Understanding global resource allocation in fission yeast through data analysis and coarse-grained mathematical modelling

Thesis submitted in partial fulfilment of the requirements for obtaining the degree of Doctor of Philosophy (PhD)

Istvan Tariq Kleijn

Department of Mathematics
Imperial College London

Declaration of originality

All work described in this thesis was either performed by myself, or referenced appropriately.

A paper describing the results in chapter 2 has been published (Kleijn et al. 2022). I also adapted parts of the introduction from this preprint for inclusion into the introductory chapter 1.

Copyright statement

Istvan Kleijn, 2022

Abstract

Unicellular organisms can grow in a large variety of environments. Even in those supporting robust growth, cellular resources are limited and their relative allocation to gene expression programmes determines physiological states and global properties such as the growth rate and the cell size. I have approached this topic from two angles, namely a comprehensive analysis of a gene expression data set and the construction of coarse-grained resource allocation models (C-GRAMs).

First, I studied a combined data set of protein and transcript abundances during growth of the fission yeast *Schizosaccharomyces pombe* on various abundant nitrogen sources. Approximately half of gene expression was significantly correlated with the growth rate, and this came alongside wide-spread nutrient-specific expression. Genes positively correlated with the growth rate participated in protein production, whereas those negatively correlated mainly belonged to the environmental stress response programme. Critically, the expression of metabolic enzymes was mainly condition specific.

Second, C-GRAMs are simple models of single cells, where large components of the macromolecular composition are abstracted into single entities. The dynamics and steady-state behaviour of such models can then be easily explored. A minimal C-GRAM with nitrogen and carbon pathways converging on biomass production described the effects of the uptake of sugars, ammonium, and/or compound nutrients such as amino acids on the translational resource allocation towards proteome sectors that maximised the growth rate. Prompted by new observations that the relation between cell volume and the growth rate was identical for both carbon and nitrogen perturbations, but that the surface-to-volume ratio was elevated in low-nitrogen conditions, I extended this to a C-GRAM that additionally accounted for the cell cycle, cell division, cell wall biosynthesis, and the effect of molecular crowding on the ribosomal efficiency.

Acknowledgements

My experience as a PhD student has been defined by the interaction with my two great supervisors. You supported me in difficult times and shared in the successes, leaving me free to pursue my interests but guiding me towards what's important. Sam, you taught me how to make sense of a scientific story. Vahid, your patient support always kept me engaged. Thank you and cheers to you both.

I would like to thank the Wellcome Trust for their generous financial support. 203968/Z/16/Z checking in.

Many thanks also to my colleagues in the Marguerat and Shahrezaei groups for your feedback on this thesis and earlier manuscripts, and for all our many discussions about science and otherwise. Luca, Yuxuan, it has been a privilege supervising you and developing our ideas together. Xi-Ming, Amalia, François, Lorenzo, Miles, Anthony, Wenhao, Lucie, Ben, you define what it means to be a scientist.

A special thanks go to all the people I have been fortunate enough to share a London flat with. Andrea, Francesc, Urbain, Tom, Edd, Alex, Julia, Paul, you were examples of persistence, developing professionalism, and scientific integrity.

Mama, papa, bedankt dat jullie me altijd hebben gesteund om uitdagingen te zoeken en mijn nieuwsgierigheid achterna te gaan.

Kathryn, thank you for always being there with me, and for everything else.

Now it would be very remarkable if any system existing in the real world could be exactly represented by any simple model. However, cunningly chosen parsimonious models often do provide remarkably useful approximations. (...) For such a model there is no need to ask the question "Is the model true?". If "truth" is to be the "whole truth" the answer must be "No". The only question of interest is "Is the model illuminating and useful?"

(G.E.P. Box (1979), in: Robustness in the Strategy of Scientific Model Building, pp. 202–203)

Contents

List of Figures

List of Tables

I Introduction

The smallest unit of life—the cell—comes in many varieties. From hyperthermophile archaea capable of surviving in boiling hot water around hydrothermal vents, to biofilms of collaborating and competing bacteria and yeasts, via small multicellular worms all the way to the many different types of tissue present in human beings, cellular life is varied and complex.

Many types of cells are remarkably fast at copying themselves: lab strains of commonly used bacteria routinely duplicate within 15-20 minutes under favourable circumstances, and laboratory yeasts can reproduce asexually in mere hours. Moreover, many unicellular organisms thrive in highly variable circumstances. Although cells might not be optimised for growth in each individual condition (Towbin et al. 2017), they are nevertheless under selective pressure to grow fast or risk being outcompeted by rival colonies or species (Dekel and Alon 2005; Lynch et al. 2014). However, cell cultures in fixed environments are faced with internal constraints: the concentrations of the building blocks and machinery required for gene expression are bounded. In other words, each atom can only be used in one molecule of biomass, and each macromolecular complex can only process one reaction at a time. Fast-growing cells must therefore possess an ability to allocate these limited resources in varied environments (Koch 1988; López-Maury et al. 2008; Molenaar et al. 2009; Berkhout et al. 2013; Bruggeman et al. 2020).

Although the complexity involved in the regulation of this behaviour is enormous, strikingly simple observations have been made. For instance, in prokaryotes, such as the bacterium *Escherichia coli*, it has been established that the abundances of many large classes of protein are linearly correlated with the cellular growth rate across a large variety of environmental conditions (Schaechter et al. 1958; Neidhardt et al. 1990; Scott et al. 2010; Hui et al. 2015; A. Schmidt et al. 2016). This is true in particular for the abundance of ribosomes, an observation termed the "ribosomal growth law". A similar result states that the logarithm of cell size and the growth rate are linearly correlated across conditions with varying nutrients (Vadia and Levin 2015). Such results show that the cell somehow

crystallises its complexity into simple behaviour rules, which should in principle be understood in terms of simple phenomenological, optimisation-based, and/or mechanistical approaches.

Despite the famous assertion, what is true for *E. coli* is not necessarily true for the elephant (not even when adding Monod's "except more so", see Friedmann (2004) for a historical perspective). The work describing allocation in prokaryotes needs to be extended and expanded, applying the concepts to more complex eukaryotic cells. Both the budding yeast *Saccharomyces cerevisiae* and the fission yeast *Schizosaccharomyces pombe* have been studied extensively. Although initial research in the resource allocation field has primarily focused on *S. cerevisiae* (Waldron and Lacroute 1975; Brauer et al. 2008; Metzl-Raz et al. 2017), a great deal is known about *S. pombe*'s physiology and regulation of its cell cycle (Mitchison and Lark 1962; Fantes and Nurse 1977), and many experimental procedures have been established for its growth in different environments (Carlson et al. 1999; Petersen and Russell 2016). The Marguerat lab has established *S. pombe* as a model organism to study the interplay between (stochastic) gene expression, growth, and cell size across conditions (Saint et al. 2019; Sun et al. 2020). A better understanding of fission yeast resource allocation will inform further work investigating the relationships between its growth and allocation states and the processes governing the cell cycle, cell division, and the regulation of cell size homeostasis. For these reasons, we chose *S. pombe* as our model eukaryote.

1.1 General concepts and observations

1.1.1 Balanced growth

Fission yeast is thought to thrive in a large variety of environments. For example, Jeffares et al. (2015) identified 161 natural isolates of *S. pombe* obtained from more than 20 countries. Its natural habitat is unclear though, as all strains had been influenced by human cultivation. This does point out that yeast typically grows in environments with a complex composition—think, for example, about all the different nutrients present in fruits or brewers' vats. On top of this, like most organisms, yeasts have adapted to life in ever changing circumstances. Crucially, external cues such as nutrient quality, stressing agents, or growth factors influence the growth rate—the relative rate of biomass accumulation.

Counter to the complexity in the "natural" environment, experimental and theoretical approaches for studying cells have focused on so-called "balanced" growth,

where cells have been cultured in the same environment during many generations (Schaechter 2006). In light of the previous paragraph, this is perhaps surprising, but this state of growth has the advantage of reproducibility, both *in vivo* and *in silico*. Moreover, it is easily studied *in formula* as the steady-state solution to a suitably defined set of ordinary differential equations. By focusing on balanced growth, we are able to study internal patterns of gene expression, and the extent to which these patterns are optimised for fast growth. This is particularly true in cultures where external resources are in excess, exemplified by traditional batch cultures and turbidostats, the latter setup being more reproducible than the former. In batch and turbidostat cultures, the growth state is fully determined by internal allocation, rather than by externally imposed nutrient concentrations. As the first step towards a full understanding of cell growth, we therefore focus on gene expression allocation under balanced growth in defined environments.

1.1.2 The ribosomal growth law

The growth rate is a fundamental property of the cell that impacts most aspects of cell physiology (Schaechter et al. 1958; Mitchison and Lark 1962; Waldron and Lacroute 1975; Fantes and Nurse 1977; Neidhardt et al. 1990; Bremer and Dennis 2008). Extensive quantitative experiments and modelling have explored this relationship, as reviewed by Klumpp and Hwa (2014), Shahrezaei and Marguerat (2015), Jun et al. (2018) and Bruggeman et al. (2020). One of the most important observations about balanced growth is the linear correlation between the growth rate and the abundance of ribosomes (both rRNA and ribosomal proteins) relative to the total biomass. The effect can be explained by considering ribosomes as fully utilised self-replicating machineries (Scott et al. 2014). In these circumstances, ribosomes will replicate themselves at a rate proportional to the fraction of time they can spend on it. In balanced growth, this fraction will be equal to the fraction of the biomass that comprises the ribosomes themselves. Therefore, a larger ribosomal fraction and a faster growth rate go hand in hand when all ribosomes are translating at all times. This linear correlation is sometimes called the R-line.

From the data, however, it is clear that the R-line is extrapolated to have a nonzero intersect with the ordinate, i.e. a significant amount of ribosomal proteins would be expressed even at zero growth. Quantitatively, the ribosomal proteome fraction ϕ_R may be expressed as

$$\phi_R = \phi_{R,0} + A\mu \tag{1.1}$$

with μ the growth rate and $\phi_{R,0}$ the aforementioned intercept; the inverse of the slope A can be interpreted as the ribosomal efficiency (a higher slope means more ribosomes are required to obtain the same increase in growth rate). Direct evidence of $\phi_{R,0}$ is found in the fact that the ribosomal proteins make up significant fractions of both the *E. coli* and *S. pombe* proteome in quiescent cells (Marguerat et al. 2012; A. Schmidt et al. 2016). Cells may keep ribosomes in non-translating states unless required explicitly. This has been shown to be beneficial when a cell encounters switches between environments (Metzl-Raz et al. 2017). It is an open question how this strategy is implemented and to what extent it exists in different organisms.

1.1.3 Proteome allocation sectors

The derivation in the section above is formulated in terms of the fractional abundance of ribosomal proteins. If this is to increase with the growth rate, the fractional abundance of other proteins is to decrease with the growth rate. After all, the fractional abundance of all proteins together must by definition add to one. Indeed, the expression of many individual genes and proteins have been found to correlate with the growth rate as well (treated in detail in section 2.1.1). It was shown that about half of the total protein mass in Escherichia coli responded to growth modulations by nutrient limitation and translational inhibition (Scott et al. 2010; You et al. 2013). These observations were formalised in a phenomenological model separating the proteome into three broad sectors based on their growth rate correlations (Scott et al. 2010). In this view, proteins that are positively correlated with the cellular growth rate during nutrient limitation and negatively during translational inhibition comprise the R-sector, whereas proteins showing the opposite behaviour comprise the P-sector. Proteins that do not respond to the growth rate belong to the Q-sector.

The Q-sector proteins are generally considered to be housekeeping genes, that always need to be expressed for the cell to survive. This suggests that they are not actively regulated in response to the growth rate. However, for such constitutively expressed genes to be uncorrelated with growth rate, the regulation of the P- and R-sector must have the same amplitude. Otherwise, the residual global response will mean that constitutively expressed proteins will still be correlated with the growth rate. If the R-sector is generally bigger than the P-sector, constitutive proteins will be negatively correlated with growth, whereas the correlation will be positive if the P-sector exceeds the R-sector.

The naive phenomenological model furthermore does not account for proteins with multiple layers of regulation. I have already remarked that the expression of ribosomal proteins contains a considerable component that is independent of growth. Such correlations with the growth rate that do not span the entire dynamic range are common throughout the proteome, and titrations of different limiting nutrients may have opposite effects (Hui et al. 2015; A. Schmidt et al. 2016). Additionally, proteins might have baseline expression levels that are uncorrelated with growth, but be induced in one or more conditions (A. Schmidt et al. 2016). If such condition-dependent proteins are highly abundant in some conditions, their expression will come at the expense of proteins involved in biosynthesis. The resulting passive regulation may even be sufficient to explain the R-line (Barenholz et al. 2016). In summary, the phenomenological separation of the proteome in P-, Q-, and R-sectors is a helpful conceptual tool, but a complete view of proteome allocation should expect expression patterns outside of this paradigm too.

1.1.4 Mechanisms associated with the regulation of allocation

The molecular mechanisms behind the observed correlations between gene expression and growth rate are not well understood. R-sector proteins are universally involved in translation and ribosome biogenesis, whereas P-sector proteins are more diverse and often involved in metabolic adaptation and stress response (Brauer et al. 2008; You et al. 2013; Hui et al. 2015; A. Schmidt et al. 2016). In *E. coli*, the master regulator CRP-cAMP has been proposed to control the P-sector assignments of carbon catabolism enzymes when growth rate was modulated by the quality of abundant carbon sources (You et al. 2013). Furthermore, evidence is mounting for the role of the stringent response, mediated by guanosine tetraphosphate (ppGpp), in regulating the expression of the *E. coli* ribosomes under nutrient limitation (M. Zhu et al. 2019; Irving et al. 2020). Whether these conclusions are transferrable to other growth modulations and/or other organisms is an open question.

In eukaryotes, growth-rate-related gene expression is thought to depend to some extent on the target-of-rapamycin complex 1 (TORC1), which is widely conserved (Weisman 2016; González and Hall 2017; Morozumi and Shiozaki 2021). TORC1 activity is affected by a variety of stressors including nutrient starvation. Upstream of TORC1, the adenosine monophosphate kinase AMPK has been proposed to mediate the response to nitrogen starvation, and intriguingly, the two complexes can inhibit each other (Davie et al. 2015; Ling et al. 2020). In fission yeast,

the TORC1 pathway is a key regulator of the balance between stress- and growth-related gene expression (López-Maury et al. 2008; Rallis et al. 2013, 2014), with downstream targets including eukaryotic initiation factor 2 subunit alpha (eIF2α) (Valbuena et al. 2012), the SAGA complex (Laboucarié et al. 2017), and the rate of fermentation through Greatwall and PP2AB55δ (Watanabe et al. 2019).

1.1.5 The interplay between metabolism, growth, and gene expression

The cell metabolism is an exquisitely complex network of interconnected processes and perturbation of single pathways can have wide-spread systemic effects (Chubukov et al. 2014). Central carbon metabolism (CCM) relies on three pathways: glycolysis, the pentose phosphate pathway, and the tricarboxylic acid (TCA) cycle (Alberts et al. 2015, ch. 2). Together, these generate energy in the form of ATP, in a process mediated by reducing agents such as NADH, and produce building blocks for biosynthesis. ATP can be generated anaerobically via fermentation—a process which consists of glycolysis and the subsequent degradation of pyruvate, or aerobically via respiration, which requires the TCA cycle and subsequent oxidative phosphorylation (OXPHOS). The extent of fermentative versus respiratory metabolism affects the NAD+/NADH redox balance and vice versa, as NAD+ reduction during glycolysis and the TCA cycle must be balanced by NADH oxidation occurring during pyruvate degradation and OXPHOS (Vemuri et al. 2007; van Hoek and Merks 2012; Campbell et al. 2018).

In eukaryotes, these reactions are compartmentalised between the cytoplasm and the mitochondria, with the latter housing the respiratory enzymes and functioning as hubs that connect diverse metabolic pathways including CCM and amino acid metabolism (Spinelli and Haigis 2018). For instance, amino acid degradation enables the assimilation of nitrogen as ammonium or glutamate via de- or transamination reactions. The remaining carbon backbone is recycled into the cell's biomass or excreted, and the associated metabolites affect carbon metabolism (Godard et al. 2007). Importantly, mitochondrial intermediates are required for amino acid biosynthesis even during fermentative energy generation (Malecki et al. 2020). In fission yeast, a single point mutation in the pyruvate kinase Pyk1, affecting its activity, has been shown to rebalance the fluxes through the fermentation and respiration pathways alongside shifts in the transcriptome and proteome composition (Kamrad et al. 2020), giving a prime example of how the cell co-

adjusts perturbations in metabolic fluxes and expression burdens. Taken together, shifts in the metabolic demand propagate throughout the cell, as most metabolic pathways are tightly interlinked.

The expression levels of CCM enzymes, and therefore the fluxes through the CCM pathways, depend on the growth environment, in particular on the nutrient conditions and presence of stressing agents. For example, during rapid growth on glucose, yeast utilises the fermentative pathway alongside the TCA cycle even in the presence of oxygen, a phenomenon known as aerobic glycolysis or the Crabtree effect (Shimizu and Matsuoka 2018). Aerobic glycolysis is also a characteristic of tumour cells, for which it is known as the Warburg effect (Vander Heiden et al. 2009). This strategy appears counterintuitive as fermentation generates fewer molecules of ATP per glucose molecule than respiration. Several hypotheses have been proposed to resolve this paradox. All require a second growth-limiting constraint besides glucose uptake which would be specific to respiro-fermentative growth (de Groot et al. 2019). Examples include the cytoplasmic density of macromolecules (Vazquez et al. 2008; Goelzer et al. 2015), total proteome allocation (Basan et al. 2015), and membrane area availability (Szenk et al. 2017).

A considerable amount of information on metabolism has been collected at the whole-cell level, and for many organisms a full mapping of the metabolic reactions is widely available (Kanehisa et al. 2017; Keseler et al. 2021). Using constraint-based approaches such as flux balance analysis, considerable progress has been made in the understanding of metabolic fluxes through the network of reactions (Orth et al. 2010). However, such approaches typically do not account for the considerable burden associated with producing proteins. Thus, a whole-cell understanding of cellular trade-offs between multiple constraints must take into account gene expression alongside metabolic maps (Goelzer and Fromion 2017; Yang et al. 2018; Dahal et al. 2020). Resource allocation constraints have been successfully introduced into genome-wide metabolic models of several organisms as more high-quality expression data has become available (O'Brien et al. 2013; Sánchez et al. 2017; Y. Chen et al. 2021). In summary, quantitative surveys of the gene expression cost of metabolic pathways are key to understanding cell physiology.

1.2 Coarse-grained modelling of unicellular organisms

The concept of proteome sectors has been the basis of several phenomenological and coarse-grained mechanistic models relating optimal resource allocation to pro-

tein abundance and cellular growth rates (Molenaar et al. 2009; Scott et al. 2014; Maitra and Dill 2015; Weiße et al. 2015; Maitra and Dill 2016; Pandey and Jain 2016; Liao et al. 2017; Bertaux et al. 2020; Hu et al. 2020). In coarse-grained models of proteome allocation, large sectors of the proteome are abstracted into a single protein, whose kinetics are explicitly described. Existing models include various levels of detail, particularly in the extent of the gene expression machinery that is taken into account. Here, I will introduce the main ideas implemented in five such models together with their main applications. In chapter 3, I will provide a methodological basis with which to understand the modelling performed in chapters 4 and 5.

1.2.1 Brief summaries of five coarse-grained resource allocation models

First, Maitra and Dill (2015) formulated a minimal model of cells with a single variable representing global energy levels, and dividing biomass between metabolic enzymes and ribosomal proteins. The same authors later extended this to include the effect of ribosomal inhibitors (Maitra and Dill 2016). They considered the energy efficiency (growth rate divided by energy production flux) in steady-state growth, which they showed to depend on the allocation of ribosomes; the optimal efficiencies were close to known experimental observations. It was assumed that ribosomal proteins were stable, but non-ribosomal proteins were degraded; with this assumption the nonzero intercept of the ribosomal growth law was proportional to the degradation rate. Under slow growth, the cell spent many resources replenishing proteins that were being degraded, but the efficiency increased with the growth rate as increasing levels of external sugar lead to a larger allocation towards ribosomal proteins.

Next, Weiße et al. (2015) constructed a model incorporating three important trade-offs in cellular growth, coupling gene expression to growth. The trade-offs considered were incorporated by means of finite pools of (i) intracellular energy, utilised in all biochemical processes contained in the model; (ii) ribosomes, which produced all proteins; and (iii) cell mass, meaning that production of one type of protein came with an opportunity cost. This model included both transcription and translation, and the availability of intracellular energy levels was assumed to determine transcription and translation rates. Most parameters were taken from the *E. coli* literature, except for a handful of gene expression parameters. These came from a Bayesian (Markov-chain Monte Carlo) fit to the ribosomal growth law

data by Scott et al. (2010). The model provided a good balance of interpretability and detail, allowing the authors to explain gene dosage compensation, the interplay between a synthetic circuit and its host cell dynamics, and the evolutionary stability of metabolic gene expression regulation. One notable interpretation of this model was that of the ATP-equivalent intracellular energy variable. This was consumed by translation and was therefore alike to amino acids. Furthermore, a good fit to the ribosomal growth law required ribosomal mRNA transcription to be induced at energy levels two orders of magnitude less than non-ribosomal mRNA.

A detailed but still coarse-grained model of resource allocation, growth, and gene expression in bacteria was developed by Liao et al. (2017). They described the dynamics of three protein species, including ribosomal proteins, with their corresponding transcripts, as well as rRNA and tRNA, ATP and free amino acid levels, RNA polymerase levels, and the abundance of the master regulator ppGpp. The model was used to investigate the burden of synthetic constructs, provided a single-cell description of a noncooperative positive feedback loop, and the dynamics of a toggle switch at single-cell, population, and ecological scales. The authors used a combination of widely available data, parameters inferred in earlier studies, and additional Bayesian parameter inference to infer all of 97 parameters.

As seen by this example, complex models accounting for many biological processes also require many parameters, whose inference is nontrivial. This is especially important for research into organisms other than *E. coli*, because fewer experimental results are available. In contrast, a more minimal approach was pioneered by Molenaar et al. (2009), who limited themselves to describing protein abundances and considered the proteome allocation that maximised the growth rate instead of implementing explicit regulatory processes. They described a two-step carbon catabolism with two parameterizations representing metabolically and catabolically efficient metabolism, nitrogen uptake, and ribosomal consumption of both to supply protein production. Molenaar et al. (2009) showed that both growth-rate-correlated gene expression and discontinuous switches between metabolic strategies at specific growth rates could result from simple optimisation of ribosomal allocation for maximal growth rate.

Such a minimal, protein-only, approach was also taken in our group by Bertaux et al. (2020), from which the coarse-grained models described in this thesis are a further development. The model by Bertaux et al. (2020) will be described in detail in section §3.4. It introduced explicit cell cycle proteins to describe the dependence of cell size as a function of growth rate under three types of modulation: nutrient

limitation, expression of useless proteins, and translational inhibition. One innovation was the inclusion of metabolites in the system size. Furthermore, unlike the growth-rate maximization employed by Molenaar et al. (2009), Bertaux et al. (2020) considered a simple regulation of ribosomal allocation as proportional to their substrate and showed that this strategy obtained near-optimal growth rates.

1.2.2 Consensus model behaviours and important differences

A consensus result of all these models is a hyperbolic dependence of the growth rate on the concentration of external nutrients. This is called the Monod law and it is one of the first global experimental observations of bacterial cultures (Monod 1949). The models also all reproduce the ribosomal growth law, and in fact both the Monod law and the ribosomal growth law have been used to quantitatively fit the model parameters. As mentioned above, the allocation to ribosomes contains a sizeable offset when extrapolated down to zero growth. In the model by Maitra and Dill (2015), this result relies on the turnover of degraded non-ribosomal proteins, whereas in the models by Weiße et al. (2015) and Liao et al. (2017) this behaviour emerges as ribosomes not actively translating mRNAs are more abundant at slow growth.

While these five coarse-grained resource allocation models all provide a better understanding of growth-rate-correlated protein abundance, the role of non-metabolic "housekeeping" proteins is less clear. Maitra and Dill (2015) and Molenaar et al. (2009) disregarded these proteins entirely. Weiße et al. (2015) explicitly include negative autoregulation of housekeeping proteins to approximate constant expression levels across conditions. For Liao et al. (2017), the housekeeping proteins are expressed constitutively and their proteome fraction is only approximately constant. Lastly, Bertaux et al. (2020) fix the housekeeping allocation manually to a constant fraction.

1.3 Outline of the thesis

In light of the above considerations, the following questions arise: How does the growth environment influence the eukaryotic cell's allocation of macromolecular resources, such as ribosomes and polymerases, towards various gene expression programmes? What is the interplay between growth, gene expression, the cell cycle, and allocation? To what extent are findings replicated across the tree of life,

and which phenomena are specific to eukaryotes? The main question explored in this thesis is, then: *how does fission yeast allocate resources to biomass production across conditions that support different growth rates*?

I approached these questions by two very different methods: (i) data analysis of genome-wide expression data, and (ii) coarse-grained mathematical modelling of growth-optimised cells. In the remainder of this thesis, the two approaches are described separately. In chapter 2, I describe my analysis of a multi-omics data set that describes the relative abundances of both protein and RNA levels across growth conditions with varying nitrogen sources. The data gathering and analysis was a large joint effort and has recently been published (Kleijn et al. 2022). In the remaining chapters, I describe the coarse-grained mathematical modelling. First, chapter 3 gives a general description of the main methodology. Then, chapter 4 describes successively more complicated models of ribosomal resource allocation, applied to conditions with varying carbon and nitrogen concentrations as well as a representation of the growth media with varying nitrogen sources from the experimental data. Finally, chapter 5 postulates a whole-cell model that additionally accounts for the size and shape of the cell.

2 Data analysis of growth-rate correlated and medium-specific gene expression in fission yeast

2.1 Introduction

As explained in the general introduction, the expression of large classes of proteins (proteome sectors) correlates with the growth rate in both bacteria and budding yeast. This is particularly important for the ribosomes, whose connection with the growth rate has been particularly well studied. With this in mind I analysed a large data set comprising both proteomics and transcriptomics data of *S. pombe* grown across a range of growth rates. While my principal objective was to study the extent of growth-correlated gene expression in fission yeast, I also performed a thorough exploratory analysis of the data set.

As I will explain in this chapter, my analysis highlighted the coexistence of both considerable growth-rate correlations and medium-specific gene expression. In order to fairly assess the correlations in the presence of medium-specific outliers, I used robust linear models that have not been previously used in this context. As expected, protein synthesis was strongly positively correlated with growth, whereas many genes negatively correlated with the growth rate were involved in the environmental stress response. Furthermore, I found that metabolic enzymes were under considerable condition-specific regulation, which is notable because they represented up to 70% of the proteome by mass. Lastly, I proposed signatures of medium-specific gene expression after subtraction of the growth rate effect, using a custom normalisation procedure.

2.1.1 Proteomics and transcriptomics

Transcriptomics and proteomics analyses have been widely used to study the interplay between growth and gene expression, and some important results have already been touched upon in chapter 1. Here I give some further background on important omics studies in the field.

Recall that the "ribosomal growth law" refers to the observation that ribosomal abundance is linearly and positively correlated with the growth rate. This obser-

vation was confirmed in genome-wide proteomics studies in *E. coli* by Peebo et al. (2015), who studied continuous cultures limited by carbon availability; by Hui et al. (2015) under titrations of carbon and nitrogen, and under translational inhibition, and most convincingly by A. Schmidt et al. (2016) in an extensive study of 22 growth conditions. An early transcriptomics study in *S. cerevisiae*, using microarrays, showed strong correlations between hundreds of transcripts with the chemostat dilution rate across six nutrient titrations (Brauer et al. 2008). Many ribosomal transcripts were positively correlated with the dilution rate, in agreement with the ribosomal growth law; furthermore, genes involved in stress response and metabolism were generally negatively correlated with the dilution rate. The *S. cerevisiae* proteome also exhibited the ribosomal growth law, as determined by Metzl-Raz et al. (2017), who combined existing data sets of cultures grown in a variety of carbon sources (Paulo et al. 2015, 2016) with their own data obtained under nitrogen and phosphorus limitation. In summary, proteomics and transcriptomics data have provided a very direct way to quantify the extent of the ribosomal growth law in particular and of growth-rate-correlated gene expression in general.

Simple correlations with the growth rate are by no means the only feature of these genome-wide omics data sets though. Besides being the first to directly confirm the ribosomal growth law in the eukaryote proteome, Metzl-Raz et al. (2017) also observed a sizeable allocation towards ribosomes when the ribosomal growth law was extrapolated to zero growth. To explain this, they proposed the presence of a pool of non-translating ribosomes, which can serve as a buffer during changing growth conditions. This strategy has also been observed in prokaryotes (Dai et al. 2016; Mori et al. 2017; Kohanim et al. 2018). The idea that proteins are expressed in excess of the minimally required capacity has been reported for metabolic pathways as well, including glucose catabolism (R. Yu et al. 2020). Further omics studies in *S. cerevisiae* have defined additional characteristics of resource allocation such as reallocation of proteome mass from amino acid biosynthesis to protein translation upon amino acid supplementation (Björkeroth et al. 2020), or the respective contribution of transcription and translation to different allocation strategies (R. Yu et al. 2021). Altogether, genome-wide omics experiments have been key to understand resource allocation in both *E. coli* and *S. cerevisiae*. The study presented in this chapter adds to this debate both by its different choice of model organism, as well as its treatment of the dichotomy between growth-rate correlations and medium-specific expression.

2.2 Methods

2.2.1 Contributions

In this section I explain the methods used in the processing of the omics data and the exploratory analysis and in-depth treatment of the growth rate models, as well as the details of the functional analyses. While I performed some of the experiments, I did not substantially contribute to their design, and I gratefully acknowledge the work of Amalia Martínez-Segura, François Bertaux, Malika Saint, and Holger Kramer in this regard. The work on growing fission yeast on nitrogen sources to obtain different growth rates was initiated in the Marguerat lab by Malika. François developed and calibrated the turbidostat setup; he also grew the cultures in the turbidostats for six of the nitrogen sources (Trp, Gly, Phe, Ser, Pro, and Amm). I grew the cultures for the Ile and Glu conditions myself. Amalia extracted the omics samples. The proteomics experiments were run by Holger from the MRC LMS proteomics facility, and the RNA-Seq experiments were performed by the MRC LMS sequencing facility. Amalia and myself jointly worked on the initial stages of the analysis. Readers interested in the details of the experimental procedures or the initial treatment of the RNA-Seq and LC-MS proteomics data are referred to the published paper (Kleijn et al. 2022). All the figures shown in this chapter are my own work.

2.2.2 Normalisation and omics data processing

RNA-Seq

After sequencing, demultiplexing, and genome mapping, the RNA-Seq pipeline yielded raw counts c_{ijk} for each gene i, growth medium j, and biological replicate k. These were normalised using the DESeq2 *estimateSizeFactors* function, yielding size factors S_{jk} for each sample (Love et al. 2014). Using this, normalised counts are defined as

$$n_{ijk} = \frac{c_{ijk}}{S_{jk}}. \tag{2.1}$$

Unless otherwise noted, the downstream transcriptome analyses were performed using these normalised counts, which enable between-sample comparison of the expression of genes or sets of genes.

Proteomics

LC-MS proteomics quantifies peptides, not proteins directly. In sample preparation, proteins are digested into peptides, (for our data, in a two-step process involving Lyx-C and trypsin) and the abundance of the peptides in the digest is what is determined by the analysis. The abundance of the peptides is then aggregated to determine protein abundance. Many peptides, however, cannot be uniquely assigned to a single protein. This happens when two (or more) proteins are close homologues, such that the same peptide could have resulted from either of them. There are a few options to resolve this situation: (i) ignore the peptide entirely in the quantification; (ii) include it for both proteins; (iii) include it for only one protein. The major downside of options (i) and (ii) is that it not only biases the abundance estimate of the individual proteins, but also of *classes* of proteins that the ambiguous ones happen to belong to. This is especially undesirable because most close homologues are functionally similar, such that they often occur together in the downstream analysis. For gene classes that contain many homologues, such as ribosomal proteins, options (i) and (ii) will severely under- or overestimate the total expression of the gene class. With ribosomal proteins so central to the resource allocation problem, we chose option (iii). We assigned the ambiguous peptide to the protein that was most abundant based on the quantification of the non-ambiguous peptides. This process is called "razoring". With each "razor" peptide assigned to a single protein, the sum of the close homologue's abundances is not affected by the ambiguity.

