute protein amount is estimated by the number of sequenced peptides per protein. The abundance of each protein is an average of all experiments and detected in at least one biological replicate.
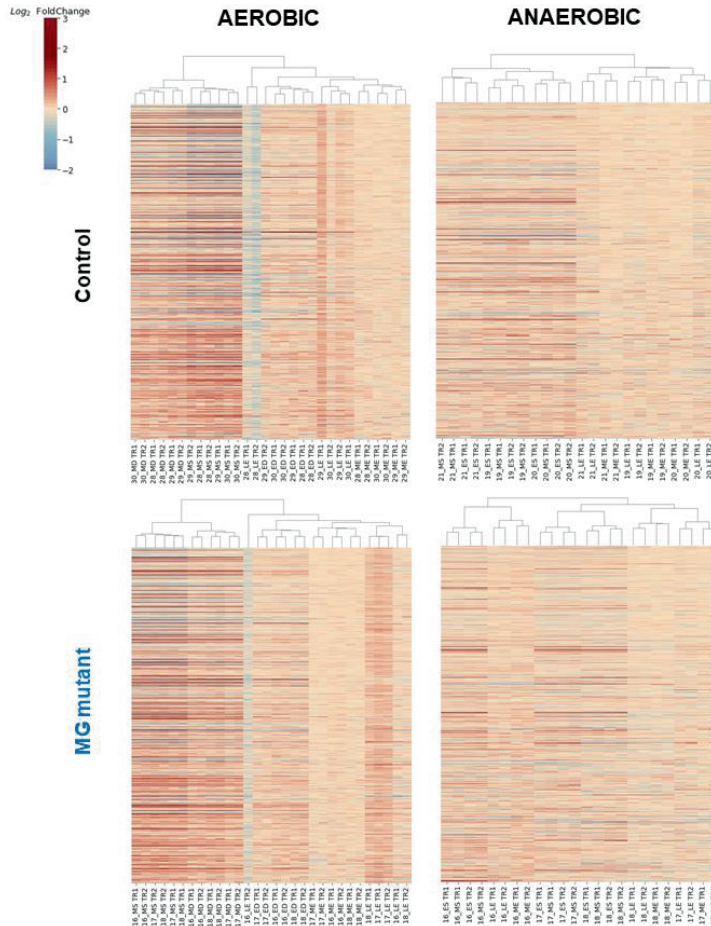
SI Figure 2. Cluster analysis of abundance changes of proteins from the control and MG yeast strain under aerobic and anaerobic conditions. The replicates were clustered based on Euclidean distances and with the average linkage method. The protein fold changes were normalised to the average of all the ME-phase experiments. The log2 of the fold changes was calculated and plotted so that no fold change (=0) is coloured white, a negative fold change is coloured blue and a positive fold change red. The columns labels represent the reactor number - indicating the biological replicate - the growth phase, and the technical replicate (TR) number. ME: mid-exponential, LE: late-exponential, ED: early-diauxic, MD: mid-diaux, ES: early-stationary, MS: mid-stationary.
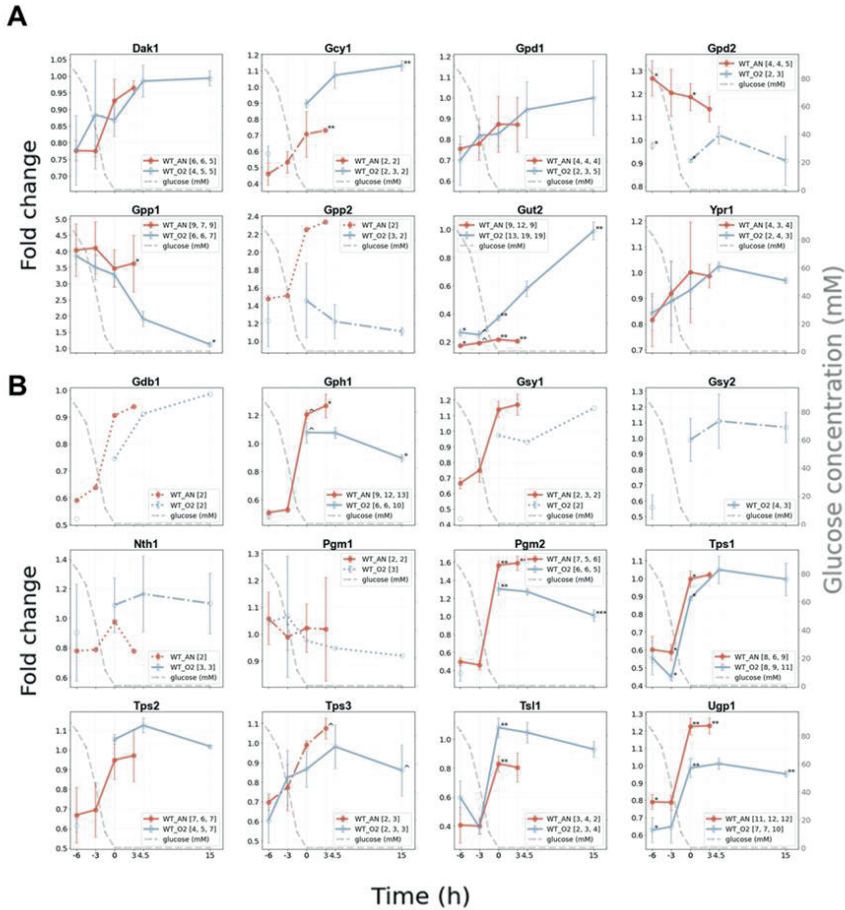
SI Figure 3. Protein profiles of CEN.PK113-7D in glycerol, glycogen and trehalose metabolism during aerobic and anaerobic growth. The biological-replicate-averaged fold-change values were plotted against the time relative to glucose depletion in hours for proteins involved in glycerol (A) and glycogen and trehalose (B) metabolism. The different colours of the line graphs represent: orange the control strain under anaerobic conditions (WT_AN), light blue: the control strain under aerobic conditions (WT_O2). The error bars show the standard deviation of the mean of the three biological replicates. The grey dashed line represents the glucose concentration in time in millimolar (mM), shown on the secondary y-axis. The number of quantified peptides per biological replicate are given in brackets. Asterisks (*) and circumflexes (ˆ) indicate the significance between the aerobic and anaerobic experiments. P-values are indicated as follows: < 0.001 (***), < 0.01 (**), < 0.05 (*), and < 0.1 (ˆ).
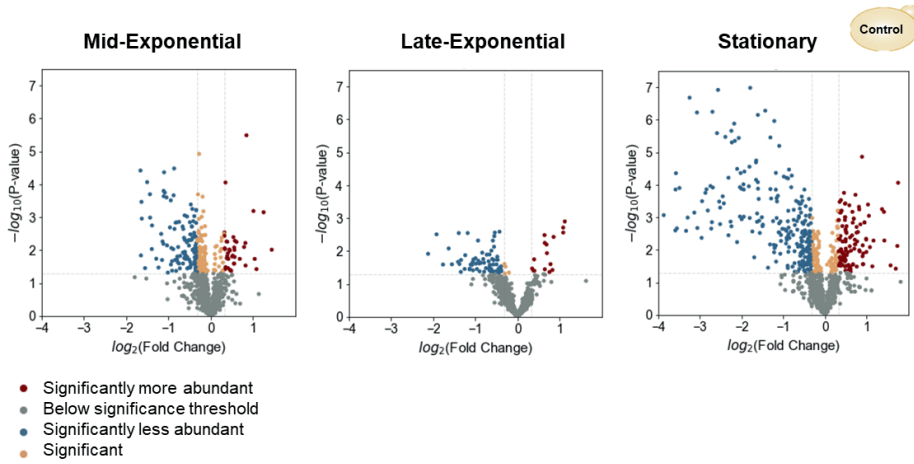
SI Figure 4. Global proteome changes between aerobically and anaerobically cultured CEN.PK113-7D yeast cells. The log2 of the abundance fold change between the two conditions (normalised to the aerobic experiments) was plotted against the -log10 of the p-value. The mid-exponential (ME), late-exponential (LE) and mid-stationary (MS) phases were compared. P-value threshold < 0.05, fold change threshold > 1.5 (log2 fold change threshold +/- 0.32).
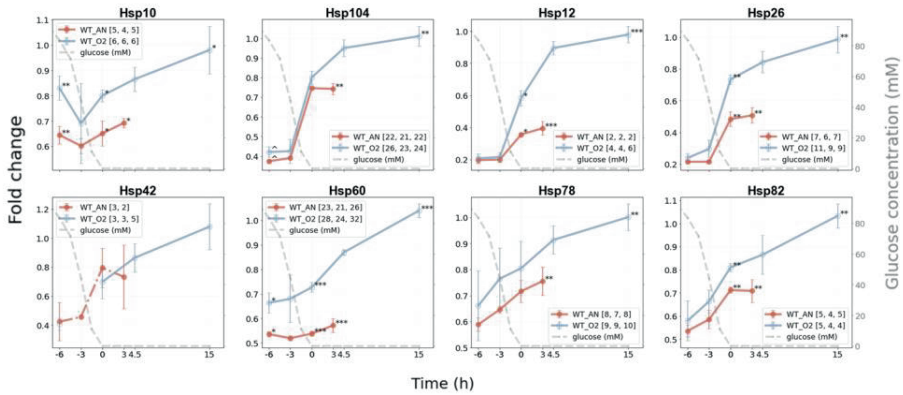
SI Figure 5. Fold changes (FC) of heat shock proteins from CEN.PK113-7D under aerobic and anaerobic conditions. The biological-replicate-averaged FCs were plotted against the time relative to glucose depletion (0h) in hours. Orange: CEN.PK113-7D under anaerobic conditions, light blue CEN.PK113-7D under aerobic conditions. The error bars show the standard deviation of the mean of the three biological replicates. In the legend, the numbers between brackets represent the number of unique peptides that were found in each biological replicate. The grey dashed line represents the glucose concentration in time in millimolar (mM), shown on the secondary y-axis. Asterisks (*) and circumflexes (^) indicate the significance between the aerobic and anaerobic experiments. Different p-values were indicated: < 0.001 (***), < 0.01 (**), < 0.05 (*), and < 0.1 (^).

Chapter 5

Towards quantifying the degree of modification in metabolic enzymes

Maxime den Ridder, Casper van der Luijt, Ramon van Valderen, Pascale Daran-Lapujade and Martin Pabst

Abstract

Post-translational modifications modulate protein properties in response to environmental or developmental changes. To date, more than one hundred different types of modifications have been identified across all domains of life. Nevertheless, the majority of modifications are rare and their biological roles have not been identified to date. On the other hand, modifications such as phosphorylation, acetylation or glycosylation, are very abundant and are known to translate important functions such as regulating enzyme activity, protein stability and degradation, signalling, or cell morphology. These modifications have also been frequently studied in many microbes such as *E. coli* and yeast. However, studies on post-translational modifications usually cover only a single or a few modifications. Furthermore, studies commonly only investigate sequence regions that are accessible via tryptic peptides. Consequently, the presence of other modifications or less assessable sequence regions are overlooked. However, measuring all possible modifications creates an enormous combinatorial sequence space that is difficult to capture by current methods. Unfortunately, there are currently no effective solutions to this challenge. Nevertheless, when studying modifications in fundamental processes such as the regulation of the metabolic flux, or metabolic diseases such as cancer, the complete repertoire of possible post-translational modifications needs to be considered. Furthermore, some modifications may represent attractive targets for engineering microbial cell factories to improve enzymatic conversions and yields. In this study, we describe and provide a proof of concept for an approach that enables to quantify the global degree of modification for individual enzymes from complex cell lysates. This approach quantifies the unmodified fraction of each peptide, and therefore considers all types of modifications, regardless their chemical nature. Quantification of the complete protein sequence is realised by the aid of a cell free synthesis produced protein standard. Finally, we exemplify this approach by monitoring the global protein modification changes of the glycolytic enzyme Pyk1 in yeast, during transition from proliferation to stationary phase under aerobic conditions. Interestingly, the majority of sequence regions showed significant changes at least at one point during the growth curve. Further development of this approach could allow to monitor complete metabolic pathways. Such experiments allow to explore the role of post-translational modifications in the regulation of metabolic pathways in yeast and beyond.

Introduction

Through billions of years of evolution, cells acquired different levels of regulation to control cellular behaviour. In particular microbial cells require to adapt rapidly to the ever changing environment. Transcriptional and translational control are arguably the best understood levels of control. However, these mechanisms are relatively slow (minutes to hours) and merely affect the protein abundance. Moreover, these types of regulation have failed to explain many of the observed responses associated with metabolic operations [133]. Nevertheless, allosteric regulation and post-translational modification (PTM) enable an additional layer of regulation. Both mechanisms can rapidly change the metabolic phenotype [186] and PTMs have attracted a lot of attention recently. PTMs are enzyme-mediated, covalent modifications of the protein backbone and can serve multiple purposes, such as regulation of enzyme activity, protein localization, or interaction with other cellular molecules [182, 201].

Proteome dynamics in *S. cerevisiae* predominantly depends on nutrient availability and drastic changes are observed in the central carbon metabolism upon substrate level changes [146, 147, 324, 325, 367]. A major challenge is to understand how these metabolic pathways are controlled by these complex regulatory mechanisms. Several recent studies confirmed that a large fraction of the yeast proteome is post-translationally modified at least once during the life time of a protein [189, 368]. For example, glycolysis, a central pathway in nearly all living cells (including yeast), is known to be highly regulated on multiple levels. Glucose is oxidized into two pyruvate molecules in this pathway to generate energy, while many intermediates provide branching points to other pathways in cell metabolism [99]. Furthermore, glycolysis is coupled to fermentative production of ethanol under conditions of oxygen limitation and/or sugar excess [100]. Glycolytic enzymes are amongst the most abundant proteins in the cell and therefore these enzymes are regulated at multiple layers. Recently, it has been demonstrated that regulation of the glycolytic flux is predominantly caused by post-translational processes [133, 158]. Interestingly, increased phosphorylation resulted in decreased enzyme activity and lower growth rates of *S. cerevisiae* under glucose-limited conditions, indicating a negative correlation between phosphorylation and activity of metabolic enzymes [142]. In addition, the shuttling between nucleus and cytoplasm of Hexokinase 2 is also regulated by phosphorylation [141]. Protein phosphorylation is one of the most frequent modifications, but also most studied PTMs in yeast, however, also other modifications such as methylation [369], acetylation [165], succinylation [252], ubiquitination [169], nitration [370] and O-mannosylation [371] have been reported for glycolytic enzymes. Outside of glycolysis, glycosylation of proteins, through O-mannosylation and N-glycosylation, can play an important role in various cellular processes including protein folding, cell-cell interaction and cell wall integrity [256, 372–375]. Interestingly, yeast shows also some rare modifications such as lipoylation, which only modifies a few enzyme complexes across the complete yeast proteome [190, 191].

5

PTMs are commonly identified directly (or after enrichment) from whole cell lysate digests using liquid chromatography coupled to high-resolution tandem mass spectrometry [180]. Advances in resolution, accuracy and quantitative properties of mass spectrometers continuously enhanced the possibilities to explore complex biological samples. While this has led to the discovery of many new modifications and modification sites across all proteins (also in yeast [189]), the modification stoichiometry and occurrence under different conditions remains generally unexplored. Knowing the degree of modification and their stoichiometry, however, is important to understand the possible biological significance of the modification for an enzyme or pathway, respectively. Detection of post-translational modifications is usually based on bottom-up approaches, in which proteins are cleaved into peptides using a protease to systematically identify modification sites [18]. The stoichiometry of the modified peptides can be measured by various approaches. The absolute amount of a modified variants, for example, can be measured by using stable isotope-labelled (modified) peptide standards [69, 376], which is also referred to as absolute quantification (AQUA) [69]. Similarly, stable isobaric labelling *e.g.* via SILAC (stable isotope labelling by amino acids in cell culture) allows to accurately determine the relative abundance changes of the modified peptide over time [377, 378]. However, as modified peptides are often low abundant, enrichment or targeting of the modified peptides (selective reaction monitoring, SRM, or parallel reaction monitoring, PRM) is employed [379, 380]. Tandem Mass Tags (TMTs) have also been used to enhance the detectability of low-abundant peptides, as the differently labelled peptides appear as single (combined) peaks in the spectra [381–383]. However, determining the modification stoichiometry from the relative signal intensity of native and modified peptide is misleading because both can be expected to differ in the ionisation efficiency. To overcome this issue, Steen *et al*. developed a label-free quantification approach to determine the 'flyability' of a phosphorylated variants and its unmodified peptide counterpart to calculate the stoichiometry [384]. Nevertheless, this method provides only an approximation and it cannot be applied to different types of PTMs. As mentioned before, methods usually focus on a single type of modification (or only a few) but neglect the possible spectrum of other modifications that may be present on the same protein or pathway [366, 385]. Moreover, it is known, that very often several different types of modifications are present on one protein simultaneously, and that modifications may interact with other modifications (often referred to as "PTM cross-talk"), which makes them functionally associated [169–171, 386]. Therefore, Mertins *et al*. developed a method which used serial enrichment to simultaneously study phosphorylation, acetylation and ubiquitination [387]. Other studies also achieved the identification of several different modifications via enrichment strategies, however, none of these groups measured the modification site stoichiometry [172, 173, 265]. An *a priori* knowledge on the presence of the types of modifications (out of all possible) is generally not available, and targeting all possibly modifications across several modification sites is not achievable. Furthermore, there is no

generic enrichment strategy for all types of modifications, and it is also not possible to produce and measure standards for all possible peptide variants.

Therefore, a method that quantifies the unmodified peptide fraction instead of the modified peptide provides an elegant solution to this challenge. Here, the quantification is done without identifying or localising the individual modifications and modification sites. If a peptide is modified by any type of modification (or several), the fraction of the unmodified peptide will decrease proportionally. This provides thus a quantitative measure of the overall degree of modification on a protein or a pathway, respectively. Changes in the degree of modification can be easily monitored even from complex cell lysates. This allows therefore to study changes in response to environmental conditions to investigate possible regulatory roles in a protein. Nevertheless, this approach requires an unmodified reference peptide for every peptide sequence of a protein [388]. Furthermore, this strategy requires a complete sequence coverage, in order to cover the complete spectrum of possible modifications sites. The production of synthetic peptides for single proteins (*e.g.* 3–5 unique peptides per protein) is relatively affordable, but it becomes increasingly expensive, when the complete protein sequence, different proteolytic fragments and complete metabolic pathways have to be considered. Secondly, accurate quantification with synthetic peptides can be challenged, because the peptides do not reflect biases from sample preparation [389]. A solution to this is the alternative synthesis of combined amino acid strings, referred to as "quantitative concatamer" (QconCAT) proteins [70]. Here, proteotypic peptides from multiple proteins are concatenated into a designer gene that is labelled with stable isotopes in cell free systems [390]. However, full sequence coverage of the target proteins is not obtained by this approach. Interestingly, the recombinant small scale production of complete proteins using cell-free protein synthesis has become popular in recent years [391–393]. Advantageously, this approach provides proteins which are fully unmodified [394]. Mair *et al.* (2016) used such an approach to quantify the degree of phosphorylation on Alzheimer's disease-associated proteins [395], while other recent studies used protein standards produced through cell free protein synthesis systems to generate reference peptides in quantitative proteome analyses [396, 397]. However, a comparable strategy has not yet been applied to monitor the (unknown) modification landscape of native metabolic enzymes in the presence of complex cell lysates.

An additional challenge when analysing metabolic enzymes is the coexistence of isoenzymes. Isoenzymes often show a large sequence homology, and they may lack unique peptide sequences for large parts of the protein. For example, in yeast glycolysis, 8 out of the 12 enzyme-catalyzed reactions are represented by multiple paralogous genes, which share similar amino acid sequences (~30–99% identity). These paralogs consist of major and minor isoforms, which resulted from a whole genome duplication event in an ancestor of *S. cerevisiae* [119, 120]. Advantageously, the recently established minimal glycolysis (MG)

5

strain [120]–in which the redundant isoenzymes have been removed–provides an ideal proteome model to study the PTM landscape of glycolytic enzymes. Additionally, no phenotypic response has been observed as a result of the deletion of the minor isoenzymes, which confirms the suitability of such a strain for studying the regulation of metabolic enzymes [120, 398].

In this study, we exemplify the quantification of the degree of modification of the key glycolytic enzyme Pyk1 from yeast. For this, we evaluated the use of cell free synthesis produced Pyk1 protein standard to provide for every peptide across the protein backbone a reference signal. We validate different quantification approaches including high-resolution fragmentation free quantification, and quantification based on reporter ions using tandem mass tag (TMT) labelling. The TMT approach generally provides improved sensitivity and accuracy for low abundant peptides which has been also demonstrated for single-cell proteomics approaches only recently [399]. Additionally, a multi-protease approach was tested to confirm that a near complete sequence coverage can be achieved in the presence of complex cell lysates. Ultimately, further developments of this approach will enable to quantify the degree of modification of complete metabolic pathways. This will allow to deconvolute the contribution of post-translational modifications in the regulation of fundamental metabolic pathways in yeast and beyond.

Materials and Methods

Yeast strains, media and storage. The MG yeast strain IMX372 (*MATa ura3-52 his3-1 leu2-3,112 MAL2-8c SUC2 glk1::SpHis5, hxk1::KlLEU2, tdh1::KlURA3, tdh2, gpm2, gpm3, eno1, pyk2, pdc5, pdc6, adh2, adh5, adh4*) used in this study shares the CEN.PK genetic background [120, 288]. Shake flask and batch cultures were grown in synthetic medium (SM) containing 5.0 g·L⁻¹ $(NH_4)SO_4$, 3.0 g·L⁻¹ KH_2PO_4, 0.5 g·L⁻¹ $MgSO_4·7H_2O$ and 1 mL·L⁻¹ trace elements in demineralized water. The medium was heat sterilized (120°C) and supplemented with 1 mL·L⁻¹ filter sterilized vitamin solution and 20 g·L⁻¹ heat sterilized (110 °C) glucose (SMG) [289]. The bioreactor medium was supplemented with 0.2 g·L⁻¹ antifoam Pluronic PE 6100 (BASF, Ludwigshafen, Germany). *E. coli* cultures were grown in Luria-Bertani (LB) liquid medium or on LB solid plates containing 100 μg·mL-1 ampicillin. Frozen stocks of *E. coli* or *S. cerevisiae* cultures were prepared by the addition of glycerol (30% v/v) in 1 ml aliquots for storage at -80 °C.

Bioreactor cultures and proteome sampling. MG yeast proteome samples were taken from aerobic batch cultures in the study by den Ridder *et al*. [398] or aerobic chemostat cultures in den Ridder *et al.* (2022) [327]. Briefly, triplicate aerobic batch cultures of MG yeast were performed and sampled for proteome analysis in different growth phases of the aerobic yeast growth curve. Proteome samples (approx. 5 g·L⁻¹ dry weight) were taken at 6, 9, 12, 16.5, 27

hours to reflect the proteome in mid-exponential, late-exponential, early-diauxic, mid-diauxic and stationary growth phase. For the chemostat samples, proteome samples (approx. 3.6 g·L^{-1} dry weight) were taken at steady state conditions. In both cases, proteome samples were collected in multifold in trichloroacetic acid (TCA) (Merck Sigma, Cat. No. T0699) with a final concentration of 10%. Samples were centrifuged at 4,000 g for 5 min at 4°C. Cell pellets were frozen at -80°C [272].

Cell free production of complete protein standards. The *HXK2, PFK1, PFK2, PGK1* and *PYK1* genes from *S. cerevisiae* were transferred to the pT7-ptT7 backbone to allow for expression in the cell-free synthesis system (New England Biolabs, Bioke, Leiden, The Netherlands). PCR amplification with Phusion® Hot Start II Polymerase of pUD1079 with primers 16233 and 16234 was performed to create a linearized backbone (**SI table 1** and **2**). The *HXK2, PFK1, PFK2, PGK1* and *PYK1* genes were PCR amplified from pUDE767, pUDE769, pUDE770, pUDE774 and pUDE777 with primers 16235 and 16236, 16829 and 16830, 16831 and 16832, 18189 and 18190, 16833 and 16834, respectively. PCR products were separated by electrophoresis on 1% (w/v) agarose gels in 1 x TAE buffer (Thermo Fisher Scientific) at 100 V for 30 minutes. The PCR products were incubated with 0.5 μL (10 U·μL-1) DpnI (Thermo Fisher Scientific) for 1 hour at 37 ˚C to remove template DNA. DNA fragments obtained by PCR were purified using the GenElute™ PCR Clean-Up Kit (Sigma-Aldrich). When unspecific PCR products were observed, the fragments were excised on a Safe Imager™ 2.0 Blue-Light transilluminator (Thermo Fisher Scientific) and purified using the Zymoclean Gel DNA Recovery Kit (Zymo Research, Irvine, CA, USA) according to the manufacturer's instructions. The pT7-ptT7 linearized backbone was then Gibson assembled using NEBuilder HiFi DNA Assembly (New England Biolabs) with the PCR amplified *HXK2, PFK1, PFK2, PGK1* or *PYK1* gene, resulting in plasmids pUD1081, pUD1107, pUD1108, pUD1184 and pUD1109, respectively. Restriction analysis was performed with FastDigest™ restriction enzymes (Thermo Fisher Scientific) according to the manufacturer's protocol with one minor adjustment. The restriction mixture was incubated for 1 hour at 37°C. Plasmid DNA was isolated from *E. coli* using the Sigma GenElute Plasmid kit (Sigma-Aldrich) following the manufacturer's protocol, while yeast plasmids were isolated using the Zymoprep Yeast Plasmid Miniprep II Kit (Zymo Research). Constructed plasmids were transformed to DH5α *E. coli* electrocompetent cells or XLI-blue chemically competent cells. Chemical transformation was done according to the manufacturer's protocol (Zymo Research). Sanger sequencing was performed by BaseClear (Leiden, the Netherlands) or Macrogen Europe (Breda, the Netherlands). For diagnostic purposes, PCR amplification was performed with DreamTaq polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions. Vectors containing glycolytic genes were expressed in the PURExpress E6840 (New England Biolabs) according to the manufacturer's instructions

5

with minor changes. Briefly, 250 ng template DNA was used for a single protein synthesis reaction. The reaction was incubated at 25°C for 3 hours and frozen directly at -20°C.

Yeast cell lysis, protein extraction and proteolytic digestion. Cell pellets of the (an)aerobic cultures were resuspended in lysis buffer composed of 100 mM TEAB containing 1% SDS and phosphatase/protease inhibitors [327]. Yeast cells were lysed by glass bead milling and thus shaken 10 times for 1 minute with a bead beater alternated with 1 min rest on ice. Proteins were reduced by addition of 5 mM DTT and incubated for 1 hour at 37°C. Subsequently, the proteins were alkylated for 60 min at room temperature in the dark by addition of 50 mM acrylamide. Protein precipitation was performed by addition of four volumes of ice-cold acetone (-20°C) and proceeded for 1 hour at -20°C. The proteins were solubilized using 100 mM ammonium bicarbonate. Proteolytic digestion was performed by Trypsin (Promega, Madison, WI), 1:100 enzyme to protein ratio, and incubated at 37°C overnight. Solid phase extraction was performed with an Oasis HLB 96-well µElution plate (Waters, Milford, USA) to desalt the mixture. Eluates were dried using a SpeedVac vacuum concentrator at 50 °C and frozen at -80 °C.

TMT labelling for quantitative proteomics experiments. Desalted peptides were reconstituted in 100 mM TEAB and TMT labels from the TMTsixplex isobaric label reagent set and TMT10plex isobaric label reagent set (Thermo Fisher Scientific) were added from stocks equilibrated in 100% anhydrous acetonitrile (ACN). Peptides were mixed with TMT labels in a 1:5 ratio (20 µg to 100 µg) and incubated for 1 hour at 25 °C and 400 rpm and the labelling reaction was stopped by addition of 5% hydroxylamine to a final concentration of 0.4% [400]. Labelled peptides from yeast cultures were then mixed in an equal manner, while peptides from protein standards were labelled and prepared for MS analysis separately. Samples were labelled as shown in **SI table 3** and **4** for TMTsix-plex and TMT10-plex experiments, respectively. Peptide solutions were diluted with water to obtain a final concentration of ACN lower than 5%. Solid phase extraction was performed to desalt the mixture. Desalted peptides were subsequently frozen at -80 °C for 1 hour and dried by vacuum centrifugation. Peptides were finally resuspended in 3% ACN/0.01% TFA prior to MS-analysis to give an approximate concentration of 500 ng per µl. Additionally, fractionation of TMT mixtures was performed for some samples using the Pierce High pH Reversed-Phase Peptide Fractionation Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The completeness of TMT labelling was confirmed by short shotgun proteomics experiments and subsequent database searching where the TMT label was used as variable modification. Here, the percentage of TMT-labelled N-termini and Lysine were translated into a total labelling efficiency (**SI Matlab code 1**). Only experiments with >95% labelling efficiency were used.

Synthetic peptide standards. Synthetic peptide standards for S. cerevisiae Pyk1 for the sequences LTSLNVVAGSDLR (>98% purity level), TSIIGTIGPK (>98% purity level), GDTYVSIQGFK (>99% purity level) and AGAGHSNTLQVSTV (>99% purity level) were purchased from Thermo Fisher Scientific (**Table 2**). Identity and purity were furthermore confirmed by separate in house mass spectrometric analysis.

Mass spectrometric analysis.

Shotgun proteomic analysis of unlabelled proteomics samples. An aliquot corresponding to approximately 0.75 µg protein digest was analysed using a one dimensional shotgun proteomics approach [290]. Briefly, the samples were analysed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL·min^{-1} over a linear gradient from 5% to 30% solvent B over 40 min, then from 30% to 60% over 15 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 60 min. Solvent A was H$_2$O containing 0.1% FA and 3% ACN, and solvent B consisted of 80% ACN in H$_2$O and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1250 m/z at 70,000 resolution in full MS mode with a maximum ion injection time (IT) of 100 ms and an automatic gain control (AGC) target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using higher-energy collisional dissociation (HCD). MS/MS scans were acquired at 17,500 resolution with an AGC target of 1E5 and IT of 100 ms, 1.2 m/z isolation width and normalized collision energy (NCE) of 28.