It was recently shown that the relative abundances, normalised using the total proteome mass, were as reproducible as those using an external standard curve (Sánchez et al. 2021). Furthermore, the proteome allocation strategy is defined in terms of the relative mass fractions. We therefore rescaled protein abundances, traditionally reported proportional to protein numbers, to the proteome mass fractions. The considerations from the previous paragraph led us to base the reported protein intensities on identified unique and razor peptides, and intensity-based absolute quantification (iBAQ) was calculated as the raw intensity/number of obtainable tryptic peptides . The data was filtered for detection in all three biological replicates. Subsequently, proteome mass fractions ϕ_{ij} were calculated for each

protein group i, sample from growth medium j and replicate k from the reported protein masses m_i, and the iBAQ quantities B_{ijk} as follows:

$$\phi_{ijk} = \frac{m_i B_{ijk}}{\sum_l m_l B_{ljk}}. \tag{2.2}$$

These proteome mass fractions were used to report protein expression, except in direct comparison to transcriptome abundances, as explained next.

Protein–mRNA ratios

With our data set of matched protein and mRNA expression, we aimed to compare the two measures, both between different conditions, and on a gene-by-gene basis. However, the reported units (normalised counts and proteome mass fractions) do not agree between the two data sets. To compare protein and mRNA expression, we used the relative number fractions for both experiments. These were calculated in the following way. For each protein group or gene i, and sample from growth medium j and replicate k, the proteome number fraction was calculated from the the iBAQ quantities B_{ijk} as

$$\psi^P_{ijk} = \frac{B_{ijk}}{\sum_l B_{ljk}}, \tag{2.3}$$

and the transcriptome number fraction was calculated from the normalised counts n_{ijk} and the transcript lengths l_i as

$$\psi^M_{ijk} = \frac{\frac{n_{ijk}}{l_i}}{\sum_l \frac{n_{ljk}}{l_l}}. \tag{2.4}$$

These two number fractions are directly comparable. Note that the ratio $\frac{\psi^P_{ijk}}{\psi^M_{ijk}}$ is proportional to the ratio of absolute protein and mRNA numbers per cell (denoted N^P_{ijk} and N^M_{ijk}, respectively) with a sample-specific normalisation:

$$\frac{\psi^P_{ijk}}{\psi^M_{ijk}} = \frac{N^P_{ijk}}{N^M_{ijk}} \frac{\mathcal{M}_{jk}}{\mathcal{P}_{jk}}, \tag{2.5}$$

where $\mathcal{M}_{jk}$ and $\mathcal{P}_{jk}$ denote the total number of transcripts and proteins per cell, respectively.

The protein-to-mRNA ratio of a given gene is heavily dependent on the average expression level. To perform meaningful analyses of between-sample protein-to-

mRNA ratio differences for each gene, this effect was removed in the following way (Franks et al. 2017). The residual log-transformed protein-to-mRNA ratio was calculated as

$$\operatorname{resid}\log_2\left(\frac{\psi^P_{ijk}}{\psi^M_{ijk}}\right) = \log_2\left(\frac{\psi^P_{ijk}}{\psi^M_{ijk}}\right) - \operatorname{median}_{m,n}\left[\log_2\left(\frac{\psi^P_{imn}}{\psi^M_{imn}}\right)\right]. \tag{2.6}$$

This way, the across-sample variation in protein-to-mRNA ratios could be compared between genes with wildly varying average expression levels.

2.2.3 Exploratory analysis and visualisation

Spearman correction

Noise in observations causes observed correlations between these observations to under-represent true underlying correlations. This effect, and a proposed method to mitigate it, was already described by Spearman (1904); a level-headed treatment of its meaning and arguments for and against its use is given by Muchinsky (1996). The Spearman correction has previously been applied to omics data (Csárdi et al. 2015; Franks et al. 2017) and we reported both corrected and uncorrected estimates of protein-mRNA correlations. Central to the method are the reliabilities r^P_j and r^M_j of the protein and mRNA data. For our data, these are defined for each condition j as the geometric mean of observed pairwise Pearson correlations between the three biological replicates:

$$r^P_j = \sqrt[3]{\rho^P_{j,1:2}\rho^P_{j,1:3}\rho^P_{j,2:3}} \tag{2.7}$$

and

$$r^M_j = \sqrt[3]{\rho^M_{j,1:2}\rho^M_{j,1:3}\rho^M_{j,2:3}} \tag{2.8}$$

Here $\rho^P_{j,k_1:k_2}$ represents the Pearson correlation between the observed log2-transformed proteome number fractions in condition j for replicates k_1 and k_2, and similarly $\rho^M_{j,k_1:k_2}$ for log2-transformed transcriptome number fractions. Likewise, a first, uncorrected, estimate of the protein-mRNA correlation is calculated as the geometric mean of observed pairwise protein–mRNA correlations ($\rho_{j,k_1:k_2}$with the above index convention):

$$\hat{\rho}_j = \sqrt[6]{\rho_{j,1:1}\rho_{j,1:2}\rho_{j,1:3}\rho_{j,2:2}\rho_{j,2:3}\rho_{j,3:3}}. \tag{2.9}$$

The final, corrected, estimate of the protein-mRNA correlation is given by

$$R_j = \frac{\hat{\rho}_j}{\sqrt{r_j^P r_j^M}}. \tag{2.10}$$

Hierarchical clustering

In order to compare the between-sample variance of gene expression across genes, we computed *z*-scores to normalise for variations in absolute expression levels. For each gene or protein group i in the sample with medium j and replicate k, this *z*-score was calculated as

$$z_{ijk} = \frac{y_{ijk} - \mu_i}{\sigma_i} \tag{2.11}$$

from the expression values y_{ijk}, where μ_i and σ_i are the mean and standard deviation of the expression of gene i across all samples.

These z-scores were used to perform hierarchical clustering (HC) and principal component analysis (PCA). The analysis was performed only on genes or protein groups that were detected across all 24 samples. HC on genes/protein groups was performed using the Euclidean distance and Ward linkage ("ward.D2"), using the 'hclust' implementation of the R statistical language (v.3.5.3). PCA was performed using the 'prcomp' function, also from R, with no further zero-centering or scaling applied.

In the transcriptome analysis, separate dendrograms were constructed for coding and non-coding RNAs, using the protein-coding list from PomBase and selecting ncRNAs from the presence of "NCRNA." in the systematic IDs.

Barcode plots

To visually compare the expression patterns of multiple proteins across conditions, I developed a "barcode" plot, such as shown in Figures 2.26 and 2.29. The directed length l_{ij} of the bar for protein i and medium j was calculated from the median proteome mass fractions across the three biological replicates:

$$x_{ij} = \underset{k=1,2,3}{\text{median}}\, \phi_{ijk}, \tag{2.12}$$

and the median across all samples,

$$M_i = \underset{j,k}{\text{median}}\, \phi_{ijk}, \tag{2.13}$$

in the following way:

$$l_{ij} = \frac{x_{ij} - M_i}{M_i}.$$

Missing data was imputed to $x_{ij} = 0$, and the scale was capped at $-1 \leqslant l_{ij} \leqslant 2$.

2.2.4 Growth rate correlations as quantified by repeated-median linear models

Repeated-median linear models

For many genes, differential expression was observed between two or more growth conditions in addition to the underlying growth-rate correlations. This affected the fit quality of the standard ordinary least squares (OLS) linear model fits. It was especially noticeable when differential expression occurred in growth media with high leverage, i.e. in the conditions with the lowest growth rate (Trp and Gly) or the highest growth rate (Amm). Some examples showing this effect are shown in Figure 2.6D and F. We therefore needed to use a method that was able to estimate linear regression slopes and intercepts in the presence of outliers.

When estimating the average of a stochastic variable across a population when outliers are present, the median is a more appropriate measure than the mean. Similarly, when fitting a linear model to outlier-ridden data, median-based methods are more appropriate than the OLS fit. Alternatively, methods may be chosen that minimise something other than the least-squares difference between fit and data, though these require tuning of hyperparameters (de Menezes et al. 2021); this was not further explored. Two median-based linear models were conveniently implemented in the R package 'mblm' (Komsta 2019): the Theil–Sen estimator (Theil 1950; Sen 1968) and the repeated-median linear model (RMLM) (Siegel 1982). The former appears to be more commonly used, but the latter has the advantage that it is robust when up to 50% of outliers are present in the data, as proven by Siegel (1982)—the Theil–Sen estimator is robust in the presence of up to ~29% of outliers. A probable reason for the low popularity of the RMLM estimator is its relatively high computational complexity when the number of data points increases. However, this was not important for our data, because only 24 data points were involved in each fit. We therefore settled on the RMLM method for computing an outlier-robust linear model.

The RMLM fitting procedure works as follows. Consider the data as pairs of an explanatory variable x (in our application, the growth rate) and a response variable

y (a measure of gene expression); the aim is to find parameters A and B for a linear fit

$$y = Ax + B. \tag{2.14}$$

First, for each point (x_i, y_i), a line segment is drawn to each of the other $N-1$ points (μ_j, y_j) where $j \neq i$. The line between points i and j has slope

$$a_{ij} = \frac{y_j - y_i}{x_j - x_i} \tag{2.15}$$

and y-intercept

$$b_{ij} = \frac{x_j y_i - x_i y_j}{x_j - x_i}. \tag{2.16}$$

The parameters A and B are now calculated by a twice-repeated computation of the median. Each point i is associated with a median slope

$$a_i = \underset{j \neq i}{\text{median}}\, a_{ij} \tag{2.17}$$

and y-intercept

$$b_j = \underset{j \neq i}{\text{median}}\, b_{ij}. \tag{2.18}$$

The regression coefficients for the slope and y-intercept of the repeated-median linear model are now defined as the medians of those medians:

$$A = \underset{i}{\text{median}}\, a_i \tag{2.19}$$

and

$$B = \underset{i}{\text{median}}\, b_i. \tag{2.20}$$

Fold change (FC) measure

The regression slope and intercept are both proportional to the average expression level of a gene (protein). A fair comparison of the steepness of the growth-rate dependencies between proteins or transcripts with different expression levels can therefore not use the regression parameters directly. To compare the growth law shapes of protein groups with varying absolute abundances, we defined a custom measure, as a normalised spread of the data according to the best-fit RMLM. It resembles a fold-change difference between the extreme ends of the data and the

median estimate. We therefore named it the fold-change FC and it is defined from the RMLM predictions $y(\mu)$ as the ratio

$$FC = \frac{y\,(\mu = \mu_{max}) - y\,(\mu = 0)}{y\,(\mu = 0.5\mu_{max})}, \tag{2.21}$$

with $\mu_{max} = 0.3\,h^{-1}$. This can be expressed in terms of the fitted slope a and the y-intercept b as:

$$FC = \frac{\mu_{max}}{0.5\mu_{max} + b/a}. \tag{2.22}$$

If expression is directly proportional to the growth rate without a y-intercept, $FC = 2$. Most genes have some baseline expression and a steep growth-rate correlation has $FC \geqslant 1$. Genes without growth-rate correlated expression have $FC \approx 0$; negative FC measures indicate negative correlations.

Sector assignment

To assign genes (proteins) to groups of significant positive or negative correlations, we used a false discovery rate (FDR) based method. We based this on the variance explained, or R-squared R^2,of the RMLMs. For each gene or protein group i, R-squared and the associated p-value was calculated using the 'summary.lm' method as

$$R_i^2 = 1 - \frac{\sum_{j,k} r_{ijk}^2}{\sum_{j,k} \left(y_{ijk} - \mu_i\right)^2}. \tag{2.23}$$

Here r_{ijk} denotes the residuals from the RMLM fit, y_{ijk} the expression (normalised counts or proteome fractions), μ_i the mean expression across samples, and the summation was performed across all N samples where the gene was detected. We calculated the tail-based false discovery rates (FDR, or q-values) and local false discovery rates (fdr) using the 'fdrtool' R package and the false non-discovery rate cut-off method (Strimmer 2008). Genes or protein groups were assigned to the P or R sector when their tail-based $FDR < 0.1$. R and P sector genes had positive and negative slopes, respectively, as determined by the fitted RMLM.

To assess fit quality, in addition to R^2, we used a normalised sum of squared residuals, defined as

$$SSR_{norm,i} = \frac{1}{N-1} \frac{\sqrt{\sum_{j,k} r_{ijk}^2}}{\mu_i}.$$

This resembles the calculation for the coefficient of variation CV, with the residuals replacing the observations themselves.

Removing growth-rate effects from differential expression analysis

To identify differential expression in the transcriptome on top of growth rate mediated effects, we performed an analysis using 'DESeq2' (v1.22.2) from the Bioconductor suite (v3.8) (Love et al. 2014; Huber et al. 2015), comparing the residual expression in each condition to a synthetic reference condition. The fold change obtained by this procedure can be interpreted as the ratio of observed normalised counts and the counts predicted by the RMLM, and the associated p-value provides an interpretable estimate of significance.

The DESeq2 analysis pipeline enables the introduction of per-gene, per-sample normalisation factors that are commonly used to correct for batch-dependent GC-content or length biases. We adapted this functionality to normalise the growth rate bias of each gene, by introducing factors N_{ijk} that converted between the measured *raw* counts c_{ijk} and RMLM-predicted *raw* counts q_{ijk}:

$$q_{ijk} = \frac{c_{ijk}}{N_{ijk}}, \tag{2.24}$$

in analogy to the size factors in equation (2.1). However, the fitting of RMLMs yielded per-gene, per-sample predictions p_{ijk} of the *normalised* counts so further conversion was necessary. Using the sample-dependent size factors,

$$q_{ijk} = p_{ijk} S_{jk}. \tag{2.25}$$

Therefore, the normalisation factors were calculated as

$$N_{ijk} = \frac{c_{ijk}}{S_{jk} p_{ijk}} = \frac{n_{ijk}}{p_{ijk}}. \tag{2.26}$$

We excluded genes with negative predicted raw counts and rescaled the normalisation factors across samples for each gene to have a geometric mean of 1 for numerical accuracy.

Using the RMLM-predicted raw counts, we further defined a synthetic reference condition with three biological replicates. The reference condition is defined from the median count across all growth media for each replicate k:

$$s_{ik} = \text{int}\left(\underset{j}{\text{median}}\, q_{ijk}\right). \tag{2.27}$$

These reference counts were rounded to the nearest integer, as they represent raw counts in the DESeq2 pipeline. By design, the q_{ijk} have no residual growth-rate trend, so a comparison between any growth medium and the synthetic reference gives an estimate for the medium-specific effect after accounting for the observed growth-rate correlations.

Subsequently, we applied the standard DESeq2 pipeline on the constructed data set with 9 conditions: the original 8 and the synthetic one, with each set having 3 biological replicates. We tested for pairwise differential expression between each of the original 8 growth conditions and the synthetic reference, and reported pairwise fold-changes F and the associated *p*-values (both uncorrected and adjusted, p_{adj}) Fold-changes were shrunk using the *lfcShrink* function of DESeq2, using the "apeglm" method (Love et al. 2014; A. Zhu et al. 2019). Genes were reported as differentially expressed (DE) if $p_{adj} < 0.01$ and $|\log_2 F| > 0.5$ for at least one condition.

2.2.5 Functional analysis

Enrichment of Gene Ontology

To assess hypothesised overrepresentation of defined functional categories in the clusters resulting from our hierarchical clustering analyses, and in the P- and R-sectors, we performed the following traditional enrichment analysis. We performed one-sided Fisher exact tests to assess the enrichment of DE genes across the *S. pombe* GO-slims and terms from the biological_process GO with at most 50 annotations in *S. pombe* (Gene Ontology Consortium 2019; Lock et al. 2019). From the resulting *p*-values, local false discovery rates $lfdr$ were calculated using the 'fdrtool's false non-discovery rate method (Strimmer 2008). In the enrichment plots for the GO-slim terms (Figures 2.3, 2.13, 2.14, and 2.33B), terms with $lfdr < 0.05$ were deemed significant, and the terms were ordered from top to bottom by increasing smallest lfdr to aid interpretation. For the biological_process enrichment plot (Figure 2.35), the significance threshold was $lfdr < 0.001$. The

significant terms were clustered hierarchically using the Euclidean distance and Ward linkage ("ward.D2"), using the 'hclust' implementation of the R statistical language (v.3.5.3).The terms were ordered by the smallest `lfdr` as much as possible while remaining consistent with the clustering constraint.

Gene set enrichment analysis

A multilevel gene set enrichment analyses were performed against the three *S. pombe* GO-slims (Gene Ontology Consortium 2019; Lock et al. 2019), using the 'fgsea' package (v1.16.0) (Korotkevich et al. 2019), with boundary parameter $\epsilon = 0$. Genes (protein groups in the proteomics analysis) were ranked based on the following signed measure of significance:

$$-\operatorname{sign}(a_i)\log_{10}(p_i)\,. \tag{2.28}$$

Here a_i and p_i are the slope and *p*-value associated with the RMLM growth-rate fit for gene (protein group) i, and $\operatorname{sign}(a_i)$ equals 1, 0, or -1 when a_i is positive, zero, or negative, respectively. This resulted in a list where the R-sector was at the top of the list, and the P-sector at the bottom. Unlike the traditional functional enrichment however, this analysis does not require an arbitrary cut-off point.

2.3 Results

2.3.1 Eight nitrogen sources spanned an approximately fourfold range of growth rates

S. pombe 972h- prototroph wild-type cells were grown in turbidostats with set maximal optical densities $OD_{600} \approx 0.4$ (Figure 2.1A). Starter cultures were inoculated at approximately $OD_{600} \approx 0.1$, left to reach the set point, and held in the turbidostat for many generations (Figure 2.1B). Approximately halfway, the cultures were diluted twofold and the growth rate was measured by a linear fit of an exponential growth model to the log-transformed OD_{600}. Samples were harvested at the end of turbidostat growth. The whole procedure was repeated in biological triplicates. The harvested samples were subjected to label-free proteomics and RNA-Seq experiments to quantify the proteome and transcriptome in each condition and replicate.

In addition to standard Edinburgh minimal media (EMM2, Petersen and Russell 2016), whose sole nitrogen source is 93.5 mM of ammonium chloride (NH_4Cl, re-

ferred to as Amm) we used seven variations that each contained 20 mM of a single amino acid (Trp, Gly, Phe, Ser, Ile, Pro, Glu) (Fantes and Nurse 1977; Carlson et al. 1999). The growth variables of the cultures are summarised in Figure 2.1C–E. The steady-state growth rates obtained ranged from ca. 0.05 h^{-1} for Trp and Glu media to ca. 0.28 h^{-1} for Amm/EMM2. The cultures were grown for approximately 2–6 days depending on the nitrogen source.

2.3.2 ANOVA results show that many *S. pombe* genes/proteins are differentially expressed

We first asked whether the fission yeast proteome composition differed significantly between the eight growth conditions. Strikingly, ~45% of the 1988 protein groups robustly detected in all samples were significantly more variable across conditions than among biological replicates ($p^{ANOVA}_{adj\ (Holm)} < 0.05$). This pervasive level of gene regulation was also apparent at the transcriptome level where ~52% of mRNAs showed significant variability. These results indicate that the composition of the proteome and transcriptome are both strongly affected by conditions that change the growth rate.

2.3.3 Hierarchical clustering of z-score transformed expression levels distinguished between expression programmes

To investigate this variability further, we used the z-score transformed protein fraction of each gene for hierarchical clustering (Figure 2.2A, Methods). This treatment enabled normalisation for protein expression levels across the proteome while preserving the variation of each protein between conditions. We defined 10 clusters that revealed two major features of the datasets (Figure 2.2A-C). First, all clusters showed clear differences in protein expression across one or more conditions. Second, the expression of several proteins was not strictly condition-specific but instead showed a coordinated linear increase with growth rate (clusters 7 and 8). Interestingly, the total baseline expression of the condition-specific clusters was positively (clusters 7, 8, and 9), or negatively (clusters 1, 2, 3, and 6) correlated with the growth rate. Apart from clusters 1, 4, and 10, clusters were enriched for defined functional categories, indicating that the shifting balance between condition-specific regulation and growth rate regulation may have physiological consequences related to the enriched functions (Figures 2.2A and 2.3).

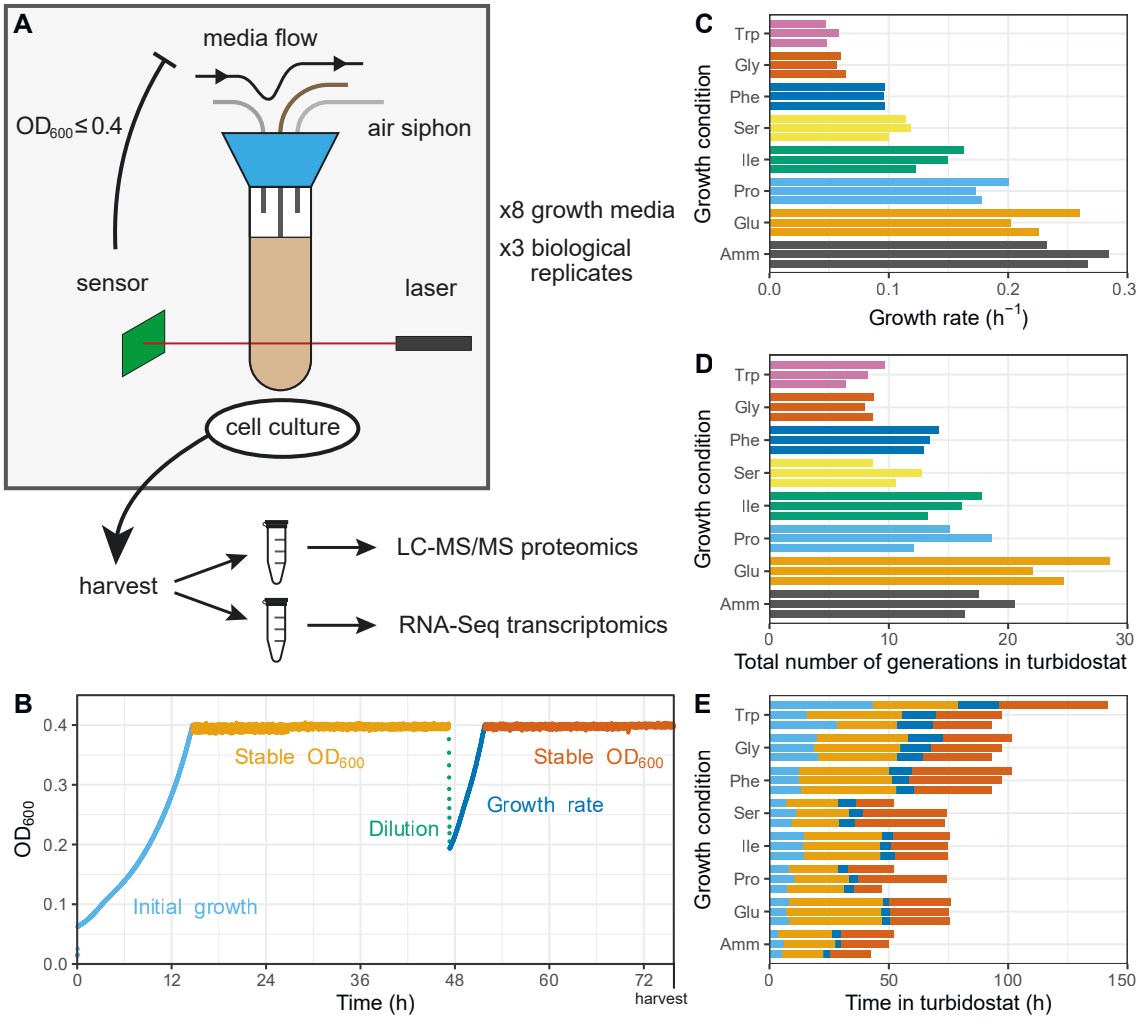
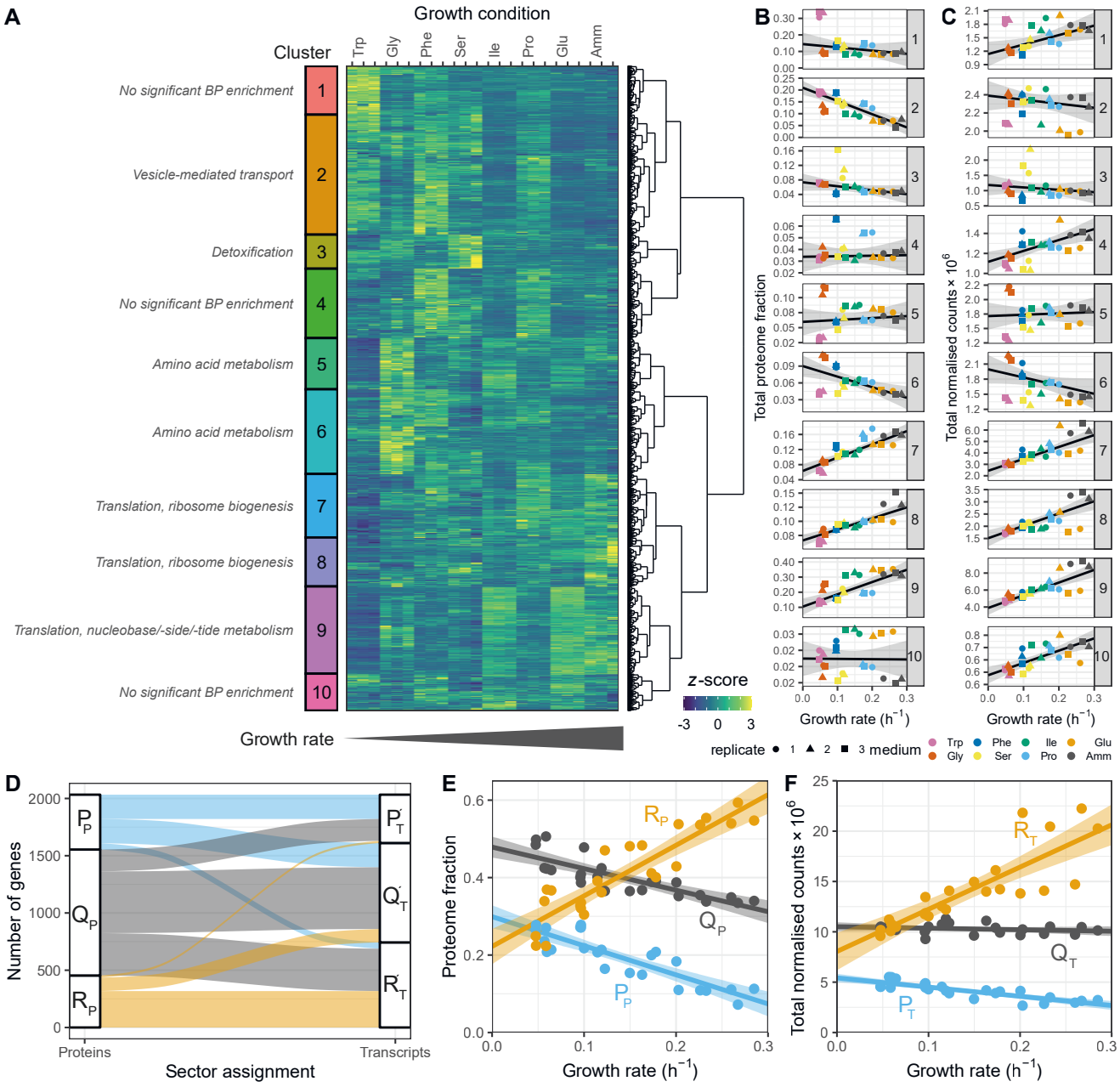


Figure 2.1: Quantification of *S. pombe* cultures grown in eight minimal media with varying nitrogen sources. **A.** Cartoon of the turbidostat culture chamber indicating the control flow and the analysis pipeline. **B.** One of the growth curves (Ile replicate 2) with the turbidostat growth strategy indicated on the plot. **C.** Growth rates μ estimated from the two-fold dilution and regrowth cycle for the 8 growth media and 3 biological replicates. **D.** Total number of generations N_G experienced by each culture. **E.** Total time T in hours that each culture spent in a turbidostat, with the duration of individual growth phases coloured as in **B.** Note that, with $T_d = \frac{\ln 2}{\mu}$ the doubling time, $N_G = \frac{T}{T_d}$.



(Caption on the following page.)

Figure 2.2: Initial exploration of growth-rate dependent and nutrient-specific components in the fission yeast proteome. Figure printed on the preceding page. **A.** Heatmap of hierarchical clustering analysis performed on proteome expression z-scores for 1988 protein groups detected across all eight conditions, which are ordered by increasing growth rate. The clusters are manually annotated with enriched categories, which is shown in more detail in Figure 2.3. The colour scale was truncated to $-3 \leqslant z \leqslant 3$. **B.** Total proteome mass fractions; **C.** total DESeq2-normalised counts for the clusters in A. together with repeated-median linear model (RMLM) fits and symmetric 95% confidence interval (CI). **D.** Flow chart of 2030 proteins detected across all conditions detailing their assignment to the proteome (left) and transcriptome (right) P-, Q-, and R-sector. **E.** Total proteome mass fractions; **F.** Total DESeq2-normalised counts for the proteome (E.) and transcriptome (F.) P-, Q-, and R-sector in orange, grey, and blue, respectively, with ordinary least squares (OLS) best fit and 95% CI.

Both modes of regulation were also apparent in the transcriptome data for coding and non-coding RNA (ncRNA) (Figures 2.2C, 2.4, and 2.5). Interestingly, most ncRNAs showed clear and reproducible condition-specific expression between replicates, suggesting the presence of active regulation, consistent with analyses using different genetic and physiological conditions (Figure 2.5AB) (Atkinson et al. 2018). To test this hypothesis, we compared the expression patterns of ncRNA from each cluster with the expression of their flanking coding genes (Figure 2.5CD). We found that, apart from the growth-rate correlated cluster 1, ncRNA expression patterns were not generally mirrored by the neighbouring mRNA, although the total expression of each cluster's neighbouring genes reflected that of the ncRNA. This indicates that many ncRNA are subjected to some level of independent regulation, but that highly abundant transcription is mirrored between coding and non-coding regions. In summary, we find that regulation of gene expression programmes across conditions that affect the growth rate has two components; one which is condition-specific and another which is coordinated with growth rate.

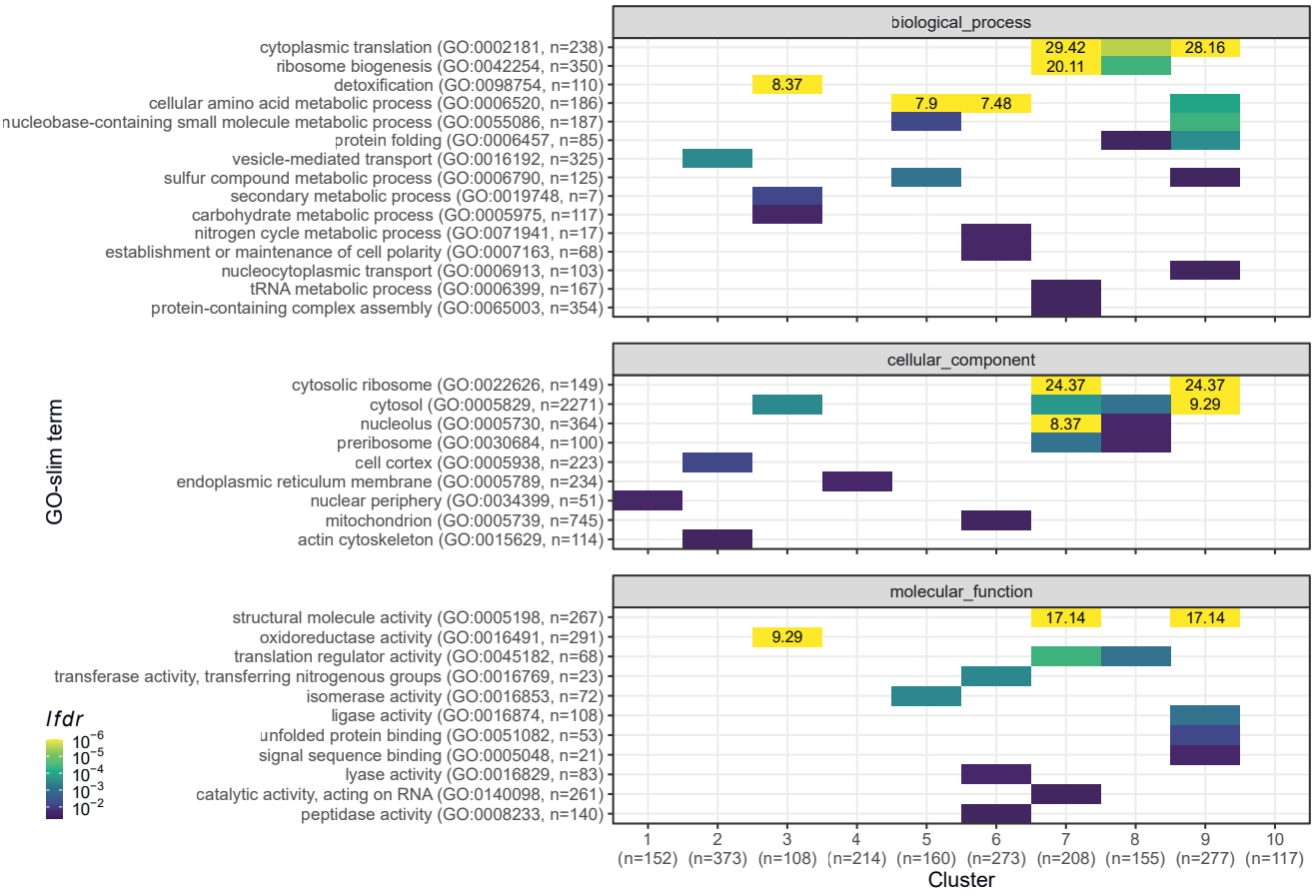


Figure 2.3: GO-slim enrichment analysis of the proteome clusters from Figure 2.2A. Log-transformed local false discovery rates lfdr are shown for the three Gene Ontology namespaces and capped at $\mathrm{lfdr} = 1.0 \times 10^{-6}$ (printed if capped). Overrepresentation analyses were performed as one-sided Fisher exact tests and only lists with $\mathrm{lfdr} < 0.05$ are shown; the numbers of genes in each GO-category and cluster are indicated.