Shotgun proteomic analysis of TMT labelled proteomics samples. An aliquot corresponding to approximately 1 µg protein digest was analysed using a one dimensional shotgun proteomics approach [290]. Briefly, the samples were analysed using a nano-liquid-chromatography system consisting of an EASY nano LC 1200, equipped with an Acclaim PepMap RSLC RP C18 separation column (50 µm x 150 mm, 2 µm, Cat. No. 164568), and a QE plus Orbitrap mass spectrometer (Thermo Fisher Scientific, Germany). The flow rate was maintained at 350 nL·min^{-1} over a linear gradient from 5% to 25% solvent B over 100 min, then from 25% to 55% over 25 min, followed by back equilibration to starting conditions. Data were acquired from 5 to 130 min. Solvent A was H$_2$O containing 0.1% FA and 3% ACN, and solvent B consisted of 80% ACN in H$_2$O and 0.1% FA. The Orbitrap was operated in data-dependent acquisition (DDA) mode acquiring peptide signals from 385–1450 m/z at 70,000 resolution in full MS mode with a maximum IT of 50 ms and an AGC target of 3E6. The top 10 precursors were selected for MS/MS analysis and subjected to fragmentation using HCD. MS/MS scans were acquired at 17,500 and 35,000 resolution for

5

TMTsixplex™ and TMT10plex™, respectively, with AGC target of 2E5 and IT of 75 ms, 0.7 m/z isolation width and NCE of 32.

Parallel reaction monitoring. Targeted analysis was performed using the Parallel Reaction Monitoring (PRM) modus of the QE plus Orbitrap mass spectrometer. The m/z and retention time of peptide targets were first obtained through a shotgun approach and subsequently targeted in a ±2.5 or 4 min retention time window using 32 normalized collision energy in positive mode. Inclusion lists can be found in **SI Table 5 – 8.**

High-resolution label free quantification. An aliquot of approximately 100 ng protein standard was injected for LC-MS analysis. The flow rate was maintained at 350 nL min⁻¹ over a linear gradient from 5% to 25% solvent B over 40 minutes, followed by a gradient to 50% B over 15 min. Solvent A was H_2O containing 0.1% FA and 3% ACN, and solvent B consisted of 80% acetonitrile in H_2O and 0.1% FA. The Orbitrap was operated in MS1 mode only over a mass range of 350-1250 m/z at 140K resolution.

Processing of mass spectrometric raw data.

Database searching for shotgun proteomics experiments using PEAKS (native and TMT labelled). Data were analysed against the proteome database from *Saccharomyces cerevisiae* (Uniprot, strain ATCC 204508 / S288C, Tax ID: 559292, July 2020) using PEAKS Studio X (Bioinformatics Solutions Inc., Waterloo, Canada). For confirming the quality of the protein standard from the pure system, a combined database comprised of the *S. cerevisiae* and *E. coli* proteome was used (TrEMBL, Tax ID: 204508 and 83333, respectively), excluding the minor redundant glycolytic isoenzymes of *S. cerevisiae*. As contaminant database, cRAP, the common Repository of Adventitious Proteins (GPM) was used. The search allowed for 20 ppm parent ion and 0.02 m/z fragment ion mass error, 3 missed cleavages, acrylamide as fixed cysteine modification and methionine oxidation and N/Q deamidation as variable modifications. In the case of a ¹⁵N-stable isotopically labelled sample, the mass increase of nitrogen from its standard atomic weight to ¹⁵N (+0.9970 Da) is set as fixed PTM for each nitrogen atom, to a maximum of 4 for Arginine. For TMT labelled samples, the TMT label (+229.16) Da was considered as additional fixed modification for all N-termini and lysine (Lys) residues. Peptide spectrum matches were filtered against 1% false discovery rates (FDR) and identifications with ≥2 unique peptides. Mass spectrometric raw data from TMT experiments were additionally analysed using PEAKS Q for changes in protein abundances between different time points. Auto normalization was used for quantitative analysis of the proteins, in which the global ratio was calculated from the total intensity of all labels in all quantifiable peptides. Quantitative analysis was performed using protein identifications containing at least 2 unique peptides, which peptide identifications were filtered against 1% FDR. The significance method for

evaluating the observed abundance changes was set to PEAKS Q and the significance score is calculated as the -10log10 of the significance testing p-value.

Analysis of targeted (PRM) experiments. To perform quantitative analysis of the peptide ratios, a MATLAB pipeline was developed that essentially calculated ratios of internal standard to sample for proteins of interest. Briefly, the raw MS data were converted to '.mzXML' formats using the msConvertGUI tool (ProteoWizard) [293] which were further imported into the Matlab environment using the function "mzxmlread()". The target fragments of each selected peptide ('targets_annotated.xlsx') for quantification of identified MS2 scans were obtained from 'PSM ion.csv' files, exported from PEAKS DB. Peptide ratios were determined by selecting the three most abundant fragments per peptide from the MS2 data using the 'get_fragments' function. All ions for these fragments were collected using the 'get_peak_max' function, both from the sample and in the internal standards. Here, the maximum of the extracted ion current (XIC) peak area for each peptide was determined and the associated scan ± max. two scans (if the area is more than 50% of the maximum peak area) were used for collection of the fragment intensities. Dividing the fragment intensities between the intensity of the native yeast sample and the protein standard provided an array of peptide ratios. For TMT-labelled samples, the ratios were determined based on the associated label fragments. The established codes can be found in **SI Matlab code 2** and **3**.

5

Analysis of high-resolution experiments. For integration of peak areas, Skyline version 4.2 was used [401]. In Skyline, data were analysed against the proteome database from *S. cerevisiae* and *E. coli*. The search allowed for 0.015 m/z fragment ion mass error, one missed cleavage, carbamidomethylation and ^{15}N (+ 0.9970 Da) as fixed cysteine and amino acid modification. The spectral library was built using peptide spectrum matches obtained from PEAKS DB (.pep.xml) from LC-MS/MS scan.

Data availability. The established Matlab data processing pipeline and mass spectrometric raw data have been deposited at 4TU research server and are available at https://doi.org/10.34894/WOSYOT.

Results and Discussion

The cell free production of (complete) protein standards

Our approach aims to quantify the unmodified peptide fraction across the protein backbone to provide an overall view on the degree of modification (**Figure 1**). This approach therefore does not require the identification of individual modifications. However, to cover the complete amino acid sequence in order to consider all possible modification sites, we required to produce a complete protein standard rather than a few synthetic peptides. A similar procedure has been descried recently by Mair *et al*., which however was not expanded to complex samples such as metabolic enzymes from whole cell lysates or complete pathways [71, 395]. Unfortunately, chemical synthesis of peptides chains is not suitable for producing larger proteins. However, the inverse approach only requires small quantities of protein reference material. Therefore, we investigated the suitability of cell free expression to provide analytical quantities of complete protein standards. Commercial cell free expression systems such as the PURExpress system, contain all necessary components for the *in vitro* transcription and translation of proteins (**Figure 2a**). This approach has also shown to successfully produce complete proteins with >100 kDa, only recently [402]. Therefore, we tested the cell free production of some key enzymes in the glycolytic pathway: Hxk2, Pfk1, Pfk2, Pgk1 and Pyk1 (**Figure 2b**). To facilitate the expression, plasmids that carry the glycolytic genes in combination with *E. coli* expression requirements (such as the T7 promotor and terminator) were constructed. The backbone of the control plasmid was used as template DNA for Gibson Assembly (**Figure 2a**). SDS-PAGE analysis of the produced enzymes already confirmed the successful expression of Hxk2 (54 kDa) and Pyk1 (55 kDa) (**Figure 2b**), however, the protein yields for Pfk1 (109 kDa), Pfk2 (106 kDa) and Pgk1 (45 kDa) were not sufficient to confirm successful synthesis (data not shown).

Purification of the synthesized proteins was tested following the manufacturer's instructions for Hxk2 [327]. However, the purification provided only low yields, and the purified material contained still a large number of *E. coli* background proteins. Additionally, the purification procedure risked inducing chemical post-isolation artefacts, which could compromise our quantification experiments.
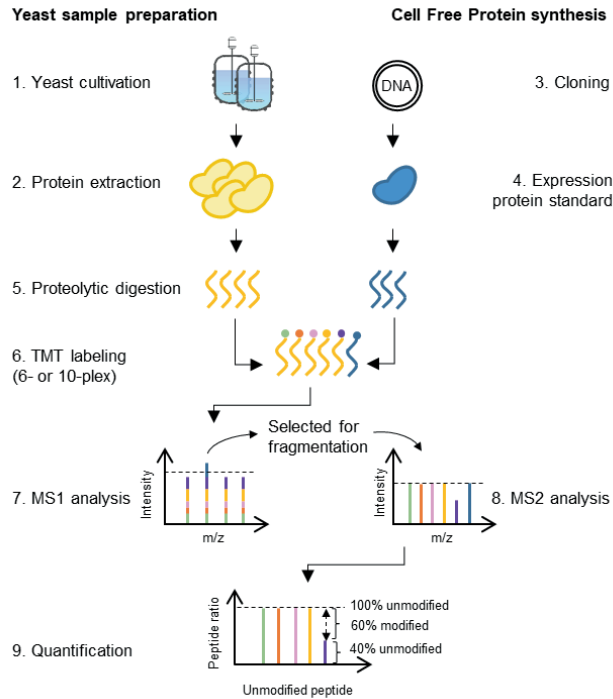
Figure 1. Schematic overview of the inverse quantification approach of the modification degree of metabolic enzymes. 1. Yeast cells are cultivated and sampled under highly controlled conditions. 2. Proteins are subsequently extracted from the cells using standardized sample preparation protocols. 3. At the same time, cell free synthesis is used to provide analytical quantities of unmodified protein reference standards. To this end, a plasmid is constructed that carries the gene of interest in combination with *E. coli* expression requirements such as the T7 promotor and terminator. 4. Addition of the constructed plasmid to the PURE system enables synthesis of the protein standard. 5. Native yeast proteins and the generated protein standard are then separately digested using proteases including trypsin, GluC and chymotrypsin to ensure full protein sequence coverage. 6. Accurate quantification is enabled through labelling and subsequent mixing of the native yeast and protein standard peptides. 7. The peptides are quantified via a targeted analysis (PRM). The TMT labelled peptides act furthermore as reference signal and as "carrier proteome" to enable a reproducible and sensitive detection and (8) quantification based on the reporter ions. 9. Peptide rations are determined by comparing the reporter ions to the synthetic reference standard. A decrease in signal indicates thereby an increase in the degree of modification for this part of the sequence.

Therefore, we decided to use the crude mixture without purification as reference standard for our experiments. To confirm the correct production of the produced protein standards, in-solution digestion followed by a shotgun proteomic analysis was performed. This confirmed the identity of all the enzymes, albeit the sequence coverage (when using Trypsin) varied between 88–52% (Hxk2 88%, Pfk1 77%, Pfk2 52%, Pgk1 66% and Pyk1 89%). Differences in sequence coverage could be explained by protein size, as Pfk1 and Pfk2 are almost twice the size of the other glycolytic proteins. Therefore, the protein yield could be lower because more translation and transcription components are used compared to smaller proteins. This could further impact the detection of more difficult to measure peptides, and therefore reduce the achievable coverage in our proteomics experiments. Nevertheless, C-terminal peptides were obtained from all enzymes, including Pfk1 and Pfk2, indicating that a small fraction of full-length protein was produced. Surprisingly, Pgk1, the smallest glycolytic protein in this study, showed a very low protein sequence coverage (66%). However, inspection of the protein sequence revealed that a significant number of the tryptic peptides were relatively small (<7 amino acids) and therefore likely escaped the successful detection [21].

Nevertheless, a complete sequence coverage usually requires a multi-protease approach [251, 403]. Therefore, we applied a three proteases approach (Trypsin, Chymotrypsin and GluC) which subsequently increased the sequence coverage to nearly 100% for Hxk2 and Pyk1 (**Figure 2c**). Nevertheless, the achievable sequence coverage depends always on the amino acid sequence of the investigated protein (and therefore on the obtained fragments and their properties) [21]. For example, even the application of more than 5 proteases could not provide 100% sequence coverage for BSA, in a recent study [253]. Furthermore, in order to investigate spike-in experiments with heavy isotope protein standards, Hxk2 and Pyk1 were also produced using ^{15}N stable isotope labelled amino acids. The correct synthesis of these heavy protein standards was finally confirmed by in-solution digestion and shotgun proteomic analysis. This provided a sequence coverage comparable to the one observed for the light proteins (83% for ^{15}N versus 88% for ^{14}N for Pyk1 and 76% for ^{15}N versus 84% for ^{14}N for Hxk2). Interestingly, additional in-gel digestion for these proteins resulted in a similar sequence coverage as obtained for the in-solution digestion. This demonstrated that the maximum sequence coverage can be already obtained from relatively complex protein mixtures. To further confirm the comparability of the light and the heavy standards, both Hxk2 standards were mixed in equal amounts directly after synthesis (to eliminate any variance in sample preparation), digested and analysed in order to determine peptide ratios. The calculated peptide ratios (light/heavy) appeared to be constant (0.77, SD=0.044) across the complete protein backbone. The only exception was the N-terminal peptide MVHLGPK which showed an unexpected ratio of approx.1.2 (**SI Figure 1**). This was presumably caused by interference with an unrelated co-eluting peptide, but it was not further investigated in the current study.
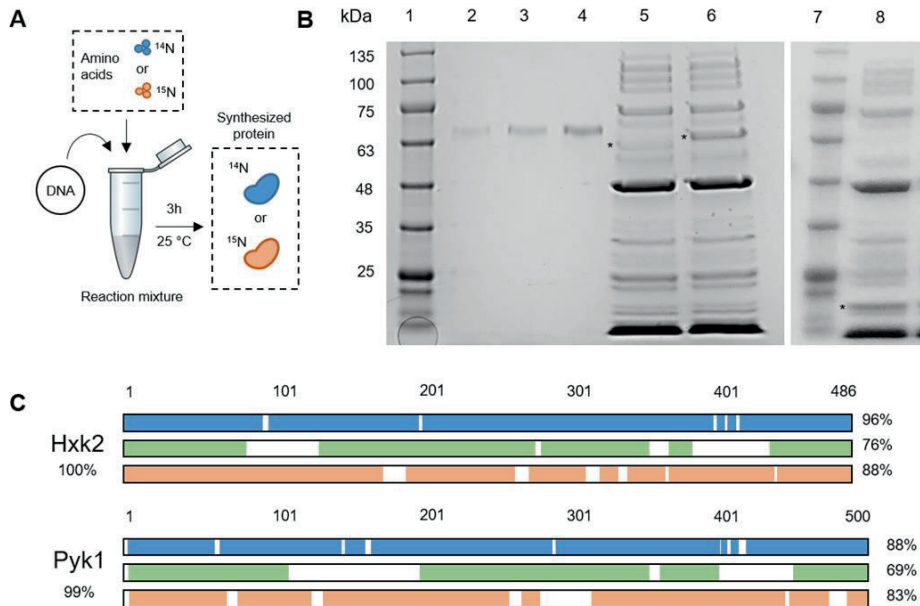
Figure 2. Cell free protein synthesis of Hxk2 and Pyk1 and analysis by SDS-PAGE and shotgun proteomics. A. Schematic of cell free protein synthesis. Template DNA carrying the gene of interest together with a T7 promotor and terminator, is required to initiate protein synthesis with E. coli transcription and translation machinery. Synthesis can be performed with ^{14}N or ^{15}N amino acids to create a light or heavy protein. *In vitro* protein synthesis was performed over 3 hours. B. SDS-PAGE analysis of ^{14}N Hxk2 (54 kDa) and Pyk1 (55 kDa) using PURExpress In Vitro Synthesis kit. Proteins were separated by electrophoresis on a 10% polyacrylamide gel in Tris-Glycine SDS buffer and stained with Coomassie Brilliant Blue. Lane 1: BLUeye Prestained Protein Ladder. Lane 2: 125 ng Bovine Serum Albumin (BSA) as a reference sample (67 kDa). Lane 3: 250 ng BSA. Lane 4: 500 ng BSA. Lane 5: reaction mixture of Hxk2 synthesis. Lane 6: reaction mixture of Pyk1 synthesis. Lane 7: BLUeye Prestained Protein Ladder. Lane 8: reaction mixture of positive control (Dhfr, 18 kDa). C. Graphical representation of Hxk2 and Pyk1 sequence coverage of ^{14}N Hxk2 and Pyk1 syntheses. Filled sections show the relative portion of the entire sequence that was observed after digestion by each enzyme. Trypsin, GluC and Chymotrypsin digestions are represented in blue, and orange, respectively.

SDS-PAGE and proteomic analysis confirmed the synthesis of complete Pyk1 and Hxk2. However, the presence of truncated proteins, perhaps at lower quantities, cannot be ruled out completely. For the case that the "full-sized protein" is the main product of the cell free synthesis, N and C-terminal peptides should be present at similar quantities. The intensity of the observed mass spectrometric signal for every peptides depends heavily on the ionization efficiency [404]. Therefore, the observed signals cannot be used to perform a quantitative comparison of different peptide sequences. An accurate quantification can only be achieved

with synthetic peptide standards. Therefore, synthetic peptides were purchased covering N and C-terminal sequence regions of Pyk1. More specifically, peptides 1 and 2 were from the N-terminus, while peptides 3 and 4 covered C-terminal parts of the protein (**Table 1**). We used the ^{15}N labelled protein standard because it allowed for direct, label free (MS1 level) comparison of the cell free protein standard and the synthesised peptides. Moreover, different quantities of the synthetic peptide standards were spiked to the protein to find an optimal ratio (**Table 1**). Surprisingly, the ratios ^{15}N/^{14}N (peptides from cell free protein standard/synthetic peptides) differed strongly across the protein backbone. For example, we obtained for peptide 1 an approx. 19-fold larger ratio compared to peptide 4 (**Table 1**). Furthermore, peptide 2 obtained an approx. 3.5-fold smaller ratio compared to peptide 1. In a second experiment, we cut the gel band representing the complete Pyk1 protein and compared it to the synthetic peptide standards with known quantities (like before). Here, the ratios ^{15}N/^{14}N (Cell free protein peptide/synthetic peptide) were very comparable ranging from 0.8 to 1.4, confirming the correctness of our testing procedure. This indicates that the protein synthesis is interrupted already at early stages in many cases. Truncated versions of the protein are not uncommon in cell free protein synthesis, and multiple reasons could cause break up or slowing down of the translation process. For example, the depletion of tRNAs through peptidyl-tRNA drop-off can halt translation [405]. Here, the use of low-usage Arg codons (AGA and CGA) proximal to the start codon could lead to premature termination of protein translation [406]. In our study, a low-use Arg codon (AGA) appears on position 3 of the *S. cerevisiae* gene, very close to the start codon, which might therefore challenge translation. This might cause the depletion of the tRNA pool in the system. Genes optimized for *E. coli* codon usage could be used to overcome this issues [405]. However, our envisaged study (monitoring modification changes of glycolytic enzymes across the yeast growth curve) does not necessarily require exactly the same amounts of peptide across the protein backbone. Only relative peptide ratio changes across a growth curve are monitored, which would indicate the presence of modification dynamics and possible functional roles.

Table 1. Full-sized cell free synthesized Pyk1 analysis. The peptide intensity of cell free synthesized Pyk1 with ^{15}N amino acids was compared to chemically synthesized ^{14}N peptide standards to assess full-sized cell free protein synthesis.

			Ratio ^{15}N/^{14}N peptide			
#	Sequence	Position	0.25 ng	1 ng	10 ng	Gel
1	LTSLNVVAGSDLR	7-19	70.0	16.7	1.4	0.8
2	TSIIGTIGPK	21-30	19.2	4.7	0.4	1.5
3	GDTYVSIQGFK	476-486	0.017	0.012	0.013	0.7
4	AGAGHSNTLQVSTV	487-500	3.8	0.9	0.1	1.4

Peptide ratio determination using tandem mass tags

Stable isotope labelled protein standards allow to spike the standard into the cell lysate for a combined digestion, detection, and quantification of the light and heavy variants in MS1. This minimizes biases in the peptide ratios because both variants undergo exactly the sample preparation procedures. However, quantification at the MS1 level requires a prior identification of the peptides via fragmentation experiments and ratios can be unexpectedly influenced by co-eluting background peptides derived from other proteins. Finally, the analysis of time course experiments requires separate experiments for every time point, and experiments can demand long measurement times. Therefore, we aimed to investigate an approach that overcomes the above-mentioned drawbacks, by making use of tandem mass tag labelling (TMT). Labelling approaches enable a more accurate quantification compared to when using label free approaches [65]. For this, the cell-free synthesized protein standards are digested and subsequently labelled with a TMT label. The TMT labelled peptides act furthermore as reference signal and as "carrier proteome" to enable a reproducible and sensitive detection and quantification based on the reporter ions. This also allows a very accurate and fast analysis of time series experiments due to its multiplexing nature. 10-plex TMT-labelling, for example, can combine up to 10 samples in a single analysis (*e.g.* 4 time points plus protein standard, each in duplicates) (**Figure 1**). More specifically, the observable peptides from the target protein are first determined by a shotgun proteomics experiment. Those create an inclusion list for the quantitative analysis using parallel reaction monitoring (PRM). Finally, an in house developed Matlab data processing pipeline is used to facilitate reproducible, automated ratio determination for the targeted peptides (**SI Matlab code 2 and 3**). The specificity and sensitivity of the PRM method however is affected by the number of peptide targets and the size of the retention time window (and therefore number of overlapping scans). The number of targets was therefore kept <100 to guarantee a sufficient number of measurement points across the chromatographic peaks.

Nevertheless, a possible challenge when using MS2 quantification for TMT labels is the reporter ion bias introduced by co-isolation of peptides. Therefore, to minimize such co-isolation related interferences, different isolation windows of 0.7, 1.2 and 1.6 m/z were tested for their quantitative performance using a targeted method (**SI Table 5**). However, similar peptide ratios and standard deviations were obtained in duplicate measurements for all three isolation windows (average peptide ratios 0.416±0.185, 0.426±0.202 and 0.437±0.197 were obtained for the isolation windows 0.7, 1.2 and 1.6 m/z; **Figure 3**). A vast majority of obtained peptide ratios was consistent between the replicates, and also no significant drop in sensitivity was observed using Spearman's rank correlation coefficient (**Figure 3**). However, to minimize the risk of co-isolation in any future experiments, the smallest isolation window (0.7 m/z) was selected.
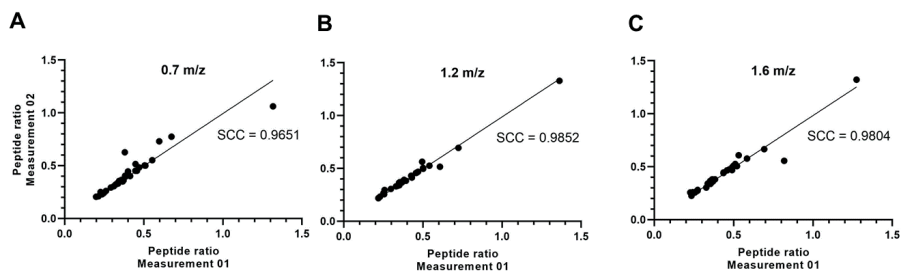
Figure 3. Isolation window optimization of the targeted (PRM) quantification approach. Spearman's correlation plots of calculated peptide ratios of detected Pyk1 peptide intensities from two different growth phases (late-exponential and stationary growth phase in aerobic MG yeast) using the PRM approach. Windows of 0.7 m/z (A), 1.2 m/z (B) and 1.6 m/z (C) were tested in duplicate PRM measurements.

The accuracy of the measured peptide ratios may also be influenced by the protein abundance ratios. Ideally, the abundance of the standard is equal to the target protein in the native yeast proteome. However, the amount of the target protein in the yeast sample may significantly change across the growth curve. Therefore, we investigated the impact of different standard to target protein ratios on the outcome. For this, we prepared a dilution series where one volume of Pyk1 standard was spiked into 16, 8 and 4 volumes of native yeast proteome. We subjected the dilution series to a targeted proteomics analysis (**SI Table 6**) and compared the differences in the number of targeted peptides and their ratios. The sequence coverage varied from 76%, 71% and 83% for ratios 16:1, 8:1 and 4:1 native yeast proteome to Pyk1 standard, respectively. Although ratios for the individual fragments varied to some degree, the mean peptide ratio obtained for all targeted peptides were in the same order of magnitude. The average peptide ratios (Pyk1 standard vs. IMX372 proteome) obtained from the different spikes 4:1, 8:1 and 16:1 resulted in rations: 1.046 ± 0.659, 0.377 ± 0.323 and 0.186 ± 0.129, respectively, which followed the expected linear decrease. However, albeit the trend was as expected, the higher dilutions (*e.g.* 16:1) resulted in a reduced number of confidently identified peptides. This indicates that the Pyk1 standard peptides also served as a carrier base level for peptide identification. Therefore, the top 10 peptides with highest intensities in each of the dilutions were verified for their linearity in the dilution series (**Figure 4a**). To compare the different slopes between the peptides, the peptide ratios were normalized to the 16:1 peptide ratios (**Figure 4b**). The slopes of the peptides were not significantly different ($p > 0.01$), indicating that a linear increase in peptide ratios relates to a linear increase in the amount of protein.
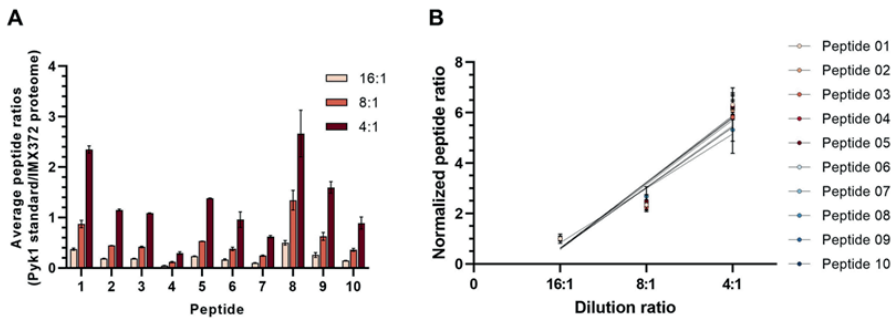
Figure 4. Evaluation of linearity and limit of detection. A. A dilution series of Pyk1 standard to yeast lysate was established (1:16, 1:8, 1:4) to study the limit of detection and the linearity of the approach. Peptide intensities were measured using the developed PRM assay for the top 10 detected peptides with the best detectability in the complete dilution series. Average Pyk1 peptide ratios of duplicate measurements and standard deviations are indicated by the error bars. B. Normalized peptide ratios of top 10 detected peptides with highest intensities and detectability in the complete dilution series. Peptide intensities were normalized against the 16:1 dilution ratio (Pyk1 reference vs. native proteome). Linear regression of the dilution series of each peptide resulted in comparable slopes for all different peptides and dilutions.

Furthermore, we aimed to assess the reproducibly of our targeted quantification approach. First, a replication series was studied, in which three technical replicates of chemostat-grown *S. cerevisiae* lysates were labelled with three different TMT labels (127, 128, 129) and mixed with the *in vitro* synthesized Pyk1 labelled with the 131 TMT. The sample was analysed in duplicates using the targeting approach (**SI Table 7**), and ratios were calculated for Pyk1 lysate peptides compared to the Pyk1 standard in channel 131). Thereby, the average peptide ratio (Pyk1 proteome vs. Pyk1 standard) from duplicate measurements were 0.21 ± 0.20, 0.23 ± 0.25 and 0.27 ± 0.25 for channel 127, 128 and 129, respectively. This illustrates that peptide ratios when using different reporter channels are fairly stable, with only small differences (possibly due to pipetting errors). However, deviations from the average peptide ratios (Pyk1 proteome vs. Pyk1 standard) showed slightly increasing ratios from channels 128, 129 to 130 (**Figure 5**). Several peptides displayed relatively large deviations from the average ratio (because of overall signal abundancies). The increased ratios indicated that these specific peptides are more abundant in the native Pyk1 proteome compared to the Pyk1 standard, which might be due to differences in sample preparation. Finally, correlation between the calculated peptide ratios obtained from the duplicate measurements was calculated using Spearman's rank correlation coefficient (SCC) to assess reproducibility of the ratios. The

average SCC of the two ratios obtained by separate analysis runs for each TMT tag was 0.96, 0.94 and 0.95 for TMT labels 127, 128 and 129, respectively.

However, finally, a full sequence coverage would be required to quantify the complete PTM landscape of Pyk1. Off-line fractionation of the proteome reduces background signals which can improve sequence coverage. Here, the Trypsin-digested TMT mixture of the IMX372 (MG) aerobic growth curve was fractionated in eight fractions using reversed phase separation based on hydrophobicity. LC-MS/MS analysis increased the number of protein identifications from 1275 to 2500, however, the sequence coverage of Pyk1 was only increased to 83%. However, an *in silico* investigation of the amino acid sequence of Pyk1 confirmed that the remaining parts do not provide peptides that can be identified by LC-MS. Previously, it has been shown that the use of multiple proteases is required to significantly increase the sequence coverage of *in vitro* synthesized Pyk1. Therefore, we performed a digestion with Trypsin, Chymotrypsin and GluC. Indeed, the Pyk1 sequence coverage was increased to approx. 95% using one-dimensional LC-MS/MS analysis runs (with individual sequence coverages of 71%, 66% and 45% for Trypsin, Chymotrypsin and GluC, respectively). The multi-protease strategy is therefore preferred over fractionation as it provides higher sequence coverage with lower instrument operation time.
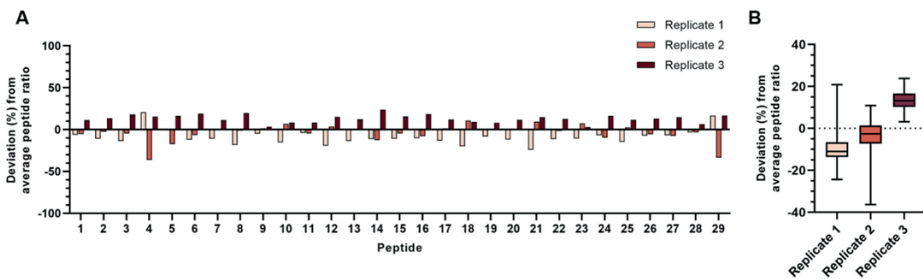
Figure 5. Evaluation the reproducibility of the inverse quantification approach. A. The reproducibility of the targeted quantification was evaluated using a series of three technical replicates of chemostat-grown yeast labelled with different TMT labels (127, 128 and 129) and mixed with the *in vitro* synthesized Pyk1 protein standard, labelled with 131. Deviations (%) from the average Pyk1 peptide ratios (native yeast vs. Pyk1 standard) of duplicate measurements are indicated by the bars. B. The boxplot shows the median ratios obtained from the different reporter ions. Except single outliners, peptide ratios were generally +/-25% of the median. The experiment shown was performed using MS2 reporter ion quantification. The accuracy is expected to be improved using MS3 reporter ion quantification.