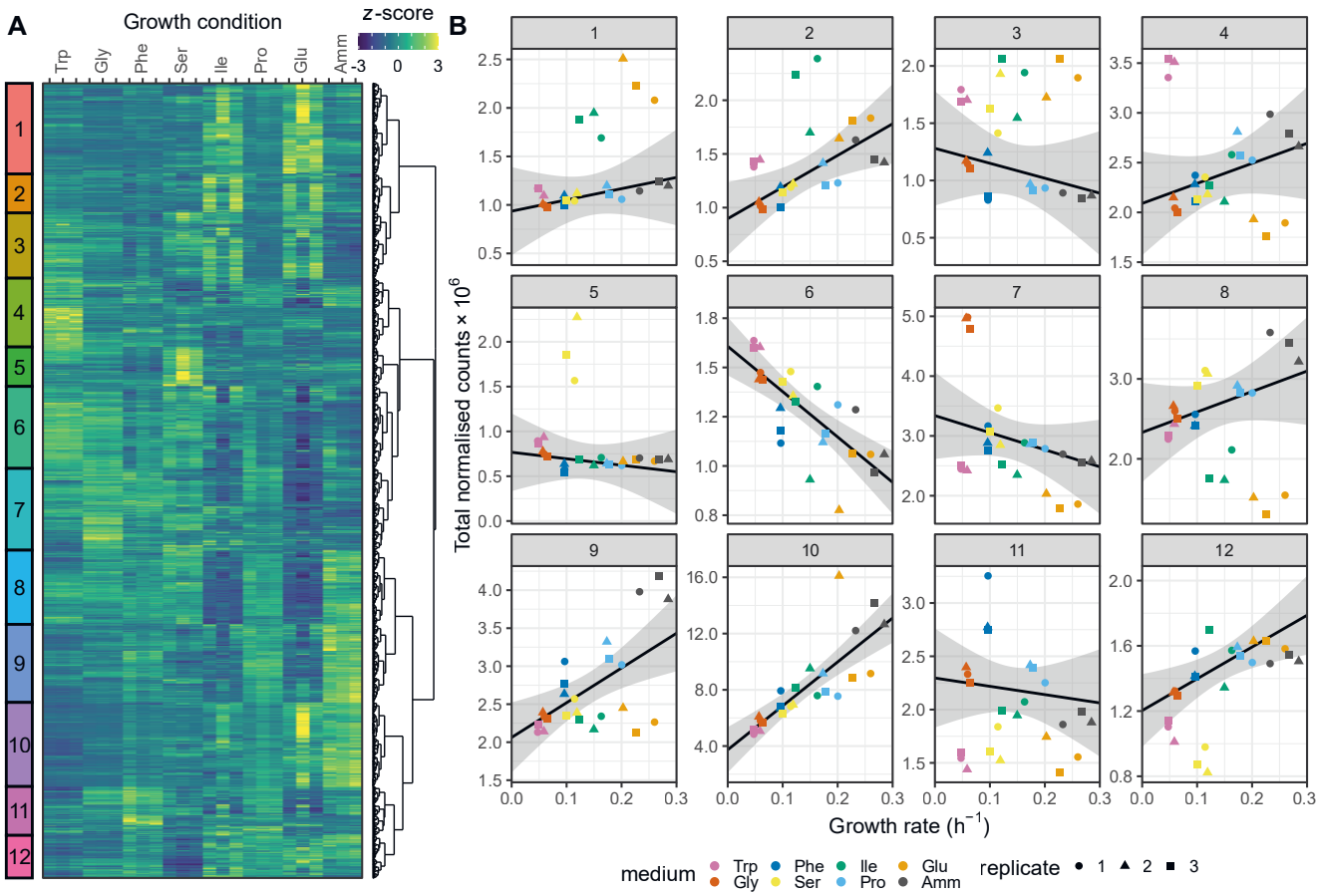
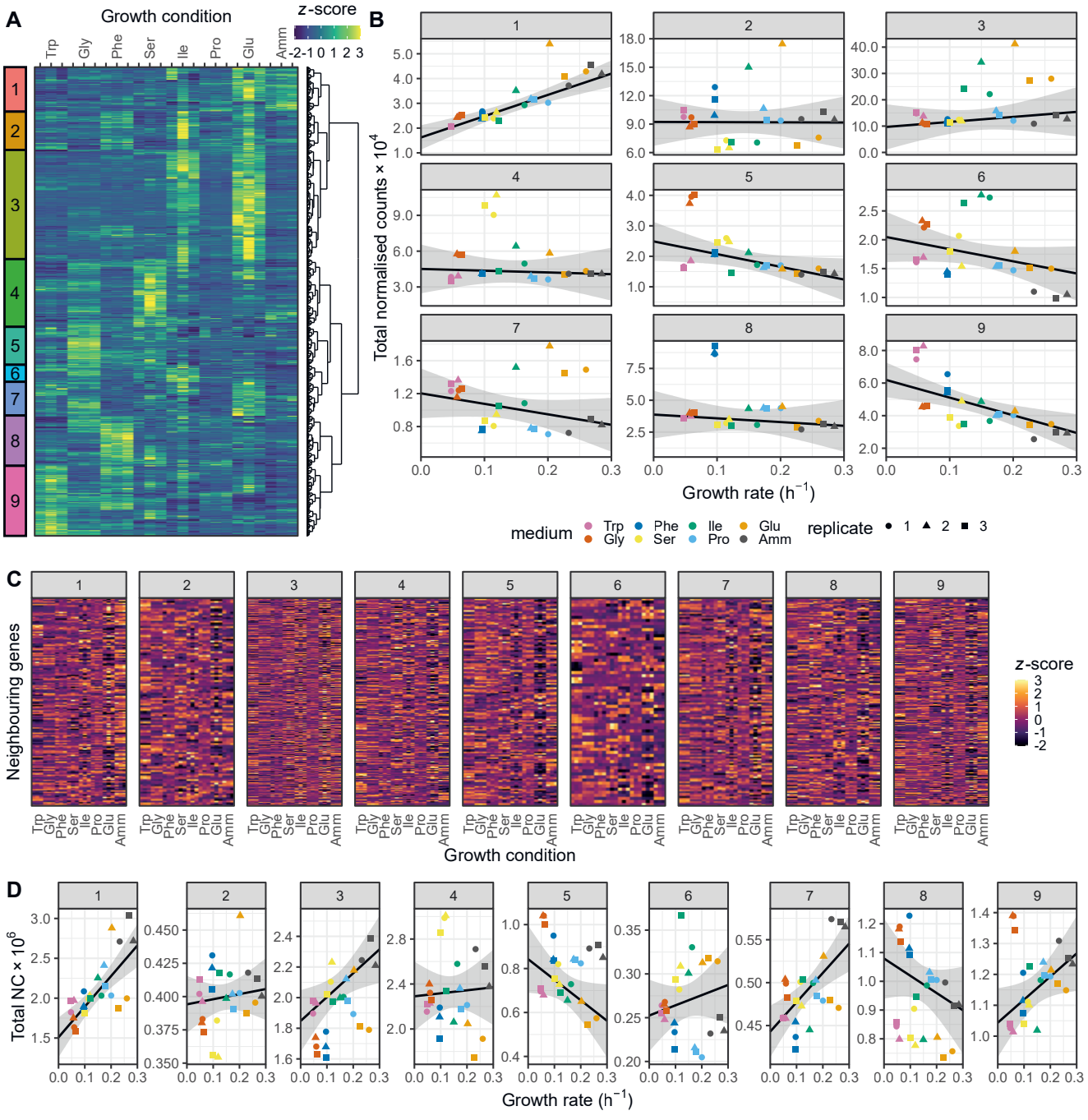


Figure 2.4: Hierarchical clustering analysis of relative mRNA abundances. A. Heatmap of hierarchical clustering analysis performed on transcriptome expression z-scores for 4979 mRNAs detected across all eight conditions, which are ordered by increasing growth rate. The colour scale was truncated to $-3 \leqslant z \leqslant 3$. **B.** Total DESeq2-normalised counts for the clusters in A. together with repeated-median linear model (RMLM) fits and symmetric 95% confidence interval (CI).



(Caption on the next page.)

Figure 2.5: Hierarchical clustering analysis of relative ncRNA abundances. Figure printed on the previous page. **A.** Heatmap of hierarchical clustering analysis performed on transcriptome expression z-scores for 1211 ncRNAs detected across all eight conditions, which are ordered by increasing growth rate. The colour scale was truncated to $-3 \leqslant z \leqslant 3$. **B.** Total DESeq2-normalised counts for the clusters in A. together with repeated-median linear model (RMLM) fits and symmetric 95% confidence interval (CI). **C.** Heatmap of z-scores for all transcripts that neighbour the ncRNAs included in A. separated by cluster from figure A. Annotations were taken from Atkinson et al. (2018). **D.** Total DESeq2-normalised counts for the neighbouring genes from C.

2.3.4 Repeated-median linear model fit parameters classified genes as growth-related

We first focused our analysis on the growth-dependent component of fission yeast gene expression. Linear correlations between the expression of individual genes and the growth rate have been observed in several organisms under different types of growth limitation (Brauer et al. 2008; Hui et al. 2015; Peebo et al. 2015; A. Schmidt et al. 2016; Metzl-Raz et al. 2017; Zavřel et al. 2019). Following the terminology used in prokaryotes, we divided proteins and mRNA into three sectors depending on whether they show a growth-dependent component that was positively (R), negatively (P), or not significantly (Q) correlated with the growth rate (Scott et al. 2010, 2014). We used repeated-median linear models to quantify the linear coordination of each protein and mRNA quantity with growth. This model fits a linear dependence in the presence of large numbers of outliers and is therefore robust to the condition-specific component of gene expression (Figure 2.6).

The linear fits generated two useful parameters. First, the slope of the linear regression is a measure of the strength of the dependence of a protein's concentration on the growth rate. Second, its y-intercept represents the fraction of the protein numbers that is not directly dependent on growth. Both parameters are directly correlated with expression levels making it difficult to disentangle the strength of the growth-rate-related regulation from an mRNA or protein from its abundance. To take this into account, we developed a normalised measure of growth dependence (Methods, Figure 2.6GH), defined as the ratio of the spread in expression level between zero and maximum growth and the median expression. It therefore denotes the "fold change" or FC of growth rate changes relative to an intermedi-

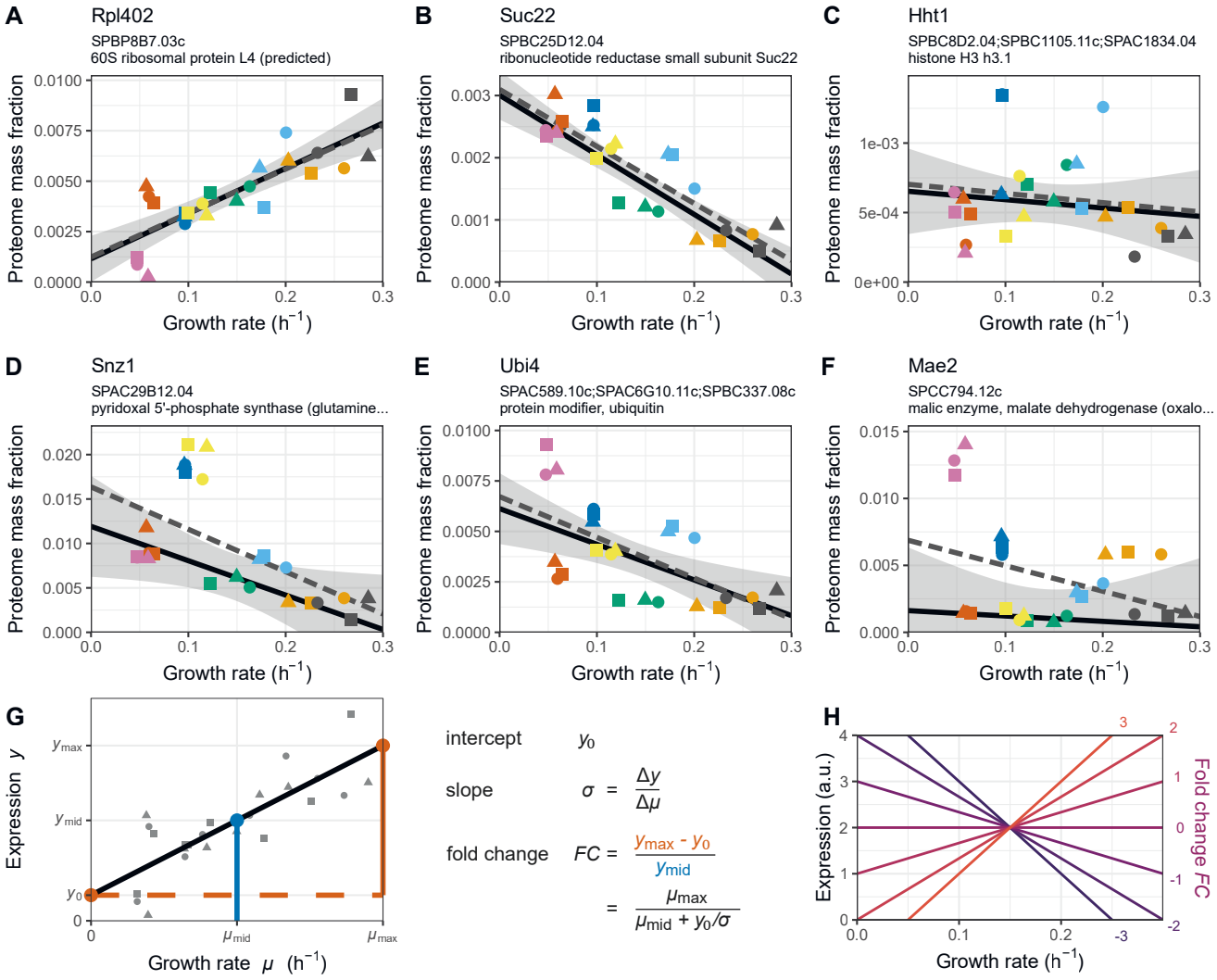


Figure 2.6: Examples of repeated-median linear model fits for selected genes. **A.–F.** Proteome expression and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey, as well as ordinary least squares (OLS) best fit in dashed grey for example protein groups Rpl402, Suc22, Hht1/Hht2/Hht3, Snz1, Ubi3/Ubi4/Ubi5, and Mae2. **G.** Graphical explanation of the fold change (FC) calculation for the example protein Rpl402, indicating the relationship between FC values, slope, and y-intercept. **H.** Illustrative growth law shapes for a series of FC values.

ate baseline. FC values are a combination of the regression slope and y-intercept which do not scale with abundance, thereby enabling a direct comparison of the growth-dependence of single genes or groups thereof.

Repeated-median linear models captured the growth-dependent component of the 10 clusters from Figure 2.2, and proteins from the R and P sectors dominated the clusters that were positively and negatively correlated with growth, respectively (Figure 2.7). Of all the genes detected in the proteome across the eight conditions examined, we found that 22% of proteins and 37% of mRNA belonged to the R sector; similarly, 24% and 21% of the proteins and mRNA belonged to the P sector, respectively. The protein and mRNA of a given gene belonged to the same sector in 53% of the cases (Figure 2.2D). When they did not, the mRNA of P or R proteins were mostly assigned to the Q sector and vice versa, with only 19 R proteins having P sector mRNA, and 55 P proteins having R sector mRNA, out of the 2033 proteins detected.

In quantitative terms, the total proteome mass fraction of the fission yeast R sector ranged between ~20% at zero growth and 55% for the fastest measured growth rate, whereas the mass fraction of the P sector similarly ranged from ~30% to 10% (Figure 2.2E). The sum of all Q sector proteins was negatively correlated with the growth rate because proteome fractions add up to one by definition. However, none of the individual proteins showed significant correlation with the growth rate. At the mRNA level, the R fraction ranged from 38% to 59% of the total normalised counts, and the P fraction from 19% to 10% (Figure 2.2F). Thus, during fast growth, over half of the gene expression burden is dedicated to factors that increase in concentration with growth rate and may therefore be limiting. Moreover, the amplitude of the variability in the concentration of fission yeast proteins and mRNA that depend on the growth rate alone is in the order of magnitude of the cut-offs that are commonly used for differential expression analysis. Therefore, differences in growth rate are important factors that affect interpretation of transcriptomics and proteomics data (R. Yu et al. 2021).

2.3.5 Growth dependent gene expression is preserved between mRNA and protein

Having obtained both transcriptomics and proteomics data for the same cultures enabled us to compare the two levels of gene expression in a unified data set. To perform a like-for-like comparison, we converted our expression measure in both data sets to relative number fractions (Balakrishnan et al. 2021). First, we explored

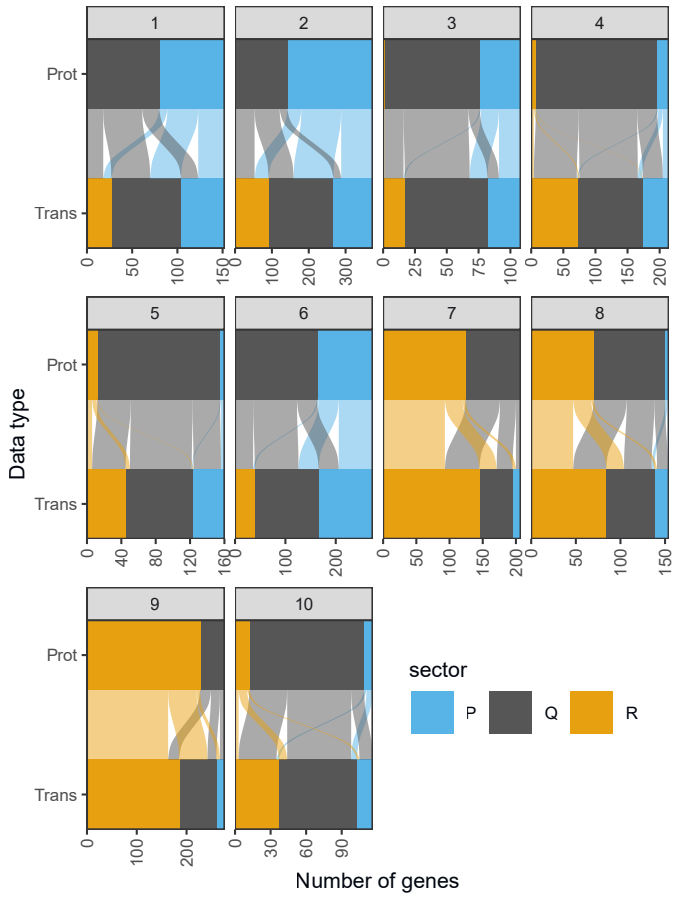


Figure 2.7: Assignment of proteins and corresponding transcripts to growth correlation categories separated by cluster. The hierarchical clusters are taken from Figure 2.2; proteins (Prot) and corresponding transcripts (Trans) are connected by lines with colours that match to the protein classification.

the correlation between protein and mRNA levels averaged across all genes, using the Spearman correction to account for the varying reproducibility of the data (Csárdi et al. 2015; Franks et al. 2017). mRNA reliabilities were in the range of 97.5-99.8% and protein reliabilities were between 92.8-97.6% (Figures 2.8 and 2.9). Spearman-corrected correlations of log-transformed relative protein and mRNA levels were ~0.8 for most conditions, with a slightly elevated correlation in EMM2 reference medium and slightly smaller correlation in Trp medium (Figures 2.10 and 2.11A). Furthermore, we found evidence of post-transcriptional amplification: the ratio of protein to mRNA generally increased with protein expression, but a plateau was reached at very high expressions (Figure 2.12). This agrees with earlier observations (Marguerat et al. 2012). In summary, our analysis indicates that mRNA expression levels are globally good predictors of proteome composition in our system.

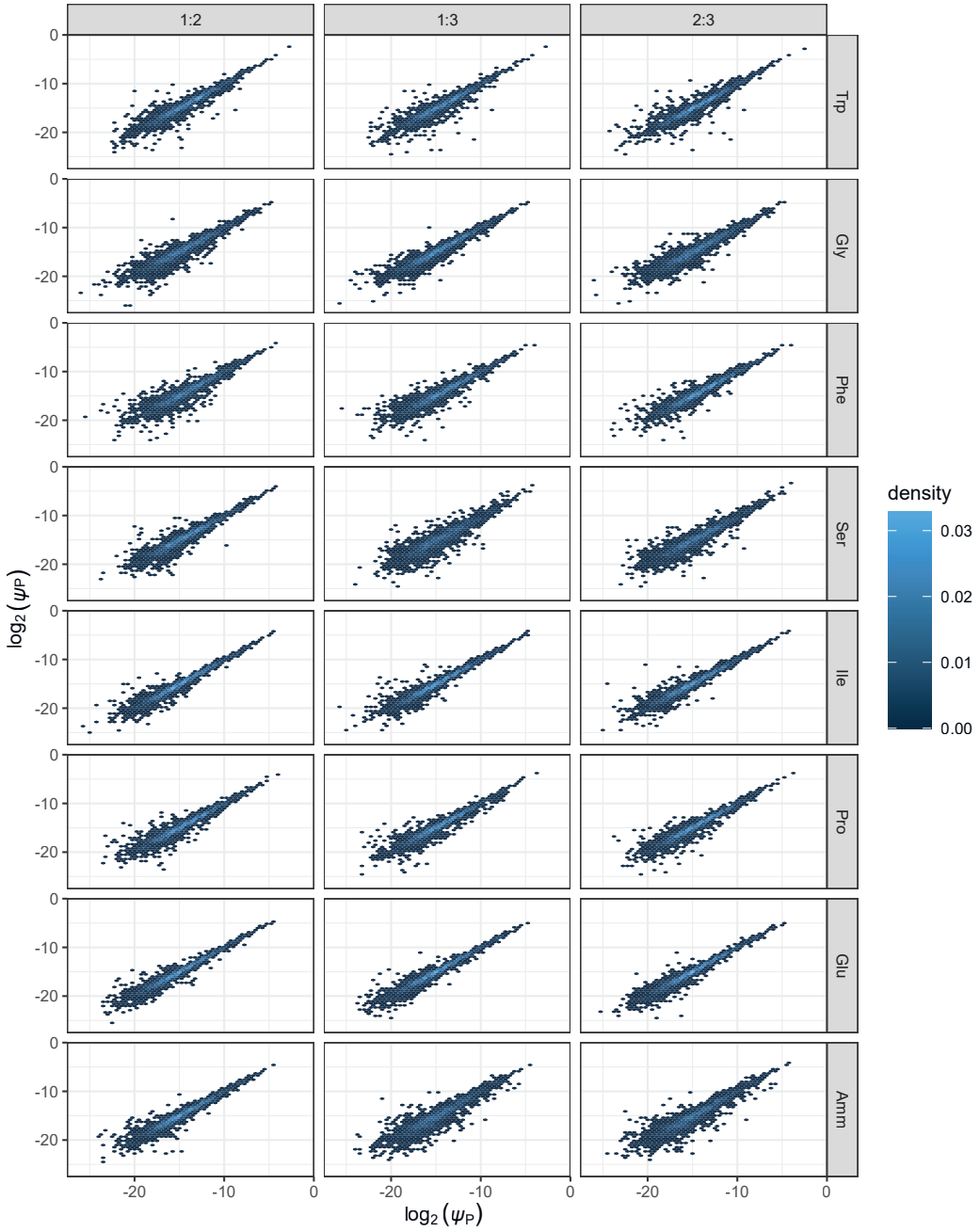


Figure 2.8: Protein number fraction densities comparing expression pairwise between the three biological replicates. Left column: replicate 1 is plotted along the x-axis and replicate 2 along the y-axis; middle column: replicate 1 on the x-axis and replicate 3 on the y-axis; right column: replicates 2 and 3.

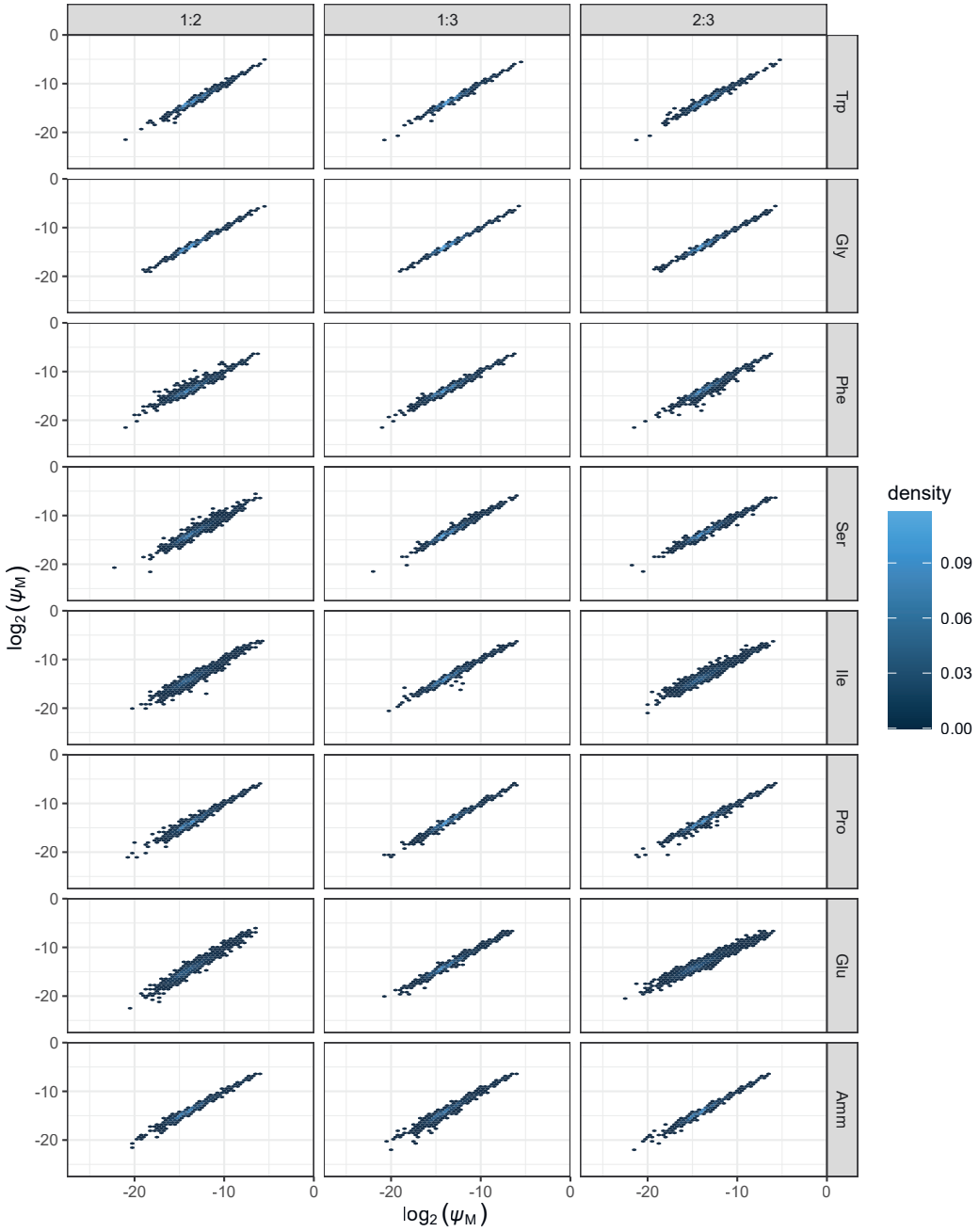


Figure 2.9: RNA number fraction densities comparing expression pairwise between the three biological replicates. Left column: replicate 1 is plotted along the x-axis and replicate 2 along the y-axis; middle column: replicate 1 on the x-axis and replicate 3 on the y-axis; right column: replicates 2 and 3.