Quantifying the degree of modification for Pyk1 across aerobic growth

Finally we wanted to exemplify the quantification approach to a growth curve of IMX372 (MG) yeast. MG yeast was cultivated in aerobic batch cultures in triplicate and samples were taken at 6, 9, 12, 16.5 and 27 hours of growth, corresponding to mid-exponential, late exponential, early-diauxic, mid-diauxic and stationary growth phases, respectively [398] (**SI Figure 2**). The changing environmental conditions are expected to induce regulatory responses on the proteome, *e.g.* via post-translational modifications. To this end, IMX372 samples from each of the 5 growth phases were TMT-labelled and each mixed with the TMT labelled peptides of the Pyk1 standard. 37 unmodified, tryptic peptides across the protein backbone were selected for targeting (**SI Table 8**). This provided a sequence coverage of nearly 70%. Peptide ratios were determined for each growth phase using the developed Matlab tool (**Figure 6a, SI Figure 3**). An average ratio across all detected peptides was calculated for each growth phase which resulted in ratios of 0.94, 0.99, 0.87, 0.79 and 0.43 for mid-exponential, late exponential, early-diauxic, mid-diauxic and stationary growth phases, respectively. The overall decreasing peptide ratios indicate an overall decrease in native Pyk1 content across the aerobic growth, which was consistent with our previous study [398]. The peptide ratios were normalized to the Pyk1 content (according to normalization factors determined by PEAKS Q) to eliminate protein abundance related influences. All ratios were further normalized to the first growth phase, 'ME' (**Figure 6b**). Peptides abundances were considered changed if the fold change was larger than ±20% compared to the base sample 'ME' (**Figure 6b** and **6c**). After normalization to the 'ME' growth phase, peptide ratios across Pyk1 were averaged at 0.92, 0.99, 1.02 and 0.87 late exponential, early-diauxic, mid-diauxic and stationary growth phases, respectively.

Interestingly, throughout the growth curve, peptides were observed with positive and negative fold changes. Most peptides contained negative fold changes, indicating that the fraction of unmodified peptide was decreased, or that the fraction of modified peptide increased (**Figure 1**). On the other hand, positive fold changes compared to the base sample indicate that the protein may be present in a modified form at the beginning of the growth curve. Only three out of 37 targeted peptides appeared to be unchanged across the growth curve, while the 34 other peptides showed changes >20% at least in one of the growth phases. The stationary growth phase resulted in the largest changes. In non-proliferating, aerobic conditions, proteins are susceptible to aging processes such as oxidation by reactive oxygen species (ROS) [239]. This may severely influence the stoichiometry of unmodified peptide fraction. Normalized to the mid-exponential phase, the early-diauxic growth phase contained the least number of modified peptides. The largest fold change was observed for the peptide in the position 69–91. It shows an increase of 166% in the mid-diauxic phase. This region of the protein is also known to be susceptible to phosphorylation, ubiquitination, acetylation and succinylation [169, 252, 407, 408]. In addition, the peptide in the position 414–425 contained the highest modified fraction in the mid-stationary phase, albeit no modifications have yet

been reported in this particular sequence. Peptide in position 205–216 also displayed a high degree of modification in the mid- and late-exponential phase (~50%). Previously, it has been demonstrated that the S213 and R216 residues are prone to phosphorylation [409] and methylation [166], respectively, although the functional roles have not been proven yet.

Well-characterized areas of Pyk1, such as the fructose-1,6-bisphosphate (FBP) binding area (T403) [410], were also quantified. FBP allosterically activates Pyk1 in the presence of glucose. Alteration of this area through PTMs might prevent FBP from binding to favor respiration following the diauxic shift as opposed to fermentation during proliferation. However, the peptide covering the sequence from 397–409) showed that apparent modification rather occurred before the diauxic shift. The residues 402–407 of Pyk1 interact with the 6'-phosphate of FBP and changes from threonine to a glutamic acid hinder FBP binding in a previous study [410]. Another important part of Pyk1 is the active site at R49, which position the phosphoryl transfer process of phosphoenolpyruvate (PEP) and ADP to ATP and pyruvate [410]. Catalytic activity of Pyk1 is abolished and yeast is unable to grow when R is changed to an A [409]. The peptide covering the position 43–49 showed modification only in the the mid-stationary phase, indicating that the unmodified active site might be important for growth. In the second step of the catalytic activity a proton is transferred by T298. The sequence region 287–312 appeared to be modified in the early-diauxic and stationary phase. During gluconeogenic growth, Pyk1 is inactivated while still present in the cell [411]. This, for example, could be accomplished via modification of the active site using PTMs.

Another interesting sequence region of Pyk1 is the low compositional complexity (LCR) area (position 369–394), which allows for reversible aggregation to preserve protein function under stress. Here, phosphorylation prevents aggregation during exponential growth, while dephosphorylation allows aggregation taking place [409]. One peptide covered this sequence region (position 370–394) and a significant degree of change (or possible modification) was observed in the late-exponential and stationary phase.

Finally, a well investigated phosphorylation site is S22, which has been shown to decrease the activity *in vitro* [412]. However, in our study, we only detected minor changes for this sequence regions (20–30, and 21–30). This observation was supported by the study of Xia *et al.* [158], in which no correlation could be detected between *in vivo* protein activity and this phosphorylation site. Phosphorylation of S22 has been demonstrated to provide a more active Pyk1 in the absence of FBP [412]. On the other hand, a study by Xu *et al.* suggests that phosphorylation is not required for Pyk1 activity, but that it is predominantly controlled by allosteric regulation [413].
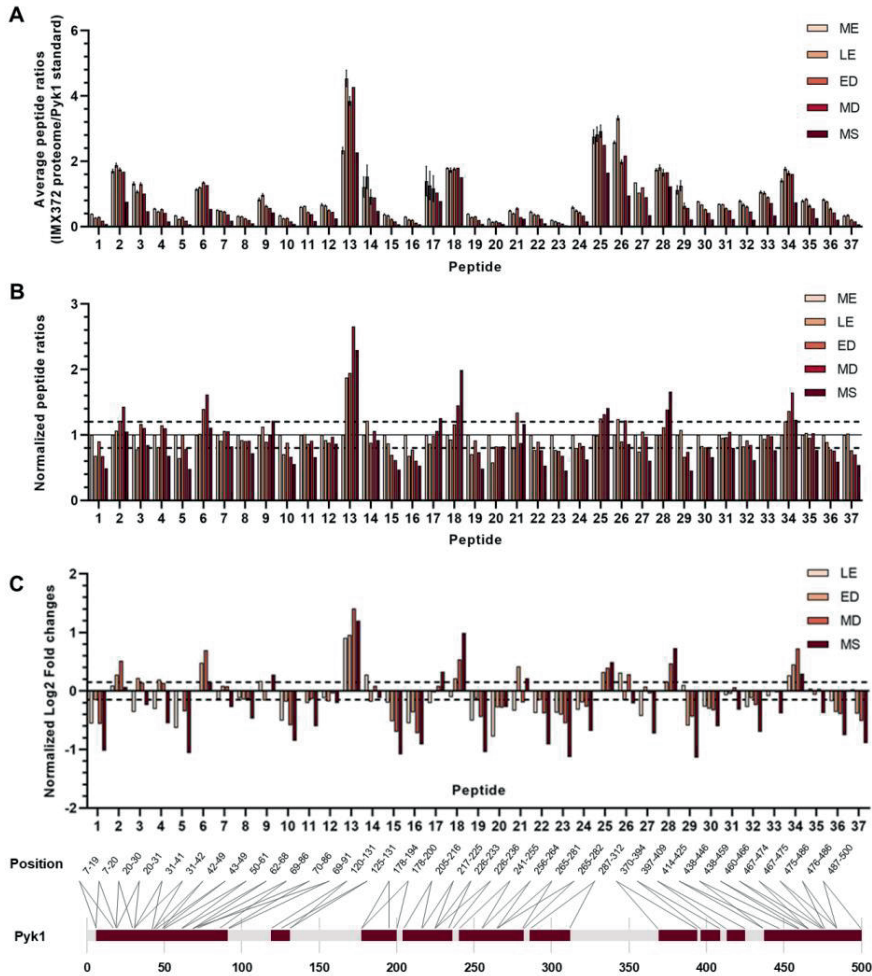
Figure 6. The degree of modification for yeast Pyk1 across anaerobic growth. A. the graph shows the average peptide ratios for native Pyk1 in yeast from exponential to stationary phase under anaerobic conditions (5 different sampling points). B. the graph shows the peptide rations from graph A normalized to the first time point (=1), C. the graph shows the normalized log2 fold change of the Pyk1 peptides, shown in graph A. The panel below shows the position of the individual tryptic peptide sequences of Pyk1 that were quantified in this experiment. Red regions were covered in this proof-of-concept experiment. Grey regions did not provide quantifiable tryptic peptide fragments. Those would require one of the alternative proteolytic enzymes (GluC or Chymotrypsin).

In summary, our approach has been exemplified for Pyk1. Nearly all peptides show significant changes in the fraction of unmodified peptides across the aerobic growth curve. Changes in abundance of the unmodified peptides could be due post-translational modification (or chemical), but it may also be due to missed cleavages (induced by adjacent natural or chemical modifications). We observed various regions of the protein that were covered by multiple peptides due to missed cleavages. Therefore, the use of multiple proteases may be necessary to provide additional confirmation for the observed changes.

Outlook and conclusions

Here, we demonstrate and evaluate an approach for the quantification of the degree of modification for complete proteins in the presence of complex cell lysates. The approach quantifies the unmodified fraction of peptides, and therefore does not require the identification and localisation of individual modifications and modification sites. Furthermore, we evaluated the production of analytical quantities of complete protein standards using cell-free synthesis. TMT labelling furthermore allowed to determine peptide rations across complete growth curves simultaneously. Finally, we established a Matlab data processing pipeline to rapidly determine and visualise peptide ratios. Nevertheless, the accuracy of the peptide ratio determination could be further improved using MS3 quantification, because the additional isolation and fragmentation of fragment ion peaks eliminates interference from unrelated peptides [68]. Nevertheless, a recent study showed a high degree of correlation between MS2 and MS3 quantification, and MS2 usually shows a higher sensitivity [140, 414].

Furthermore, in this study we only monitored the changes in the degree of modification of a single (key) enzymes in yeast glycolysis. However, given the versatility and multiplexibility of our approach, it could be expanded to complete pathways such as the 12 enzymatic reactions of glycolysis. Nevertheless, the uniqueness of the individual peptide sequences is critical to the method, which could be ensured by using multiple proteases [415]. Our study also demonstrates that native protein standards are not necessary to determine relative changes of the unmodified fraction. Relative changes between peptide abundances across different time points could simply be determined using a mix of all conditions as reference signal. This could be also used to monitor changes across complete proteomes, thereby indicating proteome fractions that are preferably regulated by PTMs. However, this would only reveal the fold change, but not the absolute value of the modified fraction.

Finally, albeit this approach has been demonstrated using yeast, it could be also employed to enzymes from other cell types such as mammalian cells. For example, post-translational modifications in human glycolysis has been associated with the Warburg effect in cancer, which increases the glycolytic flux by approx. 200 times in tumour cells compared to normal cells [91, 416].

Supplementary Information

Matlab scripts can be found online: https://doi.org/10.34894/WOSYOT
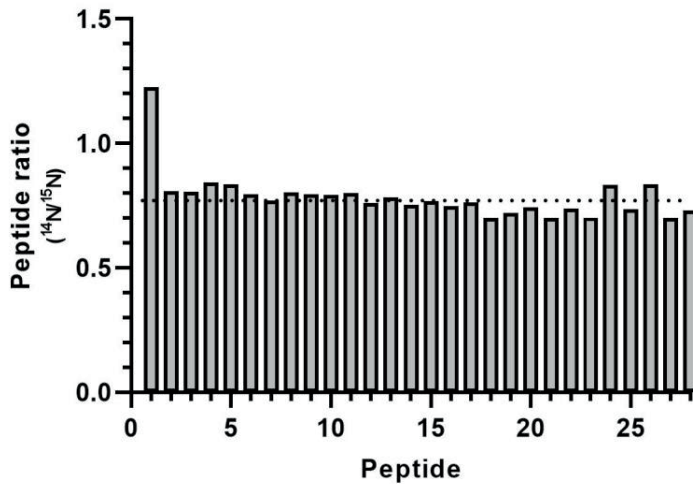
SI Figure 1. Comparison of cell free synthesized Hxk2 with light (^{14}N) and heavy (^{15}N) amino acids. Peptide intensities of the detected light Hxk2 peptides relative to their heavy counterpart resulted in peptide ratios across the Hxk2 protein. All ratios appear to be relatively constant approx. 0.77 (light to heavy ratio), with a standard deviation of 0.044, except for the N-terminal peptide (MVHLGPK) which shows an unexpected ratio of approximately 1.2.

5
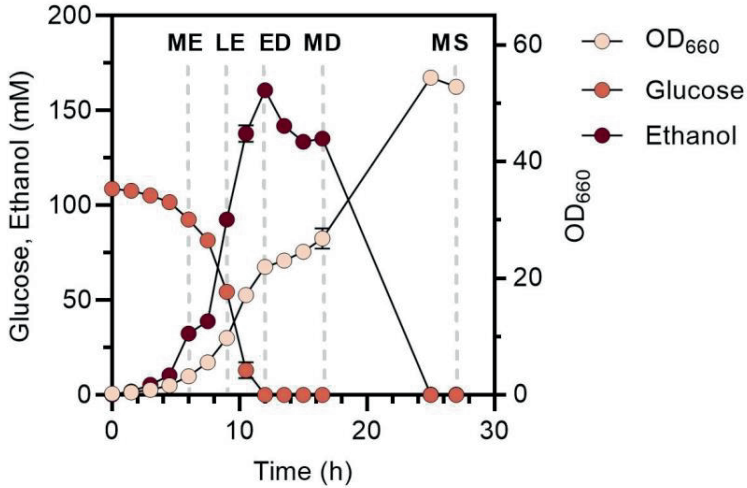
SI Figure 2. Yeast growth in aerobic batch cultures (Figure adapted from Chapter 4, **published in** Maxime den Ridder et al., bioRxiv, 2022. https://doi.org/10.1101/2022.09.23.509138). The MG yeast growth in anaerobic batch cultures was described in Chapter 4 (den Ridder et al., bioRxiv 2022), where the glucose levels are shown by the red curve, ethanol by the orange curve and the OD660 by the purple curve (three biological replicate cultures, where standard deviations are shown as error bars). Proteome samples were taken after 6, 9, 12, 16.5 and 27 hours, which corresponded to the mid-exponential ('ME'), late-exponential ('LE'), early-diauxic ('ED'), mid-diauxic ('MD') and (mid) stationary ('MS') growth phases.
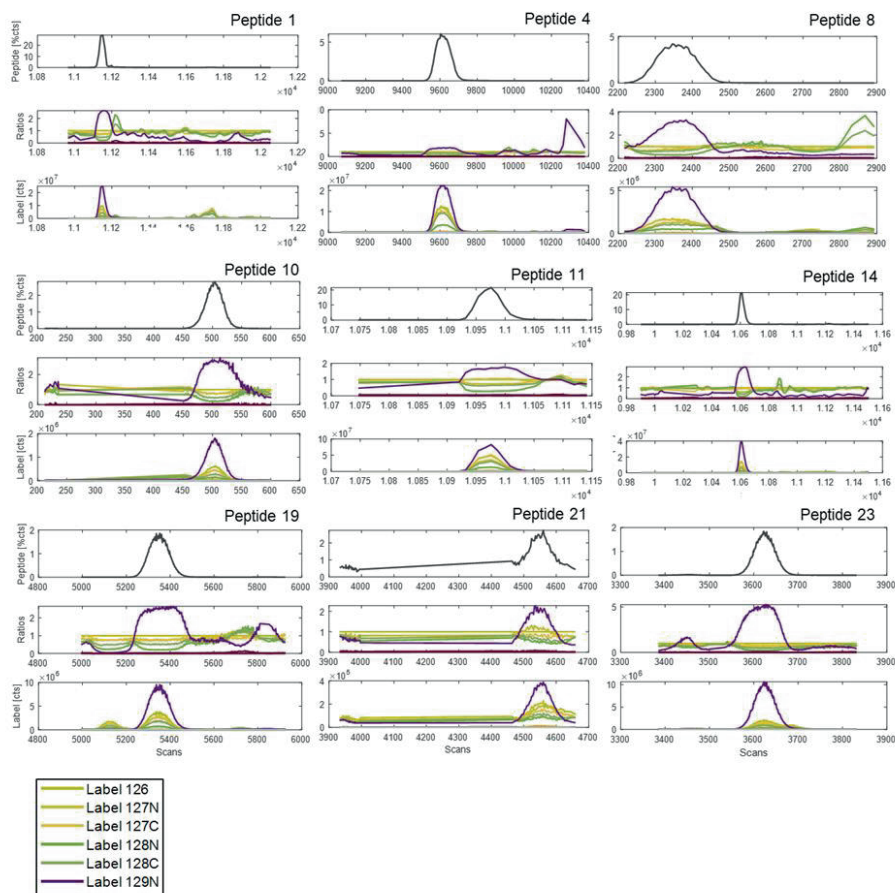
SI Figure 3. Example of Pyk1 peptide ratio determination visualized with Matlab tool. The raw MS data were imported into the Matlab environment to determine the peptide ratios (native Pyk1 vs. Pyk1 standard), exemplified here with 9 peptides. The target fragments of each selected peptides for quantification of identified MS2 scans were obtained. Peptide ratios were determined by selecting the three most abundant fragments per peptide from the MS2 data. All ions for these fragments were collected both from the sample and in the internal standards (upper panel). Here, the maximum of the extracted ion current (XIC) peak area for each peptide was determined and the associated scan ± max. two scans (if the area is more than 50% of the maximum peak area) were used for collection of the fragment intensities. Dividing the fragment intensities between the intensity of the native yeast sample and the protein standard provided an array of peptide ratios (middle panel). For TMT-labelled samples, the ratios were determined based on the associated label fragments (lower panel).

SI Table 1. Primers used in this study.

Name	Purpose	Sequence
16233	Amplification of pT7 - tT7 backbone FW	TGAGGATCCCGGGAATTCTC
16234	Amplification of pT7 - tT7 backbone RV	ATGTATATCTCCTTCTTAAAG TTAAACAAA
16235	Amplification of HXK2 FW	ATACGACTCACTATAGGGTCTAGAAATAA TTTTGTTTAACTTTAAGAAGGAGATATACAT ATGGTTCATTTAGTCCAAA
16236	Amplification of HXK2 RV	ACCACCTTAATTAAAGGCCTCCTGCAGGTT AACCTTACTCGAGAATTCCCGGGATCCTCA TTAAGCACCGATGATACCAA
16829	Amplification of PFK1 FW	ATACGACTCACTATAGGGTCTAGAAATAAT TTTGTTTAACTTTAAGAAGGAGATATACAT ATGCAATCTCAAGATTCATG
16830	Amplification of PFK1 RV	ACCACCTTAATTAAAGGCCTCCTGCAGGTT AACCTTACTCGAGAATTCCCGGGATCCTCA TCATTTGTTTTCAGCGGCTA
16831	Amplification of PFK2 FW	ATACGACTCACTATAGGGTCTAGAAATAA TTTTGTTTAACTTTAAGAAGGAGATATACA TATGACTGTTACTACTCCTTTTGTGA
16832	Amplification of PFK2 RV	ACCACCTTAATTAAAGGCCTCCTGCAGGTT AACCTTACTCGAGAATTCCCGGGATCCTCA TTAATCAACTCTCTTTCTTCCAACC
16833	Amplification of PYK1 FW	ATACGACTCACTATAGGGTCTAGAAATAA TTTTGTTTAACTTTAAGAAGGAGATATAC ATATGTCTAGATTAGAAAGATTGACCT
16834	Amplification of PYK1 RV	ACCACCTTAATTAAAGGCCTCCTGCAGGT TAACCTTACTCGAGAATTCCCGGGATCCT CATTAAACGGTAGAGACTTGCA
18189	Amplification of PGK1 FW	ATACGACTCACTATAGGGTCTAGAAATA ATTTTGTTTAACTTTAAGAAGGAGATATA CATATGTCTTTATCTTCAAAGTT
18190	Amplification of PGK1 RV	ACCACCTTAATTAAAGGCCTCCTGCAGG TTAACCTTACTCGAGAATTCCCGGGATC CTCATTATTTCTTTTCGGATAAGA

SI Table 2. Plasmids used in this study.

Name	Relevant characteristics	Reference
pUD1079	ampR pT7-EcDHFR-tT7	[402]
pUD1081	ampR pT7-ScHXK2-tT7	This study
pUD1107	ampR pT7-ScPFK1-tT7	This study
pUD1108	ampR pT7-ScPFK2-tT7	This study
pUD1109	ampR pT7-ScPYK1-tT7	This study
pUD1184	ampR pT7-ScPGK1-tT7	This study
pUDE767	ampR ScHXK2	[417]
pUDE769	ampR ScPFK1	[417]
pUDE770	ampR ScPFK2	[417]
pUDE774	ampR ScPGK1	[417]
pUDE777	ampR ScPYK1	[417]

SI Table 3. TMT labeling schemes of experiment using TMT6-plex.

Experiment	TMT label					
	126	**127**	**128**	**129**	**130**	**131**
Isolation window	-	MG - LE	-	-	MG – MS	-
Replication	-	MG - chemostat	MG - chemostat	MG - chemostat	-	Pyk1 standard

SI Table 4. TMT labeling schemes of experiment using TMT10-plex.

Experiment	TMT label									
	126	**127N**	**127C**	**128N**	**128C**	**129N**	**129C**	**130N**	**130C**	**131**
Standard to yeast ratio	-	MG – chemostat	-	-	-	-	-	Pyk1 standard	-	-
Aerobic Growth curve	MG - ME	MG - LE	MG - ED	MG - MD	MG - MS	Pyk1 standard	-	-	-	-

SI Table 5. PRM inclusion list of LC-MS/PRM analysis of the isolation window experiment.

#	Amino acid sequence	Start	End	m/z	Start (RT)	End (RT)
1	LTSLNVVAGSDLR	7	19	787.4584	40.38	48.38
2	LTSLNVVAGSDLRR	7	20	577.3416	32.76	40.76
3	TSIIGTIGPK	21	30	722.9597	34.26	42.26
4	TNNPETLVALR	31	41	728.9183	29.99	37.99
5	TNNPETLVALRK	31	42	605.3673	28.2	36.2
6	KAGLNIVR	42	49	443.6304	18.52	26.52
7	AGLNIVR	43	49	486.3138	19.22	27.22
8	MNFSHGSYEYHK	50	61	653.3260	23.48	31.48
9	SVIDNAR	62	68	502.2904	9.16	17.16
10	KSEELYPGRPLAIALDTK	69	86	672.9026	40.08	48.08
11	SEELYPGRPLAIALDTK	70	86	777.7821	44.49	52.49
12	ACDDKIMYVDYK	120	131	741.3971	36.34	44.34
13	IMYVDYK	125	131	695.3950	37.58	45.58
14	GVNLPGTDVDLPALSEK	178	194	1.092.1196	51.12	59.12
15	GVNLPGTDVDLPALSEKDKEDLR	178	200	792.9481	46.06	54.06
16	NGVHMVFASFIR	205	216	536.2953	42.6	50.6
17	TANDVLTIR	217	225	616.3632	25.86	33.86
18	IENQQGVNNFDEILK	241	255	1.110.1052	46.2	54.2
19	VTDGVMVAR	256	264	588.8331	21.15	29.15
20	GDLGIEIPAPEVLAVQK	265	281	736.4393	55.42	63.42
21	SNLAGKPVICATQMLESMTYNPRPTR	287	312	852.7037	52.75	60.75
22	AEVSDVGNAILDGADCVMLSGETAK	313	337	998.8442	66.01	74.01
23	GNYPINAVTTMAETAVIAEQAIAYLPNYDDMR	338	369	936.9655	73.7	81.7
24	NCTPKPTSTTETVAASAVAAVFEQK	370	394	1.103.9353	53.27	61.27
25	AIIVLSTSGTTPR	397	409	772.9632	32.29	40.29
26	YRPNCPIILVTR	414	425	582.3406	30.08	38.08
27	GVFPFVFEK	438	446	764.4523	56.68	64.68
28	GVFPFVFEKEPVSDWTDDVEAR	438	459	1.009.8557	63.1	71.1
29	INFGIEK	460	466	639.8940	37.23	45.23
30	AKEFGILK	467	474	797.0199	31.6	39.6
31	KGDTYVSIQGFK	475	486	677.4045	34.87	42.87
32	GDTYVSIQGFK	476	486	836.9685	39.6	47.6
33	AGAGHSNTLQVSTV	487	500	785.9225	18.46	26.46

SI Table 6. PRM inclusion list of LC-MS/PRM analysis of the yeast to standard ratio experiment.

#	Sequence	Start	End	m/z	Start (RT)	End (RT)
1	MNFSHGSYEYHK	50	61	490.2513	19.31	24.31
2	GVNLPGTDVDLPALSEK	178	194	1092.1301	40.99	45.99
3	GVNLPGTDVDLPALSEKDKEDLR	178	200	792.9557	36.99	41.99
4	EVLGEQGKDVK	226	236	630.3882	18.9	23.9
5	IENQQGVNNFDEILK	241	255	1110.1074	38.75	43.75
6	GDLGIEIPAPEVLAVQK	265	281	736.4479	43.89	48.89
7	AEVSDVGNAILDGADCVMLSGETAK	313	337	998.8536	45.97	50.97
8	GNYPINAVTTMAETAVIAEQAIAYLPNYDDMR	338	369	1.248.9556	51.5	56.5
9	NCTPKPTSTTETVAASAVAAVFEQK	370	394	828.2118	42.65	47.65
10	GVFPFVFEKEPVSDWTDDVEAR	438	459	757.6505	45.41	50.41
11	LTSLNVVAGSDLR	7	19	525.3141	32.64	37.64
12	LTSLNVVAGSDLRR	7	20	433.2631	26.99	31.99
13	TSIIGTIGPK	21	30	722.9679	27.99	32.99
14	RTSIIGTIGPK	20	30	801.0186	21.12	26.12
15	TNNPETLVALR	31	41	728.9265	24	29
16	TNNPETLVALRK	31	42	907.5564	22.8	27.8
17	KAGLNIVR	42	49	443.6350	15.36	20.36
18	AGLNIVR	43	49	486.3183	15.95	20.95
19	SVIDNAR	62	68	502.2899	7.39	12.39
20	KSEELYPGRPLAIALDTK	69	86	672.9095	32.85	37.85
21	SEELYPGRPLAIALDTK	70	86	1166.1819	35.87	40.87
22	KSEELYPGRPLAIALDTKGPEIR	69	91	648.9904	33.33	38.33
23	SEELYPGRPLAIALDTKGPEIR	70	91	961.8915	35.66	40.66
24	TGTTTNDVDYPIPPNHEMIFTTDDK	92	116	820.9146	34.78	39.78
25	TGTTTNDVDYPIPPNHEMIFTTDDKYAK	92	119	968.7561	33.92	38.92
26	IMYVDYK	125	131	695.4029	30.89	35.89
27	ACDDKIMYVDYK	120	131	1111.6069	29.68	34.68
28	IIYVDDGVLSFQVLEVVDDK	142	161	908.8496	48.68	53.68
29	GVNLPGTDVDLPALSEKDK	178	196	885.8498	37.22	42.22
30	NGVHMVFASFIR	205	216	803.9471	35.14	40.14
31	TANDVLTIR	217	225	616.3707	20.98	25.98
32	EVLGEQGK	226	233	659.3981	15.86	20.86

5

33	TANDVLTIREVLGEQGKDVK	217	236	718.9285	36.39	41.39
34	IIVKIENQQGVNNFDEILK	237	255	967.9169	43.53	48.53
35	VTDGVMVAR	256	264	392.8954	17.2	22.2
36	IIVKIENQQGVNNFDEILKVTDGVMVAR	237	264	958.3091	45.52	50.52
37	GDLGIEIPAPEVLAVQKK	265	282	641.9037	40.2	45.2
38	SNLAGKPVICATQMLESMTYNPRPTR	287	312	852.7134	42.06	47.06
39	AIIVLSTSGTTPR	397	409	515.6505	26.61	31.61
40	YRPNCPIILVTR	414	425	582.3466	25.09	30.09
41	GVFPFVFEK	438	446	509.9761	44.17	49.17
42	EPVSDWTDDVEAR	447	459	874.4291	29.73	34.73
43	INFGIEK	460	466	426.9365	30.28	35.28
44	AKEFGILK	467	474	797.0281	26.38	31.38
45	AKEFGILKK	467	475	488.3332	25.34	30.34
46	EFGILKK	469	475	508.0090	26.29	31.29
47	GDTYVSIQGFK	476	486	836.9764	32.02	37.02
48	KGDTYVSIQGFK	475	486	1015.6074	28.43	33.43
49	AGAGHSNTLQVSTV	487	500	524.2886	15.24	20.24

SI Table 7. Inclusion list of LC-MS/PRM analysis of the replication experiment.

#	Sequence	Start	End	m/z	Start (RT)	End (RT)
1	LTSLNVVAGSDLR	7	19	787.4585	39.64	44.64
2	TSIIGTIGPK	21	30	722.9612	33.54	38.54
3	TNNPETLVALR	31	41	728.9189	29.14	34.14
4	AGLNIVR	43	49	486.3138	18.38	23.38
5	KAGLNIVR	42	49	664.9419	17.39	22.39
6	SVIDNAR	62	68	502.2909	9.1	14.1
7	KSEELYPGRPLAIALDTK	69	86	896.8673	39.67	44.67
8	TGTTTNDVDYPIPPNHEMIFTTDDK	92	116	1094.2080	42.23	47.23
9	IMYVDYK	125	131	695.3965	36.78	41.78
10	IIYVDDGVLSFQVLEVVDDK	142	161	908.8405	70.96	75.96
11	GVNLPGTDVDLPALSEK	178	194	1092.1160	50.12	55.12
12	GVNLPGTDVDLPALSEKDKEDLR	178	200	792.9481	45.18	50.18
13	TANDVLTIR	217	225	411.2446	25.14	30.14
14	EVLGEQGK	226	233	659.3923	18.28	23.28
15	IENQQGVNNFDEILK	241	255	1110.1064	45.35	50.35
16	VTDGVMVAR	256	264	588.8336	20.78	25.78
17	GDLGIEIPAPEVLAVQK	265	281	1104.1521	54.52	59.52
18	SNLAGKPVICATQMLESMTYNPRPTR	287	312	852.7034	51.87	56.87
19	AEVSDVGNAILDGADCVMLSGETAK	313	337	749.3846	64.98	69.98
20	GNYPINAVTTMAETAVIAEQAIAYLPNYDDMR	338	369	936.9662	74.45	79.45
21	NCTPKPTSTTETVAASAVAAVFEQK	370	394	1103.9351	52.28	57.28
22	AIIVLSTSGTTPR	397	409	772.9642	31.67	36.67
23	YRPNCPIILVTR	414	425	582.3413	29.52	34.52
24	GVFPFVFEKEPVSDWTDDVEAR	438	459	757.6445	62.08	67.08
25	EPVSDWTDDVEAR	447	459	874.4193	35.89	40.89
26	INFGIEK	460	466	639.8949	36.43	41.43
27	AKEFGILK	467	474	797.0205	30.92	35.92
28	GDTYVSIQGFK	476	486	836.9692	39.12	44.12
29	AGAGHSNTLQVSTV	487	500	785.9220	17.5	22.5

5

SI Table 8. Inclusion list of LC-MS/PRM analysis of the IMX372 aerobic growth curve experiment.

#	Sequence	Start	End	m/z	Start (RT)	End (RT)
1	LTSLNVVAGSDLR	7	19	787.4594	108.3	118.3
2	LTSLNVVAGSDLRR	7	20	577.3424	97.3	107.3
3	RTSIIGTIGPK	20	30	534.3429	70.9	80.9
4	TSIIGTIGPK	21	30	722.9599	99.9	109.9
5	TNNPETLVALR	31	41	728.9191	86.2	96.2
6	TNNPETLVALRK	31	42	605.3674	83.4	93.4
7	KAGLNIVR	42	49	443.6313	47.1	57.1
8	AGLNIVR	43	49	486.3132	47.4	57.4
9	MNFSHGSYEYHK	50	61	490.2458	45.2	55.2
10	SVIDNAR	62	68	502.2899	16	26
11	KSEELYPGRPLAIALDTK	69	86	672.9028	107.5	117.5
12	SEELYPGRPLAIALDTK	70	86	777.7833	110.9	120.9
13	KSEELYPGRPLAIALDTKGPEIR	69	91	648.9845	107.7	117.7
14	IMYVDYK	125	131	695.3956	105.5	115.5
15	ACDDKIMYVDYK	120	131	741.399	104.5	114.5
16	GVNLPGTDVDLPALSEK	178	194	1092.12	115.7	125.7
17	GVNLPGTDVDLPALSEKDKEDLR	178	200	792.9467	111.6	121.6
18	NGVHMVFASFIR	205	216	536.2957	109.3	119.3
19	TANDVLTIR	217	225	616.3638	70.7	80.7
20	EVLGEQGK	226	233	659.3925	49.6	59.6
21	EVLGEQGKDVK	226	236	630.3834	65.3	75.3
22	IENQQGVNNFDEILK	241	255	1110.11	112.5	122.5
23	VTDGVMVAR	256	264	588.8328	53.4	63.4
24	GDLGIEIPAPEVLAVQK	265	281	1104.158	119	129
25	GDLGIEIPAPEVLAVQKK	265	282	855.5275	114.9	124.9
26	SNLAGKPVICATQMLESMTYNPRPTR	287	312	852.7064	116.5	126.5
27	NCTPKPTSTTETVAASAVAAVFEQK	370	394	1103.937	117.4	127.4
28	AIIVLSTSGTTPR	397	409	515.6449	93	103
29	YRPNCPIILVTR	414	425	582.3406	87.6	97.6
30	GVFPFVFEK	438	446	764.452	120.1	130.1
31	GVFPFVFEKEPVSDWTDDVEAR	438	459	1009.854	123.5	133.5
32	INFGIEK	460	466	639.8939	105	115
33	AKEFGILK	467	474	797.02	91.1	101.1

34	AKEFGILKK	467	475	488.3284	86.1	96.1
35	KGDTYVSIQGFK	475	486	677.4014	102.4	112.4
36	GDTYVSIQGFK	476	486	836.969	107.6	117.6
37	AGAGHSNTLQVSTV	487	500	785.9233	47.9	57.9

5

Outlook

Mass spectrometry-based cellular proteomics has become routine in many fields of research [1–4]. However, the application to industrial microbes is still in its infancy. For example, albeit the yeast *S. cerevisiae* has been subjected to many proteomic studies over the past decades, the regulation of its complex metabolism and its proteome dynamics under different substrate conditions has not been fully explored to date. Additionally, many features (*e.g.* protein modifications) have been investigated in isolation, or under poorly controlled conditions. However, predictive models and engineering efforts, *e.g.* to improve industrial strains require (quantitative) large-scale data established under highly controlled conditions. The research presented in this thesis described advanced protocols for the large-scale study of the yeast proteome. Furthermore, proteome dynamics data and methods to quantify protein modifications have been established, that support metabolic engineering efforts in order to advance yeast as cell factory and eukaryotic model organism.

In cellular proteomics, the conclusion is drawn from the identified spectra. This however leaves often large volumes of sequencing spectra uninterpreted. A fraction of these unidentified spectra derives from peptides that carry unexpected modifications. **Chapter 2** highlighted the importance of interpreting these signals, *e.g.* in order to provide a better understanding of post-translational modifications, which are potential metabolic regulators. For this, the recently advanced unrestricted (open) database searching tools provide a promising solution. However, currently, post-translational modifications are still poorly covered in cellular proteomics studies.

Additionally, when studying protein modifications or performing quantitative proteomics experiments, an effective, but also mild sample preparation protocol is required. Therefore, **Chapter 3** described a study where a matrix of different sample preparation protocols was investigated. Their performance was evaluated using a combination of orthogonal search algorithm, including open database searching to identify modifications. The majority of the unidentified sequencing spectra were predominantly of low-quality, which lack sufficient peptide fragment coverage for a confident identification. This study established a large resource of protocols and associated mass spectrometric raw data for future studies. Nevertheless, next generation algorithms have been developed only recently to increase spectra identification rates. For example, "ionbot" is a new search engine that employs machine learning to interpret complex peptide fragmentation spectra, which significantly increases the peptide identification rate [418]. Sample preparation is still expected to remain a major contributor to data variation and poor comparability between proteomics experiments [177]. Recently, more specialized protocols, such as S-trap and ultrasonic filter aided sample

preparation (FASP) digestion demonstrated an excellent proteome coverage at reduced processing times [317, 318]. Automated methods, such as immobilized enzymatic reactor (IMER) methods, are even more promising as these are directly coupled to LC-MS/MS and thereby enable automated digestion and online MS/MS analysis [419].

In addition to changes in protein modifications, protein abundance changes can be a response to the changing environment conditions (*e.g.* substrates, or substrate levels). In **Chapter 4**, *S. cerevisiae* was subjected to altering substrate (glucose) levels in aerobic and anaerobic batch cultures. The objective of this study was to capture the proteome dynamics in response to oxygen availability under dynamic glucose conditions. Interestingly, anaerobic conditions showed substantially less protein abundance changes. This supports the notion that anaerobically growing cells lack the time and resources needed to adapt to the change to carbon-depleted environment. However, the viability of the cells should be measured to support this hypothesis. In addition, aerobic stationary phase yeast have the ability to restart growth once limiting nutrients become available again [420]. Such experiments could demonstrate the readiness of yeast cells to re-enter the growth cycle after exposure to growth-arresting conditions. Stationary phase yeast cultures can function as model to explain the role of respiration and reactive oxygen species in aging, longevity and apoptosis, in particular when comparing aerobic to anaerobic conditions. Moreover, industrial-scale processes favour anaerobic environments for practical and financial reasons. Therefore, the established large-scale proteome data provide a valuable resource for the development of predictive models. In this large-scale set-up, global modification profiles of *S. cerevisiae* could be monitored over time in the (an)aerobic batch culture set-up used in **Chapter 4**, as previously it had been indicated that acetylation occurred more frequently in growth-arrested cells compared to exponentially growing yeast cells [15]. Certain modifications may be sensitive to oxygen availability and this could be explored with open search algorithms.

Quantitative proteome data were captured for a large fraction of the yeast protein biomass. Still, a deeper proteome coverage could be achieved by employing additional peptide fractionation, *e.g.* by using high pH reversed-phase fractionation, or strong cation exchange chromatography. However, in **Chapter 4**, 54 proteomes were quantified, and additional fractionation would have tremendously increased the data acquisition time, by at the same time only little additional insights regarding main metabolic routes. Additionally, studying post-translational modifications in the yeast central metabolic pathways, such as glycolysis, requires simplified strains. Therefore, **Chapter 4** included a minimal glycolysis yeast strain where the minor glycolytic paralogues were eliminated. Surprisingly, only minor protein-level differences were observed between the wild-type and the minimal glycolysis strain under different oxygen and glucose levels. However, more subtle changes could likely be resolved by employing more accurate methods, or additional statistical (clustering) approaches [324]. Nevertheless, the lack of profound proteome response (under the tested

conditions) qualifies the minimal glycolysis strain as simplified model for studying post-translational modifications in yeast glycolysis.

The potential of post-translational modifications, such as phosphorylation, to regulate enzyme activity has been many times demonstrated over the past decades. However, the spectrum of all possible modifications and their contribution to the regulation of the metabolic flux is only poorly understood, in particular for microbes. Moreover, the knowledge about the possible protein modifications in yeast has been curated from many different studies, with often very different experimental setups, culturing conditions, and analysis methods. Furthermore, the analysis of single types of modifications became standard practice, while the global analysis (and quantification) of all possible modifications remains a challenge. Therefore, **Chapter 5** provides proof of concept for a method that allows to quantify the global spectrum of protein modifications for metabolic enzymes directly from complex cell lysates. Thereby, a protein standard was produced by cell free synthesis, which aimed to quantify the unmodified peptide fraction across the protein backbone of metabolic enzymes. The standard furthermore acted as a 'carrier proteome' to enable a reproducible and sensitive quantification. The method was exemplified for the glycolytic enzyme Pyk1, across different phases of an aerobic growth. Interestingly, several regions of the protein showed changes in the degree of protein modification. Peptides were found increasingly modified in the stationary growth phase, in which proteins are susceptible to aging processes such as oxidation by reactive oxygen species, that may severely influence the stoichiometry of unmodified peptide fraction [239]. Nevertheless, the accuracy of the peptide ratio determination should be further verified using MS3 quantification, where the additional isolation and fragmentation eliminates interference from unrelated peptides [95]. Moreover, certain peptide sequences may be more sensitive to oxygen, pH and temperature conditions, and small differences in sample handling could bias the outcome (*e.g.* via increased oxidation, or deamidation products). A more complete proteome sequence coverage can also be obtained by employing the tested multi-protease approach. Additionally, the modification landscape should be also monitored under highly dynamic conditions, such as substrate perturbations, as post-translational modifications can regulate on a very short time-scale. Comparison to steady-state conditions could point towards modifications with regulatory functions. Finally, knowledge on the actually occurring modifications (and modification sites) will help to understand protein complex formation and enzyme–substrate interactions. This method is thus expected to identify effective intervention points and targets for metabolic engineering efforts.

Over the past decade, several other advanced proteomics approaches have been developed, which are expected to bring new insights into the field of microbial proteomics. For example, data-independent methods, where all peptides within a pre-selected m/z range are subjected to fragmentation, have been continuously improved over the past years [49]. This enables

O

highly reproducible and rapid quantification of large numbers of peptides across large numbers of experiments. Furthermore, advanced algorithms allow now to employ conventional database searching algorithm for the analyse of these complex fragmentation spectra [421]. Ion mobility (IM) is another technology that is increasingly employed in large-scale cellular proteomics experiments. IM enables additional peptide separation based on size and shape, more particularly, based on the peptide's collisional cross section. Ion mobility mass spectrometry already demonstrated improved selectivity, quantification accuracy and proteome coverage [36]. The application of "field asymmetric ion mobility spectrometry" (FAIMS) allowed the separation of peptides which differed only in their peptide modification site [422]. In another ion mobility technique, called trapped ion mobility that is employed in the timsTOF Pro mass spectrometer (Bruker), the ions are additionally trapped at different positions in an ion tunnel device before separation in the following mass analyser [34–36]. TIMS is also the foundation of the parallel accumulation-serial fragmentation (PASEF) technology, in which MS/MS precursor selection is synchronized with TIMS separation [37]. This process increases sequencing speed by 10-fold and consequently proteome coverage without loss of sensitivity. Shorter measurement times can be also realised by super high resolution mass spectrometers such as the scimaX "magnetic resonance mass spectrometer" (MRMS). This instrument has a mass resolution exceeding twenty million. In some applications the use of flow injection analysis enables extreme throughput, by eliminating the time-consuming chromatographic separation step. Albeit this has been hardly employed for cellular proteomics, such instruments enable the advanced chemical interpretation of (unexpected) post-translational modifications. Furthermore, quantitative proteomics using isobaric reagents can profit from real-time search algorithms. Those identify fragment spectra within milliseconds (after measurement) and only trigger the (time consuming) acquisition of quantitative spectra upon confident peptide identification [423, 424]. In comparison to standard quantitative workflows, the acquisition time was reduced by half, while obtaining the same proteomic depth.

Current mass spectrometry-based proteomics approaches require relatively large amounts of proteins and are time consuming (several hours per measurement). Consequently, proteins are extracted from many cells. This provides information only from the population average. However, cell to cell heterogeneity, and possible (metabolic) sub-populations are not resolved. Recent advancements in single-cell proteomics demonstrated the quantitative analysis of more than 1000 proteins from single (mammalian) cells [425]. However, single-cell proteomics is still in its infancy. Sample preparation and sensitivity are still limiting factors for the application to microbial cells [426]. Furthermore, current mass spectrometry-based approaches also show limitations towards detecting and quantifying proteoform heterogeneity. For example, the combination of (different) post-translational modifications on the same protein may be crucial for certain functions (protein-protein interactions). These questions are very challenging, or even impossible to address by current mass spectrometry-

based proteomics approaches. Interestingly, promising next generation proteomics approaches are currently under development. For example, single-molecule Fluorescence resonance energy transfer (FRET)-based protein fingerprinting performs tagging of specific amino acids with fluorescent reporters that are subsequently detected optically [427, 428]. Nanopore sequencing is another method where single, full-sized proteins are analysed. Thereby, the protein chain is driven through a nanopore by an external electric field or the pull of molecular motor and the amino acids could be sequenced by the change in current [429]. Promising detection of post-translational modifications has already been shown with such setups [430]. These techniques are revolutionary but are currently still in early stages of development.

O

References

1. Patterson, S.D., Aebersold, R.H.: Proteomics: the first decade and beyond. Nat. Genet. 33, 311-323. (2003)

2. Aslam, B., Basit, M., Nisar, M.A., Khurshid, M., Rasool, M.H.: Proteomics: Technologies and their applications. J. Chromatogr. Sci. 55, 182–196 (2017). https://doi.org/10.1093/chromsci/bmw167

3. Yates, J.R., Ruse, C.I., Nakorchevsky, A.: Proteomics by mass spectrometry: approaches, advances, and applications. Annu. Rev. Biomed. Eng. 11, 49–79 (2009). https://doi.org/10.1146/annurev-bioeng-061008-124934

4. Suhre, K., McCarthy, M.I., Schwenk, J.M.: Genetics meets proteomics: perspectives for large population-based studies. Nat. Rev. Genet. 22, 19–37 (2021). https://doi.org/10.1038/s41576-020-0268-2

5. Schena, M., Shalon, D., Davis, R.W., Brown, P.O.: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science (80-.). 270, 467–470 (1995). https://doi.org/10.1126/science.270.5235.467

6. Yates, J.R.: The revolution and evolution of shotgun proteomics for large-scale proteome analysis. J. Am. Chem. Soc. 135, 1629–1640 (2013). https://doi.org/10.1021/ja3094313

7. Griffiths, J.: A brief history of mass spectrometry. Anal. Chem. 80, 5678–5683 (2008). https://doi.org/10.1002/9781119377368.ch2

8. Thomson, J.J.: XL. Cathode Rays . London, Edinburgh, Dublin Philos. Mag. J. Sci. 44, 293–316 (1897). https://doi.org/10.1080/14786449708621070

9. Aston, F.W.: XLIV. The constitution of atmospheric neon . London, Edinburgh, Dublin Philos. Mag. J. Sci. 39, 449–455 (1920). https://doi.org/10.1080/14786440408636058

10. Nier, A.O.: A mass spectrometer for routine isotope abundance measurements. Sci. instruments. 11, (1940)

11. Biemann, K.: Four decades of structure determination by mass spectrometry: From alkaloids to heparin. J. Am. Soc. Mass Spectrom. 13, 1254–1272 (2002). https://doi.org/10.1016/S1044-0305(02)00441-5

12. Hunt, D.F., Buko, A.M., Ballard, J.M., Shabanowitz, J., Giordani, A.B.: Sequence analysis of polypeptides by collision activated dissociation on a triple quadrupole mass spectrometer. Biol. Mass Spectrom. 8, 397–408 (1981). https://doi.org/10.1002/bms.1200080909

13. Zubarev, R.A., Makarov, A.: Orbitrap mass spectrometry. Anal. Chem. 85, 5288–5296 (2013). https://doi.org/10.1021/ac4001223

14. Marshall, A.G., Hendrickson, C.L.: High-resolution mass spectrometers. Annu. Rev. Anal. Chem. 1, 579–599 (2008). https://doi.org/10.1146/annurev.anchem.1.031207.112945

15. Makarov, A., Denisov, E., Kholomeev, A., Balschun, W., Lange, O., Strupat, K., Horning, S.: Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. Anal. Chem. 78, 2113–2120 (2006). https://doi.org/10.1021/ac0518811

16. Mamyrin, B.A.: Time-of-flight mass spectrometry (concepts, achievements, and prospects). Int. J. Mass Spectrom. 206, 251–266 (2001). https://doi.org/10.1016/S1387-3806(00)00392-4

17. Chernushevich, I. V., Loboda, A. V., Thomson, B.A.: An introduction to quadrupole–time-of-flight mass spectrometry. J. Mass Spectrom. 36, 849–865 (2001)

R

18.   Aebersold, R., Mann, M.: Mass-spectrometric exploration of proteome structure and function. Nature. 537, 347–355 (2016). https://doi.org/10.1038/nature19949

19.   Deutsch, E.W., Csordas, A., Sun, Z., Jarnuczak, A., Perez-Riverol, Y., Ternent, T., Campbell, D.S., Bernal-Llinares, M., Okuda, S., Kawano, S., Moritz, R.L., Carver, J.J., Wang, M., Ishihama, Y., Bandeira, N., Hermjakob, H., Vizcaíno, J.A.: The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. Nucleic Acids Res. 45, D1100–D1106 (2017). https://doi.org/10.1093/nar/gkw936

20.   Wu, C.C., MacCoss, M.J.: Shotgun proteomics: Tools for the analysis of complex biological systems. Curr. Opin. Mol. Ther. 4, 242–250 (2002)

21.   Steen, H., Mann, M.: The ABC's (and XYZ's) of peptide sequencing. Nat. Rev. Mol. Cell Biol. 5, 699–711 (2004). https://doi.org/10.1038/nrm1468

22.   Aebersold, R., Mann, M.: Mass spectrometry-based proteomics. Nature. 422, 198–207 (2003). https://doi.org/10.1038/nature01511

23.   Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.C., Yates, J.R.: Protein analysis by shotgun/bottom-up proteomics. Chem. Rev. 113, 2343–2394 (2013). https://doi.org/10.1021/cr3003533

24.   Tsiatsiani, L., Heck, A.J.R.: Proteomics beyond trypsin. FEBS J. 282, 2612–2626 (2015). https://doi.org/10.1111/febs.13287

25.   Shen, Y., Smith, R.D.: Proteomics based on high-efficiency capillary separations. Electrophoresis. 23, 3106–3124 (2002). https://doi.org/10.1002/1522-2683(200209)23:18<3106::AID-ELPS3106>3.0.CO;2-Y

26.   Thakur, S.S., Geiger, T., Chatterjee, B., Bandilla, P., Fröhlich, F., Cox, J., Mann, M.: Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation. Mol. Cell. Proteomics. 10, 1–9 (2011). https://doi.org/10.1074/mcp.M110.003699

27.   Di Palma, S., Hennrich, M.L., Heck, A.J.R., Mohammed, S.: Recent advances in peptide separation by multidimensional liquid chromatography for proteome analysis. J. Proteomics. 75, 3791–3813 (2012). https://doi.org/10.1016/j.jprot.2012.04.033

28.   Fenn, J.B., Mann, M., Meng, C.K.A.I., Wong, S.F., Whitehouse, C.M.: Electrospray ionization for mass spectrometry of large biomolecules. Science (80-.). 246, (1989)

29.   Hillenkamp, F., Beavis, R.C., Brian, T.: Matrix-Assisted Laser Desorption/Ionization mass spectrometry of biopolymers. Anal. Chem. 63, (1991)

30.   Jones, A.W., Cooper, H.J.: Dissociation techniques in mass spectrometry-based proteomics. Analyst. 136, 3419–3429 (2011). https://doi.org/10.1039/c0an01011a

31.   Wang, Z., Mülleder, M., Batruch, I., Chelur, A., Textoris-Taube, K., Schwecke, T., Hartl, J., Causon, J., Castro-Perez, J., Demichev, V., Tate, S., Ralser, M.: High-throughput proteomics of nanogram-scale samples with Zeno SWATH DIA. bioRxiv. 2022.04.14.488299 (2022)

32.   Han, X., Aslanian, A., Yates, J.R.: Mass spectrometry for proteomics. Curr. Opin. Chem. Biol. 12, 483–490 (2008). https://doi.org/10.1016/j.cbpa.2008.07.024

33.   Kelstrup, C.D., Bekker-Jensen, D.B., Arrey, T.N., Hogrebe, A., Harder, A., Olsen, J. V.: Performance evaluation of the Q Exactive HF-X for shotgun proteomics. J. Proteome Res. 17, 727–738 (2018). https://doi.org/10.1021/acs.jproteome.7b00602

34.   Distler, U., Kuharev, J., Navarro, P., Tenzer, S.: Label-free quantification in ion mobility–enhanced data-independent acquisition proteomics. Nat. Protoc. 11, 795–812 (2016)

35. Baker, E.S., Livesay, E.A., Orton, D.J., Moore, R.J., Danielson, W.F., Prior, D.C., Ibrahim, Y.M., LaMarche, B.L., Mayampurath, A.M., Schepmoes, A.A., Hopkins, D.F., Tang, K., Smith, R.D., Belov, M.E.: An LC-IMS-MS platform providing increased dynamic range for high-throughput proteomic studies. J. Proteome Res. 9, 997–1006 (2010). https://doi.org/10.1021/pr900888b

36. Charkow, J., Röst, H.L.: Trapped ion mobility spectrometry reduces spectral complexity in mass spectrometry-based proteomics. Anal. Chem. 93, 16751–16758 (2021). https://doi.org/10.1021/acs.analchem.1c01399

37. Meier, F., Brunner, A.D., Koch, S., Koch, H., Lubeck, M., Krause, M., Goedecke, N., Decker, J., Kosinski, T., Park, M.A., Bache, N., Hoerning, O., Cox, J., Räther, O., Mann, M.: Online parallel accumulation–serial fragmentation (PASEF) with a novel trapped ion mobility mass spectrometer. Mol. Cell. Proteomics. 17, 2534–2545 (2018). https://doi.org/10.1074/mcp.TIR118.000900

38. Greguš, M., Kostas, J.C., Ray, S., Abbatiello, S.E., Ivanov, A.R.: Improved sensitivity of ultralow flow LC-MS-Based proteomic profiling of limited samples using monolithic capillary columns and FAIMS technology. Anal. Chem. 92, 14702–14712 (2020). https://doi.org/10.1021/acs.analchem.0c03262

39. Nesvizhskii, A.I.: A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J. Proteomics. 73, 2092–2123 (2010). https://doi.org/10.1016/j.jprot.2010.08.009

40. Chen, C., Hou, J., Tanner, J.J., Cheng, J.: Bioinformatics methods for mass spectrometry-based proteomics data analysis. Int. J. Mol. Sci. 21, (2020). https://doi.org/10.3390/ijms21082873

41. Shteynberg, D., Nesvizhskii, A.I., Moritz, R.L., Deutsch, E.W., Tu, C., Sheng, Q., Li, J., Ma, D., Shen, X., Wang, X., Shyr, Y., Yi, Z., Qu, J., Editor, S., Walker, J.M., Yu, W., Taylor, J.A., Davis, M.T., Bonilla, L.E., Lee, K.A., Auger, P.L., Farnsworth, C.C., Welcher, A.A., Patterson, S.D., Searle, B.C., Turner, M., Nesvizhskii, A.I., Dagda, R.K., Sultana, T., Lyons-Weiler, J., Alves, G., Wu, W.W., Wang, G., Shen, R.F., Yu, Y.K., Kwon, T., Choi, H., Vogel, C., Nesvizhskii, A.I., Marcotte, E.M., Agten, A., Van Houtven, J., Askenazi, M., Burzykowski, T., Laukens, K., Valkenborg, D.: Visualizing the agreement of peptide assignments between different search engines. J. Proteome Res. 7, 39–47 (2008). https://doi.org/10.1002/jms.4471

42. Searle, B.C., Turner, M., Nesvizhskii, A.I.: Improving sensitivity by probabilistically combining results from multiple MS/MS search methodologies. J. Proteome Res. 7, 245–253 (2008). https://doi.org/10.1021/pr070540w

43. Kwon, T., Choi, H., Vogel, C., Nesvizhskii, A.I., Marcotte, E.M.: MSblender: A probabilistic approach for integrating peptide identifications from multiple database search engines. J. Proteome Res. 10, 2949–2958 (2011). https://doi.org/10.1021/pr2002116

44. Alves, G., Wu, W.W., Wang, G., Shen, R.F., Yu, Y.K.: Enhancing peptide identification confidence by combining search methods. J. Proteome Res. 7, 3102–3113 (2008). https://doi.org/10.1021/pr700798h

45. Tu, C., Sheng, Q., Li, J., Ma, D., Shen, X., Wang, X., Shyr, Y., Yi, Z., Qu, J.: Optimization of search engines and postprocessing approaches to maximize peptide and protein identification for high-resolution mass data. J. Proteome Res. 14, 4662–4673 (2015). https://doi.org/10.1021/acs.jproteome.5b00536

46. Yu, W., Taylor, J.A., Davis, M.T., Bonilla, L.E., Lee, K.A., Auger, P.L., Farnsworth, C.C., Welcher, A.A., Patterson, S.D.: Maximizing the sensitivity and reliability of peptide identification in large-scale proteomic experiments by harnessing multiple search engines. Proteomics. 10, 1172–1189 (2010). https://doi.org/10.1002/pmic.200900074

47. Agten, A., Van Houtven, J., Askenazi, M., Burzykowski, T., Laukens, K., Valkenborg, D.: Visualizing the agreement of peptide assignments between different search engines. J. Mass Spectrom. 55, (2020). https://doi.org/10.1002/jms.4471

48. Dagda, R.K., Sultana, T., Lyons-Weiler, J.: Evaluation of the consensus of four peptide identification

algorithms for Tandem mass spectrometry based proteomics. J. Proteomics Bioinforma. 3, 39–47 (2010). https://doi.org/10.4172/jpb.1000119

49. Ludwig, C., Gillet, L., Rosenberger, G., Amon, S., Collins, B.C., Aebersold, R.: Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol. Syst. Biol. 14, e8126 (2018)

50. Zhang, F., Ge, W., Ruan, G., Cai, X., Guo, T.: Data-independent acquisition mass spectrometry-based proteomics and software tools: A Glimpse in 2020. Proteomics. 20, 1900276 (2020)

51. Searle, B.C., Swearingen, K.E., Barnes, C.A., Schmidt, T., Gessulat, S., Küster, B., Wilhelm, M.: Generating high quality libraries for DIA MS with empirically corrected peptide predictions. Nat. Commun. 11, 1–10 (2020). https://doi.org/10.1038/s41467-020-15346-1

52. Poulos, R.C., Hains, P.G., Shah, R., Lucas, N., Xavier, D., Manda, S.S., Anees, A., Koh, J.M.S., Mahboob, S., Wittman, M., Williams, S.G., Sykes, E.K., Hecker, M., Dausmann, M., Wouters, M.A., Ashman, K., Yang, J., Wild, P.J., deFazio, A., Balleine, R.L., Tully, B., Aebersold, R., Speed, T.P., Liu, Y., Reddel, R.R., Robinson, P.J., Zhong, Q.: Strategies to enable large-scale proteomics for reproducible research. Nat. Commun. 11, 1–13 (2020). https://doi.org/10.1038/s41467-020-17641-3