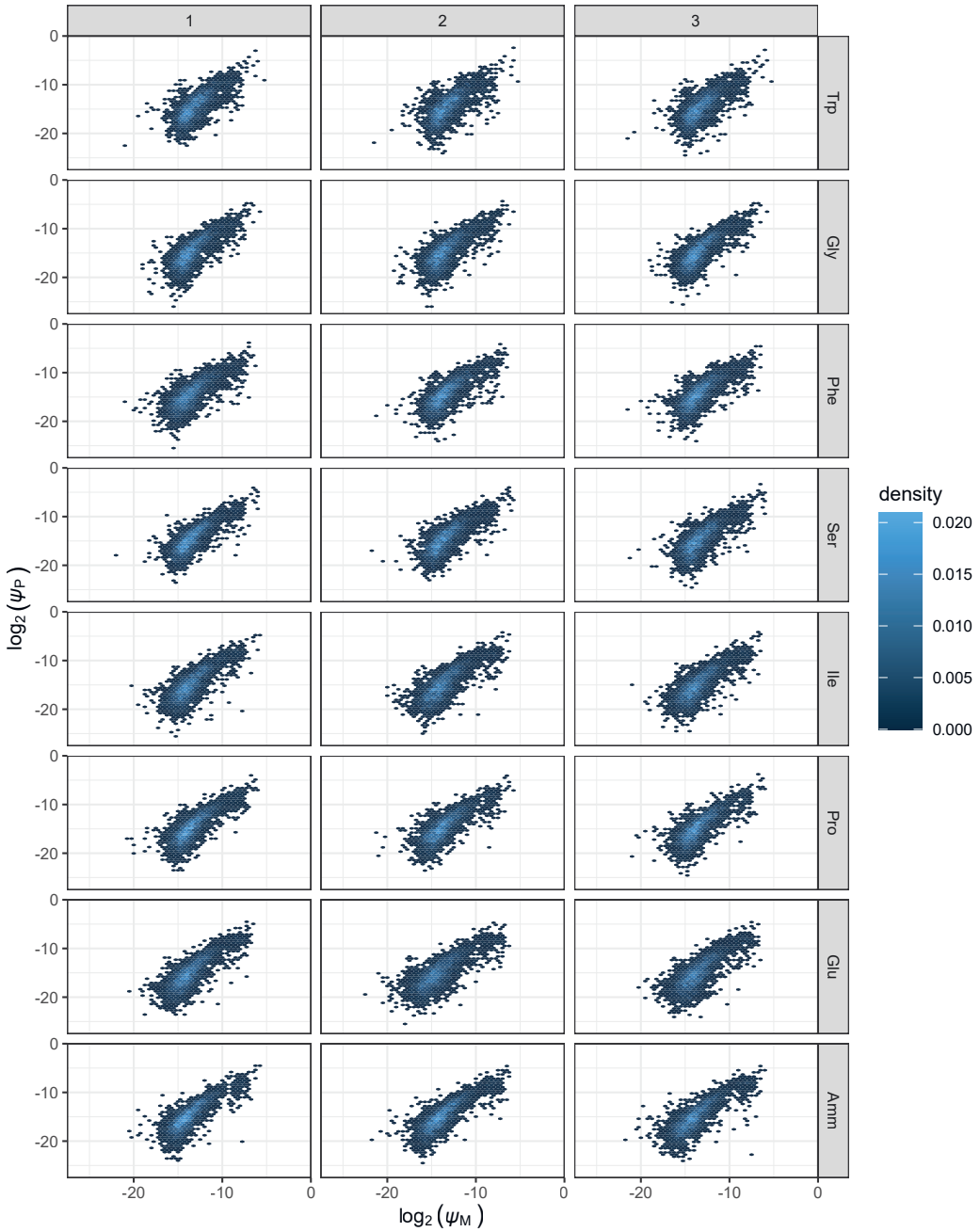
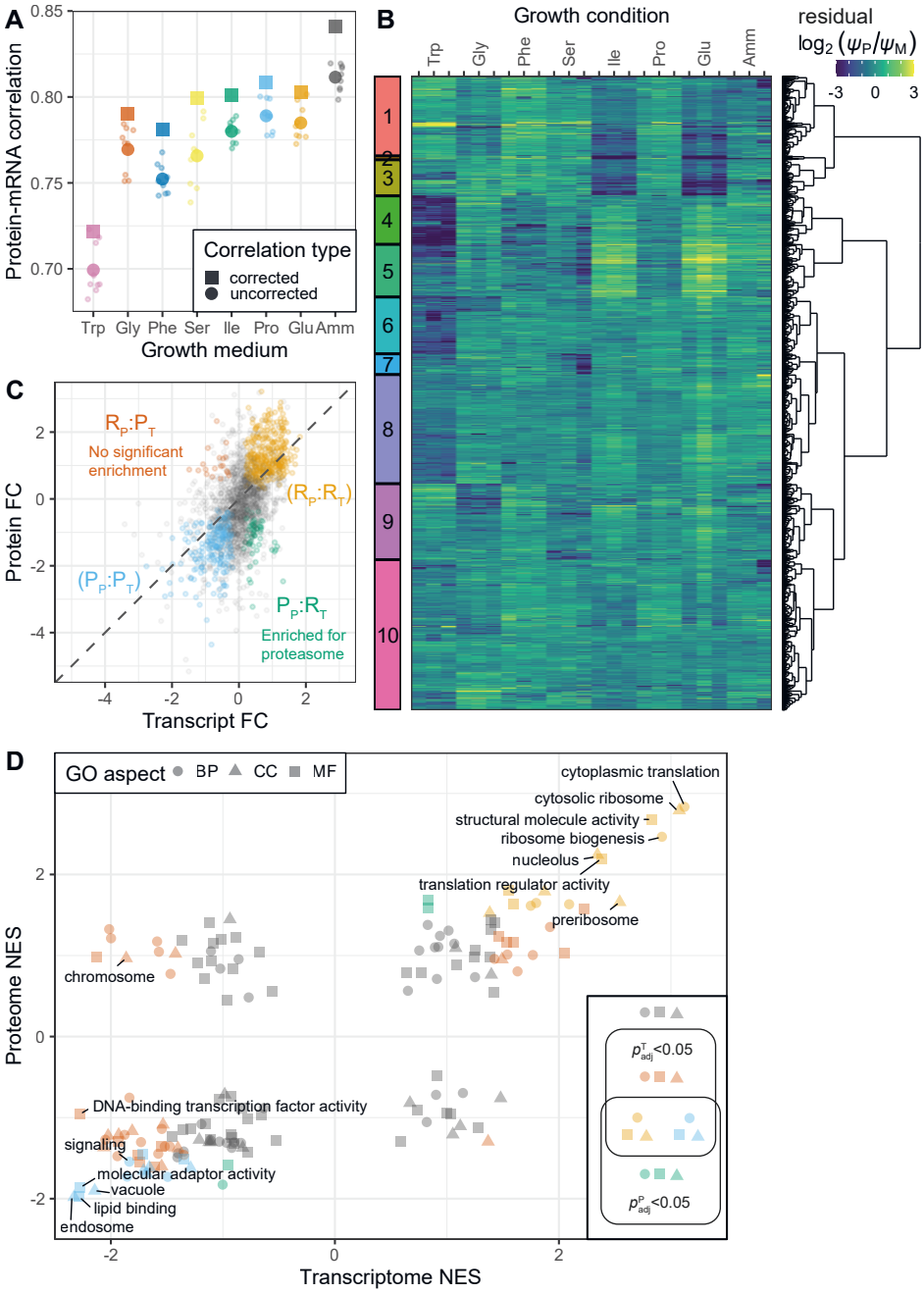


Figure 2.10: Protein/RNA number fraction density plots comparing protein and RNA expression in each sample. All eight growth conditions and three biological replicates are included.



(Caption on the following page.)

Figure 2.11: Comparison of relative protein and transcript expression levels. Figure printed on the preceding page. **A.** Pairwise correlations between mRNA and protein number fractions (small circles) with their geometric means (large circles) and Spearman-corrected best estimates (squares). **B.** Heatmap of residual log2-transformed protein-to-mRNA ratio per-gene per-sample, with genes clustered hierarchically. **C.** Transcriptome and proteome repeated-median linear model (RMLM) fold changes (FC) for 2030 genes detected in all conditions in both data sets. Proteome and transcriptome sectors are indicated by colours for genes significantly correlated with growth rate in both data sets. A functional enrichment of the off-diagonal groups is shown in Figure 2.14. **D.** Scatter plot of proteome and transcriptome normalised enrichment scores (NES) for GO-slim terms subjected to gene set enrichment analyses (GSEAs). For this analysis, genes were ranked based on the *p*-values of their RMLM fit in proteome and transcriptome. GO-slim terms are coloured based on the adjusted *p*-values of their GSEAs; terms with $p_{adj} < 10^{-9}$ in either or both of the analyses are labelled.

Second, we explored the extent of post-transcriptional regulation for each given gene in different growth media. For each gene, we calculated the log2-transformed ratio of protein and mRNA relative number fractions and subtracted from this the median ratio across conditions (Franks et al. 2017). Subsequently, we performed hierarchical clustering analysis (Figure 2.11B) and a functional enrichment of the clusters (Figure 2.13). There was little if any growth-dependent variation in the resulting residual protein-to-RNA ratios (Figure 2.11B). However, some genes showed signs of medium-specific post-transcriptional regulation, prominent in Trp, Ile and Glu (clusters 1–6). Notably, clusters 4 and 5 contained genes with elevated protein-to-mRNA ratios in Ile and Glu, but repressed ratios in Pro, Ser, Phe, and (chiefly) Trp. The enrichment analysis highlighted a moderate enrichment for metabolism in cluster 5.

Next, we compared the size of the growth-rate dependent effects between protein and mRNA by contrasting the fold change measures of genes present in both data sets. As shown in Figure 2.11C and in accordance with Figures 2.2D and 2.7, the RMLMs showed good agreement between the two types of data. Protein FC measures were generally larger than transcript FCs, again highlighting the post-transcriptional amplification. Further study of the disagreeing genes showed a

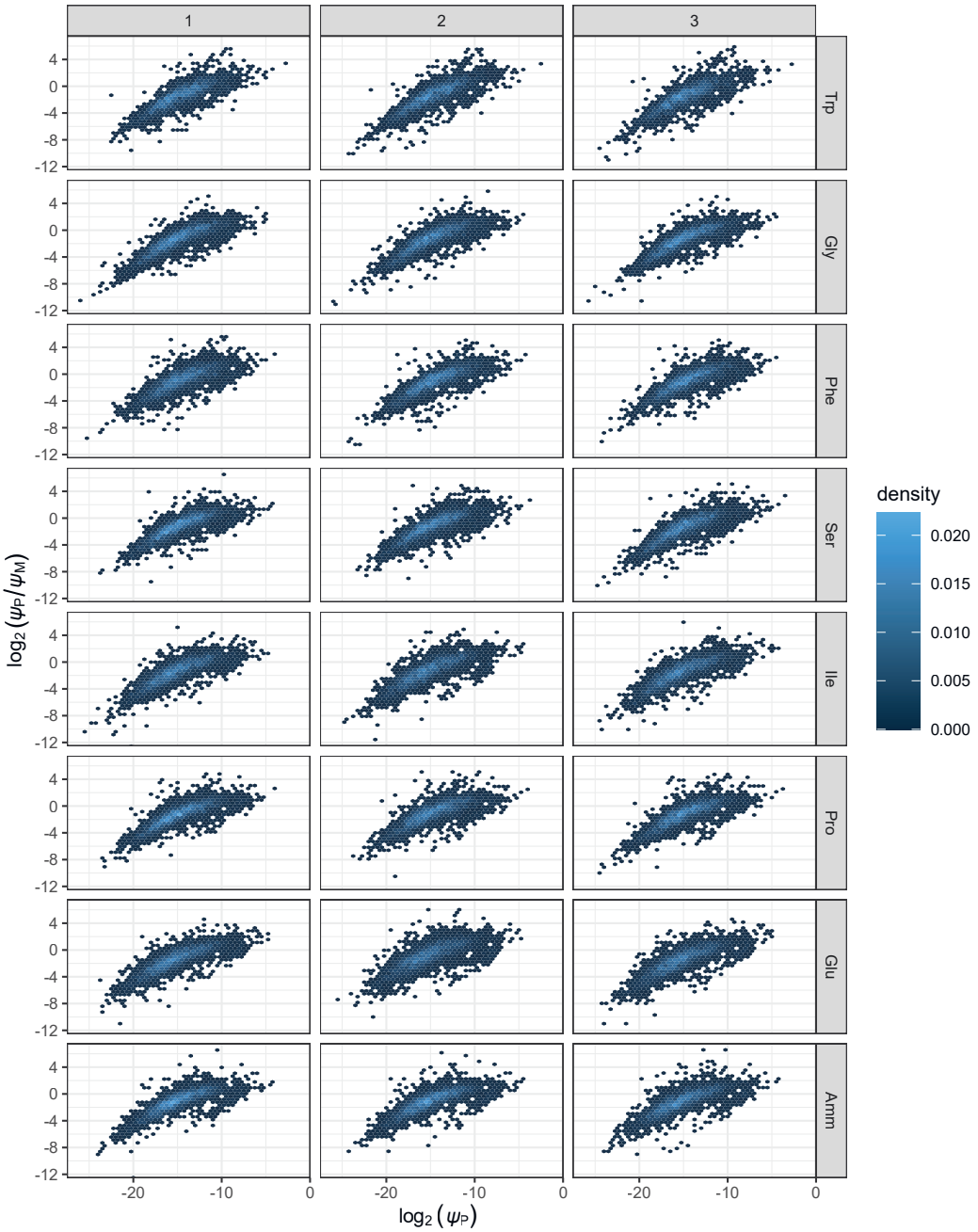


Figure 2.12: Density plots of protein number fractions versus protein-to-mRNA ratios in each sample. All eight growth conditions and three biological replicates are included.

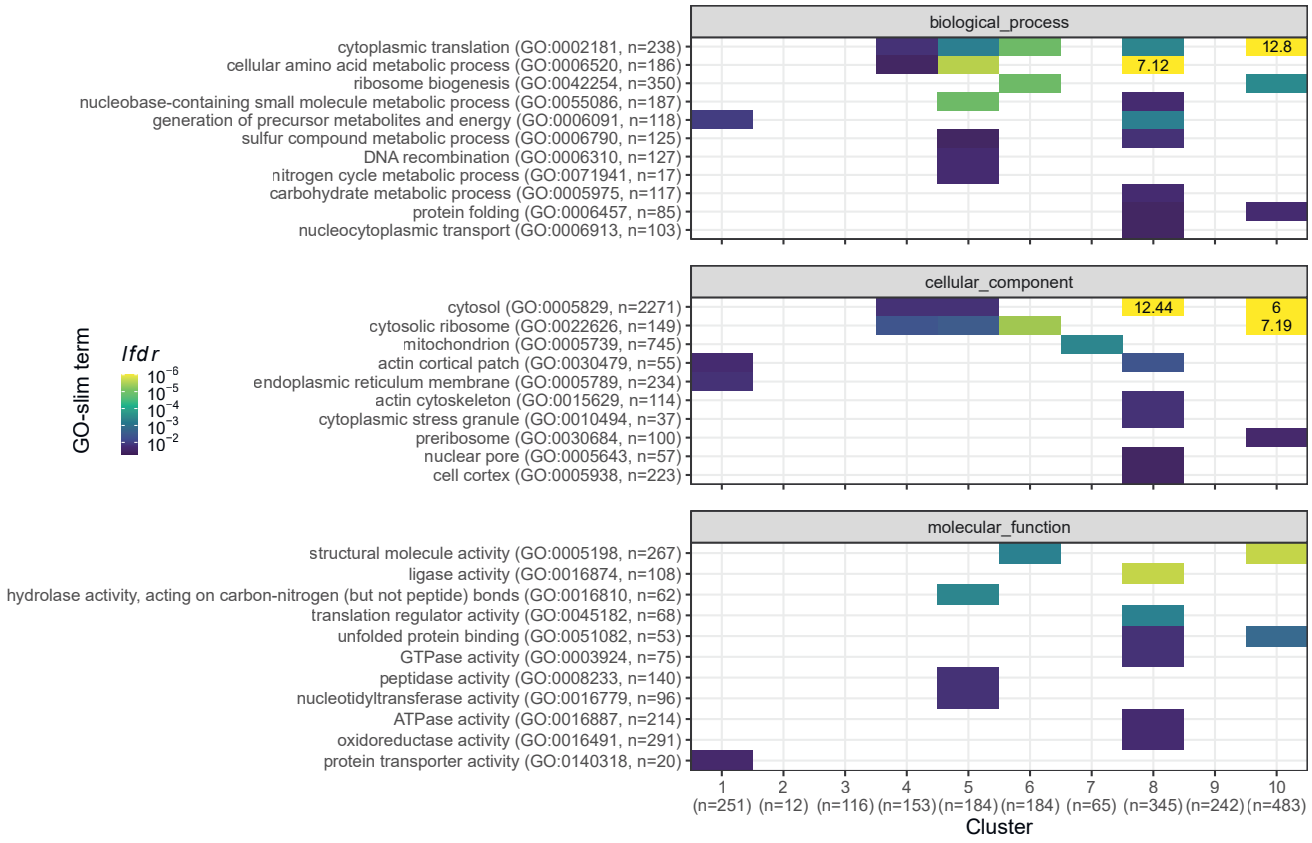


Figure 2.13: GO-slim enrichment analysis of protein-to-RNA clusters from Figure 2.11B. Log-transformed local false discovery rates lfdr are shown for the three Gene Ontology namespaces and capped at $\mathrm{lfdr} = 1.0 \times 10^{-6}$ (printed if capped). Overrepresentation analyses were performed as one-sided Fisher exact tests and only lists with $\mathrm{lfdr} < 0.05$ are shown; the numbers of genes in each GO-category and cluster are indicated.

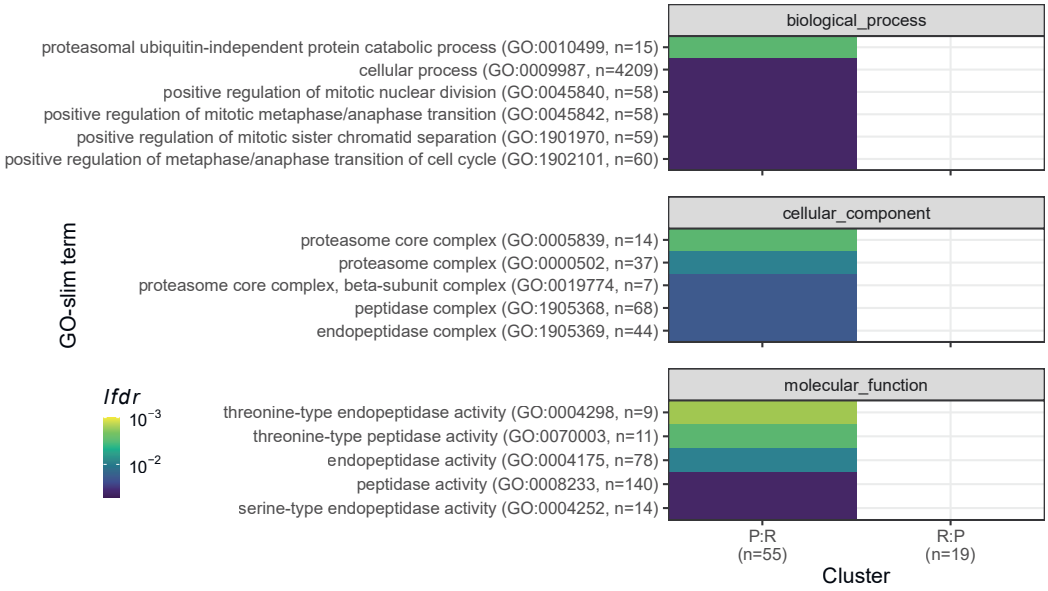


Figure 2.14: GO enrichment analysis of genes with significant and opposite growth-rate correlations between proteome and transcriptome. Log-transformed local false discovery rates $\mathtt{lfdr}$ are shown for the three Gene Ontology namespaces and capped at $\mathtt{lfdr} = 1.0 \times 10^{-3}$. Overrepresentation analyses were performed as one-sided Fisher exact tests and only lists with $\mathtt{lfdr} < 0.05$ are shown; the numbers of genes in each GO-category and gene set are indicated. For P:R genes were assigned to the P-sector in the proteome and R-sector in the transcriptome; and *vice versa* for the R:P genes.

minor enrichment of proteasomal genes in the group with negative growth-rate correlations in the proteome and positive growth-rate correlations in the transcriptome (Figure 2.14).

Finally, we compared mRNA and protein growth-related regulation for a series of functional categories using an unbiased gene set enrichment analysis, ranking genes on the signed significance measure used to determine the P- and R-sector. This showed that most categories showed a similar growth-related regulation for both mRNA and proteins (Figure 2.11D). This finding was robust to changing the ranking variable to the effect size (FC) instead of the significance (figure 2.15B). Transcripts for transcription factors, and for proteins generally bound to the chromosome, were an exception (Figure 2.11D). This was due to limited coverage in the proteomics data for these categories (Figure 2.15A). Specific functional categories will be discussed below.

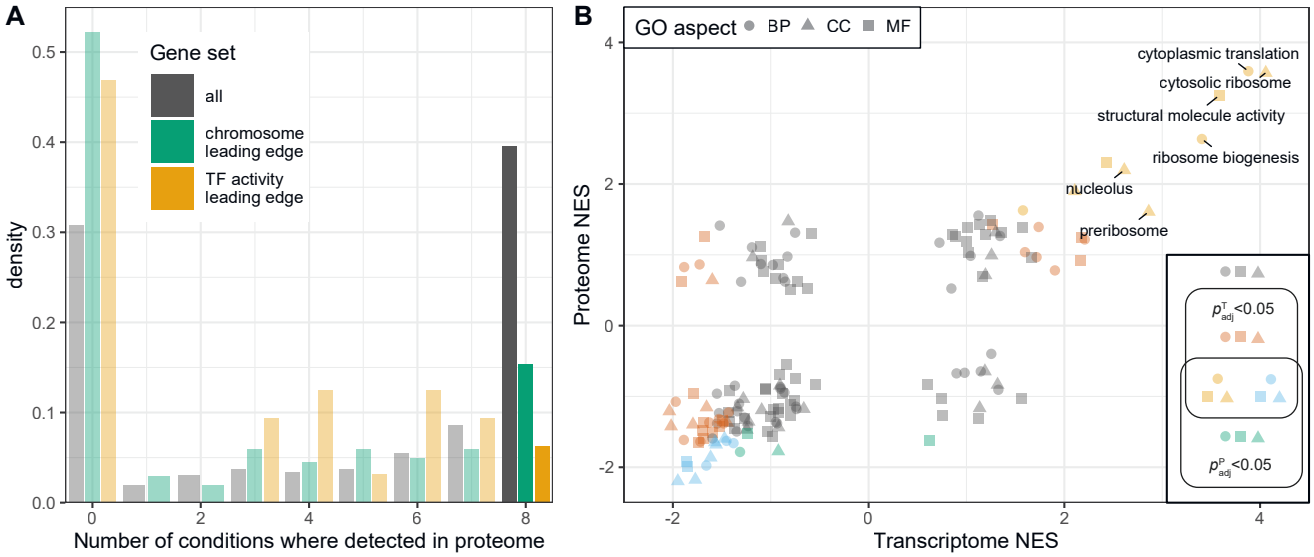
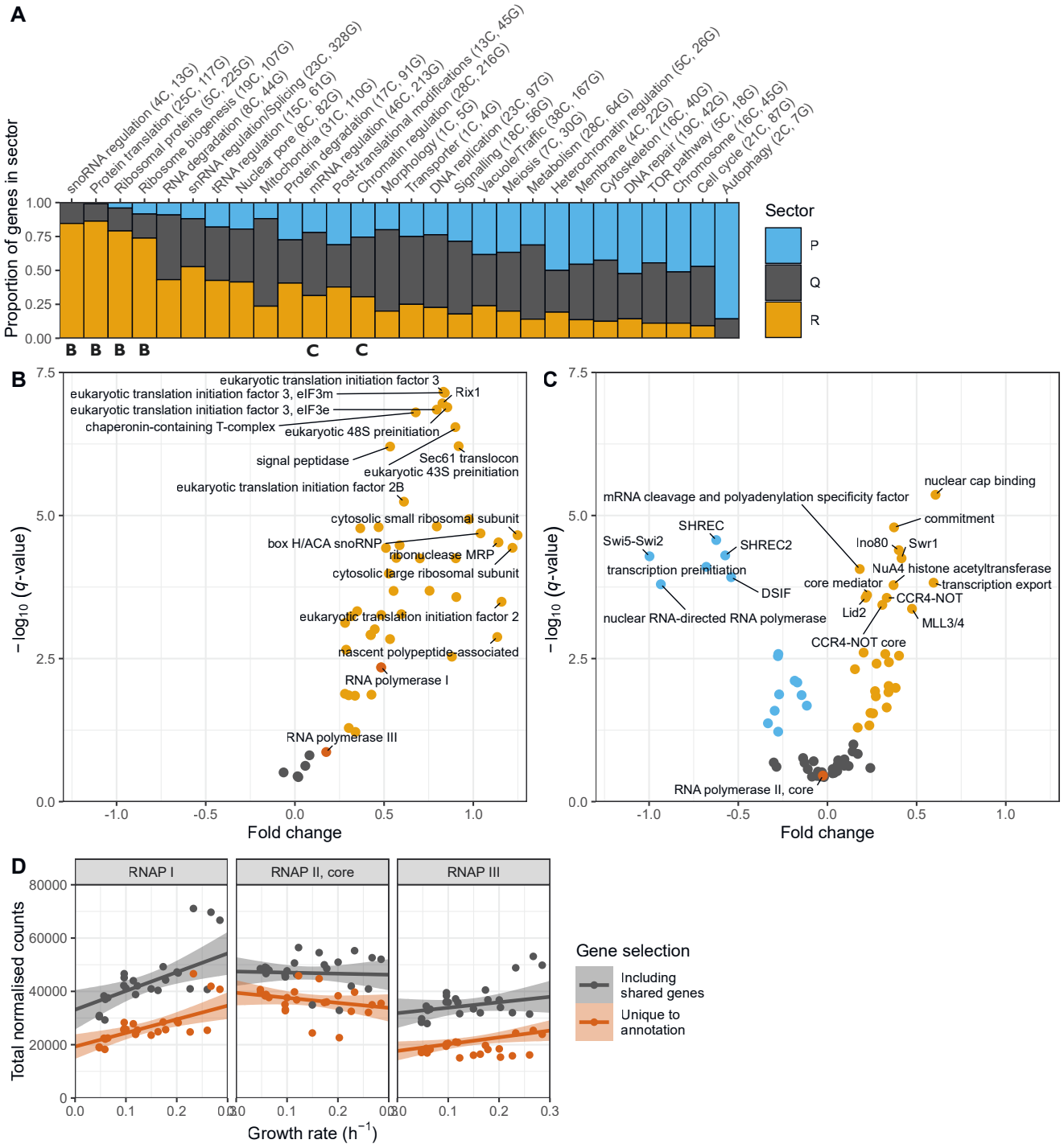


Figure 2.15: Additional analysis of gene set enrichment analysis from Figure 2.11D. A. Histogram illustrating across how many conditions protein groups were detected in the proteomics data. The behaviour of genes belonging to selected GO-slim terms is highlighted. **B.** Scatter plot of proteome and transcriptome normalised enrichment scores (NES) for GO-slim terms subjected to gene set enrichment analyses (GSEAs). For this analysis, genes were ranked based on the fold change (FC) measures from their RMLM fit in proteome and transcriptome. GO-slim terms are coloured based on the adjusted *p*-values of their GSEAs; terms with $p_{adj} < 10^{-9}$ in either or both of the analyses are labelled.

2.3.6 R sector proteins participate in all steps of the protein synthesis process

We next queried the cellular processes that had a strong R component and could therefore be either limiting for growth or regulated by it. We used a curated list of macromolecular complexes spanning most cellular processes and calculated the proportion of each complex subunit that was growth rate-dependent in each category (Figures 2.16 and 2.17) (Gene Ontology Consortium 2019; Lock et al. 2019). As observed in prokaryotes and budding yeast, the top 4 categories relying the most on R proteins belonged to a single process: the synthesis of proteins (Figure 2.16AB). Strikingly, R complexes were found at every single step of protein synthesis: the transcription of rRNAs and tRNAs and their processing, assembly and post-translational modification of the ribosome, and initiation and termination of translation (Figure 2.16B). Interestingly, expression of the chromatin-modifying complexes NuA4 and Ino80 were part of the R sector (Figure 2.16C), suggesting they may be involved in ribosome biogenesis in fission yeast as has been proposed for NuA4 in budding yeast (Uprety et al. 2015). Alternatively, these results could indicate that the chromatin structure and levels of histone modification may be limiting for growth.



(Caption on the next page.)

Figure 2.16: Functional categorisation of proteins positively correlated with growth rate. Figure printed on the previous page. **A.** Proportions of P-, Q-, and R-sector genes across manually assigned categories of protein complex annotations. Numbers of complexes (C) and genes (G) are indicated for each category. **B.** and **C.** Volcano plots summarising repeated-median linear model (RMLM) fits (of total transcriptome expression (total normalised counts) against growth rate) for each protein complex. Log-transformed q-values are plotted against fold change (FC) measures for each complex. Categories "snoRNA regulation", "protein translation", "ribosomal proteins", and "ribosome biogenesis" are shown in B., and "mRNA regulation" and "chromatin regulation" are shown in C. **D.** Aggregate normalised counts for subunits of RNA polymerases I (left), II (middle), and III (right) plotted as a function of growth rate. Total expression of subunits unique to each one of the three complexes is highlighted in vermillion whereas total expression of genes not necessarily unique is plotted in grey. RMLM fits and 95% confidence intervals are indicated.

The overall correlation between growth and the factors involved in protein synthesis had a notable exception. Although RNA polymerase (RNAP) I and specific subunits of RNAPIII were part of the R sector, RNA polymerase II specific subunits were not significantly correlated with growth rate (Figure 2.16B–D). Therefore, the number of RNAP II complexes is unlikely to be a limiting step in protein production during growth. Interestingly, RNAP II numbers were found to be limiting for the scaling of gene expression to cell size, indicating that coordination of gene expression to cell size and growth rate follow different mechanisms (Padovan-Merhar et al. 2015; Sun et al. 2020).

2.3.7 The stoichiometry of translation complexes changes with the growth rate

Differences in FC values between protein complexes indicate that their relative levels or stoichiometry changes with the growth rate. We hypothesised that these variations could provide mechanistic insights into the functioning of these complexes. To investigate this in the context of protein translation, we analysed three non-overlapping subclasses of translation proteins: the ribosomal proteins (RP), the ribosome biogenesis regulon (RiBi), and the translation initiation, elongation and termination factors (IET). The FC value for the IET class was the smallest of the three, whereas the trend line for RPs was the steepest (Figures 2.18A and 2.19A).

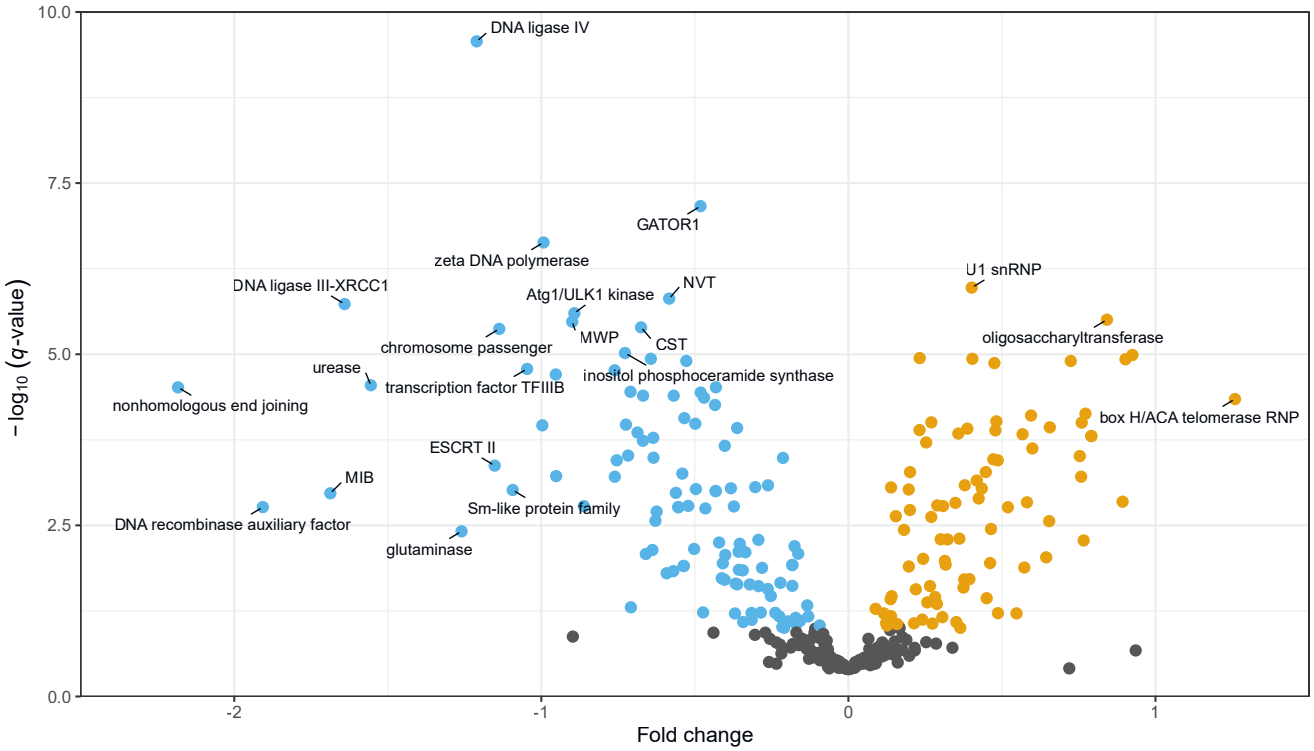
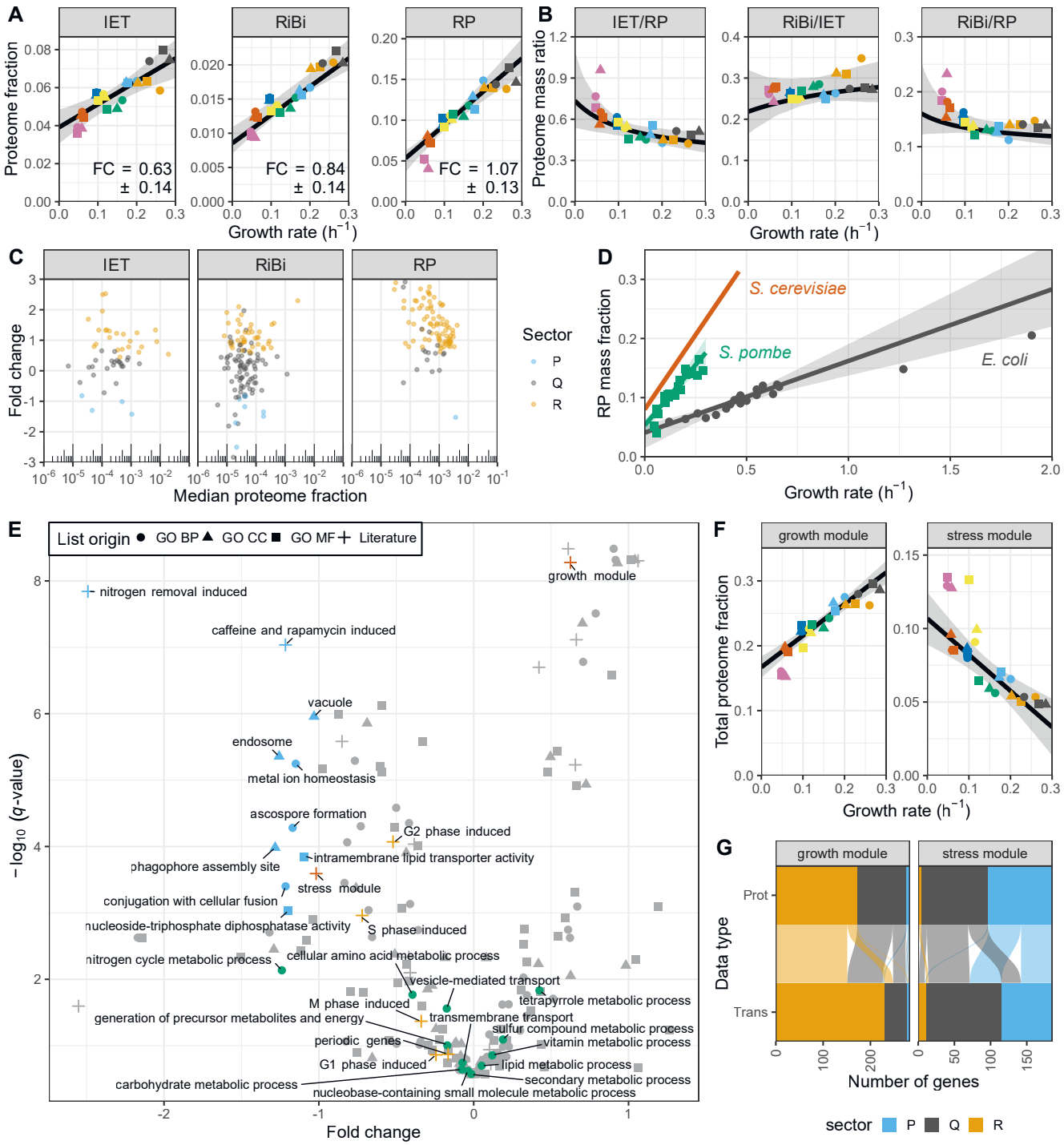


Figure 2.17: Volcano plots summarising repeated-median linear model (RMLM) fits for protein complexes not illustrated in figure 2.16B. and C. Log-transformed q-values are plotted against fold change (FC) measures for each complex.