53. Demichev, V., Szyrwiel, L., Yu, F., Teo, G.C., Rosenberger, G., Niewienda, A., Ludwig, D., Decker, J., Kaspar-Schoenefeld, S., Lilley, K.S., Mülleder, M., Nesvizhskii, A.I., Ralser, M.: dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. Nat. Commun. 13, (2022). https://doi.org/10.1038/s41467-022-31492-0

54. Lange, V., Picotti, P., Domon, B., Aebersold, R.: Selected reaction monitoring for quantitative proteomics: A tutorial. Mol. Syst. Biol. 4, (2008). https://doi.org/10.1038/msb.2008.61

55. Shi, T., Su, D., Liu, T., Tang, K., Camp II, D.G., Qian, W.-J., Smith, R.D.: Advancing the sensitivity of selected reaction monitoring-based targeted quantitative proteomics. Proteomics. 12, 1074–1092 (2012)

56. Gallien, S., Duriez, E., Crone, C., Kellmann, M., Moehring, T., Domon, B.: Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer. Mol. Cell. Proteomics. 11, 1709–1723 (2012). https://doi.org/10.1074/mcp.O112.019802

57. Peterson, A.C., Russell, J.D., Bailey, D.J., Westphall, M.S., Coon, J.J.: Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol. Cell. Proteomics. 11, 1475–1488 (2012). https://doi.org/10.1074/mcp.O112.020131

58. Ahrné, E., Muller, M., Lisacek, F.: Unrestricted identification of modified proteins using MS / MS. 671–686 (2010). https://doi.org/10.1002/pmic.200900502

59. Skinner, O.S., Kelleher, N.L.: Illuminating the dark matter of shotgun proteomics. Nat. Biotechnol. 33, 717–718 (2015). https://doi.org/10.1038/nbt.3287

60. Griss, J., Perez-Riverol, Y., Lewis, S., Tabb, D.L., Dianes, J.A., Del-Toro, N., Rurik, M., Walzer, M., Kohlbacher, O., Hermjakob, H., Wang, R., Vizcano, J.A.: Recognizing millions of consistently unidentified spectra across hundreds of shotgun proteomics datasets. Nat. Methods. 13, 651–656 (2016). https://doi.org/10.1038/nmeth.3902

61. Ishihama, Y., Oda, Y., Tabata, T., Sato, T., Nagasu, T., Rappsilber, J., Mann, M.: Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol. Cell. Proteomics. 4, 1265–1272 (2005). https://doi.org/10.1074/mcp.M500061-MCP200

62. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M.: Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a simple and accurate approach to expression proteomics. Mol. Cell. Proteomics. 1, 376–386 (2002). https://doi.org/10.1074/mcp.m200025-mcp200

63. Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G., Neumann, T., Hamon, C.:

Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal. Chem. 75, 1895–1904 (2003). https://doi.org/10.1021/ac0262560

64. Marcus, K., Eisenacher, M., Sitek, B.: Quantitative Methods in Proteomics Second Edition Methods in Molecular Biology. (2021)

65. O'Connell, J.D., Paulo, J.A., O'Brien, J.J., Gygi, S.P.: Proteome-Wide Evaluation of Two Common Protein Quantification Methods. J. Proteome Res. 17, 1934–1942 (2018). https://doi.org/10.1021/acs.jproteome.8b00016

66. Li, Z., Adams, R.M., Chourey, K., Hurst, G.B., Hettich, R.L., Pan, C.: Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ orbitrap velos. J. Proteome Res. 11, 1582–1590 (2012). https://doi.org/10.1021/pr200748h

67. Karp, N.A., Huber, W., Sadowski, P.G., Charles, P.D., Hester, S. V., Lilley, K.S.: Addressing accuracy and precision issues in iTRAQ quantitation. Mol. Cell. Proteomics. 9, 1885–1897 (2010). https://doi.org/10.1074/mcp.M900628-MCP200

68. Ting, L., Rad, R., Gygi, S.P., Haas, W.: MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat. Methods. 8, 937–940 (2011). https://doi.org/10.1038/nmeth.1714

69. Gerber, S.A., Rush, J., Stemman, O., Kirschner, M.W., Gygi, S.P.: Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc. Natl. Acad. Sci. U. S. A. 100, 6940–6945 (2003). https://doi.org/10.1073/pnas.0832254100

70. Beynon, R.J., Doherty, M.K., Pratt, J.M., Gaskell, S.J.: Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. Nat. Methods. 2, 587–589 (2005). https://doi.org/10.1038/nmeth774

71. Singh, S., Springer, M., Steen, J., Kirschner, M.W., Steen, H.: FLEXIQuant: A novel tool for the absolute quantification of proteins, and the simultaneous identification and quantification of potentially modified peptides. J. Proteome Res. 8, 2201–2210 (2009). https://doi.org/10.1021/pr800654s

72. McGovern, P.E., Zhang, J., Tang, J., Zhang, Z., Hall, G.R., Moreau, R.A., Nuñez, A., Butrym, E.D., Richards, M.P., Wang, C.S., Cheng, G., Zhao, Z., Wang, C.: Fermented beverages of pre- and proto-historic China. Proc. Natl. Acad. Sci. U. S. A. 101, 17593–17598 (2004). https://doi.org/10.1073/pnas.0407921102

73. Fuller, D.Q., Carretero, L.G.: The archaeology of neolithic cooking traditions: Archaeobotanical approaches to baking, boiling and fermenting. Archaeol. Int. 21, 109–121 (2018). https://doi.org/10.5334/ai-391

74. Chapman, A.C.: The yeast cell: What did Leeuwenhoeck see? J. Inst. Brew. (1931)

75. Pasteur, L.: La fermentation appelee lactique. Ann. Chim. Phys. 52, 404–418 (1858)

76. Goffeau, A., Barrell, B.G., Bussey, H., Davis, R.W., Dujon, B., Feldmann, H., Galibert, F., Hoheisel, J.D., Jacq, C., Johnston, M., Louis, E.J., Mewes, H.W., Murakami, Y., Philippsen, P., Tettelin, H., Oliver, S.G.: Life with 6000 Genes. Science (80-.). 274, 546–567 (1996). https://doi.org/jyu

77. EFSA: EFSA Scientific colloquium summary report. (2004)

78. Ostergaard, S., Olsson, L., Nielsen, J.: Metabolic Engineering of *Saccharomyces cerevisiae*. Microbiol. Mol. Biol. Rev. 64, 34–50 (2000). https://doi.org/10.1128/MMBR.64.1.34-50.2000.Updated

79. Mattanovich, D., Branduardi, P., Dato, L., Gasser, B., Sauer, M., Porro, D.: Recombinant gene expression in yeasts. Recomb. gene Expr. 329–358 (2012). https://doi.org/10.1007/978-1-61779-433-9

80. Nielsen, J.: Yeast Systems Biology: model organism and cell factory. Biotechnol. J. 14, 1–9 (2019).

R

https://doi.org/10.1002/biot.201800421

81. John, R.P., Anisha, G.S., Nampoothiri, K.M., Pandey, A.: Micro and macroalgal biomass: A renewable source for bioethanol. Bioresour. Technol. 102, 186–193 (2011). https://doi.org/10.1016/j.biortech.2010.06.139

82. Qin, J., Zhou, Y.J., Krivoruchko, A., Huang, M., Liu, L., Khoomrung, S., Siewers, V., Jiang, B., Nielsen, J.: Modular pathway rewiring of *Saccharomyces cerevisiae* enables high-level production of L-ornithine. Nat. Commun. 6, 1–11 (2015). https://doi.org/10.1038/ncomms9224

83. Galanie, S., Thodey, K., Trenchard, I.J., Interrante, M.F., Smolke, C.D.: Complete biosynthesis of opioids in yeast. Science (80-.). 349, (2015)

84. Nielsen, J., Keasling, J.D.: Engineering cellular metabolism. Cell. 164, 1185–1197 (2016). https://doi.org/10.1016/j.cell.2016.02.004

85. Dicarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., Church, G.M.: Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. 41, 4336–4343 (2013). https://doi.org/10.1093/nar/gkt135

86. Mans, R., van Rossum, H.M., Wijsman, M., Backx, A., Kuijpers, N.G.A., van den Broek, M., Daran-Lapujade, P., Pronk, J.T., van Maris, A.J.A., Daran, J.M.G.: CRISPR/Cas9: A molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res. 15, 1–15 (2015). https://doi.org/10.1093/femsyr/fov004

87. Swiat, M.A., Dashko, S., Den Ridder, M., Wijsman, M., Van Der Oost, J., Daran, J.M., Daran-Lapujade, P.: FnCpf1: A novel and efficient genome editing tool for *Saccharomyces cerevisiae*. Nucleic Acids Res. 45, 12585–12598 (2017). https://doi.org/10.1093/nar/gkx1007

88. Jakočiūnas, T., Jensen, M.K., Keasling, J.D.: CRISPR/Cas9 advances engineering of microbial cell factories. Metab. Eng. 34, 44–59 (2015). https://doi.org/10.1016/j.ymben.2015.12.003

89. Nielsen, J.: Yeast Systems Biology: Model Organism and Cell Factory. Biotechnol. J. 14, 1–9 (2019). https://doi.org/10.1002/biot.201800421

90. Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer, A.G., Wilke, C.O., Marcotte, E.M.: Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science (80-.). 6237, 921–925 (2015). https://doi.org/10.1016/j.physbeh.2017.03.040

91. Akram, M.: Mini-review on glycolysis and cancer. J. Cancer Educ. 28, 454–457 (2013). https://doi.org/10.1007/s13187-013-0486-9

92. Liberti, M. V, Locasale, J.W.: The Warburg effect: How does it benefit cancer cells? Trends Biochem. Sci. 41, 287 (2016). https://doi.org/10.1016/j.tibs.2016.01.004

93. Lunt, S.Y., Vander Heiden, M.G.: Aerobic glycolysis: Meeting the metabolic requirements of cell proliferation. Annu. Rev. Cell Dev. Biol. 27, 441–464 (2011). https://doi.org/10.1146/annurev-cellbio-092910-154237

94. Diaz-Ruiz, R., Rigoulet, M., Devin, A.: The Warburg and Crabtree effects: On the origin of cancer cell energy metabolism and of yeast glucose repression. Biochim. Biophys. Acta - Bioenerg. 1807, 568–576 (2011). https://doi.org/10.1016/j.bbabio.2010.08.010

95. Pronk, J.T., Steensma, H.Y., Dijken, J.P. Van: Pyruvate metabolism in *Saccharomyces cerevisiae*. Yeast. 12, 1607–1633 (1996)

96. Bouchez, C.L., Hammad, N., Cuvellier, S., Ransac, S., Rigoulet, M., Devin, A.: The Warburg effect in yeast: Repression of mitochondrial metabolism is not a prerequisite to promote cell proliferation. Front. Oncol. 10, 1–15 (2020). https://doi.org/10.3389/fonc.2020.01333

97. Janssens, G.E., Veenhoff, L.M.: Evidence for the hallmarks of human aging in replicatively aging yeast. Microb. Cell. 3, 263–274 (2016). https://doi.org/10.15698/mic2016.07.510

98. He, C., Zhou, C., Kennedy, B.K.: The yeast replicative aging model. Biochim. Biophys. Acta - Mol. Basis Dis. 1864, 2690–2696 (2018). https://doi.org/10.1016/j.bbadis.2018.02.023

99. Romano, A.H., Conway, T.: Evolution of carbohydrate metabolic pathways. Res. Microbiol. 147, 448–455 (1996). https://doi.org/10.1016/0923-2508(96)83998-2

100. De Deken, R.H.: The Crabtree effects and its relation to the petite mutation. J. Gen. Microbiol. 44, 157–165 (1966). https://doi.org/10.1099/00221287-44-2-157

101. Nevoigt, E., Stahl, U.: Osmoregulation and glycerol metabolism in the yeast *Saccharomyces cerevisiae*. FEMS Microbiol. Rev. 21, 231–241 (1997). https://doi.org/10.1016/S0168-6445(97)00058-2

102. François, J., Parrou, J.L.: Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. FEMS Microbiol. Rev. 25, 125–145 (2001). https://doi.org/10.1016/S0168-6445(00)00059-0

103. Brauer, M.J., Saldanha, A.J., Dolinski, K., Botstein, D.: Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. Mol. Biol. Cell. 16, 2503–2517 (2005). https://doi.org/10.1091/mbc.E04

104. Rosa, C.A., Péter, G.: The yeast handbook. (2006)

105. Klosinska, M.M., Crutchfield, C.A., Bradley, P.H., Rabinowitz, J.D., Broach, J.R.: Yeast cells can access distinct quiescent states. Genes Dev. 25, 336–349 (2011). https://doi.org/10.1101/gad.2011311

106. Boender, L.G.M., Almering, M.J.H., Dijk, M., van Maris, A.J.A., de Winde, J.H., Pronk, J.T., Daran-Lapujade, P.: Extreme calorie restriction and energy source starvation in *Saccharomyces cerevisiae* represent distinct physiological states. Biochim. Biophys. Acta - Mol. Cell Res. 1813, 2133–2144 (2011). https://doi.org/10.1016/j.bbamcr.2011.07.008

107. Fabrizio, P., Andrus, V.D.L.: The chronological life span of *Saccharomyces cerevisiae* to study mitochondrial dysfunction and disease. Aging Cell. 2, 73–81 (2003). https://doi.org/10.1016/j.ymeth.2008.10.004

108. Snoek, I, Steensma, H.Y.: Factors involved in anaerobic growth of *Saccharomyces cerevisiae*. Yeast. 24, 1–10 (2007). https://doi.org/10.1002/yea

109. Nurminen, T., Konttinen, K., Suomalainen, H.: Neutral lipids in the cells and cell envelope fractions of aerobic baker's yeast and anaerobic brewer's yeast. Chem. Phys. Lipids. 14, 15–32 (1975). https://doi.org/10.1016/0009-3084(75)90012-2

110. de Groot, M.J.L., Daran-Lapujade, P., van Breukelen, B., Knijnenburg, T.A., de Hulster, E.A.F., Reinders, M.J.T., Pronk, J.T., Heck, A.J.R., Slijper, M.: Quantitative proteomics and transcriptomics of anaerobic and aerobic yeast cultures reveals post-transcriptional regulation of key cellular processes. Microbiology. 153, 3864–3878 (2007). https://doi.org/10.1099/mic.0.2007/009969-0

111. Ansell, R., Granath, K., Hohmann, S., Thevelein, J.M., Adler, L.: The two isoenzymes for yeast NAD+-dependent glycerol 3-phosphate dehydrogenase encoded by GPD1 and GPD2 have distinct roles in osmoadaptation and redox regulation. EMBO J. 16, 2179–87 (1997)

112. Sun, S., Gresham, D.: Cellular quiescence in budding yeast. Yeast. 38, 12–29 (2021). https://doi.org/10.1002/yea.3545

113. Escalera-Fanjul, X., Quezada, H., Riego-Ruiz, L., González, A.: Whole-Genome duplication and yeast's fruitful way of life. Trends Genet. 35, 42–54 (2019). https://doi.org/10.1016/j.tig.2018.09.008

114. Conant, G.C., Wolfe, K.H.: Increased glycolytic flux as an outcome of whole-genome duplication in yeast.

R

Mol. Syst. Biol. 3, (2007). https://doi.org/10.1038/msb4100170

115. Breitenbach-Schmitt, I., Heinisch, J., Schmitt, H.D., Zimmermann, F.K.: Yeast mutants without phosphofructokinase activity can still perform glycolysis and alcoholic fermentation. MGG Mol. Gen. Genet. 195, 530–535 (1984). https://doi.org/10.1007/BF00341458

116. Gancedo, C., Flores, C.-L.: Moonlighting proteins in yeasts. Microbiol. Mol. Biol. Rev. 72, 197–210 (2008). https://doi.org/10.1128/mmbr.00036-07

117. Liebermeister, W., Noor, E., Flamholz, A., Davidi, D., Bernhardt, J., Milo, R.: Visual account of protein investment in cellular functions. Proc. Natl. Acad. Sci. U. S. A. 111, 8488–8493 (2014). https://doi.org/10.1073/pnas.1314810111

118. Van Den Brink, J., Canelas, A.B., Van Gulik, W.M., Pronk, J.T., Heijnen, J.J., De Winde, J.H., Daran-Lapujade, P.: Dynamics of glycolytic regulation during adaptation of *Saccharomyces cerevisiae* to fermentative metabolism. Appl. Environ. Microbiol. 74, 5710–5723 (2008). https://doi.org/10.1128/AEM.01121-08

119. Fraenkel, D.G.: The top genes: On the distance from transcript to function in yeast glycolysis. Curr. Opin. Microbiol. 6, 198–201 (2003). https://doi.org/10.1016/S1369-5274(03)00023-7

120. Solis-Escalante, D., Kuijpers, N.G.A., Barrajon-Simancas, N., van den Broek, M., Pronk, J.T., Daran, J.M., Daran-Lapujade, P.: A Minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. Eukaryot. Cell. 14, 804–816 (2015). https://doi.org/10.1128/EC.00064-15

121. Rodríguez, A., De La Cera, T., Herrero, P., Moreno, F.: The hexokinase 2 protein regulates the expression of the GLK1, HXK1 and HXK2 genes of *Saccharomyces cerevisiae*. Biochem. J. 355, 625–631 (2001). https://doi.org/10.1042/bj3550625

122. Vallari, R.C., Cook, W.J., Audino, D.C., Morgan, M.J., Jensen, D.E., Laudano, A.P., Denis, C.L.: Glucose repression of the yeast ADH2 gene occurs through multiple mechanisms, including control of the protein synthesis of its transcriptional activator, ADR1. Mol. Cell. Biol. 12, 1663–1673 (1992). https://doi.org/10.1128/mcb.12.4.1663

123. Herrero, P., Galíndez, J., Ruiz, N., Martínez-Campa, C., Moreno, F.: Transcriptional regulation of the *Saccharomyces cerevisiae* HXK1, HXK2 and GLK1 genes. Yeast. 11, 137–144 (1995). https://doi.org/10.1002/yea.320110205

124. Boles, E., Schulte, F., Miosga, T., Freidel, K., Schlüter, E., Zimmermann, F.K., Hollenberg, C.P., Heinisch, J.J.: Characterization of a glucose-repressed pyruvate kinase (Pyk2p) in *Saccharomyces cerevisiae* that is catalytically insensitive to fructose-1,6- bisphosphate. J. Bacteriol. 179, 2987–2993 (1997). https://doi.org/10.1128/jb.179.9.2987-2993.1997

125. Boer, V.M., De Winde, J.H., Pronk, J.T., Piper, M.D.W.: The genome-wide transcriptional responses of *Saccharomyces cerevisiae* grown on glucose in aerobic chemostat cultures limited for carbon, nitrogen, phosphorus, or sulfur. J. Biol. Chem. 278, 3265–3274 (2003). https://doi.org/10.1074/jbc.M209759200

126. Fauchon, M., Lagniel, G., Aude, J.C., Lombardia, L., Soularue, P., Petat, C., Marguerie, G., Sentenac, A., Werner, M., Labarre, J.: Sulfur sparing in the yeast proteome in response to sulfur demand. Mol. Cell. 9, 713–723 (2002). https://doi.org/10.1016/S1097-2765(02)00500-2

127. GANCEDO, J.M.: Carbon catabolite repression in yeast. Eur. J. Biochem. 206, 297–313 (1992). https://doi.org/10.1111/j.1432-1033.1992.tb16928.x

128. Ahuatzi, D., Herrero, P., De La Cera, T., Moreno, F.: The glucose-regulated nuclear localization of Hexokinase 2 in *Saccharomyces cerevisiae* Is Mig1-dependent. J. Biol. Chem. 279, 14440–14446 (2004). https://doi.org/10.1074/jbc.M313431200

129. Kane, P.M.: The where, when, and how of organelle acidification by the yeast vacuolar H + -ATPase . Microbiol. Mol. Biol. Rev. 70, 177–191 (2006). https://doi.org/10.1128/mmbr.70.1.177-191.2006

130. Lu, M., Ammar, D., Ives, H., Albrecht, F., Gluck, S.L.: Physical interaction between aldolase and vacuolar H+-ATPase is essential for the assembly and activity of the proton pump. J. Biol. Chem. 282, 24495–24503 (2007). https://doi.org/10.1074/jbc.M702598200

131. Decker, B.L., Wickner, W.T.: Enolase activates homotypic vacuole fusion and protein transport to the vacuole in yeast. J. Biol. Chem. 281, 14523–14528 (2006). https://doi.org/10.1074/jbc.M600911200

132. Bradley, P.H., Gibney, P.A., Botstein, D., Troyanskaya, O.G., Rabinowitz, J.D.: Minor isozymes tailor yeast metabolism to carbon availability. mSystems. 4, (2019). https://doi.org/10.1128/msystems.00170-18

133. Daran-Lapujade, P., Rossell, S., van Gulik, W.M., Luttik, M.A.H., de Groot, M.J.L., Slijper, M., Heck, A.J.R., Daran, J.-M., de Winde, J.H., Westerhoff, H. V., Pronk, J.T., Bakker, B.M.: The fluxes through glycolytic enzymes in *Saccharomyces cerevisiae* are predominantly regulated at posttranscriptional levels. Proc. Natl. Acad. Sci. 104, 15753–15758 (2007). https://doi.org/10.1073/pnas.0707476104

134. Blázquez, M.A., Lagunas, R., Gancedo, C., Gancedo, J.M.: Trehalose-6-phosphate a new regulator of yeast glycolysis that inhibits hexokinases. FEBS. 329, 51–54 (1993)

135. Nissler, K., Otto, A., Hofmann, W.S.E.: Stimulation of yeast phosphofructokinase activity by fructose 2, 6-bisphosphate. Biochem. Biophys. Res. Commun. 111, 294–300 (1983)

136. Chen, Y., Nielsen, J.: Flux control through protein phosphorylation in yeast. 1–14 (2016). https://doi.org/10.1093/femsyr/fow096

137. Rossio, V., Paulo, J.A.: Quantitative proteome and phosphoproteome datasets of DNA replication and mitosis in *Saccharomyces cerevisiae*. Data Br. 45, 108741 (2022). https://doi.org/10.1016/j.dib.2022.108741

138. Chen, M., Li, H., Zhuang, Y.-P., Nielsen, J.: System-level analysis of flux regulation of yeast show that glycolytic flux is controlled by allosteric regulation and enzyme phosphorylation. 1–24 (2022)

139. Oliveira, A.P., Ludwig, C., Picotti, P., Kogadeeva, M., Aebersold, R., Sauer, U.: Regulation of yeast central metabolism by enzyme phosphorylation. Mol. Syst. Biol. 8, (2012). https://doi.org/10.1038/msb.2012.55

140. Hogrebe, A., Von Stechow, L., Bekker-Jensen, D.B., Weinert, B.T., Kelstrup, C.D., Olsen, J. V.: Benchmarking common quantification strategies for large-scale phosphoproteomics. Nat. Commun. 9, (2018). https://doi.org/10.1038/s41467-018-03309-6

141. Fernańdez-García, P., Peláez, R., Herrero, P., Moreno, F.: Phosphorylation of yeast hexokinase 2 regulates its nucleocytoplasmic shuttling. J. Biol. Chem. 287, 42151–42164 (2012). https://doi.org/10.1074/jbc.M112.401679

142. Xia, J., Sánchez, B.J., Chen, Y., Nielsen, J.: Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation. Nat. Commun. 1–12 (2022). https://doi.org/10.1038/s41467-022-30513-2

143. Picotti, P., Clément-Ziza, M., Lam, H., Campbell, D.S., Schmidt, A., Deutsch, E.W., Röst, H., Sun, Z., Rinner, O., Reiter, L., Shen, Q., Michaelson, J.J., Frei, A., Alberti, S., Kusebauch, U., Wollscheid, B., Moritz, R.L., Beyer, A., Aebersold, R.: A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. Nature. 494, 266–270 (2013). https://doi.org/10.1038/nature11835

144. Bruckmann, A., Hensbergen, P.J., Balog, C.I.A., Deelder, A.M., Brandt, R., Snoek, I.S.I., Steensma, H.Y., van Heusden, G.P.H.: Proteome analysis of aerobically and anaerobically grown *Saccharomyces cerevisiae* cells. J. Proteomics. 71, 662–669 (2009). https://doi.org/10.1016/j.jprot.2008.11.012

R

145. Helbig, A.O., Daran-Lapujade, P., Van Maris, A.J.A., De Hulster, E.A.F., De Ridder, D., Pronk, J.T., Heck, A.J.R., Slijper, M.: The diversity of protein turnover and abundance under nitrogen-limited steady-state conditions in *Saccharomyces cerevisiae*. Mol. Biosyst. 7, 3316–3326 (2011). https://doi.org/10.1039/c1mb05250k

146. Zampar, G.G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M.: Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol. Syst. Biol. 9, (2013). https://doi.org/10.1038/msb.2013.11

147. 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). https://doi.org/10.1074/mcp.M114.045849

148. Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., Aebersold, R.: Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. Mol. Syst. Biol. 7, 1–13 (2011). https://doi.org/10.1038/msb.2010.122

149. Slavov, N., Budnik, B.A., Schwab, D., Airoldi, E.M., van Oudenaarden, A.: Constant growth rate can be supported by decreasing energy flux and increasing aerobic glycolysis. Cell Rep. 7, 705–714 (2014). https://doi.org/10.1016/j.celrep.2014.03.057

150. Davidson, G.S., Joe, R.M., Roy, S., Meirelles, O., Allen, C.P., Wilson, M.R., Tapia, P.H., Manzanilla, E.E., Dodson, A.E., Chakraborty, S., Carter, M., Young, S., Edwards, B., Sklar, L., Werner-Washburne, M.: The proteomics of quiescent and nonquiescent cell differentiation in yeast stationary-phase cultures. Mol. Biol. Cell. 22, 988–998 (2011). https://doi.org/10.1091/mbc.E10-06-0499

151. Hebert, A.S., Richards, A.L., Bailey, D.J., Ulbrich, A., Coughlin, E.E., Westphall, M.S., Coon, J.J.: The one hour yeast proteome. Mol. Cell. Proteomics. 13, 339–347 (2014). https://doi.org/10.1074/mcp.M113.034769

152. Gao, Y., Ping, L., Duong, D., Zhang, C., Dammer, E.B., Li, Y., Chen, P., Chang, L., Gao, H., Wu, J., Xu, P.: Mass-Spectrometry-Based Near-Complete Draft of the *Saccharomyces cerevisiae* Proteome. J. Proteome Res. 20, 1328–1340 (2021). https://doi.org/10.1021/acs.jproteome.0c00721

153. Webb, K.J., Xu, T., Park, S.K., Yates, J.R.: Modified MuDPIT separation identified 4488 proteins in a system-wide analysis of quiescence in yeast. J. Proteome Res. 12, 2177–2184 (2013). https://doi.org/10.1021/pr400027m

154. Werner-Washburne, M., Braun, E.L., Crawford, M.E., Peck, V.M.: Stationary phase in *Saccharomyces cerevisiae*. Mol. Microbiol. 19, 1159–1166 (1996). https://doi.org/10.1111/j.1365-2958.1996.tb02461.x

155. Bisschops, M.M.M., Vos, T., Martínez-moreno, R., De, P., Cortés, T., Pronk, J.T., Daran-lapujade, P.: Oxygen availability strongly affects chronological lifespan and thermotolerance in batch cultures of *Saccharomyces cerevisiae*. 2, 429–444 (2015). https://doi.org/10.15698/mic2015.11.238

156. Lao-Martil, D., Verhagen, K.J.A., Schmitz, J.P.J., Teusink, B., Wahl, S.A., van Riel, N.A.W.: Kinetic modeling of *Saccharomyces cerevisiae* central carbon metabolism: Achievements, limitations, and opportunities. Metabolites. 12, (2022). https://doi.org/10.3390/metabo12010074

157. Nilsson, A., Nielsen, J.: Metabolic Trade-offs in Yeast are Caused by F1F0-ATP synthase. Sci. Rep. 6, 1–11 (2016). https://doi.org/10.1038/srep22264

158. Xia, J., Sánchez, B., Chen, Y., Campbell, K., Kasvandik, S., Nielsen, J.: Proteome allocations change linearly with specic growth rate of *Saccharomyces cerevisiae* under glucose-limitation. 1–21 (2021)

159. Metzl-Raz, E., Kafri, M., Yaakov, G., Soifer, I., Gurvich, Y., Barkai, N.: Principles of cellular resource allocation revealed by condition-dependent proteome profiling. Elife. 6, 1–21 (2017). https://doi.org/10.7554/eLife.28034

160. Björkeroth, J., Campbell, K., Malina, C., Yu, R., Bartolomeo, F. Di, Nielsen, J.: Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. Proc. Natl. Acad. Sci. U. S. A. 117, 21804–21812 (2020). https://doi.org/10.1073/pnas.1921890117

161. Elsemman, I.E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S.W., van Heerden, J., Bruggeman, F.J., Bisschops, M.M.M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., Teusink, B.: Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. Nat. Commun. 13, 1–12 (2022). https://doi.org/10.1038/s41467-022-28467-6

162. Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., Nielsen, J.: Absolute yeast mitochondrial proteome quantification reveals trade-off between biosynthesis and energy generation during diauxic shift. Proc. Natl. Acad. Sci. U. S. A. 117, 7524–7535 (2020). https://doi.org/10.1073/pnas.1918216117

163. Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M.M.A., Pronk, J.T., Slijper, M., Heck, A.J.R., Elsemman, I.E., Rodriguez Prado, A., Grigaitis, P., Garcia Albornoz, M., Harman, V., Holman, S.W., van Heerden, J., Bruggeman, F.J., Bisschops, M.M.M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., Teusink, B., Huh, K., W., Falvo, V., J., Gerke, C., L., Carroll, S., A., Howson, W., R., Weissman, S., J., O'Shea, K., E., Prokisch, H., Scharfe, C., Camp, D.G., Xiao, W., David, L., Andreoli, C., Monroe, M.E., Moore, R.J., Gritsenko, M.A., Kozany, C., Hixson, K.K., Mottaz, H.M., Zischka, H., Ueffing, M., Herman, Z.S., Davis, R.W., Meitinger, T., Oefner, P.J., Smith, R.D., Steinmetz, L.M., Malina, C., Yu, R., Bjorkeroth, J., Kerkhoven, E.J., Nielsen, J., Di Bartolomeo, F., Malina, C., Campbell, K., Mormino, M., Fuchs, J., Vorontsov, E., Gustafsson, C.M., Nielsen, J., Lao-Martil, D., Verhagen, K.J.A., Schmitz, J.P.J., Teusink, B., Wahl, S.A., van Riel, N.A.W., Lu, H., Kerkhoven, E.J., Nielsen, J., Xia, J., Sánchez, B.J., Chen, Y., Nielsen, J.: Proteome allocations change linearly with the specific growth rate of *Saccharomyces cerevisiae* under glucose limitation. Nat. Commun. 2, 1–12 (2022). https://doi.org/10.1038/s41467-022-30513-2

164. Creasy, D.M., Cottrell, J.S.: Unimod: Protein modifications for mass spectrometry. Proteomics. 4, 1534–1536 (2004). https://doi.org/10.1002/pmic.200300744

165. Weinert, B.T., Iesmantavicius, V., Moustafa, T., Schölz, C., Wagner, S.A., Magnes, C., Zechner, R., Choudhary, C.: Acetylation dynamics and stoichiometry in *Saccharomyces cerevisiae*. Mol. Syst. Biol. 10, 1–12 (2014). https://doi.org/10.1002/msb.134766

166. Pang, C.N.I., Gasteiger, E., Wilkins, M.R.: Identification of arginine- and lysine-methylation in the proteome of *Saccharomyces cerevisiae* and its functional implications. BMC Genomics. 11, 1–16 (2010). https://doi.org/10.1186/1471-2164-11-92

167. Chen, Y., Nielsen, J.: Flux control through protein phosphorylation in yeast. FEMS Yeast Res. 16, 1–14 (2016). https://doi.org/10.1093/femsyr/fow096

168. Minguez, P., Parca, L., Diella, F., Mende, D.R., Kumar, R., Helmer-Citterich, M., Gavin, A.C., Van Noort, V., Bork, P.: Deciphering a global network of functionally associated post-translational modifications. Mol. Syst. Biol. 8, 1–14 (2012). https://doi.org/10.1038/msb.2012.31

169. Swaney, D.L., Beltrao, P., Starita, L., Guo, A., Rush, J., Fields, S., Krogan, N.J., Villén, J.: Global analysis of phosphorylation and ubiquitylation crosstalk in protein degradation. Nat. Methods. 10, (2013). https://doi.org/doi:10.1038/nmeth.2519

170. Venne, A.S., Kollipara, L., Zahedi, R.P.: The next level of complexity: Crosstalk of posttranslational modifications. Proteomics. 14, 513–524 (2014). https://doi.org/10.1002/pmic.201300344

171. Huang, Y., Xu, B., Zhou, X., Li, Y., Lu, M., Jiang, R., Li, T.: Systematic Characterization and Prediction of Post-Translational Modification Cross-Talk. Mol. Cell. Proteomics. 761–770 (2015). https://doi.org/10.1074/mcp

172. Separovich, R.J., Wong, M.W.M., Chapman, T.R., Slavich, E., Hamey, J.J., Wilkins, M.R.: Post-

R

translational modification analysis of *Saccharomyces cerevisiae* histone methylation enzymes reveals phosphorylation sites of regulatory potential. J. Biol. Chem. 296, 100192 (2021). https://doi.org/10.1074/JBC.RA120.015995

173. Hamey, J.J., Nguyen, A., Wilkins, M.R.: Discovery of arginine methylation, phosphorylation, and their co-occurrence in condensate-associated proteins in *Saccharomyces cerevisiae*. J. Proteome Res. 20, 2420–2434 (2021). https://doi.org/10.1021/acs.jproteome.0c00927

174. Tamayo Rojas, S.A., Schmidl, S., Boles, E., Oreb, M.: Glucose-induced internalization of the *S. cerevisiae* galactose permease Gal2 is dependent on phosphorylation and ubiquitination of its aminoterminal cytoplasmic tail. FEMS Yeast Res. 21, 1–10 (2021). https://doi.org/10.1093/femsyr/foab019

175. Separovich, R.J., Wilkins, M.R.: Ready, SET, Go: Post-translational regulation of the histone lysine methylation network in budding yeast. J. Biol. Chem. 297, 100939 (2021). https://doi.org/10.1016/j.jbc.2021.100939

176. Smith, D.L., Erce, M.A., Lai, Y.W., Tomasetig, F., Hart-Smith, G., Hamey, J.J., Wilkins, M.R.: Crosstalk of phosphorylation and arginine methylation in disordered SRGG repeats of *Saccharomyces cerevisiae* fibrillarin and its association with nucleolar localization. J. Mol. Biol. 432, 448–466 (2020). https://doi.org/10.1016/j.jmb.2019.11.006

177. Piehowski, P.D., Petyuk, V.A., Orton, D.J., Xie, F., Moore, R.J., Ramirez-Restrepo, M., Engel, A., Lieberman, A.P., Albin, R.L., Camp, D.G., Smith, R.D., Myers, A.J.: Sources of technical variability in quantitative LC-MS proteomics: Human brain tissue sample analysis. J. Proteome Res. 12, 2128–2137 (2013). https://doi.org/10.1021/pr301146m

178. Aebersold, R., Mann, M.: Mass spectrometry-based proteomics. 422, (2003)

179. Cox, J., Mann, M.: Quantitative, high-resolution proteomics for data-driven systems biology. Annu. Rev. Biochem. 80, 273–299 (2011). https://doi.org/10.1146/annurev-biochem-061308-093216

180. Oliveira, A.P., Sauer, U.: The importance of post-translational modifications in regulating *Saccharomyces cerevisiae* metabolism. FEMS Yeast Res. 12, 104–117 (2012). https://doi.org/10.1111/j.1567-1364.2011.00765.x

181. Richards, A.L., Hebert, A.S., Ulbrich, A., Bailey, D.J., Coughlin, E.E., Westphall, M.S., Coon, J.J.: One-hour proteome analysis in yeast. Nat. Protoc. 10, 701–714 (2015). https://doi.org/10.1038/nprot.2015.040

182. Walsh, C.T., Garneau-Tsodikova, S., Gatto, G.J.: Protein posttranslational modifications: The chemistry of proteome diversifications. Angew. Chemie - Int. Ed. 44, 7342–7372 (2005). https://doi.org/10.1002/anie.200501023

183. Smith, L.M., Kelleher, N.L.: Proteoform: A single term describing protein complexity. Nat. Methods. 10, 186–187 (2013). https://doi.org/10.1038/nmeth.2369

184. Daran-Lapujade, P., Jansen, M.L.A., Daran, J.M., Van Gulik, W., De Winde, J.H., Pronk, J.T.: Role of transcriptional regulation in controlling fluxes in central carbon metabolism of *Saccharomyces cerevisiae*: A chemostat culture study. J. Biol. Chem. 279, 9125–9138 (2004). https://doi.org/10.1074/jbc.M309578200

185. Nielsen, J.: It Is All about Metabolic Fluxes. J. Bacteriol. 185, 7031–7035 (2003). https://doi.org/10.1128/JB.185.24.7031–7035.2003

186. Gerosa, L., Sauer, U.: Regulation and control of metabolic fluxes in microbes. Curr. Opin. Biotechnol. 22, 566–575 (2011). https://doi.org/10.1016/j.copbio.2011.04.016

187. Oughtred, R., Stark, C., Breitkreutz, B.J., Rust, J., Boucher, L., Chang, C., Kolas, N., O'Donnell, L., Leung, G., McAdam, R., Zhang, F., Dolma, S., Willems, A., Coulombe-Huntington, J., Chatr-Aryamontri, A., Dolinski, K., Tyers, M.: The BioGRID interaction database: 2019 update. Nucleic Acids Res. 47,

D529–D541 (2019). https://doi.org/10.1093/nar/gky1079

188. Swaney, D.L., Villén, J.: Proteomic Analysis of Protein Posttranslational Modifications by Mass Spectrometry. 207–210 (2016). https://doi.org/10.1101/pdb.top077743

189. Ledesma, L., Sandoval, E., Cruz-martı, U., Moreno-a, P., Mejı, S., Garcı, E.: YAAM: Yeast Amino Acid Modifications database. Database. 1–8 (2018). https://doi.org/10.1093/database/bay033

190. Rowland, E.A., Snowden, C.K., Cristea, I.M.: Protein lipoylation: an evolutionarily conserved metabolic regulator of health and disease. Curr. Opin. Chem. Biol. 42, 76–85 (2018). https://doi.org/10.1016/j.cbpa.2017.11.003

191. Baldi, N., Dykstra, J.C., Luttik, M.A.H., Pabst, M., Wu, L., Benjamin, K.R., Vente, A., Pronk, J.T., Mans, R.: Functional expression of a bacterial α-ketoglutarate dehydrogenase in the cytosol of *Saccharomyces cerevisiae*. Metab. Eng. 56, 190–197 (2019). https://doi.org/10.1016/j.ymben.2019.10.001

192. The Uniprot Consortium: Uniprot: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169 (2017)

193. Fresques, T., Niles, B., Aronova, S., Mogri, H., Rakhshandehroo, T., Powers, T.: Regulation of ceramide synthase by casein kinase 2-dependent phosphorylation in *Saccharomyces cerevisiae*. J. Biol. Chem. 290, 1395–1403 (2015). https://doi.org/10.1074/jbc.M114.621086

194. Nakic, Z.R., Seisenbacher, G., Posas, F., Sauer, U.: Untargeted metabolomics unravels functionalities of phosphorylation sites in *Saccharomyces cerevisiae*. BMC Syst. Biol. 10, 1–15 (2016). https://doi.org/10.1186/s12918-016-0350-8

195. Oliveira, A.P., Ludwig, C., Zampieri, M., Weisser, H., Aebersold, R., Sauer, U., Papinski, D., Schuschnig, M., Reiter, W., Wilhelm, L., Barnes, C.A., Maiolica, A., Hansmann, I., Pfaffenwimmer, T., Kijanska, M., Stoffel, I., Lee, S.S., Brezovich, A., Lou, J.H., Turk, B.E., Aebersold, R., Ammerer, G., Peter, M., Kraft, C., Narita, T., Weinert, B.T., Choudhary, C., Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., Kouzarides, T., Winter, D.L., Abeygunawardena, D., Hart-Smith, G., Erce, M.A., Wilkins, M.R., Lawrence, M.B.D., Coutin, N., Choi, J.K., Martin, B.J.E., Irwin, N.A.T., Young, B., Loewen, C., Howe, L.J., Chen, Y., Nielsen, J., Huang, J., Luo, Z., Ying, W., Cao, Q., Huang, H., Dong, J., Wu, Q., Zhao, Y., Qian, X., Dai, J., Fresques, T., Niles, B., Aronova, S., Mogri, H., Rakhshandehroo, T., Powers, T., Liu, C., Yu, H., Li, L., Raguz Nakic, Z., Seisenbacher, G., Posas, F., Sauer, U.: Lysine methylation modulates the protein-protein interactions of yeast cytochrome *C* Cyc1p. PLoS Genet. 15, 1–15 (2015). https://doi.org/10.1126/scisignal.2005768

196. Papinski, D., Schuschnig, M., Reiter, W., Wilhelm, L., Barnes, C.A., Maiolica, A., Hansmann, I., Pfaffenwimmer, T., Kijanska, M., Stoffel, I., Lee, S.S., Brezovich, A., Lou, J.H., Turk, B.E., Aebersold, R., Ammerer, G., Peter, M., Kraft, C.: Early steps in autophagy depend on direct phosphorylation of Atg9 by the Atg1 kinase. Mol. Cell. 53, 471–483 (2014). https://doi.org/10.1016/j.molcel.2013.12.011

197. Memisoglu, G., Lanz, M.C., Eapen, V. V., Jordan, J.M., Lee, K., Smolka, M.B., Haber, J.E.: Mec1ATR Autophosphorylation and Ddc2ATRIP Phosphorylation Regulates DNA Damage Checkpoint Signaling. Cell Rep. 1090–1102 (2019). https://doi.org/10.1016/j.celrep.2019.06.068

198. Allfrey, V.G., Faulkner, R., Mirsky, A.E.: Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. Proc. Natl. Acad. Sci. U. S. A. 51, 786–94 (1964)

199. Lawrence, M.B.D., Coutin, N., Choi, J.K., Martin, B.J.E., Irwin, N.A.T., Young, B., Loewen, C., Howe, L.J.: Histone acetylation, not stoichiometry, regulates linker histone binding in *Saccharomyces cerevisiae*. Genetics. 207, 347–355 (2017). https://doi.org/10.1534/genetics.117.1132

200. Tessarz, P., Santos-Rosa, H., Robson, S.C., Sylvestersen, K.B., Nelson, C.J., Nielsen, M.L., Kouzarides, T.: Glutamine methylation in histone H2A is an RNA-polymerase-I-dedicated modification. Nature. 505, 564–568 (2014). https://doi.org/10.1038/nature12819

R

201. Tripodi, F., Nicastro, R., Reghellin, V., Coccetti, P.: Post-translational modifications on yeast carbon metabolism: Regulatory mechanisms beyond transcriptional control. Biochim. Biophys. Acta - Gen. Subj. 1850, 620–627 (2015). https://doi.org/10.1016/j.bbagen.2014.12.010

202. Yates, J.R., Speicher, S., Griffin, P.R., Hunkapiller, T.: Peptide mass maps: A highly informative approach to protein identification, (1993)

203. Mann, M., Wilm, M.: Error-tolerant identification of peptides in sequence databases by peptide sequence tags. Anal. Chem. 66, 4390–4399 (1994). https://doi.org/10.1021/ac00096a002

204. Washburn, M.P., Wolters, D., Yates, J.R.: Large-scale analysis of the yeast proteome by multidimensional protein identification technology. Nat. Biotechnol. 19, 242–247 (2001). https://doi.org/10.1038/85686

205. Olsen, J. V, Godoy, L.M.F. De, Li, G., Macek, B., Mortensen, P., Pesch, R., Makarov, A., Lange, O., Horning, S., Mann, M.: Parts per million mass accuracy on an Orbitrap mass spectrometer via lock mass injection into a C-trap (2005). https://doi.org/10.1074/mcp.T500030-MCP200

206. Chick, J.M., Kolippakkam, D., Nusinow, D.P., Zhai, B., Rad, R., Huttlin, E.L., Gygi, S.P.: A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. 33, (2015). https://doi.org/10.1038/nbt.3267

207. Langella, O., Valot, B., Balliau, T., Blein-Nicolas, M., Bonhomme, L., Zivy, M.: X!TandemPipeline: A tool to manage sequence redundancy for protein inference and phosphosite identification. J. Proteome Res. 16, 494–503 (2017). https://doi.org/10.1021/acs.jproteome.6b00632

208. Ma, C.W.M., Lam, H.: Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. J. Proteome Res. 13, 2262–2271 (2014). https://doi.org/10.1021/pr401006g

209. Na, S., Bandeira, N., Paek, E.: Fast multi-blind modification search through tandem mass spectrometry. Mol. Cell. Proteomics. 11, M111.010199 (2012). https://doi.org/10.1074/mcp.m111.010199

210. Ma, B., Zhang, K., Hendrie, C., Liang, C., Li, M., Doherty-Kirby, A., Lajoie, G.: PEAKS: Powerful software for peptide *de novo* sequencing by tandem mass spectrometry. Rapid Commun. Mass Spectrom. 17, 2337–2342 (2003). https://doi.org/10.1002/rcm.1196

211. Bateman, A.: UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515 (2019). https://doi.org/10.1093/nar/gky1049

212. Han, X., He, L., Xin, L., Shan, B., Ma, B.: PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. J. Proteome Res. 10, 2930–2936 (2011). https://doi.org/10.1021/pr200153k

213. Li, Q., Shortreed, M.R., Wenger, C.D., Frey, B.L., Schaffer, L. V., Scalf, M., Smith, L.M.: Global post-translational modification discovery. J. Proteome Res. 16, 1383–1390 (2017). https://doi.org/10.1021/acs.jproteome.6b00034

214. Solntsev, S.K., Shortreed, M.R., Frey, B.L., Smith, L.M.: Enhanced global post-translational modification discovery with MetaMorpheus. J. Proteome Res. 17, 1844–1851 (2018). https://doi.org/10.1021/acs.jproteome.7b00873

215. He, L., Han, X., Man, B.: *De novo* sequencing with limited number of post-translational modifications per peptide. J. Bioinform. Comput. Biol. 11, 1350007 (2013)

216. Yang, H., Chi, H., Zhou, W.J., Zeng, W.F., He, K., Liu, C., Sun, R.X., He, S.M.: Open-pNovo: *De novo* peptide sequencing with thousands of protein modifications. J. Proteome Res. 16, 645–654 (2017). https://doi.org/10.1021/acs.jproteome.6b00716

217. Tabb, D.L., Saraf, A., Yates, J.R.: GutenTag: high-throughput sequence tagging via an empirically derived

fragmentation model. Anal. Chem. 75, 6415–6421 (2003). https://doi.org/10.1038/jid.2014.371

218.   Tanner, S., Shu, H., Frank, A., Wang, L.C., Zandi, E., Mumby, M., Pevzner, P.A., Bafna, V.: InsPecT: Identification of posttranslationally modified peptides from tandem mass spectra. Anal. Chem. 77, 4626–4639 (2005). https://doi.org/10.1021/ac050102d

219.   Chi, H., Liu, C., Yang, H., Zeng, W., Wu, L., Zhou, W., Wang, R., Niu, X., Ding, Y., Zhang, Y., Wang, Z., Chen, Z., Sun, R., Liu, T., Tan, G., Dong, M., Xu, P., Zhang, P., He, S.: Comprehensive identification of peptides in tandem mass spectra using an efficient open search engine. 36, (2018). https://doi.org/10.1038/nbt.4236

220.   Kong, A.T., Leprevost, F. V, Avtonomov, D.M., Mellacheruvu, D., Nesvizhskii, A.I.: MSFragger : ultrafast and comprehensive peptide identification in mass spectrometry – based proteomics. 14, (2017). https://doi.org/10.1038/nmeth.4256

221.   Devabhaktuni, A., Lin, S., Zhang, L., Swaminathan, K., Gonzalez, C.G., Olsson, N., Pearlman, S.M., Rawson, K., Elias, J.E.: TagGraph reveals vast protein modification landscapes from large tandem mass spectrometry datasets. Nat. Biotechnol. 37, (2019). https://doi.org/10.1038/s41587-019-0067-5

222.   Bern, M., Kil, Y.J., Becker, C.: Byonic: advanced peptide and protein identification software, Curr Protoc Bioinformatics.   Curr   Protoc   Bioinforma.   Chapter   13.   Unit13   (2012). https://doi.org/10.1002/0471250953.bi1320s40.Byonic

223.   Frewen, B.E., Merrihew, G.E., Wu, C.C., Noble, W.S., MacCoss, M.J.: Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal. Chem. 78, 5678–5684 (2006). https://doi.org/10.1021/ac060279n

224.   Griss, J.: Spectral library searching in proteomics. Proteomics. 16, 729–740 (2016). https://doi.org/10.1002/pmic.201500296

225.   Lam, H., Deutsch, E.W., Eddes, J.S., Eng, J.K., King, N., Stein, S.E., Aebersold, R.: Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 7, 655–667 (2007). https://doi.org/10.1002/pmic.200600625

226.   Zhang, X., Li, Y., Shao, W., Lam, H.: Understanding the improved sensitivity of spectral library searching over sequence database searching in proteomics data analysis. Proteomics. 11, 1075–1085 (2011). https://doi.org/10.1002/pmic.201000492

227.   Bittremieux, W., Meysman, P., Noble, W.S., Laukens, K.: Fast open modification spectral library searching through approximate nearest neighbor indexing. J. Proteome Res. 17, 3463–3474 (2018). https://doi.org/10.1021/acs.jproteome.8b00359

228.   David, M., Fertin, G., Rogniaux, H., Tessier, D.: SpecOMS: A full open modification search method performing all-to-all spectra comparisons within minutes. J. Proteome Res. 16, 3030–3038 (2017). https://doi.org/10.1021/acs.jproteome.7b00308

229.   Solntsev, S.K., Shortreed, M.R., Frey, B.L., Smith, L.M.: Enhanced global post-translational modification discovery   with   MetaMorpheus.   J.   Proteome   Res.   17,   1844–1851   (2018). https://doi.org/10.1021/acs.jproteome.7b00873

230.   Yu, F., Li, N., Yu, W.: PIPI : PTM-Invariant peptide identification using coding method. 4435, (2016). https://doi.org/10.1021/acs.jproteome.6b00485

231.   Jiang, T., Zhou, X., Taghizadeh, K., Dong, M., Dedon, P.C.: N-formylation of lysine in histone proteins as a secondary modification arising from oxidative DNA damage. Proc. Natl. Acad. Sci. 104, 60–65 (2007). https://doi.org/10.1073/pnas.0606775103

232.   Zheng, S., Doucette, A.A.: Preventing N- and O-formylation of proteins when incubated in concentrated formic acid. Proteomics. 16, 1059–1068 (2016). https://doi.org/10.1002/pmic.201500366

R

233. Kollipara, L., Zahedi, R.P.: Protein carbamylation: *In vivo* modification or *in vitro* artefact? Proteomics. 13, 941–944 (2013). https://doi.org/10.1002/pmic.201200452

234. Boja, E.S., Fales, H.M.: Overalkylation of a protein digest with iodoacetamide. Anal. Chem. 73, 3576–3582 (2001). https://doi.org/10.1021/ac0103423

235. Müller, T., Winter, D.: Systematic evaluation of protein reduction and alkylation reveals massive unspecific side effects by iodine-containing reagents. Mol. Cell. Proteomics. 16, 1173–1187 (2017). https://doi.org/10.1074/mcp.m116.064048

236. Hao, P., Ren, Y., Alpert, A.J., Sze, S.K.: Detection, evaluation and minimization of nonenzymatic deamidation in proteomic sample preparation. Mol. Cell. Proteomics. 10, O111.009381 (2011). https://doi.org/10.1074/mcp.o111.009381

237. Liu, S., Zhang, C., Campbell, J.L., Zhang, H., Yeung, K.K., Han, V.K.M., Lajoie, G.A.: Formation of phosphopeptide-metal ion complexes in liquid chromatography / electrospray mass spectrometry and their influence on phosphopeptide detection. 2747–2756 (2005). https://doi.org/10.1002/rcm.2105

238. Purwaha, P., Silva, L.P., Hawke, D.H., Weinstein, J.N., Lorenzi, P.L.: An artifact in LC-MS/MS measurement of glutamine and glutamic acid: In-source cyclization to pyroglutamic acid. Anal. Chem. 86, 5633–5637 (2014). https://doi.org/10.1021/ac501451v

239. Stadtman, E.R., Van Remmen, H., Richardson, A., Wehr, N.B., Levine, R.L.: Methionine oxidation and aging. Biochim. Biophys. Acta - Proteins Proteomics. 1703, 135–140 (2005). https://doi.org/10.1016/j.bbapap.2004.08.010

240. Carapito, C., Klemm, C., Aebersold, R., Domon, B.: Systematic LC-MS analysis of labile post-translational modifications in complex mixtures. 2608–2614 (2009)

241. Olsen, J. V, Mann, M.: Status of large-scale analysis of post-translational modifications by mass spectrometry *. 3444–3452 (2013). https://doi.org/10.1074/mcp.O113.034181

242. Grassl, J., Westbrook, J.A., Robinson, A., Borén, M., Dunn, M.J., Clyne, R.K.: Preserving the yeast proteome from sample degradation. Proteomics. 9, 4616–4626 (2009). https://doi.org/10.1002/pmic.200800945

243. Herbert, B., Hopwood, F., Oxley, D., McCarthy, J., Laver, M., Grinyer, J., Goodall, A., Williams, K., Castagna, A., Righetti, P.G.: β-elimination: An unexpected artefact in proteome analysis. Proteomics. 3, 826–831 (2003). https://doi.org/10.1002/pmic.200300414

244. Li, X., Gerber, S.A., Rudner, A.D., Beausoleil, S.A., Haas, W., Villén, J., Elias, J.E., Gygi, S.P.: Large-scale phosphorylation analysis of alpha-factor-arrested *Saccharomyces cerevisiae*. J. Proteome Res. 6, 1190–1197 (2007). https://doi.org/10.1021/pr060559j

245. Han, G., Ye, M., Zhou, H., Jiang, X., Feng, S., Jiang, X., Tian, R., Wan, D., Zou, H., Gu, J.: Large-scale phosphoproteome analysis of human liver tissue by enrichment and fractionation of phosphopeptides with strong anion exchange chromatography. Proteomics. 8, 1346–1361 (2008). https://doi.org/10.1002/pmic.200700884

246. Soares, N.C., Blackburn, J.M.: Mass spectrometry targeted assays as a tool to improve our understanding of post-translational modifications in pathogenic bacteria. Front. Microbiol. 7, 1–5 (2016). https://doi.org/10.3389/fmicb.2016.01216

247. Adachi, J., Narumi, R., Tomonaga, T.: Targeted phosphoproteome analysis using Selected/Multiple Reaction Monitoring (SRM/MRM). In: Reinders, J. (ed.) Proteomics in Systems Biology: Methods and Protocols. pp. 87–100. Springer New York, New York, NY (2016)

248. Arsova, B., Watt, M., Usadel, B.: Monitoring of plant protein post-translational modifications using targeted proteomics. Front. Plant Sci. 9, 1–9 (2018). https://doi.org/10.3389/fpls.2018.01168

249. Allmer, J.: Existing bioinformatics tools for the quantitation of post-translational modifications. Amino Acids. 42, 129–138 (2012). https://doi.org/10.1007/s00726-010-0614-3

250. Guan, X., Rastogi, N., Parthun, M.R., Freitas, M.A.: Discovery of histone modification crosstalk networks by stable isotope labeling of amino acids in cell culture mass spectrometry (SILAC MS). Mol. Cell. Proteomics. 12, 2048–2059 (2013). https://doi.org/10.1074/mcp.M112.026716

251. Swaney, D.L., Wenger, C.D., Coon, J.J.: Value of using multiple proteases for large-scale mass spectrometry-based proteomics research articles. 1323–1329 (2010)

252. Henriksen, P., Wagner, S.A., Weinert, B.T., Sharma, S., Bačinskaja, G., Rehman, M., Juffer, A.H., Walther, T.C., Lisby, M., Choudhary, C.: Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. Mol. Cell. Proteomics. 11, 1510–1522 (2012). https://doi.org/10.1074/mcp.M112.017251

253. Giansanti, P., Tsiatsiani, L., Low, T.Y., Heck, A.J.R.: Six alternative proteases for mass spectrometry-based proteomics beyond trypsin. Nat. Protoc. 11, 993–1006 (2016). https://doi.org/10.1038/nprot.2016.057

254. Glatter, T., Ludwig, C., Ahrné, E., Aebersold, R., Heck, A.J.R., Schmidt, A.: Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J. Proteome Res. 11, 5145–5156 (2012). https://doi.org/10.1021/pr300273g

255. Perchey, R.T., Tonini, L., Tosolini, M., Fournié, J.J., Lopez, F., Besson, A., Pont, F.: PTMselect: optimization of protein modifications discovery by mass spectrometry. Sci. Rep. 9, 5–11 (2019). https://doi.org/10.1038/s41598-019-40873-3

256. Grünwald-Gruber, C., Altmann, F.: LC-MS Analysis of (Glyco-)Proteins of *Pichia pastoris*. In: Gasser B., Mattanovich D. (eds) Recombinant Protein Production in Yeast. Methods in Molecular Biology,. (2019)

257. Wu, S.L., Hühmer, A.F.R., Hao, Z., Karger, B.L.: On-line LC-MS approach combining collision-induced dissociation (CID), electron-transfer dissociation (ETD), and CID of an isolated charge-reduced species for the trace-level characterization of proteins with post-translational modifications. J. Proteome Res. 6, 4230–4244 (2007). https://doi.org/10.1021/pr070313u

258. Elviri, L.: ETD and ECD Mass spectrometry fragmentation for the characterization of protein post translational modifications. Tandem Mass Spectrom. - Appl. Princ. (2012). https://doi.org/10.5772/35277

259. Wuhrer, M.: Glycomics using mass spectrometry. Glycoconj. J. 30, 11–22 (2013). https://doi.org/10.1007/s10719-012-9376-3

260. Röst, H.L., Rosenberger, G., Navarro, P., Gillet, L., Miladinoviä, S.M., Schubert, O.T., Wolski, W., Collins, B.C., Malmström, J., Malmström, L., Aebersold, R.: OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat. Biotechnol. 32, 219–223 (2014). https://doi.org/10.1038/nbt.2841

261. Rosenberger, G., Liu, Y., Röst, H.L., Ludwig, C., Buil, A., Bensimon, A., Soste, M., Spector, T.D., Dermitzakis, E.T., Collins, B.C., Malmström, L., Aebersold, R.: Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. Nat. Biotechnol. 35, 781–788 (2017). https://doi.org/10.1038/nbt.3908

262. Meyer, J.G., Mukkamalla, S., Steen, H., Nesvizhskii, A.I., Gibson, B.W., Schilling, B.: PIQED: Automated identification and quantification of protein modifications from DIA-MS data. Nat. Methods. 14, 646–647 (2017). https://doi.org/10.1038/nmeth.4334

263. Chen, Y., Wang, Y., Nielsen, J.: Systematic inference of functional phosphorylation events in yeast metabolism. Bioinformatics. 33, 1995–2001 (2017). https://doi.org/10.1093/bioinformatics/btx110

R

264. Zahedi, R.P.: Joining forces: studying multiple post-translational modifications to understand dynamic disease mechanisms. Expert Rev. Proteomics. 13, 1055–1057 (2016). https://doi.org/10.1080/14789450.2016.1231577