As a result, the ratios between IET and RPs was significantly higher at slow growth (Figure 2.18B). It has been shown that RPs are held in reserve at slow growth rates (Metzl-Raz et al. 2017); these results suggest that an even larger fraction of IET and possibly RiBi proteins could be held in reserve. The relative abundances in EMM of IET:RiBi:RP were approximately 4:1:8 for the proteome mass fractions and 5:4:64 for the transcriptome RPKMs. This confirms earlier observations that the burden on transcription for RP synthesis is higher than for the rest of the proteome (M. W. Schmidt et al. 2007; Marguerat et al. 2012). The growth laws for the initiation and elongation factors were almost identical to each other, suggesting constant stoichiometry with the growth rate (Figure 2.19BC). Within the IET category, elongation factors were about three times as abundant as initiation factors, and about fifty times compared to termination factors (Figure 2.19BC). This is in line with biochemical evidence showing that translation initiation is a limiting step for protein synthesis (Aylett and Ban 2017). Taken together, we have shown how the growth law can inform on the regulation of gene expression through changes in the stoichiometry of factors with the growth rate.



(Caption on the following page.)

Figure 2.18: Analysis of translational proteins and of proteins negatively correlated with the growth rate in *S. pombe*. Figure printed on the preceding page. **A.** Total proteome fractions as a function of growth rate for proteins involved in translation initiation, elongation, and termination (IET; left panel), ribosome biogenesis (RiBi; middle panel), and ribosomal proteins (RP; right panel). Repeated-median linear model (RMLM) best fits and 95% confidence intervals (CIs) based on a bootstrapping analysis are shown in black and grey. The RMLM fold change (FC) measure and standard deviation according to the bootstrapping analysis are printed on the figure. **B.** Ratios of proteome mass fractions of the three categories. Superimposed are the best-fit ratios and 95% CIs, as determined by the bootstrapped models fitted to the data in A. **C.** FC measures for all individual protein groups assigned to the IET, RiBi, and RP categories as a function of their median expression. **D.** Ribosomal protein mass fraction as a function of growth rate for *S. pombe* (this study: Kleijn et al. 2022), *E. coli* (reanalysed from A. Schmidt et al. 2016, see also Figure 2.22), and *S. cerevisiae* (fit parameters taken from Metzl-Raz et al. 2017). **E.** Volcano plots summarising repeated-median linear model (RMLM) fits (of total proteome expression against growth rate) for each GO-slim term and literature list (Mata et al. 2002; D. Chen et al. 2003; Rustici et al. 2004; Rallis et al. 2013; Kamrad et al. 2020). Log-transformed q-values are plotted against fold change (FC) measures for each gene class. Lists with $q < 0.001$ are plotted in blue, biological_process GO-slim terms related to metabolism are in green, the stress and growth modules from D. Chen et al. (2003) in vermillion, and cell cycle induced modules from Rustici et al. (2004) are in orange. **F.** Total proteome mass fractions with RMLM best-fit and symmetric 95% CIs for the growth module (core environmental stress response (CESR) repressed) and stress module (CESR induced). **G.** Assignment of proteins and corresponding transcripts to growth correlation categories for growth and stress modules; proteins (Prot) and corresponding transcripts (Trans) are connected by lines with colours that match to the protein classification.

2.3.8 Most ribosomal proteins are coordinately expressed

Furthermore, the large burden of RPs during fast growth resulted from the coordinated growth-related expression of most individual RPs and from a growth dependence component steeper than that of IET and RiBi (Figure 2.18C). This indicates that the aggregate burden of RPs results from coordinated regulation at the level of single genes (Petibon et al. 2020). The IET and RiBi categories also contained more proteins that were assigned to the P- and Q-sectors, and whose expression data was not well explained by the robust model due to significant condition-dependent expression (Figures 2.20 and 2.21). For instance, the initiation factor eIF3E was present in sub-stoichiometric amounts relative to eIF3A. Interestingly eIF3E has been shown to selectively regulate the translation of transcripts coding for metabolic enzymes (Shah et al. 2016).

Sixteen protein groups annotated as RPs were assigned to the Q-sector because their expression was not significantly positively correlated with the growth rate and we here explore these Q-RPs further (Figure 2.21). Their relative protein abundances were slightly lower than those of RPs that did belong to the R sector (R-RPs, see Figure 2.21A). However, median transcript abundances and FC values were not significantly different between Q-RPs and R-RPs (Figure 2.21B). This opens the possibility of regulatory feedback at the post-transcriptional level. Interestingly, half of these Q-RPs (Rlp7, Rpl102, Rpl2501, Rpl35A02, Rps1502, Rps20, Rps27, and Rps2801, and Rps2802) are annotated with functions in ribosome biogenesis on the PomBase database (Lock et al. 2019). Additionally, the budding yeast orthologue of Q-RP Rps20 has been proposed to regulate RNAPIII transcription, providing a potential link between ribosomes and tRNA synthesis (Warner and McIntosh 2009). Together this suggests that Q-RPs could be attractive candidate proteins that could have additional functions outside of the ribosome.

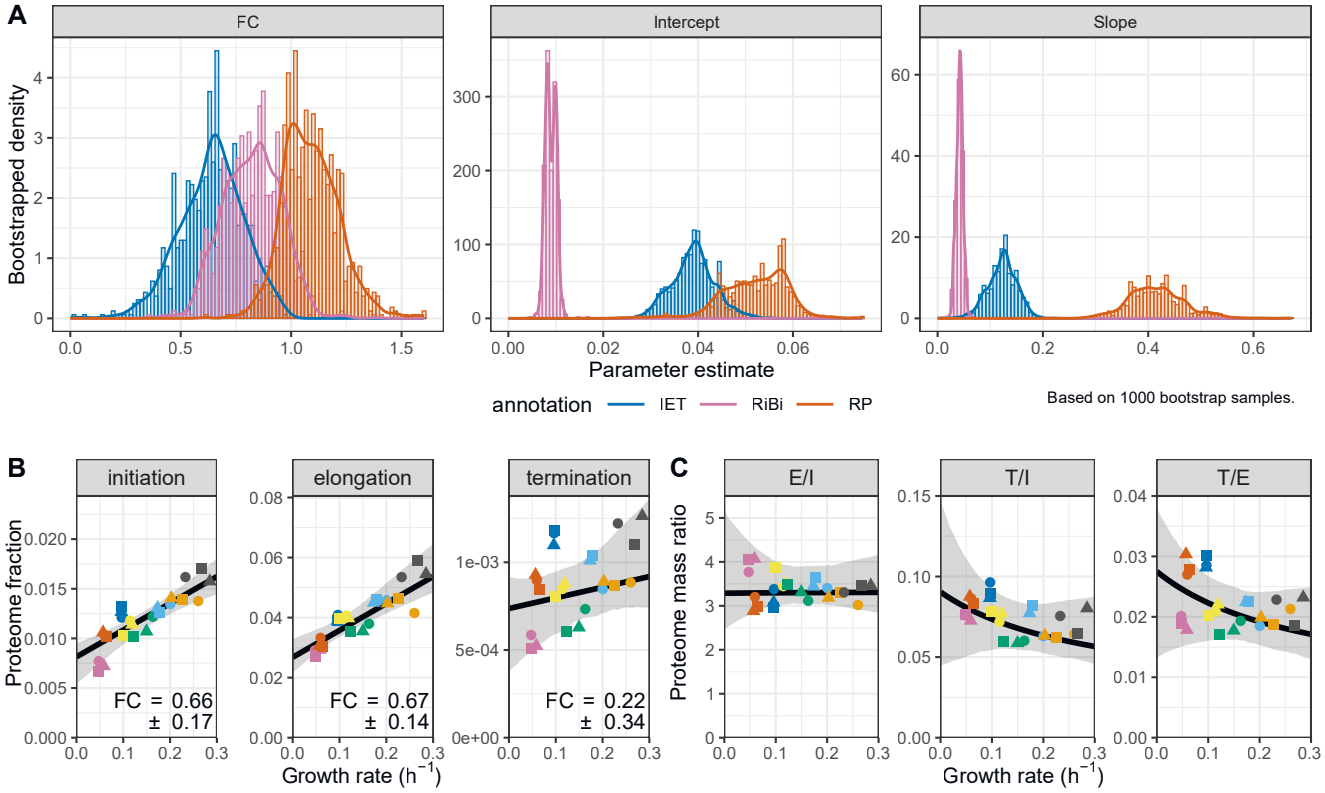
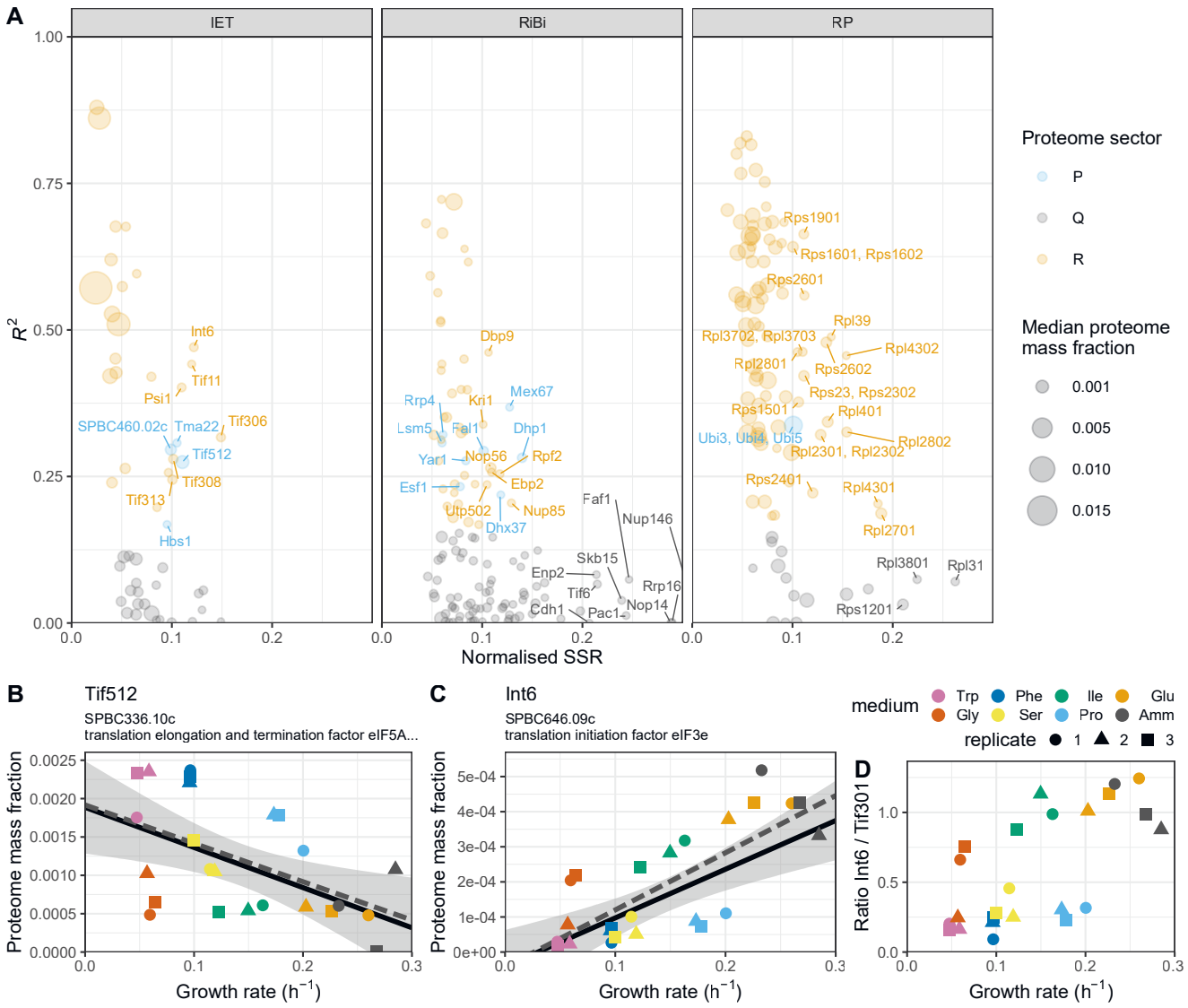
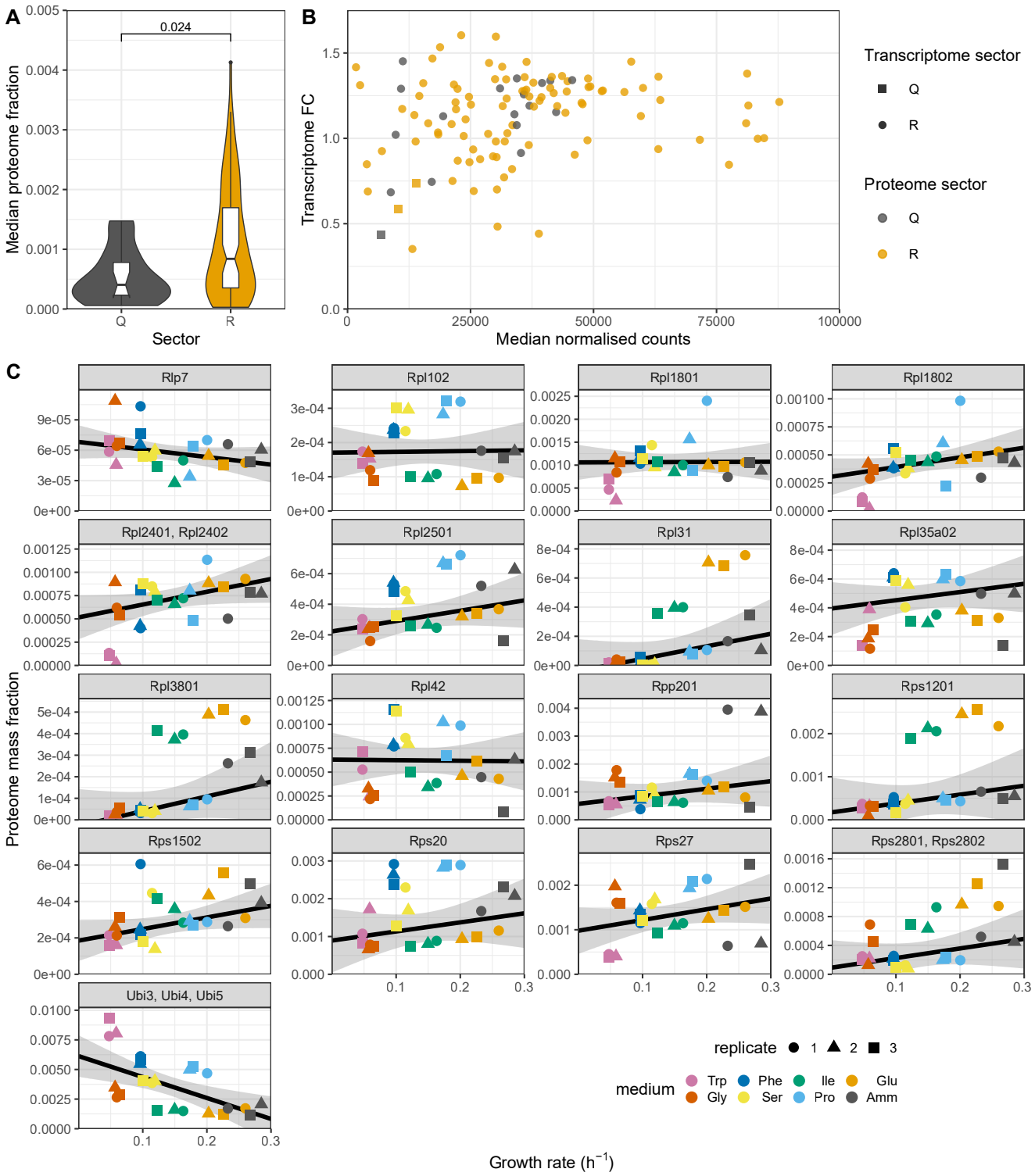


Figure 2.19: Further analysis of translational proteins in *S. pombe*. **A.** Histograms of parameters for bootstrap analysis of Figure 2.18A showing fold change (FC), intercept, and slope. **B.** Total proteome fractions as a function of growth rate for proteins involved in translation initiation (I), elongation (E), and termination (T). Repeated-median linear model (RMLM) best fits and 95% confidence intervals (CIs) based on a bootstrapping analysis are shown in black and grey. The RMLM fold change (FC) measure and standard deviation according to the bootstrapping analysis are printed on the figure. **C.** Ratios of proteome mass fractions of the three IET subcategories. Superimposed are the best-fit ratios and 95% CIs, as determined by the bootstrapped models fitted to the data in A.



(Caption on the next page.)

Figure 2.20: Further analysis of translational proteins not significantly positively correlated with growth rate. Figure printed on the previous page. **A.** Scatter plot of normalised sum of squared residuals (SSR) and coefficient of determination $R^2$of the repeated-median linear model (RMLM) best fits for protein groups annotated to proteins involved in translation initiation, elongation, and termination (IET; left panel), ribosome biogenesis (RiBi; middle panel), and ribosomal proteins (RP; right panel). Q-sector proteins were labelled if their normalised SSR was greater than 0.2; R-sector at a threshold of 0.1; all P-sector proteins were labelled. proteins involved in translation initiation, elongation, and termination (IET; left panel), ribosome biogenesis (RiBi; middle panel), and ribosomal proteins (RP; right panel). **B.-C.** Proteome expression and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey for translation elongation and termination factor eIF5A (Tif512, B.) and translation initiation factor eIF3e (Int6, C.). **D.** Observed protein mass ratios of Int6 and major eIF3 subunit Tif301 across growth conditions.



(Caption on the following page.)

Figure 2.21: Expression of ribosomal proteins (RPs) not significantly positively correlated with growth rate. Figure printed on the preceding page. **A.** Violin and box plots of median proteome mass fractions across conditions for RPs in Q- and R-sector. **B.** Scatter plot of relative transcriptome abundances (DESeq2 normalised counts) and transcriptome repeated-median linear model fold changes (FC) for RPs in Q- and R-sector. **C.** Proteome expression and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey for the 16 Q-sector RPs and the P-sector protein group comprising ubiquitin (Ubi4) and ubiquitin-RP fusion proteins Ubi3 and Ubi5.

2.3.9 The findings on translational proteins were confirmed in *E. coli*

Principles of proteome allocation are often conserved in prokaryotes and eukaryotes despite significant mechanistic differences in the way genes are transcribed and translated (Dai and Zhu 2020). Therefore, we thought to compare our findings in fission yeast with published datasets from the budding yeast *S. cerevisiae* and the bacterium *E. coli* (A. Schmidt et al. 2016; Metzl-Raz et al. 2017). We reanalysed published proteomics data for E. coli cells growing at different rates in a series of environmental conditions to extract the relative proteome fractions, and we subsequently computed the growth law parameters for translational proteins (A. Schmidt et al. 2016). For *S. cerevisiae*, we used growth law parameters of ribosomal proteins published elsewhere (Metzl-Raz et al. 2017). We found that *E. coli* could sustain a given growth rate with a smaller fraction of RPs than both yeasts, which was primarily due to a smaller growth law slope (Figure 2.18D). This suggests that the effective translation rate in the yeasts is lower than that of *E. coli*. Among the two yeasts, fission yeast used its RPs significantly more efficiently—using a smaller RP mass fraction to sustain any given growth rate—than the budding yeast trend line, but the effect could not be assigned to a significant difference in either the slope or the intercept parameter specifically. Next, we asked whether the changes in stoichiometry of translational proteins during slow growth were conserved in *E. coli*. Again, both the IET/RP and RiBi/RP ratios were higher during slower growth (Figure 2.22AB), because the individual RPs had steeper growth laws (Figure 2.22C). A steeper growth law of RPs than that of elongation factors was recently predicted by a model of E. coli that minimised the total

expression cost (Hu et al. 2020). Our results indicate that allocation strategies are conserved even though protein production differs mechanistically between the two kingdoms.

2.3.10 P sector proteins are part of the core environmental stress response programme

To complement our analysis of the R sector, we next examined fission yeast proteins from the P sector, i.e., proteins with a negative growth-dependent component. In contrast to the R sector clusters 1 and 2, we could not identify P sector clusters whose expression could be explained exclusively by a negative growth rate correlation (Figure 2.2A-C). This indicates that proteins with a strong P component are also often regulated in response to specific nitrogen sources. Moreover, the growth component for P proteins was less significant overall than for R proteins (Figure 2.23A–B). These results suggest that regulation of the R and P sectors may differ mechanistically.

Unlike R proteins, which are mostly involved in protein production, P proteins belonged to a diverse set of complexes participating in a large array of functions (Figure 2.16). As individual proteins, they showed weaker correlations than R sector complexes (Figure 2.23C–D). To determine if this diverse set of P proteins was participating in a common higher-level functional programme we analysed the fission yeast GO-slims alongside 21 lists covering fission yeast physiology and environmental responses (Figures 2.18E and 2.24A) (Mata et al. 2002; D. Chen et al. 2003; Rustici et al. 2004; Rallis et al. 2013; Kamrad et al. 2020). Functional classes with strong P-sector components included vacuole biology, endosome and phagosome, transport and genes induced in the adaptation to nitrogen removal, and/or after treatment with caffeine and rapamycin. The latter two classes, which had the strongest response, are thought to be controlled by TORC1 (Mata et al. 2002; Rallis et al. 2013). This suggests that nitrogen sources supporting slower growth rates trigger a form of metabolic stress response. Accordingly, the total expression of the fission yeast core environmental stress response programme up-regulated genes (CESR up) was negatively correlated with the growth rate (Figure 2.18F). This stress module comprises genes induced in response to a wide range of environmental and genetic perturbations (D. Chen et al. 2003; Pancaldi et al. 2010). Conversely, genes down-regulated as part of the CESR response (CESR down, also called growth module) belonged to the R sector (Figures 2.18F–G and 2.24B). This finding validates the long-standing hypothesis that the balanced

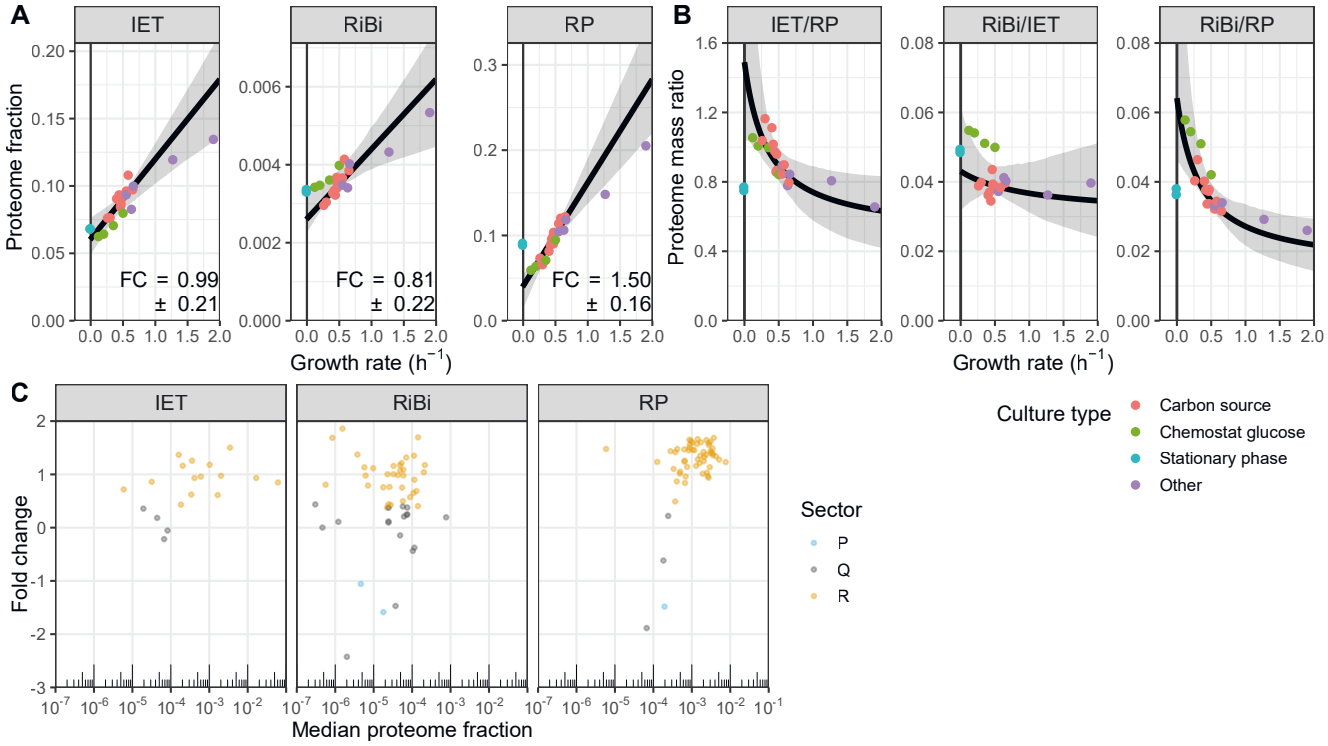


Figure 2.22: Analysis of translational proteins in *E. coli*. Data from A. Schmidt et al. (2016) was reanalysed to generate this figure. **A.** Total proteome fractions as a function of growth rate for proteins involved in translation initiation, elongation, and termination (IET; left panel), ribosome biogenesis (RiBi; middle panel), and ribosomal proteins (RP; right panel). Repeated-median linear model (RMLM) best fits and 95% confidence intervals (CIs) based on a bootstrapping analysis are shown in black and grey. The RMLM fold change (FC) measure and standard deviation according to the bootstrapping analysis are printed on the figure. **B.** Ratios of proteome mass fractions of the three categories. Superimposed are the best-fit ratios and 95% CIs, as determined by the bootstrapped models fitted to the data in A. **C.** FC measures for all individual protein groups assigned to the IET, RiBi, and RP categories as a function of their median expression.

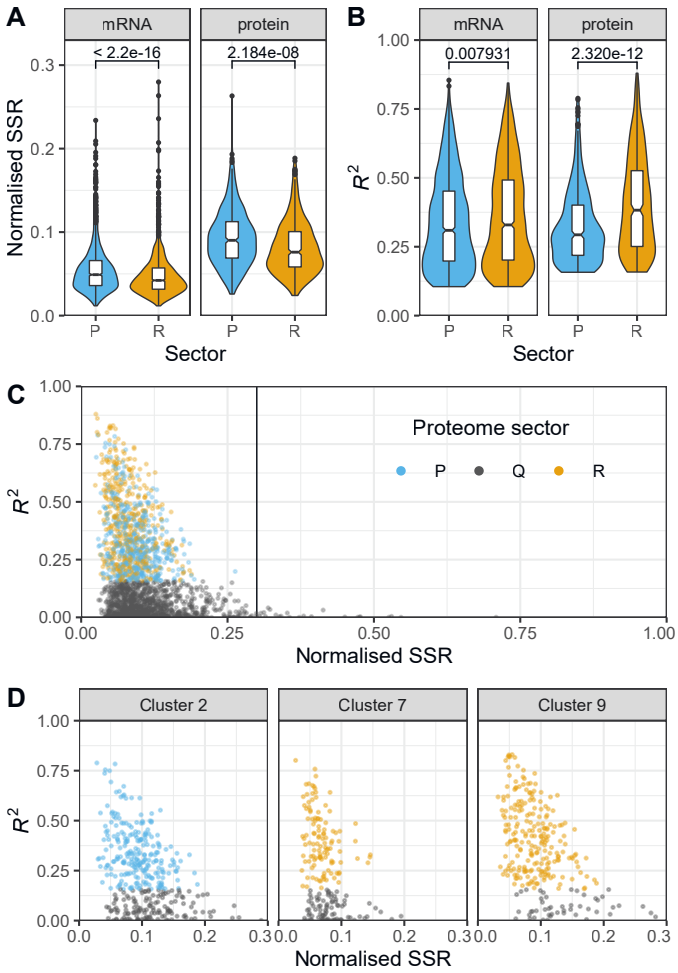
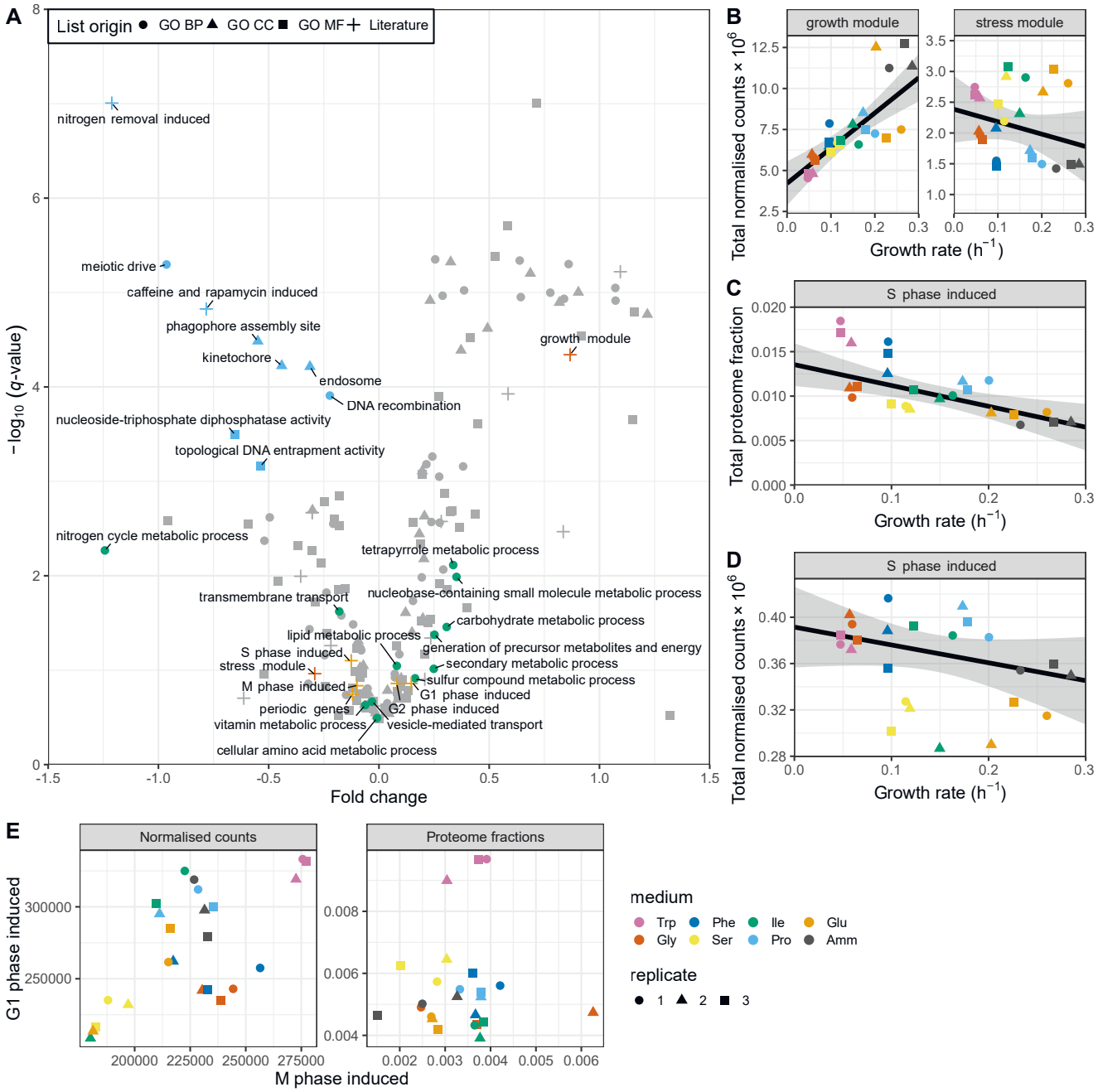


Figure 2.23: Analysis of residuals to repeated-median linear models (RMLMs) for R- and P-sector proteins. **A.** Violin and box plots of normalised sum of squared residuals (SSR) of the repeated-median linear model (RMLM) best fits for protein groups belonging to the P- or R-sector. **B.** As A. for the RMLM coefficients of determination R^2. **C.** Scatter plot of normalised SSR and R^2 for all protein groups coloured by their proteome sector. **D.** Subset of C. for protein groups belonging to clusters 2, 7, and 9 from the analysis first plotted in Figure 2.3A.

expression of the fission yeast stress response is quantitatively connected with the growth rate (López-Maury et al. 2008). Additionally, P proteins were enriched for factors regulated during the S phases of the cell cycle, which is consistent with evidence that the cell-cycle phase length differs between nitrogen sources, in particular growth on Trp (Figures 2.18F and 2.24C–D) (Carlson et al. 1999; Rustici et al. 2004). Notably, we did not observe a simple relationship between the expression of cell cycle markers and the growth rate (Figure 2.24E). This is in line with earlier flow cytometry and microscopy data, which did not find a straightforward relationship between the length of cell cycle phases and the growth rate upon growth on different nitrogen sources (Carlson et al. 1999).



(Caption on the next page.)

Figure 2.24: Further analysis of genes negatively correlated with the growth rate in *S. pombe*. Figure printed on the previous page. **A.** Volcano plots summarising repeated-median linear model (RMLM) fits (of total transcriptome expression against growth rate) for each GO-slim term and literature list (Mata et al. 2002; D. Chen et al. 2003; Rustici et al. 2004; Rallis et al. 2013; Kamrad et al. 2020). Log-transformed *q*-values are plotted against fold change (FC) measures for each gene class. Lists with $q < 0.001$ are plotted in blue, biological_process GO-slim terms related to metabolism are in green, the stress and growth modules from D. Chen et al. (2003) in vermillion, and cell cycle induced modules from Rustici et al. (2004) are in orange. **B.** Total transcriptome expression with RMLM best-fit and symmetric 95% CIs for the growth module (core environmental stress response (CESR) repressed) and stress module (CESR induced). **C.** and **D.** Total proteome (C.) and transcriptome (D.) expression with RMLM best-fit and symmetric 95% CIs for S-phase induced periodic genes (Rustici et al. 2004). **E.** Scatter plot of total transcriptome (left) and proteome (right) expression levels for M-phase versus G1-phase induced periodic genes (Rustici et al. 2004).

2.3.11 The burden of specific metabolic pathways is principally condition-dependent

Notably, the functional classes involved in metabolism were not strongly negatively correlated with the growth rate (Figure 2.18E) and the fission yeast P sector was only marginally enriched in proteins involved in central and energy metabolism (Figure 2.25). This contrasts with previous data from *E. coli* and *S. cerevisiae* where metabolic genes have been reported to be important components of the P sector (Hui et al. 2015; A. Schmidt et al. 2016; Metzl-Raz et al. 2017). However, when considered globally, the sum of protein mass fractions dedicated to metabolic enzymes was clearly anti-correlated with growth in fission yeast, ranging from ~70% of the proteome in poor nitrogen sources to ~55% in the fastest media (Figure 2.26A). This indicates that in our system which does not rely on titration of a limiting nutrient to modulate the growth rate, the total protein burden on metabolism is linked to the growth rate, whereas allocation to specific enzymes is not. Therefore, the global anti-correlation of metabolic enzymes with growth rate observed in our data may be a manifestation of the trade-off between metabolism and translation, and not the result of the direct quantitative regulation of metabolic enzymes expression with the growth rate.

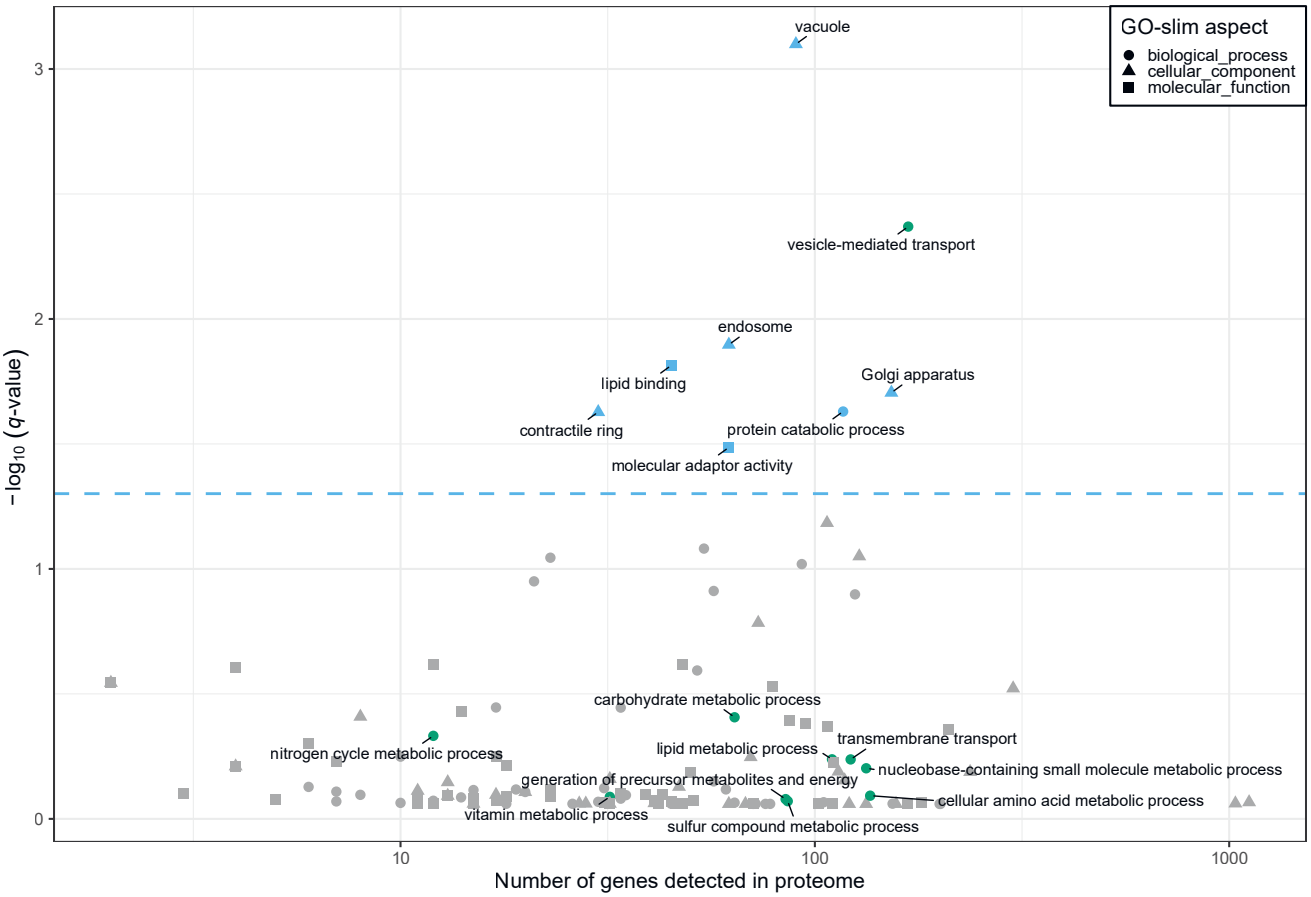
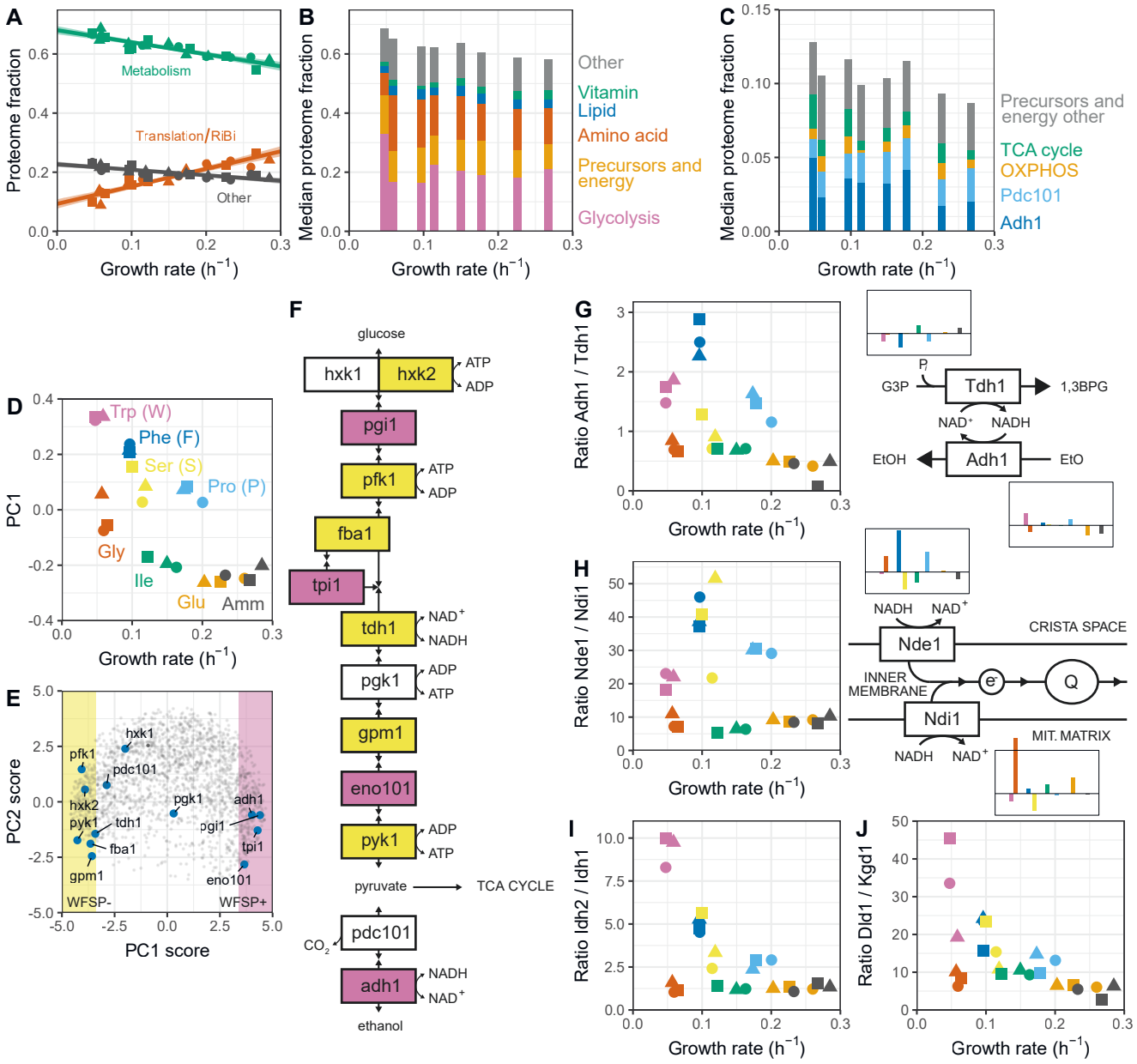


Figure 2.25: Scatter plot visualisation of GO-slim enrichment analysis of proteins negatively correlated with growth rate. Log-transformed *q*-values plotted against the number of genes detected across all conditions in the *S. pombe* proteome for each GO-slim term. GO-slim terms with $q < 0.05$ are plotted in blue and the threshold is highlighted in the figure; the biological_process GO-slim terms related to metabolism are plotted in green.



(Caption on the following page.)

Figure 2.26: Exploratory analysis of metabolic enzyme expression. Figure printed on the preceding page. **A.** Total proteome fractions as a function of growth rate for proteins involved in translation and ribosome biogenesis, energy metabolism and transport, and all other genes. **B.** Bar graphs of median aggregate proteome mass fraction stacked by metabolic protein category, separated by condition and plotted as a function of growth rate. **C.** As B. for proteins in oxidative phosphorylation (OXPHOS) and TCA cycle, proteins annotated with "generation of precursor metabolites and energy" and none of the other 4 categories from B., as well as the Adh1 and Pdc101 proteins making up the ethanol fermentation pathway. **D.** Principal component 1 (PC1) contribution to protein expression across growth conditions as a function of growth rate. **E.** PCA biplot of protein groups detected across all growth conditions. Areas with greater than 50% of expression variance explained by PC1 are shaded in yellow (negative correlation with PC1, WFSP- list) and pink (positive correlation, WFSP+ list). Enzymes comprising glycolysis and ethanol fermentation are indicated as blue points. **F.** Illustration of glycolysis and ethanol fermentation pathways, with WFSP proteins coloured as in E. **G.** Observed protein mass ratios of Adh1 and Tdh1 proteins across growth conditions, with cartoon indicating their functions. **H.** As G. for NADH dehydrogenase proteins Nde1 and Ndi1. **I.–J.** Observed protein mass ratios of isocitrate dehydrogenase subunits Idh2/Idh1 (I.) and alpha-ketoglutarate dehydrogenase subunits Dld1/Kgd1.

On top of the growth-dependent components, many fission yeast proteins showed clear condition-specific gene regulation (Figure 2.2ABC). Functional analysis indicated an enrichment of these genes for functions related to metabolism. This is consistent with the adoption of distinct metabolic allocation strategies in response to growth with different nitrogen sources (Alam et al. 2016; Mülleder et al. 2016). We classified metabolic genes into six non-overlapping classes based on the following GO terms: canonical glycolysis (GO:0061621), generation of precursors and energy (GO:0006091), cellular amino acid metabolic process (GO:0006520, which includes the interconversion of ammonium, glutamate, and glutamine), lipid metabolic process (GO:0006629), vitamin metabolic process (GO:0006766), and all other metabolic pathways (including transport of metabolites) (Figures 2.26B, 2.27). To avoid overestimating the burden of gene expression by double-counting genes assigned to multiple terms, each protein was assigned only to the first of these GO-terms it was annotated with. The relative allocation to each class was

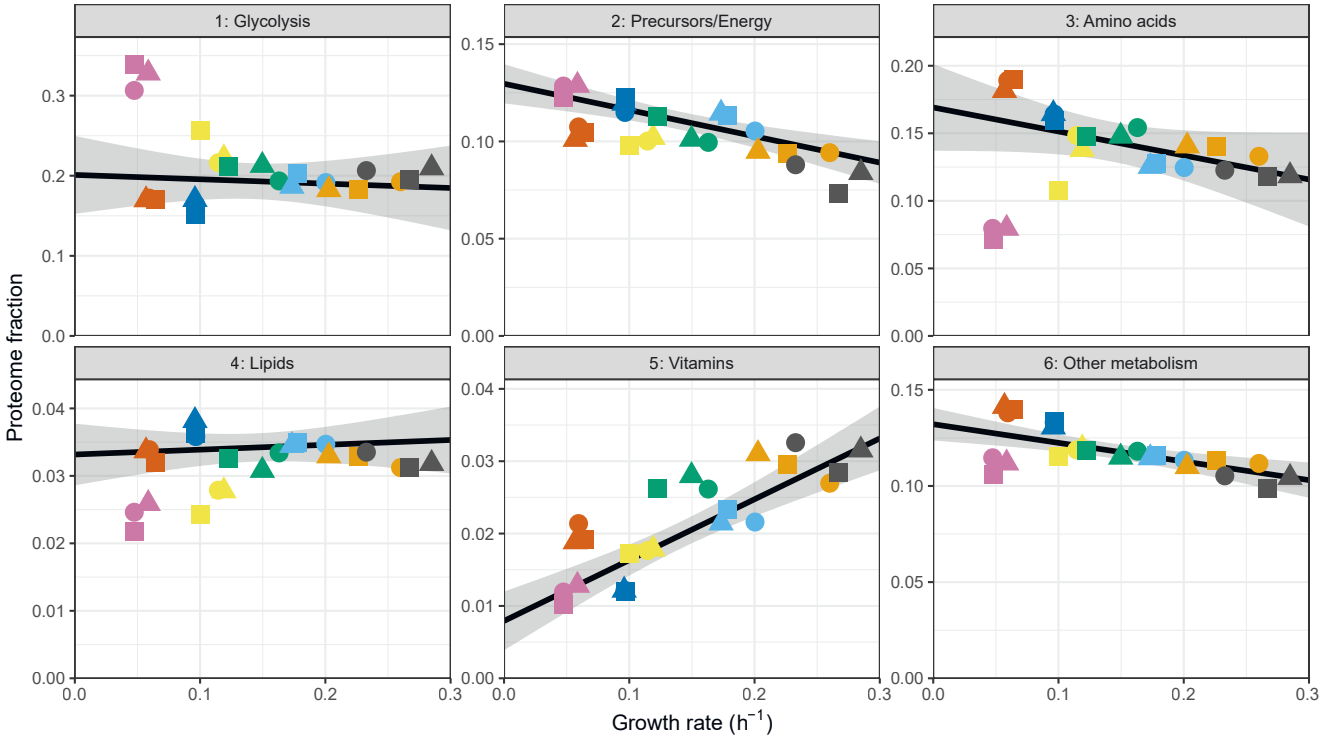


Figure 2.27: Aggregate growth rate models of metabolic proteins. Total proteome mass fractions and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey for 6 categories of proteins covering metabolism (see Figure 2.26B).

condition-specific, indicating that metabolic states rely differentially on specific pathways (Figure 2.26B). We note that similar growth rates can be supported by different allocation strategies, as in the case of the Trp and Gly containing media in which cells channelled resources preferentially towards glycolysis (Trp) or amino acid metabolism (Gly) (Figures 2.26B and 2.27). The growth-related components of those categories were weak, except for the vitamin metabolism proteins which belonged to the R sector and the precursor/energy proteins that showed a significant P component (see below, Figure 2.27). Most coenzymes are stable molecules synthesised only as much as necessary to support growth (Hartl et al. 2017). The strong positive correlation of vitamin metabolism expression with growth rate suggests that cells also minimise the translation burden of vitamin metabolic enzymes. In summary expression of metabolic enzymes in our system, although connected to the growth rate, is mainly condition- and pathway-specific.

We next took a closer look at the energy metabolism pathways and their negative correlation with the growth rate. Nutrient quality, cell growth, and energy metabolism are intimately connected. The generation of ATP through fermentation is often favoured in conditions that support faster growth whereas slow-growing cells in limiting conditions tend to switch to respiratory metabolism (Vander Heiden et al. 2009; Shimizu and Matsuoka 2018). Therefore, we asked whether protein allocation to either energy metabolism pathway was correlated with the nitrogen sources used and/or growth rate. To this end, we split the non-glycolytic generation of precursors and energy category into the fermentation enzymes pyruvate decarboxylase (Pdc101) and alcohol dehydrogenase (Adh1), and the respiration process into tricarboxylic acid cycle (TCA, GO: 0006099) and oxidative phosphorylation (OXPHOS, GO:0006119) enzymes (Figures 2.26C and 2.28). Surprisingly, none of the categories were consistently correlated with the growth rate. Instead, condition-specific expression was dominant, and a clear repression of all OXPHOS complexes upon growth on serine was observed (Figure 2.29). A recent report showed that serine catabolism generates high levels of reactive oxygen species (ROS) in *S. pombe*, suggesting that respiration may be repressed upon growth on serine to avoid a further increase in ROS (Kanou et al. 2020). Notably, expression of the fermentation enzymes Adh1 and Pdc101, although variable between conditions, was consistently higher than the total expression of the respiratory enzymes. Moreover, respiratory enzymes were not induced in nitrogen sources supporting slow growth. Taken together, the expression balance between fermentation and respiratory enzymes was not quantitatively connected to the growth rate, but depended on the nutrient properties.

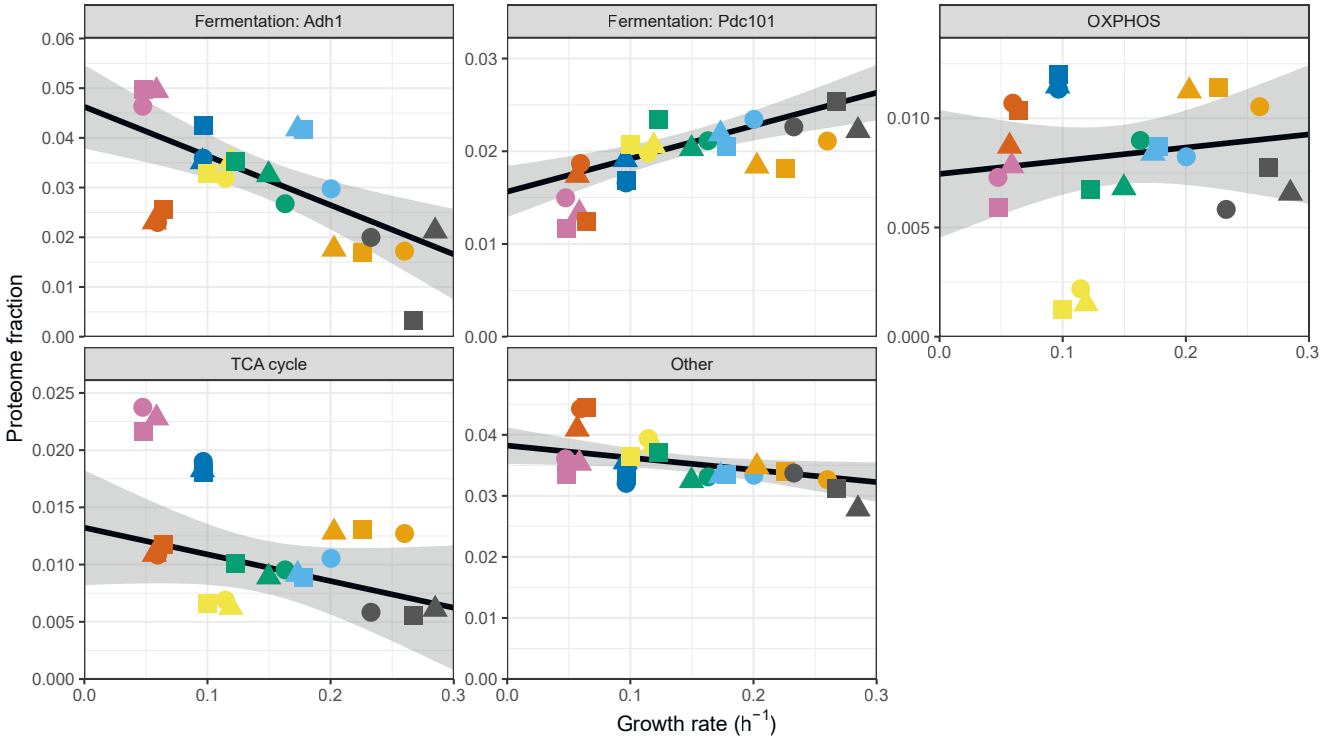
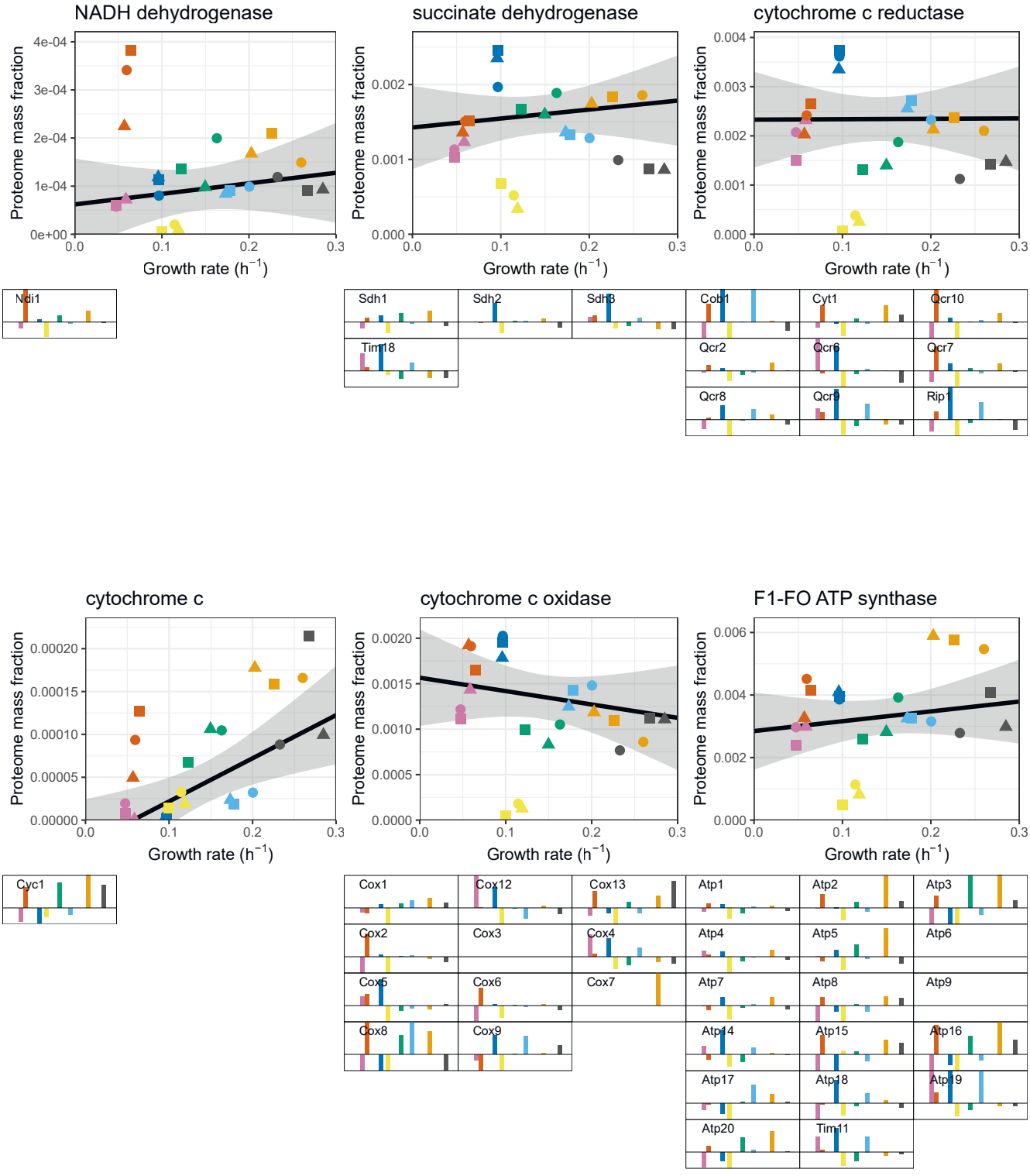


Figure 2.28: Aggregate growth rate models of energy metabolism proteins. Total proteome mass fractions and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey for 5 subcategories of energy metabolism proteins (see Figure 2.26C).



(Caption on the next page.)

Figure 2.29: Aggregate growth rate models and individual bar code plots for proteins comprising the complexes involved in the respiratory electron transport chain and proton pumps. Repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval are indicated in grey. Figure printed on the previous page.

2.3.12 Two groups of four conditions each were characterised by different expression patterns, chiefly in glycolysis and NAD-dependent metabolism

To complement this analysis, we searched for condition-specific patterns of protein expression that were not related to the growth rate in our proteomics dataset using principal component analysis (PCA) (Figure 2.30). The first principal component (PC1) explained 29% of the total variance and split the culture conditions in two irrespective of the growth rate with Trp (W), Phe (F), Ser (S), and Pro (P) in one group (from here on termed the WFSP media) and Gly (G), Ile (I), Glu (E) and Amm in the other (Figure 2.26D). Strikingly 24% (474/1988) of proteins had more than 50% of their variance explained by PC1. We defined two large classes of protein based on their response to this component: i) WFSP+ consisting of 259 proteins that were positively correlated with PC1 and therefore induced in the WFSP media; ii) WFSP- characterised by 215 proteins with expression negatively correlated with PC1 and therefore repressed in the WFSP media. Interestingly, no single principal component was dominated by growth rate correlation (Figure 2.30E), reinforcing the point that nutrient-specific and growth-dependent components of gene expression coexist for many proteins.

Glycolytic and NAD-dependent enzymes were the two major classes of proteins overrepresented in the WFSP lists. First, most glycolytic enzymes belonged to one of the two WFSP classes (Figures 2.30E and 2.31). These enzymes were highly expressed across conditions, amounting to ~15%–30% of the total proteome mass (Figures 2.30B and 2.27). Therefore, the total gene expression burden of cellular metabolism across the WFSP conditions was heavily affected by the abundance of a small number of enzymes. Second, the two enzymes glyceraldehyde-3-phosphate (G3P) dehydrogenase Tdh1 and alcohol dehydrogenase Adh1 were assigned to opposing WFSP lists, and the ratio of Adh1/Tdh1 expression was highly elevated in the WFSP conditions (Figures 2.30E and 2.32). Fermentation of a single molecule of glucose generates two molecules of ethanol and carbon dioxide. During the process, Tdh1 reduces two NAD+ molecules and Adh1 oxidises two NADH

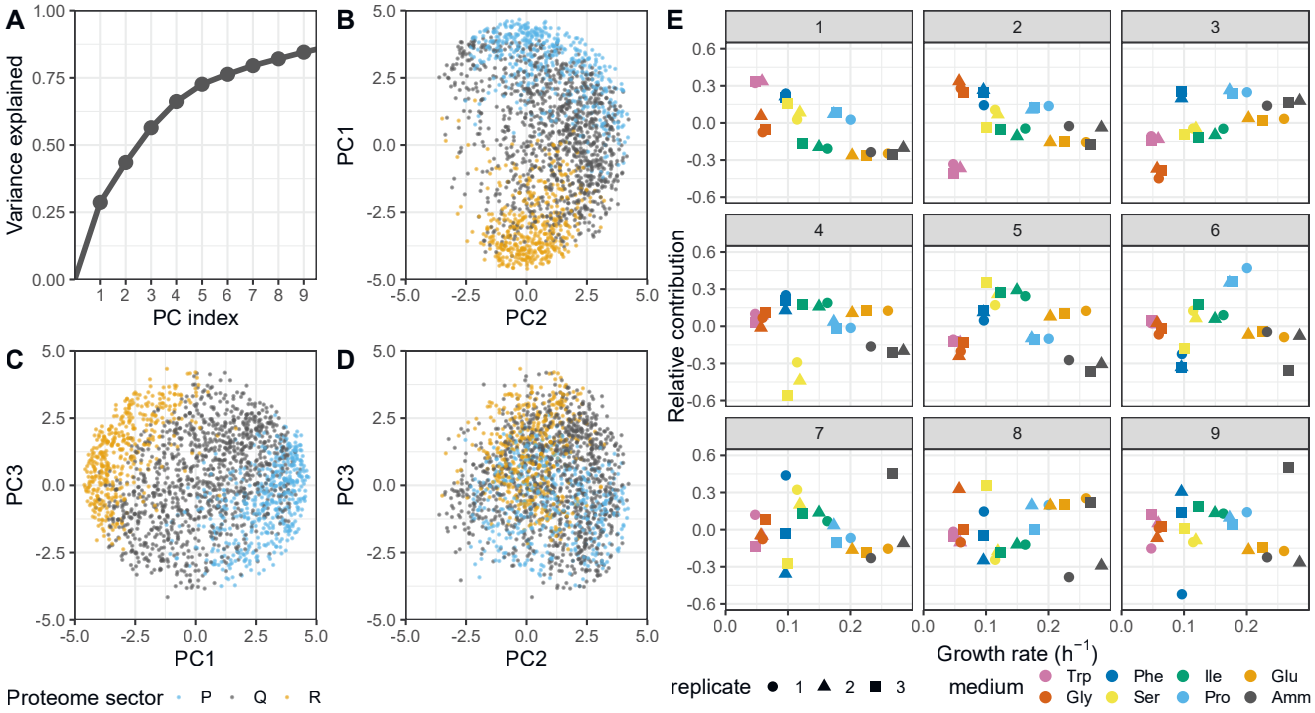


Figure 2.30: Principal component analysis (PCA) of proteome expression data. **A.** Cumulative variance of proteome expression data explained by the first 9 principal components (PCs) of the PCA. **B.–D.** Scatter plots of principal components for all protein groups detected across all eight growth conditions and their proteome sector assignments (R: orange, P: blue, Q: grey). **E.** PC contributions to protein expression across growth conditions as a function of growth rate.

molecules. Therefore, the elevated Adh1/Tdh1 balance exerts a pressure on the NAD+/NADH equilibrium towards the NAD+ side. The induction of Adh1 and repression of Tdh1 proteins may be a controlled response to maintain homeostasis under disruptions to the NAD+/NADH redox balance. This way, differential resource allocation towards the NAD-cycling glycolytic–fermentation pathway may indicate that the metabolic rewiring invoked by the WFSP nitrogen sources could result from changes in the cell redox balance.