265. Telekawa, C., Boisvert, F.M., Bachand, F.: Proteomic profiling and functional characterization of post-translational modifications of the fission yeast RNA exosome. Nucleic Acids Res. 46, 11169–11183 (2018). https://doi.org/10.1093/nar/gky915

266. Šoštarić, N., O, F.J., Giansanti, P., R Heck, A.J., Gavin, A.-C., van Noort, V.: Effects of acetylation and phosphorylation on subunit interactions in three large eukaryotic complexes. 10, 1–63 (2019)

267. Duan, G., Walther, D.: The Roles of Post-translational Modifications in the Context of Protein Interaction Networks. PLoS Comput. Biol. 11, 1–23 (2015). https://doi.org/10.1371/journal.pcbi.1004049

268. Yachie, N., Saito, R., Sugiyama, N., Tomita, M., Ishihama, Y.: Integrative features of the yeast phosphoproteome and protein-protein interaction map. PLoS Comput. Biol. 7, (2011). https://doi.org/10.1371/journal.pcbi.1001064

269. Ning, K., Fermin, D., Nesvizhskii, A.I.: Computational analysis of unassigned high-quality MS/MS spectra in proteomic data sets. Proteomics. 10, 2712–2718 (2010). https://doi.org/10.1002/pmic.200900473

270. Nesvizhskii, A.I., Roos, F.F., Grossmann, J., Vogelzang, M., Eddes, J.S., Gruissem, W., Baginsky, S., Aebersold, R.: Dynamic spectrum quality assessment and iterative computational analysis of shotgun proteomic data: Toward more efficient identification of post-translational modifications, sequence polymorphisms, and novel peptides. Mol. Cell. Proteomics. 5, 652–670 (2006). https://doi.org/10.1074/mcp.M500319-MCP200

271. Shteynberg, D., Nesvizhskii, A.I., Moritz, R.L., Deutsch, E.W.: Combining results of multiple search engines in proteomics. Mol. Cell. Proteomics. 12, 2383–2393 (2013). https://doi.org/10.1074/mcp.R113.027797

272. Kanshin, E., Tyers, M., Thibault, P.: Sample collection method bias effects in quantitative phosphoproteomics. J. Proteome Res. 14, 2998–3004 (2015). https://doi.org/10.1021/acs.jproteome.5b00404

273. Suttapitugsakul, S., Xiao, H., Smeekens, J., Wu, R.: Evaluation and optimization of reduction and alkylation methods to maximize peptide identification with MS-based proteomics. Mol. Biosyst. 13, 2574–2582 (2017). https://doi.org/10.1039/c7mb00393e

274. Güray, M.Z., Zheng, S., Doucette, A.A., Niu, L., Zhang, H., Wu, Z., Wang, Y., Liu, H., Wu, X., Wang, W., Simpson, D.M., Beynon, R.J., Jiang, L., He, L., Fountoulakis, M.: Mass spectrometry of intact proteins reveals +98 u chemical artifacts following precipitation in acetone. J. Proteome Res. 16, 444–450 (2018). https://doi.org/10.1021/acs.jproteome.6b00841

275. Santa, C., Anjo, S.I., Manadas, B.: Protein precipitation of diluted samples in SDS-containing buffer with acetone leads to higher protein recovery and reproducibility in comparison with TCA/acetone approach. Proteomics. 16, 1847–1851 (2016). https://doi.org/10.1002/pmic.201600024

276. Jiang, L., He, L., Fountoulakis, M.: Comparison of protein precipitation methods for sample preparation prior to proteomic analysis. J. Chromatogr. A. 1023, 317–320 (2004). https://doi.org/10.1016/j.chroma.2003.10.029

277. Simpson, D.M., Beynon, R.J.: Acetone precipitation of proteins and the modification of peptides. J. Proteome Res. 9, 444–450 (2010). https://doi.org/10.1021/pr900806x

278. Crowell, A.M.J., Wall, M.J., Doucette, A.A.: Maximizing recovery of water-soluble proteins through acetone precipitation. Anal. Chim. Acta. 796, 48–54 (2013). https://doi.org/10.1016/j.aca.2013.08.005

279. Wiśniewski, J.R., Zougman, A., Nagaraj, N., Mann, M.: Universal sample preparation method for proteome analysis. Nat. Methods. 6, 359–362 (2009). https://doi.org/10.1038/nmeth.1322

280. Manza, L.L., Stamer, S.L., Ham, A.J.L., Codreanu, S.G., Liebler, D.C.: Sample preparation and digestion for proteomic analyses using spin filters. Proteomics. 5, 1742–1745 (2005). https://doi.org/10.1002/pmic.200401063

281. Lin, Z., Ren, Y., Shi, Z., Zhang, K., Yang, H., Liu, S., Hao, P.: Evaluation and minimization of nonspecific tryptic cleavages in proteomic sample preparation. Rapid Commun. Mass Spectrom. 34, (2020). https://doi.org/10.1002/rcm.8733

282. Cañas, B., Piñeiro, C., Calvo, E., López-Ferrer, D., Gallardo, J.M.: Trends in sample preparation for classical and second generation proteomics. J. Chromatogr. A. 1153, 235–258 (2007). https://doi.org/10.1016/j.chroma.2007.01.045

283. Klont, F., Bras, L., Wolters, J.C., Ongay, S., Bischoff, R., Halmos, G.B., Horvatovich, P.: Assessment of sample preparation bias in mass spectrometry-based proteomics. Anal. Chem. 90, 5405–5413 (2018). https://doi.org/10.1021/acs.analchem.8b00600

284. Lenčo, J., Khalikova, M.A., Švec, F.: Dissolving Peptides in 0.1% Formic Acid Brings Risk of Artificial Formylation. J. Proteome Res. 19, 993–999 (2020). https://doi.org/10.1021/acs.jproteome.9b00823

285. Kollipara, L., Zahedi, R.P.: Protein carbamylation: *In vivo* modification or *in vitro* artefact? Proteomics. 13, 941–944 (2013). https://doi.org/10.1002/pmic.201200452

286. Kuznetsova, K.G., Levitsky, L.I., Pyatnitskiy, M.A., Ilina, I.Y., Bubis, J.A., Solovyeva, E.M., Zgoda, V.G., Gorshkov, M. V., Moshkovskii, S.A.: Cysteine alkylation methods in shotgun proteomics and their possible effects on methionine residues. J. Proteomics. 231, 104022 (2021). https://doi.org/10.1016/j.jprot.2020.104022

287. Wiśniewski, J.R., Zettl, K., Pilch, M., Rysiewicz, B., Sadok, I.: 'Shotgun' proteomic analyses without alkylation of cysteine. Anal. Chim. Acta. 1100, 131–137 (2020). https://doi.org/10.1016/j.aca.2019.12.007

288. Entian, K.-D., Kötter, P.: 25 Yeast genetic strain and plasmid collections. Methods Microbiol. 36, 629–666 (2007). https://doi.org/10.1016/S0580-9517(06)36025-4

289. Verduyn, C., Postma, E., Scheffers, W.A., Van Dijken, J.P.: Effect of benzoic acid on metabolic fluxes in yeasts: A continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast. 8, 501–517 (1992). https://doi.org/10.1002/yea.320080703

290. Köcher, T., Pichler, P., Swart, R., Mechtler, K.: Analysis of protein mixtures from whole-cell extracts by single-run nanolc-ms/ms using ultralong gradients. Nat. Protoc. 7, 882–890 (2012). https://doi.org/10.1038/nprot.2012.036

291. Kleikamp, H.B.C., Pronk, M., Tugui, C., Guedes da Silva, L., Abbas, B., Lin, Y.M., van Loosdrecht, M.C.M., Pabst, M.: Database-independent *de novo* metaproteomics of complex microbial communities. Cell Syst. 12, 375-383.e5 (2021). https://doi.org/10.1016/j.cels.2021.04.003

292. Mellacheruvu, D., Wright, Z., Couzens, A.L., Lambert, J.P., St-Denis, N.A., Li, T., Miteva, Y. V., Hauri, S., Sardiu, M.E., Low, T.Y., Halim, V.A., Bagshaw, R.D., Hubner, N.C., Al-Hakim, A., Bouchard, A., Faubert, D., Fermin, D., Dunham, W.H., Goudreault, M., Lin, Z.Y., Badillo, B.G., Pawson, T., Durocher, D., Coulombe, B., Aebersold, R., Superti-Furga, G., Colinge, J., Heck, A.J.R., Choi, H., Gstaiger, M., Mohammed, S., Cristea, I.M., Bennett, K.L., Washburn, M.P., Raught, B., Ewing, R.M., Gingras, A.C., Nesvizhskii, A.I.: The CRAPome: A contaminant repository for affinity purification-mass spectrometry data. Nat. Methods. 10, 730–736 (2013). https://doi.org/10.1038/nmeth.2557

293. Chambers, M.C., MacLean, B., Burke, R., Amodei, D., Ruderman, D.L., Neumann, S., Gatto, L., Fischer, B., Pratt, B., Egertson, J., Hoff, K., Kessner, D., Tasman, N., Shulman, N., Frewen, B., Baker, T.A.,

Brusniak, M.Y., Paulse, C., Creasy, D., Flashner, L., Kani, K., Moulding, C., Seymour, S.L., Nuwaysir, L.M., Lefebvre, B., Kuhlmann, F., Roark, J., Rainer, P., Detlev, S., Hemenway, T., Huhmer, A., Langridge, J., Connolly, B., Chadick, T., Holly, K., Eckels, J., Deutsch, E.W., Moritz, R.L., Katz, J.E., Agus, D.B., MacCoss, M., Tabb, D.L., Mallick, P.: A cross-platform toolkit for mass spectrometry and proteomics. Nat. Biotechnol. 30, 918–920 (2012). https://doi.org/10.1038/nbt.2377

294.   Pabst, M., Grouzdev, D., Lawson, C.E., Kleikamp, H.B.C., de Ram, C., Louwen, R., Lin, Y., Lücker, S., van Loosdrecht, M.C.M., Laureni, M.: A general approach to explore prokaryotic protein glycosylation reveals the unique surface layer modulation of an anammox bacterium. ISME J. (2021). https://doi.org/10.1038/s41396-021-01073-y

295.   Van Rossum, G., & Drake, F.L.: Python 3 Reference Manual. CreateSpace. (2009)

296.   Carbon, S., Douglass, E., Good, B.M., … et al. .: The Gene Ontology resource: Enriching a GOld mine. Nucleic Acids Res. 49, D325–D334 (2021). https://doi.org/10.1093/nar/gkaa1113

297.   Nickerson, J.L., Doucette, A.A.: Rapid and quantitative protein precipitation for proteome analysis by mass spectrometry. J. Proteome Res. 19, 2035–2042 (2020). https://doi.org/10.1021/acs.jproteome.9b00867

298.   Pérez-Rodriguez, S., Ramírez, O.T., Trujillo-Roldán, M.A., Valdez-Cruz, N.A.: Comparison of protein precipitation methods for sample preparation prior to proteomic analysis of Chinese hamster ovary cell homogenates. Electron. J. Biotechnol. 48, 86–94 (2020). https://doi.org/10.1016/j.ejbt.2020.09.006

299.   Rajalingam, D., Loftis, C., Xu, J.J., Kumar, T.K.S.: Trichloroacetic acid-induced protein precipitation involves the reversible association of a stable partially structured intermediate. Protein Sci. 18, 980–993 (2009). https://doi.org/10.1002/pro.108

300.   Manadas, B.J., Vougas, K., Fountoulakis, M., Duarte, C.B.: Sample sonication after trichloroacetic acid precipitation increases protein recovery from cultured hippocampal neurons, and improves resolution and reproducibility in two-dimensional gel electrophoresis. Electrophoresis. 27, 1825–1831 (2006). https://doi.org/10.1002/elps.200500757

301.   Lin, Y., Zhou, J., Bi, D., Chen, P., Wang, X., Liang, S.: Sodium-deoxycholate-assisted tryptic digestion and identification of proteolytically resistant proteins. Anal. Biochem. 377, 259–266 (2008). https://doi.org/10.1016/j.ab.2008.03.009

302.   Hustoft, H.K., Reubsaet, L., Greibrokk, T., Lundanes, E., Malerod, H.: Critical assessment of accelerating trypsination methods. J. Pharm. Biomed. Anal. 56, 1069–1078 (2011). https://doi.org/10.1016/j.jpba.2011.08.013

303.   Scheerlinck, E., Dhaenens, M., Van Soom, A., Peelman, L., De Sutter, P., Van Steendam, K., Deforce, D.: Minimizing technical variation during sample preparation prior to label-free quantitative mass spectrometry. Anal. Biochem. 490, 14–19 (2015). https://doi.org/10.1016/j.ab.2015.08.018

304.   Canelas, A.B., Ras, C., ten Pierick, A., van Dam, J.C., Heijnen, J.J., van Gulik, W.M.: Leakage-free rapid quenching technique for yeast metabolomics. Metabolomics. 4, 226–239 (2008). https://doi.org/10.1007/s11306-008-0116-4

305.   Schmidt, S.A., Jacob, S.S., Ahn, S.B., Rupasinghe, T., Krömer, J.O., Khan, A., Varela, C.: Two strings to the systems biology bow: Co-extracting the metabolome and proteome of yeast. Metabolomics. 9, 173–188 (2013). https://doi.org/10.1007/s11306-012-0437-1

306.   Ludwig, K.R., Schroll, M.M., Hummon, A.B.: Comparison of in-Solution, FASP, and S-Trap based digestion methods for bottom-up proteomic studies. J. Proteome Res. 17, 2480–2490 (2018). https://doi.org/10.1021/acs.jproteome.8b00235

307.   Tanca, A., Biosa, G., Pagnozzi, D., Addis, M.F., Uzzau, S.: Comparison of detergent-based sample preparation workflows for LTQ-Orbitrap analysis of the *Escherichia coli* proteome. Proteomics. 13, 2597–

2607 (2013). https://doi.org/10.1002/pmic.201200478

308. Bennion, B.J., Daggett, V.: The molecular basis for the chemical denaturation of proteins by urea. Proc. Natl. Acad. Sci. U. S. A. 100, 5142–5147 (2003). https://doi.org/10.1073/pnas.0930122100

309. Kolkman, A., Daran-Lapujade, P., Fullaondo, A., Olsthoorn, M.M.A., Pronk, J.T., Slijper, M., Heck, A.J.R.: Proteome analysis of yeast response to various nutrient limitations. Mol. Syst. Biol. 2, (2006). https://doi.org/10.1038/msb4100069

310. Ho, B., Baryshnikova, A., Brown, G.W.: Unification of protein abundance datasets yields a quantitative *Saccharomyces cerevisiae* proteome. Cell Syst. 6, 192-205.e3 (2018). https://doi.org/10.1016/j.cels.2017.12.004

311. Richards, A.L., Merrill, A.E., Coon, J.J.: Proteome sequencing goes deep. Curr. Opin. Chem. Biol. 24, 11–17 (2015). https://doi.org/10.1016/j.cbpa.2014.10.017

312. Sun, S., Zhou, J.Y., Yang, W., Zhang, H.: Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. Anal. Biochem. 446, 76–81 (2014). https://doi.org/10.1016/j.ab.2013.10.024

313. Tarasova, I.A., Chumakov, P.M., Moshkovskii, S.A., Gorshkov, M. V.: Profiling modifications for glioblastoma proteome using ultra-tolerant database search: Are the peptide mass shifts biologically relevant or chemically induced? J. Proteomics. 191, 16–21 (2019). https://doi.org/10.1016/j.jprot.2018.05.010

314. Geoghegan, K.F., Hoth, L.R., Tan, D.H., Borzilleri, K.A., Withka, J.M., Boyd, J.G.: Cyclization of N-terminal S-carbamoylmethylcysteine causing loss of 17 Da from peptides and extra peaks in peptide maps. J. Proteome Res. 1, 181–187 (2002). https://doi.org/10.1021/pr025503d

315. Güray, M.Z., Zheng, S., Doucette, A.A.: Mass spectrometry of intact proteins reveals +98 u chemical artifacts following precipitation in acetone. J. Proteome Res. 16, 889–897 (2017). https://doi.org/10.1021/acs.jproteome.6b00841

316. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuen, D., Zhang, W., Zhang, Z., Lajoie, G.A., Ma, B.: PEAKS DB: *De novo* sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics. 11, 1–8 (2012). https://doi.org/10.1074/mcp.M111.010587

317. Hailemariam, M., Eguez, R.V., Singh, H., Bekele, S., Ameni, G., Pieper, R., Yu, Y.: S-Trap, an ultrafast sample-preparation approach for shotgun proteomics. J. Proteome Res. 17, 2917–2924 (2018). https://doi.org/10.1021/acs.jproteome.8b00505

318. Carvalho, L.B., Capelo-Martínez, J.L., Lodeiro, C., Wiśniewski, J.R., Santos, H.M.: Ultrasonic-based filter aided sample preparation as the general method to sample preparation in proteomics. Anal. Chem. 92, 9164–9171 (2020). https://doi.org/10.1021/acs.analchem.0c01470

319. Paddon, C.J., Keasling, J.D.: Semi-synthetic artemisinin: A model for the use of synthetic biology in pharmaceutical development. Nat. Rev. Microbiol. 12, 355–367 (2014). https://doi.org/10.1038/nrmicro3240

320. Ro, D.K., Paradise, E.M., Quellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., Chang, M.C.Y., Withers, S.T., Shiba, Y., Sarpong, R., Keasling, J.D.: Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature. 440, 940–943 (2006). https://doi.org/10.1038/nature04640

321. Nielsen, J., Larsson, C., van Maris, A., Pronk, J.: Metabolic engineering of yeast for production of fuels and chemicals. Curr. Opin. Biotechnol. 24, 398–404 (2013). https://doi.org/10.1016/j.copbio.2013.03.023

322. Van Dijken, J.P., Bauer, J., Brambilla, L., Duboc, P., Francois, J.M., Gancedo, C., Giuseppin, M.L.F., Heijnen, J.J., Hoare, M., Lange, H.C., Madden, E.A., Niederberger, P., Nielsen, J., Parrou, J.L., Petit, T.,

R

Porro, D., Reuss, M., Van Riel, N., Rizzi, M., Steensma, H.Y., Verrips, C.T., Vindeløv, J., Pronk, J.T.: An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. Enzyme Microb. Technol. 26, 706–714 (2000). https://doi.org/10.1016/S0141-0229(00)00162-9

323. Zampar, G.G., Kümmel, A., Ewald, J., Jol, S., Niebel, B., Picotti, P., Aebersold, R., Sauer, U., Zamboni, N., Heinemann, M.: Temporal system-level organization of the switch from glycolytic to gluconeogenic operation in yeast. Mol. Syst. Biol. 9, (2013). https://doi.org/10.1038/msb.2013.11

324. Costenoble, R., Picotti, P., Reiter, L., Stallmach, R., Heinemann, M., Sauer, U., Aebersold, R.: Comprehensive quantitative analysis of central carbon and amino-acid metabolism in *Saccharomyces cerevisiae* under multiple conditions by targeted proteomics. Mol. Syst. Biol. 7, (2011). https://doi.org/10.1038/msb.2010.122

325. Picotti, P., Bodenmiller, B., Mueller, L.N., Domon, B., Aebersold, R.: Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. Cell. 138, 795–806 (2009). https://doi.org/10.1016/j.cell.2009.05.051

326. Helbig, A.O., De Groot, M.J.L., Van Gestel, R.A., Mohammed, S., De Hulster, E.A.F., Luttik, M.A.H., Daran-Lapujade, P., Pronk, J.T., Heck, A.J.R., Slijper, M.: A three-way proteomics strategy allows differential analysis of yeast mitochondrial membrane protein complexes under anaerobic and aerobic conditions. Proteomics. 9, 4787–4798 (2009). https://doi.org/10.1002/pmic.200800951

327. den Ridder, M., Knibbe, E., van den Brandeler, W., Daran-Lapujade, P., Pabst, M.: A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications. J. Proteomics. 261, 104576 (2022). https://doi.org/10.1016/j.jprot.2022.104576

328. Luo, Z., Yu, K., Xie, S., Monti, M., Schindler, D., Fang, Y., Zhao, S., Liang, Z., Jiang, S., Luan, M., Xiao, C., Cai, Y., Dai, J.: Compacting a synthetic yeast chromosome arm. Genome Biol. 22, 1–18 (2021). https://doi.org/10.1186/s13059-020-02232-8

329. Postma, E.D., Couwenberg, L.G.F., van Roosmalen, R.N., Geelhoed, J., de Groot, P.A., Daran-Lapujade, P.: Top-down, knowledge-based genetic reduction of yeast central carbon metabolism. bioRxiv. 2021.08.24.457526 (2021)

330. Waskom, M.: Seaborn: Statistical Data Visualization. J. Open Source Softw. 6, 3021 (2021). https://doi.org/10.21105/joss.03021

331. Kanehisa, M., Susumu Goto: KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28, 27–30 (2000). https://doi.org/10.3892/ol.2020.11439

332. Cock, P.J.A., Antao, T., Chang, J.T., Chapman, B.A., Cox, C.J., Dalke, A., Friedberg, I., Hamelryck, T., Kauff, F., Wilczynski, B., De Hoon, M.J.L.: Biopython: Freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 25, 1422–1423 (2009). https://doi.org/10.1093/bioinformatics/btp163

333. Virtanen, P., Gommers, R., Oliphant, T.E., … *et al.* : SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat. Methods. 17, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

334. Bedre, R.: Reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit, (2020)

335. Jensen, L.J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., von Mering, C.: STRING 8 - A global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 37, 412–416 (2009). https://doi.org/10.1093/nar/gkn760

336. Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., García-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., Wertz, J., Sun, Z., Kawano, S., Okuda, S., Watanabe, Y., Hermjakob, H., Maclean, B., Maccoss, M.J., Zhu, Y., Ishihama, Y., Vizcaíno, J.A.: The ProteomeXchange consortium in 2020: Enabling "big data" approaches in proteomics. Nucleic Acids Res. 48, D1145–D1152 (2020). https://doi.org/10.1093/nar/gkz984

337. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., Walzer, M., Wang, S., Brazma, A., Vizcaíno, J.A.: The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. 50, D543–D552 (2022). https://doi.org/10.1093/nar/gkab1038

338. Ghaemmaghami, S., Huh, W.K., Bower, K., Howson, R.W., Belle, A., Dephoure, N., O'Shea, E.K., Weissman, J.S.: Global analysis of protein expression in yeast. Nature. 425, 737–741 (2003). https://doi.org/10.1038/nature02046

339. Fuge, E.K., Braun, E.L., Werner-Washburne, M.: Protein synthesis in long-term stationary-phase cultures of *Saccharomyces cerevisiae*. J. Bacteriol. 176, 5802–5813 (1994). https://doi.org/10.1128/jb.176.18.5802-5813.1994

340. Choder, M.: A general topoisomerase I-dependent transcriptional repression in the stationary phase in yeast. Genes Dev. 5, 2315–2326 (1991). https://doi.org/10.1101/gad.5.12a.2315

341. Valcourt, J.R., Lemons, J.M.S., Haley, E.M., Kojima, M., Demuren, O.O., Coller, H.A.: Staying alive: Metabolic adaptations to quiescence. Cell Cycle. 11, 1680–1696 (2012). https://doi.org/10.4161/cc.19879

342. Paalme, T., Elken, R., Vilu, R., Korhola, M.: Growth efficiency of *Saccharomyces cerevisiae* on glucose/ethanol media with a smooth change in the dilution rate (A-stat). Enzyme Microb. Technol. 20, 174–181 (1997). https://doi.org/10.1016/S0141-0229(96)00114-7

343. Özcan, S., Johnston, M.: Function and Regulation of Yeast Hexose Transporters. Microbiol. Mol. Biol. Rev. 63, 554–569 (1999). https://doi.org/10.1128/mmbr.63.3.554-569.1999

344. Ozcan, S., Johnston, M.: Three different regulatory mechanisms enable yeast hexose transporter (HXT) genes to be induced by different levels of glucose. Mol. Cell. Biol. 15, 1564–1572 (1995). https://doi.org/10.1128/mcb.15.3.1564

345. Thomson, J.M., Gaucher, E.A., Burgan, M.F., Kee, D.W. de, Li, T., Aris, J.P., Benner, S.A.: Resurrecting ancestral alcohol dehydrogenases from yeast. Nat. Genet. 37, 630–635 (2005). https://doi.org/10.1038/ng1553.Resurrecting

346. Camarasa, C., Faucet, V., Dequin, S.: Role in anaerobiosis of the isoenzymes for *Saccharomyces cerevisiae* fumarate reductase encoded by OSM1 and FRDS1. Yeast. 391–401. (2007). https://doi.org/10.1002/yea

347. Påhlman, A.K., Granath, K., Ansell, R., Hohmann, S., Adler, L.: The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative Stress. J. Biol. Chem. 276, 3555–3563 (2001). https://doi.org/10.1074/jbc.M007164200

348. van Dijken, J.P., Scheffers, W.A.: Redox balances in the metabolism of sugars by yeasts. FEMS Microbiol. Lett. 32, 199–224 (1986). https://doi.org/10.1016/0378-1097(86)90291-0

349. Herrero, E., Ros, J., Bellí, G., Cabiscol, E.: Redox control and oxidative stress in yeast cells. Biochim. Biophys. Acta - Gen. Subj. 1780, 1217–1235 (2008). https://doi.org/10.1016/j.bbagen.2007.12.004

350. Han, D., Williams, E., Cadenas, E.: Mitochondrial respiratory chain-dependent generation of superoxide anion and its release into the intermembrane space. Biochem. J. 353, 411–416 (2001). https://doi.org/10.1042/0264-6021:3530411

351. Kwast, K.E., Burke, P. V, Poyton, R.O.: Oxygen sensing and the transcriptional regulation of oxygen-responsive genes in yeast. J. Exp. Biol. 201, 1177–1195 (1998)

352. Bisschops, M.M.M., Vos, T., Martínez-Moreno, R., Cortés, P. de la T., Pronk, J.T., Daran-Lapujade, P.: Oxygen availability strongly affects chronological lifespan and thermotolerance in batch cultures of *Saccharomyces cerevisiae*. Microb. Cell. 2, 429–444 (2015). https://doi.org/10.15698/mic2015.11.238

R

353.   Wilson, K., McLeod, B.J.: The influence of conditions of growth on the endogenous metabolism of *Saccharomyces cerevisiae*: effect on protein, carbohydrate, sterol and fatty acid content and on viability. Antonie Van Leeuwenhoek. 42, 397–410 (1976). https://doi.org/10.1007/BF00410171

354.   Herman, P.K.: Stationary phase in yeast. Curr. Opin. Microbiol. 5, 602–607 (2002). https://doi.org/10.1016/S1369-5274(02)00377-6

355.   Malina, C., Di Bartolomeo, F., Kerkhoven, E.J., Nielsen, J.: Constraint-based modeling of yeast mitochondria reveals the dynamics of protein import and iron-sulfur cluster biogenesis. iScience. 24, 103294 (2021). https://doi.org/10.1016/j.isci.2021.103294

356.   Lu, H., Kerkhoven, E.J., Nielsen, J.: Multiscale models quantifying yeast physiology: towards a whole-cell model. Trends Biotechnol. 40, 291–305 (2022). https://doi.org/10.1016/j.tibtech.2021.06.010

357.   Elsemman, I.E., Prado, A.R., Grigaitis, P., Albornoz, M.G., Harman, V., Holman, S.W., Heerden, J. van, Bruggeman, F.J., Bisschops, M.M.M., Sonnenschein, N., Hubbard, S., Beynon, R., Daran-Lapujade, P., Nielsen, J., Teusink, B.: Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. bioRxiv. 2021.06.11.448029 (2021)

358.   Regueira, A., Lema, J.M., Mauricio-Iglesias, M.: Microbial inefficient substrate use through the perspective of resource allocation models. Curr. Opin. Biotechnol. 67, 130–140 (2021). https://doi.org/10.1016/j.copbio.2021.01.015

359.   Chen, Y., Li, F., Mao, J., Chen, Y., Nielsen, J.: Yeast optimizes metal utilization based on metabolic network and enzyme kinetics. Proc. Natl. Acad. Sci. U. S. A. 118, (2021). https://doi.org/10.1073/pnas.2020154118

360.   Keng, T.: HAP1 and ROX1 form a regulatory pathway in the repression of HEM13 transcription in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 12, 2616–2623 (1992). https://doi.org/10.1128/mcb.12.6.2616-2623.1992

361.   Amillet, J.M., Buisson, N., Labbe-Bois, R.: Positive and negative elements involved in the differential regulation by heme and oxygen of the HEM13 gene (coproporphyrinogen oxidase) in *Saccharomyces cerevisiae*. Curr. Genet. 28, 503–511 (1995). https://doi.org/10.1007/BF00518161

362.   Jordá, T., Puig, S.: Regulation of ergosterol biosynthesis in *Saccharomyces cerevisiae*. Genes (Basel). 11, 1–18 (2020). https://doi.org/10.3390/genes11070795

363.   Rintala, E., Toivari, M., Pitkänen, J.P., Wiebe, M.G., Ruohonen, L., Penttilä, M.: Low oxygen levels as a trigger for enhancement of respiratory metabolism in *Saccharomyces cerevisiae*. BMC Genomics. 10, 461 (2009). https://doi.org/10.1186/1471-2164-10-461

364.   Van Hoek, P., Van Dijken, J.P., Pronk, J.T.: Effect of specific growth rate on fermentative capacity of baker's yeast. Appl. Environ. Microbiol. 64, 4226–4233 (1998). https://doi.org/10.1128/aem.64.11.4226-4233.1998

365.   Abramova, N.E., Cohen, B.D., Sertil, O., Kapoor, R., Davies, K.J.A., Lowry, C. V.: Regulatory mechanisms controlling expression of the DAN/TIR mannoprotein genes during anaerobic remodeling of the cell wall in *Saccharomyces cerevisiae*. Genetics. 157, 1169–1177 (2001). https://doi.org/10.1093/genetics/157.3.1169