To follow up on this observation, we further investigated the burden of NAD-dependent pathways. NADH is oxidised by NADH dehydrogenases that are situated in the inner mitochondrial membrane; the enzyme transfers two electrons per NADH molecule to the electron transport chain to power ATP synthesis. On the other hand, NAD+ is reduced several times during each iteration of the TCA cycle by the α-ketoglutarate (αKG) dehydrogenase complex (KGDHC), the isocitrate dehydrogenase (IDH) complex, and the malic enzymes. Fission yeast is thought to have two separate NADH dehydrogenase enzymes, Ndi1 and Nde1, with the NAD-binding domain of Ndi1 facing the mitochondrion and Nde1 facing the cytosol. We examined the expression burden of these enzymes in our data and found that, although neither belonged to one of the WFSP lists, the ratio of Nde1/Ndi1 expression was strongly elevated in the WFSP conditions (Figures 2.30H and 2.32). The IDH complex comprises the two subunits Idh1 and Idh2, and KGDHC consists of four subunits: Kgd1, Kgd2, Ymr31, and Dld1, the latter being part of multiple complexes. Dld1 and Idh2 were part of the WFSP+ class, unlike any of the other subunits. As above, the ratio of protein abundances for Dld1/Kgd1 and Idh2/Idh1 were elevated in the WFSP conditions (Figures 2.30IJ and 2.32). Therefore, the response to the WFSP nitrogen sources altered the stoichiometry of NAD-dependent enzymatic complexes. Importantly, these signatures were not detected in our transcriptomics data, suggesting a role for post-transcriptional regulation. In line with this, ubiquitin and its related pathways, as well as the translation factors eIF3e and eIF5A, showed strong WFSP patterns suggesting a role for protein stability (Figures 2.6 and 2.20B–D). In summary, we identified two distinct cellular states that differed in the expression of enzymes involved in fermentation and the cell's redox balance that were not correlated with the growth rate.

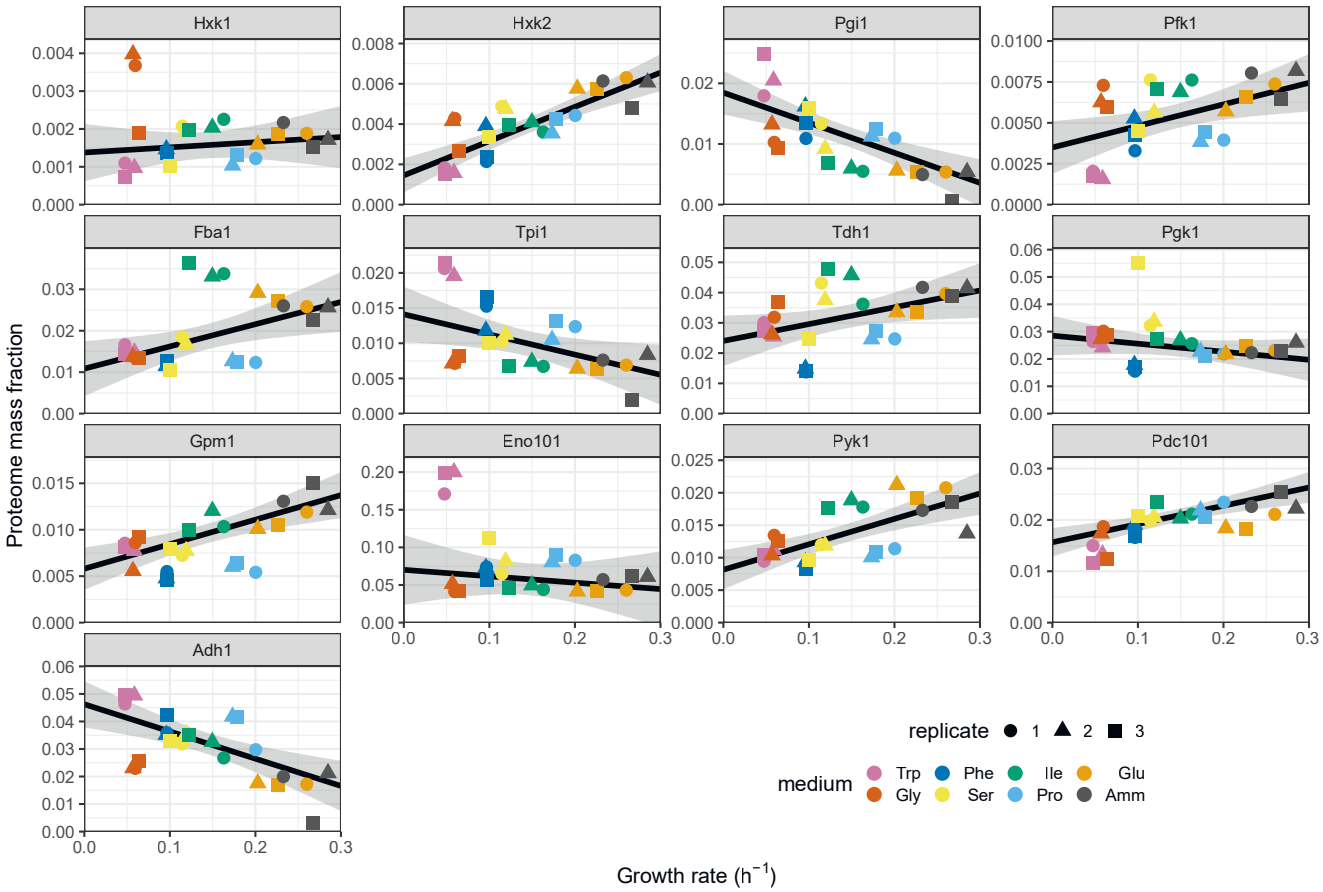


Figure 2.31: Growth rate models of enzymes comprising the glycolysis and ethanol fermentation pathways. Proteome expression and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey for the enzymes comprising glycolysis and ethanol fermentation.

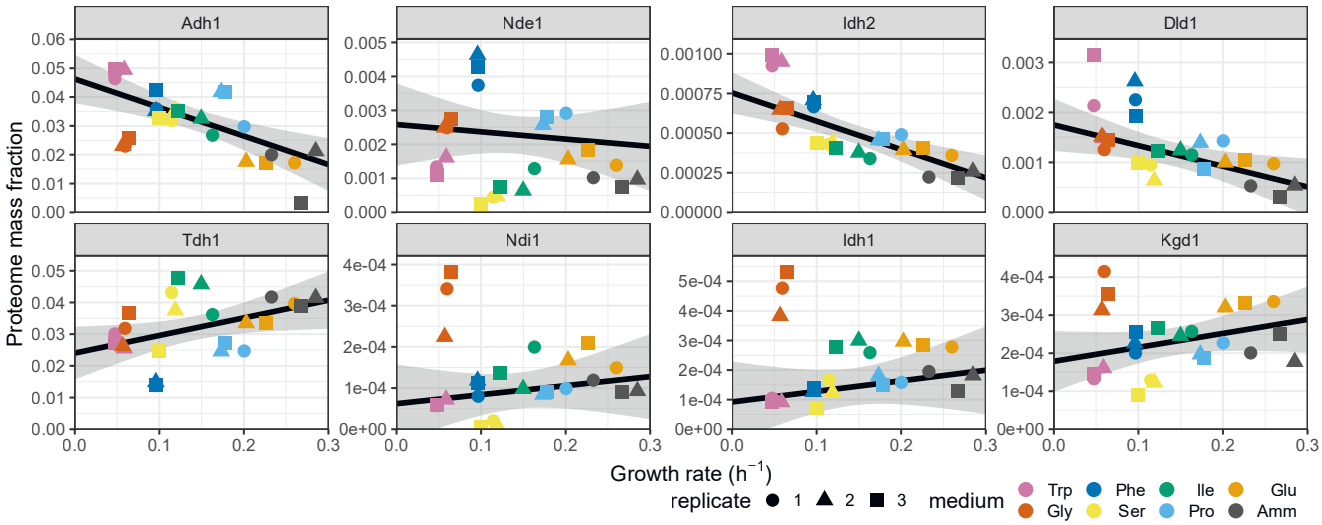
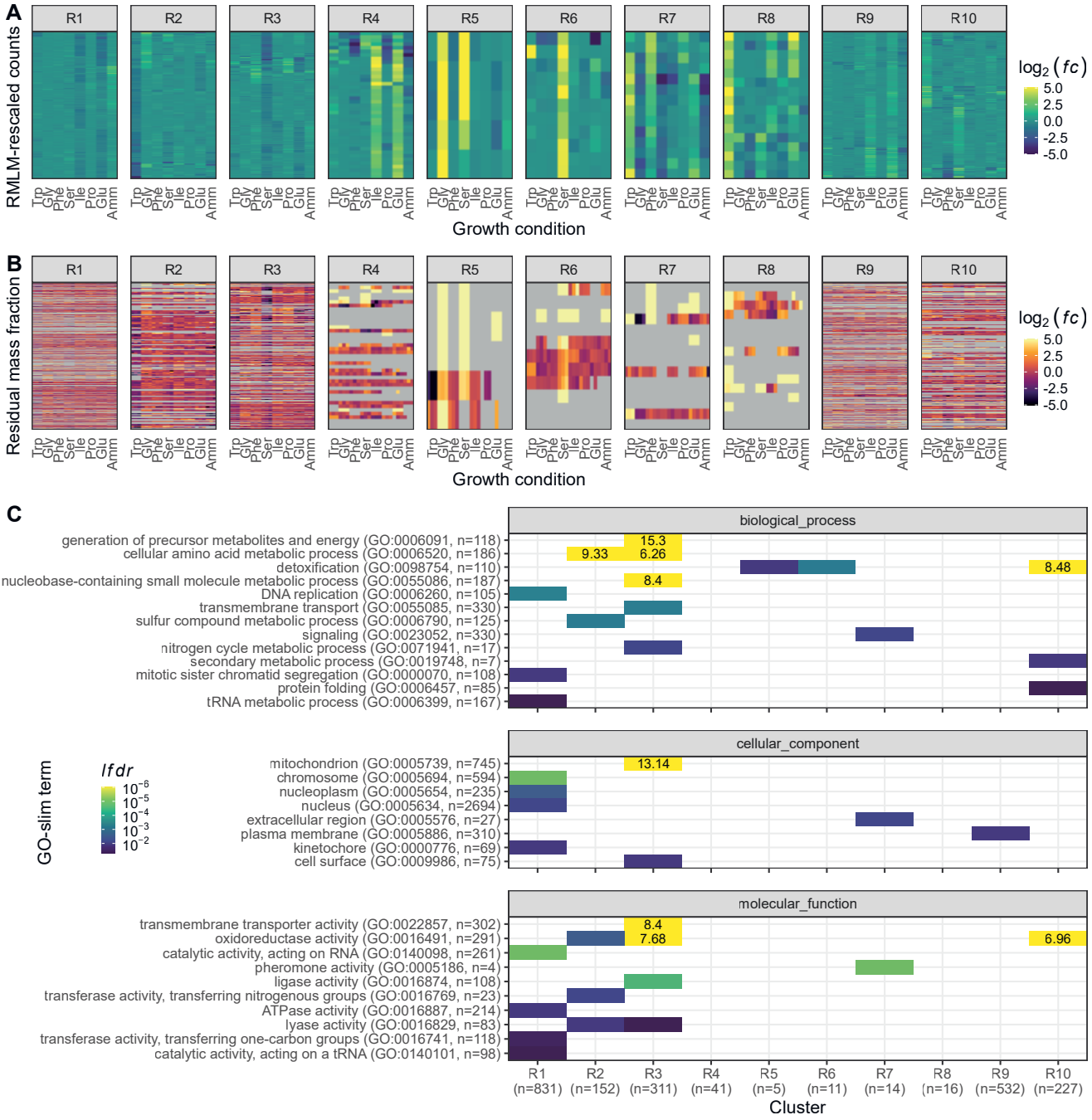


Figure 2.32: Growth rate models of selected enzymes requiring the NAD cofactor. Proteome expression and repeated-median linear model (RMLM) best fit in solid black and 95% confidence interval in grey.

2.3.13 Correcting for growth-rate dependence revealed additional transcriptional signatures of growth on single amino acid sources

Defining the heterogeneity of metabolic states is key to a mechanistic understanding of cell population evolution, but this requires disentangling the gene signatures that depend on the growth rate from those that are purely nutrient specific. Our dataset has the unique capacity to achieve this. We performed differential expression analysis on our RNA-Seq dataset, by comparing each growth condition to a reference transcriptome obtained via averaging all the conditions, and corrected for the growth-dependent component of gene expression (Methods). We defined 10 signatures (termed R1–R10) by clustering the log2-transformed fold change ratios with respect to the synthetic reference of all genes that were significantly enriched in at least one condition (Figures 2.33A and 2.34).



(Caption on the following page.)

Figure 2.33: Exploration of nitrogen-source-specific transcriptome signatures after correcting for growth rate dependence. Figure printed on the preceding page. **A.** DESeq2 log2-transformed fold changes (*fc*) relative to the repeated-median linear model (RMLM) predicted synthetic reference (after shrinkage) for 10 transcriptome signatures R1–R10 with scales capped at ± 5. The growth conditions in the columns were ordered by their growth rate; the genes in the rows according to a hierarchical clustering analysis (Figure 2.34). **B.** log2-transformed ratios of observed and RMLM-predicted proteome fractions for the same clusters and gene ordering as shown in A. Grey tiles indicate missing data in the proteomics. **C.** GO-slim enrichment analysis of clusters from A. Log-transformed local false discovery rates $\mathtt{lfdr}$ are shown for the three Gene Ontology namespaces and capped at $\mathtt{lfdr} = 1.0 \times 10^{-6}$ (printed if capped). Overrepresentation analyses were performed as one-sided Fisher exact tests and only lists with $\mathtt{lfdr} < 0.05$ are shown; the numbers of genes in each GO-category and cluster are indicated.

The 10 signatures covered the differential expression of 2140 genes in total, representing ~43% of the fission yeast transcriptome. Five signatures (R2, R3, R5, R6, and R8) were also visible at the proteome level (Figure 2.33B). About 67% of the mRNA present in the transcriptomic signatures were quantified in at least one condition in the proteome and ~38% were detected in all conditions, indicating that this relatively limited agreement was not due to the lower coverage of the proteomics data.

We next performed functional enrichment analyses of the transcriptomics clusters (Methods), using Gene Ontology annotations (Gene Ontology Consortium 2019; Lock et al. 2019). Broader functional categories were captured using GO-slim analysis (Figure 2.33C), and specific pathways using terms from the biological_process ontology with at most 50 annotations. List overlap analyses (Figure 2.35) as well as gene set enrichment analyses (GSEA), ranking genes based on their log2 fold change over the synthetic reference after shrinkage, were performed for each growth medium (Figure 2.36). In agreement with our observation that respiratory genes were repressed in Ser medium, the Ser repressed cluster R3 was strongly enriched for genes related to mitochondrial metabolism. Additionally, genes from clusters R6 and R10, which were induced in Ser medium, were enriched for detoxification; this was confirmed in the GSEA as well. The Ser response also contained oxidoreductases and proteins involved in metal ion homeostasis, which is compat-

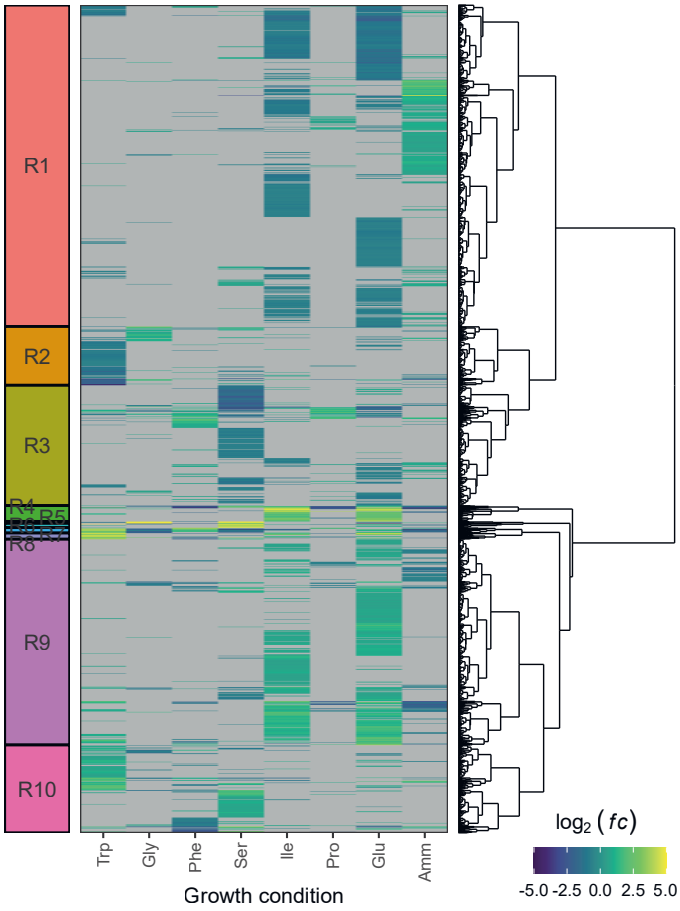
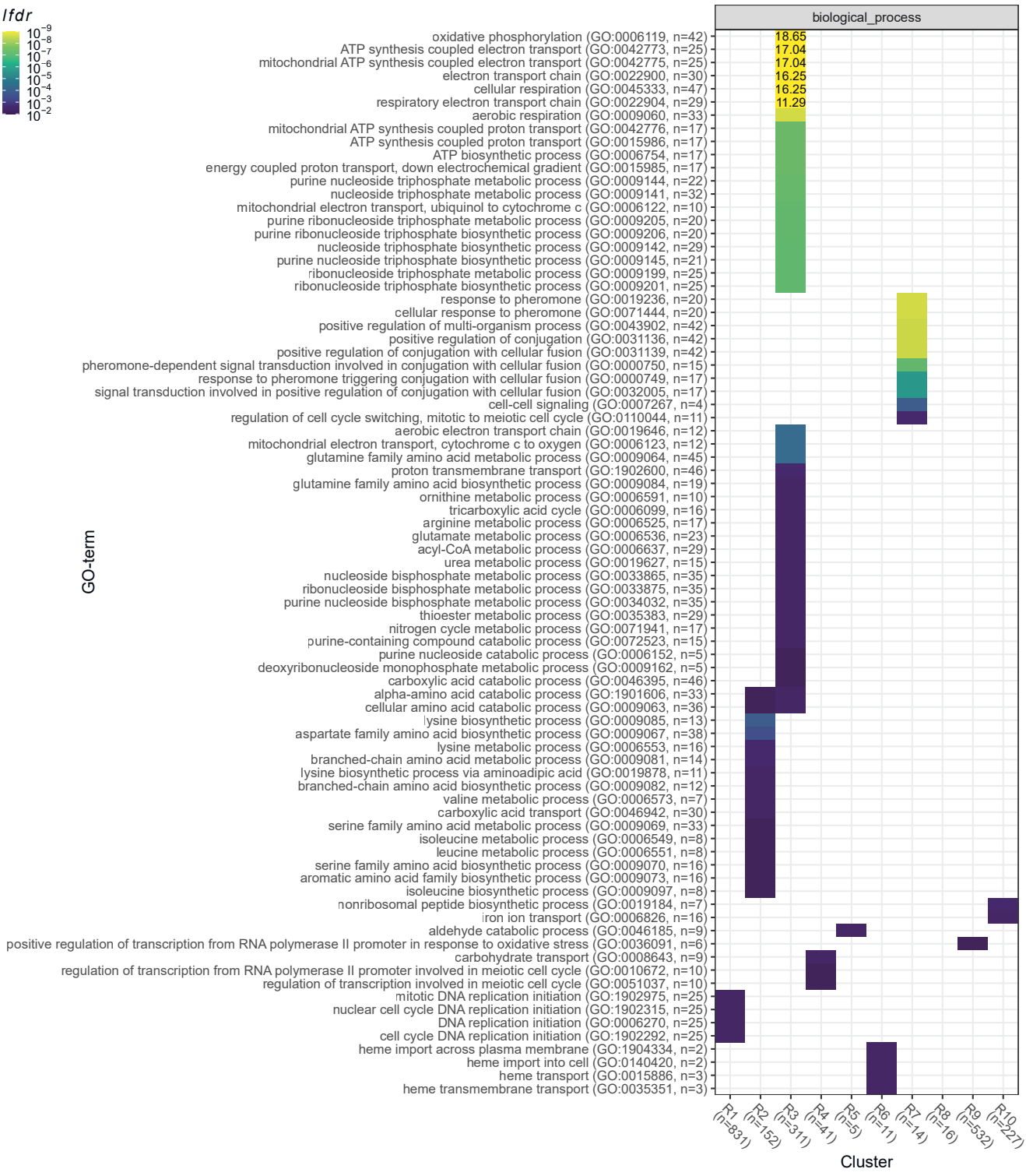


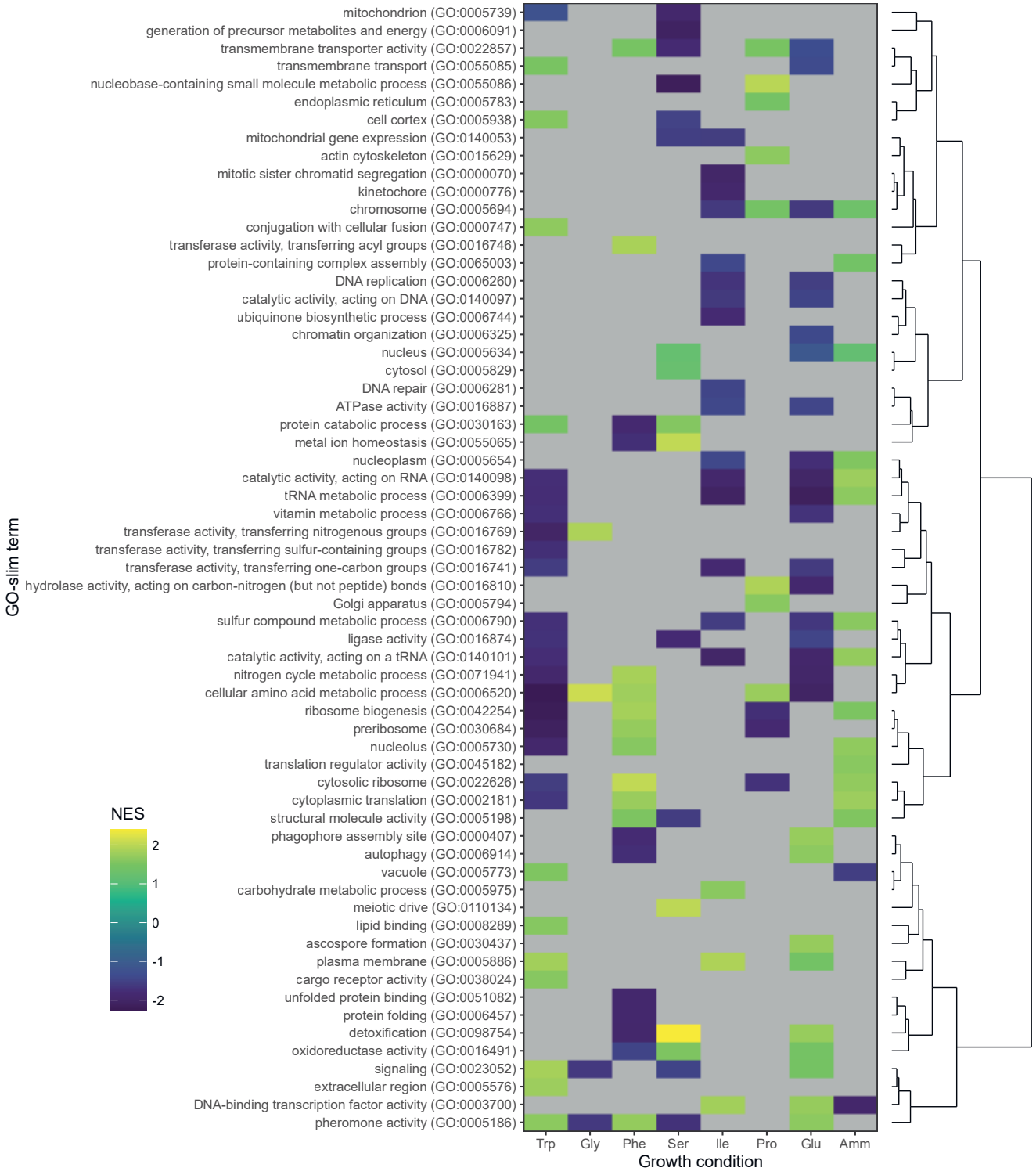
Figure 2.34: Hierarchical clustering analysis of transcriptome signatures after correcting for growth rate dependence. DESeq2 log2-transformed fold changes (*fc*) relative to the repeated-median linear model (RMLM) predicted synthetic reference (after shrinkage) with scales capped at ± 5. Only genes with at least one condition meeting abs$(\log_2 \mathrm{fc})$ and $p_{adj} < 0.01$ were coloured in. Gene-condition pairs not meeting this threshold are shown in grey. The 10 clusters further explored are indicated by the coloured bar on the left.

ible with the recently reported high levels of ROS generated by serine catabolism (Kanou et al. 2020). The Trp repressed cluster R2 was enriched for genes related to amino acid metabolism (Figure 2.35) and the corresponding GO-term also had a negative NES value (Figure 2.36), again suggesting that the slow growth sustained by the Trp medium was not due to any additional expression burden of disrupted amino acid synthesis. The small cluster R7 was enriched for genes related to pheromone activity (M-factor precursors), signalling, and the induction of meiosis (Figure 2.35). Interestingly, the signature expression across conditions for these genes (induced in Trp, Phe, Pro, and Glu containing media) mirrored that of Mae2 (Figure 2.6F), which removes excess carbon from the TCA cycle. As meiosis is usually induced by nitrogen starvation (Petersen and Russell 2016), this result suggests that the state of central carbon metabolism may also play a role in the meiotic transition, as (elemental) nitrogen was abundant in all growth media used. Altogether, we identified a rich set of metabolic signatures that were not dependent on the growth rate, but exclusively reflect changes in external nutrients.



(Caption on the next page.)

Figure 2.35: Gene Ontology overlap analyses of transcriptome signatures after correcting for growth rate dependence. The analysis was restricted to GO-terms belonging to the "biological_process" ontology with at most 50 annotations. Log-transformed local false discovery rates $\mathtt{lfdr}$ are shown for the three Gene Ontology namespaces and capped at $\mathtt{lfdr} = 1.0 \times 10^{-9}$ (printed if capped). Overrepresentation analyses were performed as one-sided Fisher exact tests and only lists with $\mathtt{lfdr} < 0.01$ are shown; the numbers of genes in each GO-category and cluster are indicated. Figure printed on the previous page.



(Caption on the following page.)

Figure 2.36: Gene set enrichment analyses (GSEAs) of medium-specific expression. Heatmap and associated dendrogram of normalised enrichment scores (NES) associated with eight GSEAs of GO-slim terms. For each growth condition, one GSEA was performed: ranking transcripts by their estimated and shrinkage-corrected log fold change relative to the synthetic reference. Terms with at least one significant enrichment ($p_{adj} < 0.05$) are shown; term-condition pairs not meeting this threshold are shown in grey. Figure printed on the preceding page.

2.4 Discussion

In this chapter I showed the quantitative proteome and transcriptome of the fission yeast *S. pombe* grown in eight defined media that provided a range of growth rates. Each medium contained a single source of nitrogen, which was present in abundance. Therefore, the observed variations in gene expression were determined by system-level shifts in resource allocation rather than by the response of a single pathway to the titration of a limiting nutrient. In earlier studies, which focused on *S. cerevisiae*, a specific limiting nutrient was typically used to perturb resource allocation (Brauer et al. 2008; Hui et al. 2015; R. Yu et al. 2020). In chemostats the growth rate is affected externally by the nutrient quantity, and the same growth rate can be obtained by limiting several different nutrients (Airoldi et al. 2016). The turbidostat cultures used for the analysis presented here can be understood as continuous flask cultures, in which the growth rate is determined by internal allocation in response to the nutrient quality.

The results from this chapter suggest that shifts in resource allocation can trigger two layers of gene expression regulation, with the first layer resulting in significant growth-rate correlations and the second layer resulting in condition-specific effects. Many proteins and mRNAs showed a combination of both layers of regulation. This suggests that condition-specific responses occur on top of global gene regulation that is coordinated with the growth rate (Shahrezaei and Marguerat 2015). Importantly, the growth-rate-dependent component complicates the interpretation of condition-specific responses. Where condition-specific responses are seen together with (expected) differences in growth rate, the growth-rate dependent component of gene expression regulation should be accounted for (Airoldi et al. 2009; Pancaldi et al. 2010; R. Yu et al. 2021). The subtraction of a robust linear model, as proposed here, is one way to do so.

Importantly, the global layer of regulation discussed here affects relative abundances of proteins and of mRNAs. The growth rate is also known to affect the absolute abundance of macromolecules, but indirectly: scaling of gene expression with cell size generally ensures a constant concentrations of mRNA with only few exceptions (Padovan-Merhar et al. 2015; Chávez et al. 2016). This effect is mediated by a size-dependent increase of RNA polymerase II (RNAPII) initiation rates, such that RNAPII abundance is limiting for growth (Sun et al. 2020). The choice of nitrogen source does affect the cell size but not in a manner coordinated with the growth rate, as confirmed in an analysis of which more details are provided in chapter 5. In line with this, we found that RNAPII expression was not increasing with the growth rate suggesting that, unlike for gene expression scaling to cell size, its numbers are not limiting for the rate of growth.

Curiously, the variably-expressed enzymes were primarily involved in carbon metabolism, whereas the growth media differed only in their nitrogen sources and glucose was abundantly present in each of them. The broad features of the catabolism of the nitrogen source and its possible effects on growth-optimal gene expression are studied in the remainder of the thesis using coarse-grained mathematical modelling. A more detailed view of the interplay between nitrogen and carbon metabolism at the level of individual enzymes might be given by future genome-scale models of fission yeast that integrate metabolism and gene expression. The data generated for this thesis and the published paper (Kleijn et al. 2022) might be used to constrain and inform the construction of such models (O'Brien et al. 2013; Sánchez et al. 2017; Y. Chen et al. 2021).