366.   Ridder, M. Den, Daran-Lapujade, P., Pabst, M.: Shot-gun proteomics : why thousands of unidentified signals matter. (2018). https://doi.org/10.1093/femsyr/foz088

367.   Slavov, N., Budnik, B.A., Schwab, D., Airoldi, E.M., van Oudenaarden, A.: Constant growth rate can be supported by decreasing energy flux and increasing aerobic glycolysis. Cell Rep. 7, 705–714 (2014). https://doi.org/10.1016/j.celrep.2014.03.057

368.   den Ridder, M., Daran-Lapujade, P., Pabst, M.: Shot-gun proteomics: why thousands of unidentified

signals matter. FEMS Yeast Res. 20, 1–9 (2020). https://doi.org/10.1093/femsyr/foz088

369. Wang, K., Zhou, Y.J., Liu, H., Cheng, K., Mao, J., Wang, F., Liu, W., Ye, M., Zhao, Z.K., Zou, H.: Proteomic analysis of protein methylation in the yeast *Saccharomyces cerevisiae*. J. Proteomics. 114, 226–233 (2015). https://doi.org/10.1016/j.jprot.2014.07.032

370. Kang, J.W., Lee, N.Y., Cho, K.C., Lee, M.Y., Choi, D.Y., Park, S.H., Kim, K.P.: Analysis of nitrated proteins in *Saccharomyces cerevisiae* involved in mating signal transduction. Proteomics. 15, 580–590 (2015). https://doi.org/10.1002/pmic.201400172

371. Neubert, P., Halim, A., Zauser, M., Essig, A., Joshi, H.J., Zatorska, E., Larsen, I.S.B., Loibl, M., Castells-Ballester, J., Aebi, M., Clausen, H., Strahl, S.: Mapping the O-Mannose glycoproteome in *Saccharomyces cerevisiae*. Mol. Cell. Proteomics. 15, 1323–1337 (2016). https://doi.org/10.1074/mcp.M115.057505

372. Loibl, M., Strahl, S.: Protein O-mannosylation: What we have learned from baker's yeast. Biochim. Biophys. Acta - Mol. Cell Res. 1833, 2438–2446 (2013). https://doi.org/10.1016/j.bbamcr.2013.02.008

373. Poljak, K., Selevsek, N., Ngwa, E., Grossmann, J., Losfeld, M.E., Aebi, M.: Quantitative profiling of N-linked glycosylation machinery in yeast *Saccharomyces cerevisiae*. Mol. Cell. Proteomics. 17, 18–30 (2018). https://doi.org/10.1074/mcp.RA117.000096

374. Gentzsch, M., Tanner, W., Seidl, T.: Protein-O-glycosylation in yeast : protein-specific mannosyltransferases Kre9p Pir2p Ctsip Aga2p Barlp Kex2p Ggplp. Glycobiology. 7, 481–486 (1997)

375. Ecker, M., Mrsa, V., Hagen, I., Deutzmann, R., Strahl, S., Tanner, W.: O-mannosylation precedes and potentially controls the N-glycosylation of a yeast cell wall glycoprotein. EMBO Rep. 4, 628–632 (2003). https://doi.org/10.1038/sj.embor.embor864

376. Wu, R., Haas, W., Dephoure, N., Huttlin, E.L., Zhai, B., Sowa, M.E., Gygi, S.P.: A large-scale method to measure absolute protein phosphorylation stoichiometries. Nat. Methods. 8, 677–683 (2011). https://doi.org/10.1038/nmeth.1636

377. Lim, M.Y., O'Brien, J., Paulo, J.A., Gygi, S.P.: Improved method for determining absolute phosphorylation stoichiometry using bayesian statistics and isobaric labeling. J. Proteome Res. 16, 4217–4226 (2017). https://doi.org/10.1021/acs.jproteome.7b00571

378. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M.L., Jensen, L.J., Gnad, F., Cox, J., Jensen, T.S., Nigg, E.A., Brunak, S., Mann, M.: Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. 3, (2010). https://doi.org/10.1126/scisignal.2000475

379. Lyu, J., Wang, Y., Mao, J., Yao, Y., Wang, S., Zheng, Y., Ye, M.: Pseudotargeted MS method for the sensitive analysis of protein phosphorylation in protein complexes. Anal. Chem. 90, 6214–6221 (2018). https://doi.org/10.1021/acs.analchem.8b00749

380. Mayya, V., Rezual, K., Wu, L., Fong, M.B., Han, D.K.: Absolute quantification of multisite phosphorylation by selective reaction monitoring mass spectrometry. Mol. Cell. Proteomics. 5, 1146–1157 (2006). https://doi.org/10.1074/mcp.T500029-MCP200

381. Chua, X.Y., Mensah, T., Aballo, T., Mackintosh, S.G., Edmondson, R.D., Salomon, A.R.: Tandem mass tag approach utilizing pervanadate BOOST channels delivers deeper quantitative characterization of the tyrosine phosphoproteome. Mol. Cell. Proteomics. 19, 730–743 (2020). https://doi.org/10.1074/mcp.TIR119.001865

382. Yi, L., Tsai, C.F., Dirice, E., Swensen, A.C., Chen, J., Shi, T., Gritsenko, M.A., Chu, R.K., Piehowski, P.D., Smith, R.D., Rodland, K.D., Atkinson, M.A., Mathews, C.E., Kulkarni, R.N., Liu, T., Qian, W.J.: Boosting to Amplify Signal with Isobaric Labeling (BASIL) Strategy for comprehensive quantitative phosphoproteomic characterization of small populations of cells. Anal. Chem. 91, 5794–5801 (2019). https://doi.org/10.1021/acs.analchem.9b00024

R

383.   Tsai, C.F., Ogata, K., Sugiyama, N., Ishihama, Y.: Motif-centric phosphoproteomics to target kinase-mediated signaling pathways. Cell Reports Methods. 2, 100138 (2022). https://doi.org/10.1016/j.crmeth.2021.100138

384.   Steen, H., Jebanathirajah, J.A., Springer, M., Kirschner, M.W.: Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. Proc. Natl. Acad. Sci. U. S. A. 102, 3948–3953 (2005). https://doi.org/10.1073/pnas.0409536102

385.   Prus, G., Hoegl, A., Weinert, B.T., Choudhary, C.: Analysis and interpretation of protein post-translational modification site stoichiometry. Trends Biochem. Sci. 44, 943–960 (2019). https://doi.org/10.1016/j.tibs.2019.06.003

386.   Leutert, M., Entwisle, S.W., Villén, J.: Decoding post-translational modification crosstalk with proteomics. Mol. Cell. Proteomics. 20, 0–11 (2021). https://doi.org/10.1016/j.mcpro.2021.100129

387.   Mertins, P., Qiao, J.W., Patel, J., Udeshi, N.D., Clauser, K.R., Mani, D.R., Burgess, M.W., Gillette, M.A., Jaffe, J.D., Carr, S.A.: Integrated proteomic analysis of post-translational modifications by serial enrichment. Nat. Methods. 10, 634–637 (2013). https://doi.org/10.1038/nmeth.2518

388.   Worboys, J.D., Sinclair, J., Yuan, Y., Jørgensen, C.: Systematic evaluation of quantotypic peptides for targeted analysis of the human kinome. Nat. Methods. 11, 1041–1044 (2014). https://doi.org/10.1038/nmeth.3072

389.   Brownridge, P., Beynon, R.J.: The importance of the digest: Proteolysis and absolute quantification in proteomics. Methods. 54, 351–360 (2011). https://doi.org/10.1016/j.ymeth.2011.05.005

390.   Simpson, D.M., Beynon, R.J.: QconCATs: Design and expression of concatenated protein standards for multiplexed protein quantification. Anal. Bioanal. Chem. 404, 977–989 (2012). https://doi.org/10.1007/s00216-012-6230-1

391.   Narumi, R., Masuda, K., Tomonaga, T., Adachi, J., Ueda, H.R., Shimizu, Y.: Cell-free synthesis of stable isotope-labeled internal standards for targeted quantitative proteomics. Synth. Syst. Biotechnol. 3, 97–104 (2018). https://doi.org/10.1016/j.synbio.2018.02.004

392.   Silverman, A.D., Karim, A.S., Jewett, M.C.: Cell-free gene expression: an expanded repertoire of applications. Nat. Rev. Genet. 21, 151–170 (2020). https://doi.org/10.1038/s41576-019-0186-3

393.   Khambhati, K., Bhattacharjee, G., Gohil, N., Braddick, D., Kulkarni, V., Singh, V.: Exploring the potential of cell-free protein synthesis for extending the abilities of biological systems. Front. Bioeng. Biotechnol. 7, 1–16 (2019). https://doi.org/10.3389/fbioe.2019.00248

394.   Shimizu, Y., Kuruma, Y., Ying, B.W., Umekage, S., Ueda, T.: Cell-free translation systems for protein engineering, (2006)

395.   Mair, W., Muntel, J., Tepper, K., Tang, S., Biernat, J., Seeley, W.W., Kosik, K.S., Mandelkow, E., Steen, H., Steen, J.A.: FLEXITau: Quantifying post-translational modifications of Tau protein *in vitro* and in human disease. Anal. Chem. 88, 3704–3714 (2016). https://doi.org/10.1021/acs.analchem.5b04509

396.   Narumi, R., Shimizu, Y., Ukai-Tadenuma, M., Ode, K.L., Kanda, G.N., Shinohara, Y., Sato, A., Matsumoto, K., Ueda, H.R.: Mass spectrometry-based absolute quantification reveals rhythmic variation of mouse circadian clock proteins. Proc. Natl. Acad. Sci. U. S. A. 113, E3461–E3467 (2016). https://doi.org/10.1073/pnas.1603799113

397.   Xian, F., Zi, J., Wang, Q., Lou, X., Sun, H., Lin, L., Hou, G., Rao, W., Yin, C., Wu, L., Li, S., Liu, S.: Peptide biosynthesis with stable isotope labeling from a cell-free expression system for targeted proteomics with absolute quantification. Mol. Cell. Proteomics. 15, 2819–2828 (2016). https://doi.org/10.1074/mcp.O115.056507

398.   Ridder, M. Den, Brandeler, W. Van Den, Altiner, M., Daran-lapujade, P., Pabst, M.: Proteome dynamics

during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast. 1–20 (2022)

399. Budnik, B., Levy, E., Harmange, G., Slavov, N.: SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. 19, 1–12 (2018). https://doi.org/10.1186/s13059-018-1547-5

400. Zecha, J., Satpathy, S., Kanashova, T., Avanessian, S.C., Kane, M.H., Clauser, K.R., Mertins, P., Carr, S.A., Kuster, B.: TMT labeling for the masses: A robust and cost-efficient, in-solution labeling approach. Mol. Cell. Proteomics. 18, 1468–1478 (2019). https://doi.org/10.1074/mcp.TIR119.001385

401. MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., MacCoss, M.J.: Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 26, 966–968 (2010). https://doi.org/10.1093/bioinformatics/btq054

402. New England Biolabs: PURExpress® *in vitro* protein synthesis. (2017)

403. Meyer, B., Papasotiriou, D.G., Karas, M.: 100% protein sequence coverage: A modern form of surrealism in proteomics. Amino Acids. 41, 291–310 (2011). https://doi.org/10.1007/s00726-010-0680-6

404. Tabb, D.L., Vega-Montoto, L., Rudnick, P.A., Variyath, A.M., Ham, A.J.L., Bunk, D.M., Kilpatrick, L.E., Billheimer, D.D., Blackman, R.K., Cardasis, H.L., Carr, S.A., Clauser, K.R., Jaffe, J.D., Kowalski, K.A., Neubert, T.A., Regnier, F.E., Schilling, B., Tegeler, T.J., Wang, M., Wang, P., Whiteaker, J.R., Zimmerman, L.J., Fisher, S.J., Gibson, B.W., Kinsinger, C.R., Mesri, M., Rodriguez, H., Stein, S.E., Tempst, P., Paulovich, A.G., Liebler, D.C., Spiegelman, C.: Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. J. Proteome Res. 9, 761–776 (2010). https://doi.org/10.1021/pr9006365

405. Doerr, A., De Reus, E., Van Nies, P., Van Der Haar, M., Wei, K., Kattan, J., Wahl, A., Danelon, C.: Modelling cell-free RNA and protein synthesis with minimal systems. Phys. Biol. 16, (2019). https://doi.org/10.1088/1478-3975/aaf33d

406. Cruz-Vera, L.R., Magos-Castro, M.A., Zamora-Romo, E., Guarneros, G.: Ribosome stalling and peptidyl-tRNA drop-off during translational delay at AGA codons. Nucleic Acids Res. 32, 4462–4468 (2004). https://doi.org/10.1093/nar/gkh784

407. Weinert, B.T., Schölz, C., Wagner, S.A., Iesmantavicius, V., Su, D., Daniel, J.A., Choudhary, C.: Lysine succinylation is a frequently occurring modification in prokaryotes and eukaryotes and extensively overlaps with acetylation. Cell Rep. 4, 842–851 (2013). https://doi.org/10.1016/j.celrep.2013.07.024

408. Zhou, X., Li, W., Liu, Y., Amon, A.: Cross-compartment signal propagation in the mitotic exit network. Elife. 10, 1–30 (2021). https://doi.org/10.7554/eLife.63645

409. Saad, S., Cereghetti, G., Feng, Y., Picotti, P., Peter, M., Dechant, R.: Reversible protein aggregation is a protective mechanism to ensure cell cycle restart after stress. Nat. Cell Biol. 19, 1202–1213 (2017). https://doi.org/10.1038/ncb3600

410. Jurica, M.S., Mesecar, A., Heath, P.J., Shi, W., Nowak, T., Stoddard, B.L.: The allosteric regulation of pyruvate kinase by fructose-1,6-bisphosphate. Structure. 6, 195–210 (1998). https://doi.org/10.1016/S0969-2126(98)00021-5

411. Cortassa, S., Aon, J.C., Aon, M.A.: Fluxes of carbon, phosphorylation, and redox intermediates during growth of *Saccharomyces cerevisiae* on different carbon sources. Biotechnol. Bioeng. 47, 193–208 (1995). https://doi.org/10.1002/bit.260470211

412. Portela, P., Moreno, S., Rossi, S.: Characterization of yeast pyruvate kinase 1 as a protein kinase A substrate, and specificity of the phosphorylation site sequence in the whole protein. Biochem. J. 396, 117–126 (2006). https://doi.org/10.1042/BJ20051642

R

413. Xu, Y.-F., Zhao, X., Glass, D.S., Absalan, F., Perlman, D.H., Broach, J.R., Rabinowitz, J.D.: Regulation of yeast pyruvate kinase by ultrasensitive allostery independent of phosphorylation. Mol. Cell. 48, 1–3 (2012). https://doi.org/10.1016/j.molcel.2012.07.013.Regulation

414. Huang, F.K., Zhang, G., Lawlor, K., Nazarian, A., Philip, J., Tempst, P., Dephoure, N., Neubert, T.A.: Deep coverage of global protein expression and phosphorylation in breast tumor cell lines using TMT 10-plex isobaric labeling. J. Proteome Res. 16, 1121–1132 (2017). https://doi.org/10.1021/acs.jproteome.6b00374

415. Miller, R.M., Millikin, R.J., Hoffmann, C. V., Solntsev, S.K., Sheynkman, G.M., Shortreed, M.R., Smith, L.M.: Improved protein inference from multiple protease bottom-up mass spectrometry data. J. Proteome Res. 18, 3429–3438 (2019). https://doi.org/10.1021/acs.jproteome.9b00330

416. Heiden, M.G.V., Cantley, L.C., Thompson, C.B.: Understanding the warburg effect: The metabolic requirements of cell proliferation. Science (80-.). 324, 1029–1033 (2009). https://doi.org/10.1126/science.1160809

417. Postma, E.D., Dashko, S., Van Breemen, L., Taylor Parkins, S.K., Van Den Broek, M., Daran, J.M., Daran-Lapujade, P.: A supernumerary designer chromosome for modular *in vivo* pathway assembly in *Saccharomyces cerevisiae*. Nucleic Acids Res. 49, 1769–1783 (2021). https://doi.org/10.1093/nar/gkaa1167

418. Degroeve Sven, Gabriels Ralf, Velghe Kevin, Bouwmeester Robbin, Tichshenko Natalia, Martens Lennart: ionbot: a novel, innovative and sensitive machine learning approach to LC-MS/MS peptide identification. (2021)

419. Yuan, H., Zhang, S., Zhao, B., Weng, Y., Zhu, X., Li, S., Zhang, L., Zhang, Y.: Enzymatic reactor with trypsin immobilized on graphene oxide modified polymer microspheres to achieve automated proteome quantification. Anal. Chem. 89, 6324–6329 (2017). https://doi.org/10.1021/acs.analchem.7b00682

420. Laporte, D., Jimenez, L., Gouleme, L., Sagot, I.: Yeast quiescence exit swiftness is influenced by cell volume and chronological age. Microb. Cell. 5, 104–111 (2018). https://doi.org/10.15698/mic2018.02.615

421. Zhou, C., Schulz, B.L.: Glycopeptide variable window SWATH for improved data independent acquisition glycoprotein analysis. Anal. Biochem. 597, 113667 (2020). https://doi.org/10.1016/j.ab.2020.113667

422. Adoni, K.R., Cunningham, D.L., Heath, J.K., Leney, A.C.: FAIMS enhances the detection of PTM crosstalk sites. J. Proteome Res. 21, 930–939 (2022). https://doi.org/10.1021/acs.jproteome.1c00721

423. Schweppe, D.K., Eng, J.K., Yu, Q., Bailey, D., Rad, R., Navarrete-Perea, J., Huttlin, E.L., Erickson, B.K., Paulo, J.A., Gygi, S.P.: Full-featured, real-time database searching platform enables fast and accurate multiplexed quantitative proteomics. J. Proteome Res. 19, 2026–2034 (2020). https://doi.org/10.1021/acs.jproteome.9b00860

424. Erickson, B.K., Mintseris, J., Schweppe, D.K., Navarrete-Perea, J., Erickson, A.R., Nusinow, D.P., Paulo, J.A., Gygi, S.P.: Active instrument engagement combined with a real-time database search for improved performance of sample multiplexing workflows. J. Proteome Res. 18, 1299–1306 (2019). https://doi.org/10.1021/acs.jproteome.8b00899

425. Kelly, R.T.: Single-cell Proteomics: Progress and Prospects. Mol. Cell. Proteomics. 19, 1739–1748 (2020). https://doi.org/10.1074/mcp.R120.002234

426. Cheung, T.K., Lee, C.Y., Bayer, F.P., McCoy, A., Kuster, B., Rose, C.M.: Defining the carrier proteome limit for single-cell proteomics. Nat. Methods. 18, 76–83 (2021). https://doi.org/10.1038/s41592-020-01002-5

427. Swaminathan, J., Boulgakov, A.A., Hernandez, E.T., Bardo, A.M., Bachman, J.L., Marotta, J., Johnson, A.M., Anslyn, E. V., Marcotte, E.M.: Highly parallel single-molecule identification of proteins in

zeptomole-scale mixtures. Nat. Biotechnol. 36, 1076–1091 (2018). https://doi.org/10.1038/nbt.4278

428.   Van Ginkel, J., Filius, M., Szczepaniak, M., Tulinski, P., Meyer, A.S., Joo, C.: Single-molecule peptide fingerprinting. Proc. Natl. Acad. Sci. U. S. A. 115, 3338–3343 (2018). https://doi.org/10.1073/pnas.1707207115

429.   Ouldali, H., Sarthak, K., Ensslen, T., Piguet, F., Manivet, P., Pelta, J., Behrends, J.C., Aksimentiev, A., Oukhaled, A.: Electrical recognition of the twenty proteinogenic amino acids using an aerolysin nanopore. Nat. Biotechnol. 38, 176–181 (2020). https://doi.org/10.1038/s41587-019-0345-2

430.   Restrepo-Pérez, L., Wong, C.H., Maglia, G., Dekker, C., Joo, C.: Label-free detection of post-translational modifications with a Nanopore. Nano Lett. 19, 7957–7964 (2019). https://doi.org/10.1021/acs.nanolett.9b03134

R

Acknowledgements

This dissertation would not have been here today without the support of many people. I would like to express my gratitude here to all of you (in no particular order).

First of all, I would like to thank **Martin** for offering me the opportunity to work on this project and your guidance throughout my project. I did not know anything about proteomics when I started this project, but your knowlegde and patience helped me to become the manager of my own project. You encouraged me to step out of my comfort zone and therefore I gained skills in many different research fields. **Pascale**, thank you for fulfilling the role of promotor in my project. Your advice and help in science, but most of all, in personal development was appreciated by me. I am grateful for your view as outsider on the proteomics work and your indispensable expertise in yeast physiology.

This work would not have been possible without the help of my students, I really enjoyed collaborating on this project with you. **Ramon**, **Jelle**, **Meryem**, **Casper** and **Wiebeke**, your contributions provided the fundaments of this thesis and I learned a lot from you.

I would like to give a big thank you to my dear paranymphs. **Aafke**, I first got to know you during some overnight experiments. My experiments would have definitely failed (more often) without your help. I admire your perseverance, I don't know how you do it (all those batch experiments)! One of the highlights of my PhD was definitely the conference in Vancouver, I had a lot of fun with you! **Claudia**, I am happy you decided to pursue a PhD after your internship in our lab. Our walking breaks have kept me sane in the final stages of my PhD, I am happy you did not mind me complaining all the time. Your ability to stand up for yourself and to set boundaries are an inspiration to me.

I was also very grateful for my office mates during my PhD. **Carla**, I was so happy we moved from that big student office to our big (broomcloset) office. Thank you for your support and friendship during our time together in the office. Traveling together in Brasil was certainly highlight for me! **Agi**, it was a lot of fun sharing the office with you. Your risk-taking character encouraged me to go on more adventures. **Yanfang**, I am happy you were assigned to my office after I was flying solo in the office for a while. I am really impressed by your discipline and kindness, thanks for being such an amazing office mate. **Shree**, we are both the odd (proteomics) ducks at IMB, so I am happy we are in the same boat. Your advice and kindness were highly appreciated by me. And, thanks again to Yanfang and Shree, for listening to my complaints in the final stages of making the thesis.

During my PhD, I was lucky enough to be part of two different research groups, CSE and IMB. However, the core of my project was performed in the MS lab. **Carol**, you were a tremendous help in the lab, but also a fun lunch partner, it was great to talk to you, especially in covid times (and beyond!). **Hugo**, thank you for introducing me to coffee walks, those were very appreciated. **Dita**, I enjoyed working together in the MS lab, which is definitely more organized since you started here. **Jitske**, I really appreciated your help in the Aplagon project.

The research groups would not function without the technicians. Before I started this project, I never thought I would be working with bioreactors because it really did not like this type of work during my studies. However, **Erik** and **Christiaan**, your support during these bioreactor experiments was essential for its success and also for making me enthusiastic about this type of work. **Marijke**, I really appreciated your eye for detail while checking my thesis and your genuine interest in my well-being. I would also like to thank **Apilena**, **Astrid** and **Jannie** for your help with the preparations for my fermentations. It was more an exception than a rule that my preparations were done in time for autoclaving, but you always waited for me.

The first two years of my PhD were performed in CSE. I was warmly welcomed into the group by everyone and I enjoyed our lunches, dinners, drinks and muffin Thursdays together. **Florence**, thank you for kindness and introducing me to the BT PhD committee and AS PhD council, which I never would have joined without your encouragement. A quick shoutout to my fellow members of the BT PhD committee (**Aafke**, **Alexandra**, **Marina**, **Philipp** and **Sergio**), it was a lot of fun keeping the social life of BT alive during the pandemic. **Koen**, you were a few months ahead in the PhD timeline, so I really appreciated your help through the PhD milestones. A special thanks to **Eleni,** for your incredible recommendations for Greek islands, **Mariana**, for our amazing trip in Brasil and **Leonor**, **Yaya**, **Jinrui**, **Joan**, **Timmy** and **Walter** for the fun times at CSE!

Mid-pandemic, I joined IMB and I would like to thank **Anna**, **Céline**, **Charlotte**, **Clara**, **Denzel**, **Eline**, **Flip**, **Ilja**, **Jean-Marc**, **Marcel**(s), **Maria**, **Marieke**, **Miriam**, **Nicole**, **Rinke**, **Robert**, **Rosalie**, **Rozanne**, **Sagarika**, **Steven**, **Susan**, **Tobias** and all others for welcoming me into the group and for the enjoyable times during lunch, drinks and conferences. A special thanks to **Ewout**, thanks for running your first chemostat so that I would have material for my proteomics experiments. I would also like to thank **Jack**, because without your recommendation to Martin, I would not have started (and finished) this PhD project.

Of course, I would also like to thank my all my friends for the wonderful times in life. The PhD became part of my identity, but I am so grateful for the times when I did not have to think of work. A special thanks to **Anna** and **Maud**, for dedicating an evening to coming up

with propositions! I would also like to thank **Sophie** and **Dyantha**, although you both quit LST after one year, I made it to the end of the LST careerpath in Delft with you both still in my life! A specials thanks to my fellow PhD girls outside of Delft. **Dyantha**, I am truly impressed by your work ethic and your ability to put the PhD in perspective, I really needed that at times. **Suzanne**, it was amazing to visit you in Montreal and see you in the place where you successfully obtained a PhD degree (your natural habitat). I treasure our discussions on a life as a woman in science and I am sure you will become a great professor one day.

My family deserves the biggest acknowlegdement. **Papa**, **mama**, **Valerie**, **Denise** and **Jeroen**, I am so grateful for our amazing bond and your unconditional love and support. None of this work would have been possible without your encouragement. Ranging from borrowing the car for my overnight experiments (which lasted for several months apparently) to designing the cover and coming up with my propositions. Of course, I am even more grateful for the good times in Brabant without a mention of the PhD.

Finally, I would like to thank myself for not giving up during this project. It has been a hell of a ride, but I could not be happier to move on to the next chapter of my life, whatever that may be.

A

Curriculum vitae

Maxime den Ridder was born on August 10th 1992 in Zevenbergen, the Netherlands. In 2011, she started with the Bachelor's programme Life Science & Technology at the Delft University of Technology (TU Delft) and Leiden University. During her Bachelor, Maxime studied at the University of Oslo for five months to obtain her minor in biosciences. She then joined the Chirlmin Joo Lab at the Bionanoscience department to complete her bachelor research project on RNA interference under the supervision of Dr. Mohamed Fareh. After a half year internship working on miRNA interfence at the Chirlmin Joo lab, she started her Master's programme Life Science & Technology in 2015 with specialization Cell Factory. In addition to the regular curriculum, Maxime participated in an industrial group project with DSM as part of fulfillment of the Honour's Programme of the Applied Sciences faculty. In the second year of the Master's programme, she joined the Industrial Microbiology group for her Master thesis. Here, she worked on developing a new CRISPR genome editing tool for *Saccharomyces cerevisiae*, under the supervision of Michal Świat and Prof. Pascale Daran-Lapujade. The Master was completed in 2018 with an company internship at Evolva in Basel, Switzerland, in which metabolic engineering was used for optimization of industrial yeast production strains. In May 2018, Maxime started her PhD project in the Cell Systems Engineering section at TU Delft, with Dr. Martin Pabst and Prof. Pascale Daran-Lapujade as promotors. The results of this PhD project on yeast proteome dynamics using mass spectrometry are described in this thesis.

C

List of publications

den Ridder, M., Daran-Lapujade, P., & Pabst, M. (2022). Proteome dynamics during transition from exponential to stationary phase under aerobic and anaerobic conditions in yeast. *bioRxiv*.

Rodrigues, C. I. S., **den Ridder, M.,** Pabst, M., Gombert, A. K., & Wahl, S. A. (2022). Comparative proteome analysis of different *Saccharomyces cerevisiae* strains during growth on sucrose and glucose. *Scientific Reports*, 13(1), 2126.

Holub, M., Birnie, A., Japaridze, A., van der Torre, J., **den Ridder, M.**, ... & Dekker, C. (2022). Extracting and characterizing protein-free megabasepair DNA for *in vitro* experiments. *Cell Reports Methods* 2(12): 100366.

Boonekamp, F. J., Knibbe, E., Vieira-Lara, M. A., Wijsman, M., Luttik, M. A., van Eunen, K., **den Ridder, M**., ... & Daran-Lapujade, P. (2022). Full humanization of the glycolytic pathway in *Saccharomyces cerevisiae*. *Cell Reports*, 39(13), 111010.

den Ridder, M., Knibbe, E., van den Brandeler, W., Daran-Lapujade, P., & Pabst, M. (2022). A systematic evaluation of yeast sample preparation protocols for spectral identifications, proteome coverage and post-isolation modifications. *Journal of Proteomics*, 261, 104576.

den Ridder, M., Daran-Lapujade, P., & Pabst, M. (2020). Shot-gun proteomics: Why thousands of unidentified signals matter. *FEMS yeast research*, 20(1), foz088.

Boonekamp, F. J., Dashko, S., Duiker, D., Gehrmann, T., van den Broek, M., **den Ridder, M**., ... & Daran-Lapujade, P. (2020). Design and experimental evaluation of a minimal, innocuous watermarking strategy to distinguish Near-Identical DNA and RNA sequences. *ACS synthetic biology*, 9(6), 1361-1375

Świat, M. A., Dashko, S., **den Ridder, M.**, Wijsman, M., van der Oost, J., Daran, J. M., & Daran-Lapujade, P. (2017). Fn Cpf1: a novel and efficient genome editing tool for *Saccharomyces cerevisiae*. *Nucleic acids research*, 45(21), 12585-12598.

Selected conference contributions

den Ridder M, van den Brandeler W, Daran-Lapujade P, Pabst M. Temporal proteome dynamics of (minimal glycolysis) yeast during aerobic and anaerobic growth. ISSY36, Vancouver, Canada. 2022 (Oral presentation)

den Ridder M, Daran-Lapujade P, Pabst M. A systematic study to maximise proteome coverage while minimizing artefacts through sample preparation procedures in yeast large-scale proteomics, ICAP 2021, Lisbon, Portugal, 2021. (online symposium, Oral presentation)