As expected based on earlier studies in other model organisms, the main feature of the data was a trade-off between proteins involved in protein production and stress and metabolism. All steps of protein production were represented in the R sector, and almost all ribosomal proteins were strongly positively correlated with the growth rate. Interestingly the growth law parameters were not significantly different between fission yeast and budding yeast, suggesting there might be a biophysical constraint on the translation rate. Related to this, the chemical composition of ribosomes—many dozens of small proteins bound to a small handful of large RNA molecules—is optimal for fast self-replication (Reuveni et al. 2017). The slope of the growth law in *E. coli* was smaller than both yeasts; this quantitative difference may be explained by fundamental differences between the eukaryotic and prokaryotic ribosome. Notably, the main qualitative features of the growth

law were present both in the data presented in this chapter and in comparable data on *E. coli*, suggesting that the main underlying principles are conserved far across the tree of life.

Unlike the expression of ribosomal proteins, the expression of individual metabolic enzymes was strongly condition-specific. Yet, the aggregate of this condition-specific regulation still induced a large change in the total proteome burden of metabolism, as the glycolytic proteins and enzymes related to NAD metabolism are generally highly abundant. In this light, it is interesting to note that the relative activity levels of many promoters, including ones of genes involved in translation, were constant across a variety of environments in both *E. coli* and *S. cerevisiae* (Keren et al. 2013). The induction of medium-specific gene expression can have a large passive effect on the expression of such constitutively expressed proteins, such that they appear as part of the R sector even in the absence of additional regulation (Barenholz et al. 2016). In this interpretation, a small allocation to condition-specific enzymes will allow for a larger R-sector allocation and thereby induce faster growth. On the other hand, explicit regulation of growth-rate correlated expression, e.g. mediated by the TOR pathway, could similarly decrease the upper bound on the expression of condition-specific enzymes. Steady-state expression data does not provide evidence to decide which regulation is more prominent and full understanding of the interplay between growth and gene expression must account for both of these effects simultaneously.

In summary, resource allocation in fission yeast across media with different nitrogen sources is characterised by two distinct expression programmes: (i) a growth-rate-correlated trade-off between protein production and stress response, and (ii) medium-specific responses to the growth conditions with a large effect on the expression burden of metabolism. To better understand the interplay between growth and gene expression, I constructed coarse-grained models of cellular resource allocation, which will be explored in the remainder of this thesis.

3 Methodology of coarse-grained resource allocation models (C-GRAMs)

3.1 Introduction

In many unicellular organisms, both cell size and the abundances of large protein classes are correlated with the growth rate across environmental conditions or upon translational inhibition. To explain these observations, I constructed coarse-grained resource allocation models (C-GRAMs), which can provide a general understanding of the correlations between protein abundance, cell size, and growth. In chapter 4, I describe how minimal C-GRAMs with nitrogen and carbon pathways converging on biomass production explained the effects of the uptake of sugars, ammonium, and/or compound nutrients such as amino acids on the translational resource allocation towards proteome sectors that maximised the growth rate. In chapter 5, I describe how I extended the minimal model to a C-GRAM that additionally accounted for the cell cycle, cell division, cell wall biosynthesis, and the effect of molecular crowding on the ribosomal efficiency. In this chapter, I give a general overview of the methodology used to construct and analyse both of these C-GRAMs, as well as some insights that should be more generally applicable to this type of modelling. C-GRAMs codify scientific understanding in an organism-agnostic manner, thereby affording the study of growth rate maximisation and the interplay between cell size, resource allocation, and other aspects of cell physiology across species in a unified framework.

3.2 General concepts

3.2.1 The dynamics of relative abundances

Several excellent reviews of the basic concepts underlying mechanistic whole-cell models have been published in the last few years, notably by De Jong et al. (2017), Bruggeman et al. (2020), and Dourado and Lercher (2020). These models describe the growing cell as a well-mixed solution of proteins, sometimes transcripts, and intermediate metabolites. Such approaches can be used to describe genome-wide

as well as coarse-grained whole-cell models. In this thesis we limit ourselves only to models of the latter type. What follows is a synthesis of concepts with a focus on these coarse-grained models.

Consider the vector $\mathbf{X}$ of absolute abundances associated with all the molecular species present in the cell. These abundances can be measured in multiple ways; three common choices are outlined in Table 3.1. We now consider the system size

$$\Omega = \sum_i X_i. \tag{3.1}$$

It is often thought of as cell volume, but this makes certain assumptions that will be explored further in section 3.2.2. For cultures in balanced growth, such as those obtained in the turbidostats of chapter 2, the growth rate

$$\mu = \frac{1}{\Omega}\frac{d\Omega}{dt} \tag{3.2}$$

is well-defined and constant. Notably, in this formalism, the growth rate is an emergent property of the system.

The relative abundances

$$\mathbf{x} = \frac{\mathbf{X}}{\Omega} \tag{3.3}$$

follow the concentration constraint

$$\sum_i x_i = 1. \tag{3.4}$$

The dynamics of any given cellular constituent i, written in terms of its relative abundance x_i, is calculated as

$$\dot{x}_i = \frac{\dot{X}_i}{\Omega} - \frac{\dot{\Omega} X_i}{\Omega^2} \equiv \nu_i(\mathbf{x}) - \mu x_i, \tag{3.5}$$

where the net production rate

$$\nu_i(\mathbf{x}) = \frac{\dot{X}_i}{\Omega} \tag{3.6}$$

has been written in terms of the relative abundances to create a closed system of equations.

Table 3.1: Common choices for the abundance measure in whole-cell models. Note that molar concentrations $c = \frac{N}{\Omega_V}$ and mass concentrations $\rho = \frac{M}{\Omega_V}$ are composite measures.

absolute abundance X	system size Ω	relative abundance x
molecule number N	total number of molecules Ω_N	molar fraction
mass of molecular species M	cell mass Ω_M	mass fraction ϕ
occupied volume	cell volume Ω_V	volume fraction

Balanced growth, introduced earlier in section 1.1.1, is characterized by the steady state of these equations, i.e. $\dot{x} = 0$, such that for each molecular species i, its production rate is balanced by the dilution due to growth:

$$v_i(\mathbf{x}) = \mu x_i. \tag{3.7}$$

In other words, the relative abundances are constant during balanced growth, because all constituents in the vector $\mathbf{X}$ grow exponentially with the same rate μ. Furthermore, any well-defined choice of abundance measure must by its construction result in the same growth rate μ.

3.2.2 The dynamics of absolute abundances and concentrations

Sometimes, reaction fluxes are determined not in terms of relative abundances, but in terms of absolute abundances. That is, instead of $v_i(\mathbf{x})$ being directly known, the dynamics of X_i are given in terms of

$$\dot{X}_i \equiv V_i(\mathbf{X}). \tag{3.8}$$

It may sometimes be assumed that the flux vector $\mathbf{V}$ is *extensive*, meaning that it satisfies the requirement

$$\mathbf{V}(\omega\mathbf{X}) = \omega\mathbf{V}(\mathbf{X}) \tag{3.9}$$

for any constant ω. In that case, relative and absolute fluxes can be readily converted into each other, because

$$\mathbf{V}(\mathbf{X}) = \frac{1}{\Omega}\mathbf{V}(\frac{1}{\Omega}\mathbf{X}) = \frac{1}{\Omega}\mathbf{V}(\mathbf{x}) \equiv \mathbf{v}(\mathbf{x}). \tag{3.10}$$

However, this assumption is not usually valid when metabolite abundances are included in the vector $\mathbf{X}$—see section 3.3.1 for the reason why.

Furthermore, the system can also be described in terms of concentrations, where the system size is explicitly linked to cell volume, but such a description is subtly different from the above. For example, Bruggeman et al. (2020) consider the molar concentrations, $\mathbf{c} = \frac{\mathbf{N}}{\Omega_V}$, with $\mathbf{N}$ the vector of molecule numbers and Ω_V the cell volume. A similar argument applies to the mass concentrations, as used by De Jong et al. (2017). The basic form of the equation describing the dynamics of the concentration vector $\mathbf{c}$ takes the same form as equation (3.5), namely

$$\dot{\mathbf{c}} = \frac{\dot{\mathbf{N}}}{\Omega_V} - \frac{\dot{\Omega}_V \mathbf{N}}{\Omega_V^2}. \tag{3.11}$$

Note that $\mu_V \equiv \frac{1}{\Omega_V}\frac{d\Omega_V}{dt}$ can be defined in terms of the total volume, whereas $\dot{\mathbf{N}} = \boldsymbol{V}(\mathbf{N})$ is initially given in terms of molecule numbers.

In general, for systems described by concentration dynamics such as equation (3.11), a system size must be specified to normalise the abundances. Here this role is played by the total volume Ω_V. Similarly, the total mass Ω_M could be considered, or even a compound measure such as the total cell buoyant density Ω_M/Ω_V. This contrasts with systems described by the relative abundance dynamics of equation (3.5), for which the absolute abundances were sufficient.

3.2.3 The interpretation of coarse-grained models

The key choices in formulating a model, then, are the biological interpretations of the molecules $\mathbf{x}$ and the functional forms that the fluxes $\boldsymbol{v}(\mathbf{x})$ assume. In a coarse-grained model, few molecules are included—in other words, the dimensionality of $\mathbf{x}$ is small, typically not much greater than 10 and sometimes as small as 3 or 4. These models usually focus on protein abundances, with the metabolites there to regulate the protein production. Some models additionally include transcription and translation of mRNA. A very minimal example model was initially proposed by Maitra and Dill (2015) and expanded to include ribosome inactivation (Maitra and Dill 2016). Two models including both transcription and translation are those by Weiße et al. (2015) and Liao et al. (2017). Importantly, molecular species in a coarse-grained resource allocation model do not map one-to-one to molecules that might be found in the cell. Rather, they represent the total abundance of whole classes of proteins.

3.2.4 The stoichiometry matrix

The flux vector $\boldsymbol{v}(\boldsymbol{x})$ can be written in terms of the stoichiometry matrix S and a vector $\mathbf{j}(\boldsymbol{x})$ that contains the conversion rates associated with each reaction (Heinrich and Schuster 1996):

$$\boldsymbol{v}(\boldsymbol{x}) = S\mathbf{j}(\boldsymbol{x}). \tag{3.12}$$

In this matrix S, each column corresponds to a molecular species, and each row to a reaction. Negative values correspond to the consumption of the molecule and positive values correspond to production. If reaction j is balanced, this means that the j-th column of S has sum zero:

$$\sum_{i=1}^{n} S_{ij} = 0. \tag{3.13}$$

In coarse-grained models, not all reactions are balanced, because the system describes the state of the cell, or of the whole culture. Therefore, reactions exist that grow the cell (or culture) by consuming external nutrients, thereby increasing biomass. For such reactions,

$$\sum_{i=1}^{n} S_{ik} = 1, \tag{3.14}$$

representing that each "firing" of the reaction k increases the total abundance Ω by one unit.

3.2.5 The distinction between proteins and metabolites

At this point, it is helpful to distinguish metabolites from macromolecules. For proteins, v_i typically represents only the production rate, because many proteins are stable for many cell generations, and their degradation rates are negligible compared to the dilution rate μ that is due to growth. For metabolites, v_i represents the net metabolic flux, incorporating both their production and utilisation. For metabolites, it can be appropriate to neglect the growth-dilution term when their dynamics are dominated by fast turnover rates. For notation purposes, we split the concentration vector $\boldsymbol{x}$ into three parts, representing amino acids, other metabolites, and proteins:

$$\boldsymbol{x} = \begin{pmatrix} \mathbf{a} \\ \mathbf{m} \\ \mathbf{p} \end{pmatrix}. \tag{3.15}$$

The proteins $\mathbf{p}$ are the main constituents of the total biomass and are produced from the amino acids a by a single reaction, catalysed by the ribosomes. The remaining metabolites $\mathbf{m}$ are intermediates. Likewise, the flux vector is decomposed as

$$\mathbf{j}(\mathbf{x}) = \begin{pmatrix} \hat{\mathbf{j}}(\mathbf{x}) \\ j_R(\mathbf{x}) \end{pmatrix} \tag{3.16}$$

and the flux through the ribosomes j_R is separated from the remaining fluxes $\hat{\mathbf{j}}$.

The protein production reactions are taken to be catalysed by the ribosomes, themselves represented as proteins in this formalism. Importantly, they consume the amino acids a. The relative protein production rates are determined by an allocation vector $\mathbf{f}$, which has the same number of components as $\mathbf{p}$. Considering a stable protein i that does not change or degrade after its production, its time evolution is given by

$$\dot{p}_i = f_i j_R(\mathbf{x}) - \mu p_i. \tag{3.17}$$

If required, this equation can be supplemented with a degradation term, but we typically assume that proteins are stable. The ribosomal production reactions satisfy mass conservation typified by equation (3.13). In other words, the consumed substrate a must be balanced with the proteins produced in the ribosomal flux $j_R(\mathbf{x})$. In practice, this means that

$$\sum_i f_i = 1. \tag{3.18}$$

We refer to this as the allocation constraint.

If we assume that all proteins are produced by ribosomes, and do not change at all afterwards, the stoichiometry matrix S takes a particular form. In a system

with n conversion reactions, one amino-acid component, m metabolites, and p proteins, S is an $(1+m+p)$ by n matrix with the following shape:

$$S = \begin{pmatrix} S_1^A & S_2^A & \cdots & S_n^A & -1 \\ S_{1,1} & S_{1,2} & \cdots & S_{1,n} & 0 \\ S_{2,1} & S_{2,2} & \cdots & S_{2,n} & 0 \\ \vdots & \vdots & \ddots & \vdots & 0 \\ S_{m,1} & S_{m,2} & \cdots & S_{m,n} & 0 \\ 0 & 0 & \cdots & 0 & f_1 \\ 0 & 0 & \cdots & 0 & f_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & f_p \end{pmatrix} = \begin{pmatrix} \left(\mathbf{S}^A\right)^T & -1 \\ \hat{S} & \mathbf{0} \\ \emptyset & \mathbf{f} \end{pmatrix}. \tag{3.19}$$

We call the m by n matrix $\hat{S}$ the metabolic stoichiometry matrix; the (row) vector $\mathbf{S}^A$ quantifies the stoichiometry of the reactions that produce the amino acids.

In the end, then, the concentration dynamics of coarse-grained models with a single biomass production reaction catalysed by ribosomes and stable proteins can generally be described as:

$$\dot{\mathbf{x}} = \begin{pmatrix} \dot{a} \\ \dot{\mathbf{m}} \\ \dot{\mathbf{p}} \end{pmatrix} = S\mathbf{j}(\mathbf{x}) - \mu\mathbf{x} = \begin{pmatrix} \mathbf{S}^A \cdot \hat{\mathbf{j}}(\mathbf{x}) - j_R(\mathbf{x}) - \mu a \\ \hat{S}\hat{\mathbf{j}}(\mathbf{x}) - \mu\mathbf{m} \\ j_R(\mathbf{x})\mathbf{f} - \mu\mathbf{p} \end{pmatrix}. \tag{3.20}$$

This decomposition is a helpful tool to understand the models of chapters 4 and 5, but is is not always applicable. In particular, it does not apply to models such as that by Bertaux et al. (2020) and explored in section §3.4, which incorporate conversions between different protein classes after their production.

3.3 Enzyme kinetics

3.3.1 Michaelis–Menten kinetics for single-substrate reactions

The functional forms of the fluxes in coarse-grained models are inspired by models of enzyme kinetics, for example through Michaelis–Menten rate dependences. For a Michaelis-Menten enzyme X with concentration e_X, the production rate $j^{(X)}$ of a

metabolite by that enzyme through conversion of a substrate with concentration x_i is given by

$$j^{(X)}(e_X; x_i) = k^{(X)} e_X \frac{x_i}{x_i + K_M^{(X)}}, \tag{3.21}$$

with $k^{(X)}$ and $K_M^{(X)}$ constants that can be related to the enzyme kinetic rates. While there is no direct justification why single-enzyme models can be directly applied to the abstracted enzyme classes of coarse-grained models, the Michaelis-Menten rate law has three useful and important properties, namely it is (i) linear in the enzyme concentration, (ii) linear in the substrate concentration at low concentrations, while it (iii) saturates with increasing substrate concentrations.

Empirically, the growth rate of bacterial cultures follows a similar dependency on the availability of a single limiting nutrient source, the concentration of the nutrient source in the environment taking the place of the substrate concentration in the cell (Monod 1949). The dependency from equation (3.21) is then known as the Monod curve; it can be explained by assuming Michaelis–Menten kinetics of a limiting enzyme (De Jong et al. 2017). However, the empirical observation is strong: the enzyme kinetics described by equation (3.21) can be used to describe whole pathways, even though pathway fluxes are not always limited by the kinetics of a single enzyme.

It is quickly seen that the Michaelis–Menten enzyme flux is not extensive. Let ω a constant, then it follows from equation (3.21) that

$$\begin{aligned} j_i(\omega e_X; \omega x_i) &= k^{(X)} \omega e_X \frac{\omega x_i}{\omega x_i + K_M^{(X)}} && = \\ &= \omega k^{(X)} e_X \frac{x_i}{x_i + \frac{K_M^{(X)}}{\omega}} && \neq \omega j_i(e_X; x_i). \end{aligned} \tag{3.22}$$

Regardless, the hyperbolic dependence on x_j is preferred, because the alternative linear dependence would mean that high enzyme fluxes can be sustained by unrealistically high metabolite concentrations. However, the non-extensivity of the flux function means that it is important to choose the exact abundance measure in which the fluxes are expressed, because the assumptions required by the equivalence equation (3.10) are not met. In chapter 4, the system is described in terms of mass fractions. The model in chapter 5 additionally describes the cell size using multiple size measures, so there the model is recast in terms of absolute abundances per cell.

3.3.2 Two-substrate reactions

For reactions involving more than one substrate, the functional dependency of the catalytic rate has to be chosen too. We opted to use a functional form that can be derived using similar assumptions to those underlying Michaelis–Menten kinetics. This meant that the the three useful properties (i)–(iii) noted in the previous section were maintained.

The flux equation that best describes two-substrate enzyme kinetics depends on the reaction mechanism. The main assumptions and derivations associated with each mechanism have been explained in detail by Marangoni (2003, ch. 7). A main distinction exists between the ping-pong mechanism, where a product is released before the second substrate binds to the enzyme, and sequential mechanisms, where both substrates bind sequentially, either in a random or an ordered fashion, before any product is released. A general form for the conversion rate j_i, catalysed by enzyme X with concentration e_X, taking two substrates with concentrations x_j and x_k, is given by

$$j_i(e_X; x_j, x_k) = k^{(X)} e_X \frac{x_j x_k}{x_j x_k + a x_j + b x_j + c}. \tag{3.23}$$

Here $k^{(X)}$, a, b, and c are free constants whose values can be expressed in terms of kinetic rate parameters. Depending on the reaction mechanism, a, b, or c may be equal to zero. It can be seen that the rate function for two-substrate reactions reverts to a Michaelis–Menten dependency when one substrate concentration is held fixed regardless of the reaction mechanism.

So as to directly compare the parametrizations of one- and two-substrate enzymes, we recast the formula in terms of different parameters as

$$j_i(e_X; x_j, x_k) = k^{(X)} e_X \frac{x_j x_k}{(x_j + K_j^{(X)})(x_k + K_k^{(X)}) - \Delta^{(X)}}. \tag{3.24}$$

In the coarse-grained models described in this thesis, I chose the simple $\Delta^{(X)} = 0$, such that the rate law is simply the product of two Michaelis–Menten rate laws. This symmetrical case corresponds to a mechanism where the equilibrium constant for the binding of each of the two substrates does not depend on the presence or absence of the other. In coarse-grained modelling, the reaction mechanism is not well defined, as the two-substrate enzyme typically represents one or more en-

tire pathways. This choice was therefore made because it treated both substrates symmetrically, so the topology of the full metabolic network did not need to be assessed.

3.4 An example model of coarse-grained metabolism, translation, growth, and division

A recent effort from the Shahrezaei group served as the baseline for the modelling presented in this thesis (Bertaux et al. 2020). The model was used to describe the interplay between cell growth and size regulation in *E. coli*, both deterministically and stochastically. It consisted of a minimal number of components, making it interpretable and applicable to multiple model systems. We preferred this approach to previous efforts describing the dynamics of protein and mRNA abundances in *E. coli*, because the latter contained many parameters (Weiße et al. 2015; Liao et al. 2017), most of which are as yet undetermined for *S. pombe*. Modifying these models would have required a great deal of parameter inference. An earlier model with a suitably minimal approach to the number of variables required ad-hoc assumptions about the degradation of certain enzymes (Maitra and Dill 2015, 2016), which we considered to be inappropriate. A closely related modelling approach to ours was described by Molenaar et al. (2009), and direct comparisons are indicated where appropriate, both in this chapter and the rest of this thesis. Here, I will briefly explain the baseline model as it was originally proposed by Bertaux et al. (2020); any extensions will be described in the remainder of this thesis.

First, metabolism is represented by a single class of enzymes E that catabolise an external nutrient into an internal precursor metabolite A. Active ribosomes R_a then convert this metabolite into proteins, which in addition to E and R include a sizeable fraction of housekeeping proteins Q and a small fraction of cell-cycle regulators X. Furthermore, the model allows for the presence of "useless" protein U and inactive ribosomes R_i, the latter due to a translation-inhibiting drug such as chloramphenicol, representing experiments performed by Scott et al. (2010).

The original model (Bertaux et al. 2020) had extensions to account for both the cell cycle and the stochasticity of cellular reactions. However, I here focus on the steady-state behaviour of the ordinary differential equations (ODEs) correspond-

ing to the relative abundances associated with the aforementioned molecules. In the formalism from equation (3.20), the constituent concentrations are given as

$$\mathbf{x} = \begin{pmatrix} a \\ \mathbf{p} \end{pmatrix} = \begin{pmatrix} a \\ \begin{pmatrix} e & r_a & r_i & q & x & u \end{pmatrix}^T \end{pmatrix}, \tag{3.25}$$

and the allocation vector as

$$\mathbf{f} = (f_E, f_R, 0, f_Q, f_X, f_U)^T. \tag{3.26}$$

No metabolites besides a are present. Note that ribosomes are produced in the active state. The system size is calculated using all abundances, including of the metabolite A, such that its concentration a is included in the constraint

$$\sum_i x_i = a + e + r_a + r_i + q + x + u = 1. \tag{3.27}$$

In contrast, earlier coarse-grained whole-cell models calculated the cell volume or mass based on the protein constituents only (Molenaar et al. 2009; Maitra and Dill 2015; Weiße et al. 2015).

The rate of nutrient uptake is determined by the environment as well as by the kinetic parameters associated with the enzyme E. These effects are assimilated into a single parameter, leading to the (concentration-normalised) flux through E being given by

$$j_E = k_E e, \tag{3.28}$$

with k_E the parameter that describes the nutrient quality. The translation flux through active ribosomes with substrate A is modelled as a Michaelis–Menten rate law, with flux

$$j_R = \sigma \frac{a}{a + a_{sat}} r_a, \tag{3.29}$$

Active ribosomes are inactivated, and *vice versa*, with rates k_{on}^{cm} and k_{off}^{cm}, respectively. All proteins are assumed to be stable. Therefore, the flux vector $\mathbf{j}$ is given by

$$\mathbf{j}(\mathbf{x}) = \begin{pmatrix} k_E e \\ k_{on}^{cm} r_a \\ k_{off}^{cm} r_i \\ \sigma \frac{a}{a+a_{sat}} r_a \end{pmatrix}, \tag{3.30}$$

and the stoichiometry matrix by

$$S = \begin{pmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & f_E \\ 0 & -1 & 1 & f_R \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & f_Q \\ 0 & 0 & 0 & f_X \\ 0 & 0 & 0 & f_U \end{pmatrix}. \tag{3.31}$$

Note that the interconversion between active and inactive ribosomes violates the assumption that all proteins are stable after their production, so the stoichiometry matrix decomposition of equation (3.19) is not valid.

The following set of equations describes the system:

$$\dot{a} = ke - \sigma \frac{a}{a + a_{sat}} r_a - \mu a, \tag{3.32}$$

$$\dot{e} = f_E \sigma \frac{a}{a + a_{sat}} r_a - \mu e, \tag{3.33}$$

$$\dot{r}_a = f_R \sigma \frac{a}{a + a_{sat}} r_a - k_{on}^{cm} r_a + k_{off}^{cm} r_i - \mu r_a, \tag{3.34}$$

$$\dot{r}_i = k_{on}^{cm} r_a - k_{off}^{cm} r_i - \mu r_i, \tag{3.35}$$

$$\dot{q} = f_Q \sigma \frac{a}{a + a_{sat}} r_a - \mu q, \tag{3.36}$$

$$\dot{x} = f_X \sigma \frac{a}{a + a_{sat}} r_a - \mu x, \tag{3.37}$$

$$\dot{u} = f_U \sigma \frac{a}{a + a_{sat}} r_a - \mu u. \tag{3.38}$$

It has been observed in *E. coli* that the translation elongation rate robustly depends on the RNA/protein ratio via a Michaelis-Menten-like hyperbolic function (Dai et al. 2016). In the model, this translates to the requirement that the ribosomal saturation fraction itself follows the following dependence:

$$\frac{a}{a + a_{sat}} \propto \frac{f_R}{f_R + K} \tag{3.39}$$

for some constant K, as the former represents the translation rate and the RNA/protein ratio is a proxy for the ribosome content. We imposed the simple dynamic regulation

$$f_R = \min\left(\delta a, 1 - f_Q - f_X - f_U\right) \tag{3.40}$$

to satisfy the allocation constraint equation (3.18). Transiently, the constraint sometimes needed to be invoked, but the steady-state concentration of metabolic precursor always satisfied

$$a < \frac{1 - f_Q - f_X - f_U}{\delta}. \tag{3.41}$$

This situation, where the metabolic precursor represents a small fraction of the total system size, is also biologically required, and suitable parametrizations all satisfied the restriction from equation (3.41). In this model, then, the allocation fractions f_Q, f_X, and f_U were imposed parameters, f_R was set as above, and the allocation constraint was satisfied by

$$f_E = 1 - f_Q - f_X - f_U - f_R. \tag{3.42}$$

3.5 Optimising ribosomal allocation to maximise growth rate

3.5.1 Maximising growth rate is an effective strategy to minimise the number of parameters

While the model described above in section §3.4 contained few parameters, making it interpretable, the models developed in the remainder of this thesis extended the scope considerably. In particular, I introduced pathways that processed metabolites in parallel to one another. Such pathways must balance their fluxes, as they jointly produce an end molecule. Many examples of regulatory processes have been found that can achieve this, as reviewed by Chubukov et al. (2014). On short time scales, covalent or allosteric interactions between metabolites and proteins govern the response, for example through end-product inhibition of linear pathways. On longer time scales, persistent imbalances are counteracted by broad and pervasive gene expression regulation such as explored in chapter 2. Choosing the right level of coarse-graining here is difficult: both of these types of regulation are required for cells to grow robustly, but introducing multiple levels of regulation would necessarily require many parameters.

We side-stepped this problem by focussing not on the implementation of the regulatory pathways, but by studying the steady-state under the assumption that the cell is able to regulate its expression in a manner that maximised its growth rate. Precise optimality conditions have been formulated for the *E. coli* carbon uptake system and expression is indeed optimal for several carbon sources (Towbin et al. 2017). Furthermore, a recent theoretical advance pointed out a general method of adapting gene expression control towards the optimum (Planqué et al. 2018), which, however, fell short of a general mechanistic implementation of the control. In a model similar in scope to ours, the optimal choice between a metabolically efficient or catabolically efficient pathway was shown to depend on the growth rate (Molenaar et al. 2009).

In practice, the allocation vector $\mathbf{f}$ from equation (3.20) was therefore chosen not as a free parameter, but rather as the result of an optimisation routine that maximised the growth rate μ. Some components of the vector were not part of this optimisation, for example when they represented an experimental constraint (like f_U in section §3.4 above) or the fraction of housekeeping proteins (like f_Q), whose allocation does not depend on growth rate. The allocation constraint equation (3.18) must of course be satisfied by the optimisation routine.

3.5.2 Numerical simulation and optimisation of allocation models

The first requirement for finding optimal allocation is to have an efficient way to numerically determine the steady-state associated with a given set of parameters. For this, I numerically evolved the set of ODEs dynamically until a steady state was found. The initial condition for the evolution could partially be guessed from the allocation fractions—the protein content in steady state is proportional to the proteome allocation—but I had to establish the initial metabolite concentrations manually. Unless otherwise noted, I used an initial guess of 0.05 for the relative concentration of each metabolite.

Software-wise, I implemented all the numerical analysis in the Julia programming language and in particular I heavily relied on the DifferentialEquations.jl ecosystem (Rackauckas and Nie 2017). Within this ecosystem, many algorithms for solving differential equations are made available with a unified interface. I used trial-and-error to find one that worked well. For solving the concentration ODEs towards steady-state, I settled on the Rodas5 algorithm provided by the ecosystem.

The optimisation problem can be defined as finding the allocation fraction $\mathbf{f}$ that maximises the growth rate μ. What complicates this dramatically is that the function $\mu(\mathbf{f})$ is only defined implicitly. Efficient numerical optimisation routines would require explicit forms of $\mu(\mathbf{f})$ and, importantly, its gradient functions $\nabla_{\mathbf{f}}\mu$. These gradients are not available, because the nonlinearity of the Michaelis–Menten kinetics prohibits an explicit solution for μ in terms of $\mathbf{f}$ in equations such as equations (3.32) to (3.38). This restricted me to using a gradient-free optimisation routine, and I settled on using the Nelder–Mead algorithm as defined in the Optim.jl package (Nelder and Mead 1965; Gao and Han 2012; Mogensen and Riseth 2018).

The allocation constraint (3.18) further complicated the optimisation procedure. While I could eliminate one element of the allocation vector by imposing the allocation constraint, the sampling strategy still allowed for situations where this constrained element became negative. As a practical solution, I evaluated the doubling time for allocation fractions whose sum exceeded the constraint as a predefined very large value.

The Nelder–Mead algorithm proved to be fast, but sensitive to some numerical inaccuracies when applied to the allocation optimisation problem. The algorithm particularly struggled when the optimal allocation fraction had elements set to zero, i.e. the optimal cell entirely lacked expression of some enzymes. In these occasions, I ran the optimisation several times, each with different allocation fractions set to zero and excluded from the optimisation. Then, each iteration resulted in one set of optimal parameter vector and associated steady state and growth rate. Out of these, the set with the largest growth rate (smallest doubling time) was selected as the optimum.

4 C-GRAMs describing steady-state growth under nitrogen and carbon catabolism

4.1 Introduction

In this chapter I describe a model constructed to describe the resource allocation of ribosomes. It includes only proteins, since these are the main constituent of the cell. In this introduction I will describe the main experimental observations that the model reflects either as an assumption or as an output. I will then describe the model in increasing levels of detail; the intermediate models were analysed themselves too. Next I will show results generated from particular instances of the model, with parameters modified using the example model by Bertaux et al. (2020) introduced in the previous chapter as a starting point. These parameter choices reflect bacterial physiology, although the basic modelling backbone was general enough to be applied to yeast as well—this will be the topic of the following chapter.

4.1.1 The abundance of most proteins is correlated with growth rate

A prominent feature of gene expression is the pervasiveness of correlations with the growth rate when the latter is perturbed by nutrient modulations. The work described in chapter 2 is the latest addition to a large body of work showing correlations between broad classes of proteins and the cellular growth rate in multiple model organisms (Hui et al. 2015; A. Schmidt et al. 2016; Metzl-Raz et al. 2017; Zavřel et al. 2019). The model described in the current chapter reflects this basic premise of unicellular physiology.

Recall how the proteome can be divided into three major sectors:

1. the R-sector of ribosomes and other proteins required for self-replication and the production of proteins. The abundance of these proteins is positively correlated with growth rate under nutrient modulations.

2. the Q-sector of “housekeeping” proteins whose expression is not correlated with growth rate.

3. the P-sector of "stress" proteins that, in *E. coli*, also contains metabolic enzymes, whose abundance is negatively correlated with growth rate. Note that, even though in our *S. pombe* data analysis (chapter 2) medium-specific expression was present to a large degree for metabolic enzymes, the total expression of all metabolic enzymes taken together was negatively correlated with the growth rate. Therefore at the coarse-grained level, where only the total abundance of proteome classes is being considered, the two observations agree.

4.1.2 Cells are capable of both respiratory and respirofermentative growth

As introduced in section 1.1.5, cells express both fermentation and respiratory pathways in conditions supporting fast growth. Then, energy (in the form of ATP) is generated primarily through fermentation. Additionally, the biosynthesis of amino acids and lipids still requires the intermediates generated by the respiratory enzymes and the TCA cycle. During these conditions, glycolytic flux is large and the excess carbon is excreted, a phenomenon called overflow metabolism. In more restrictive conditions, the fermentation pathway is not expressed and energy is solely generated by respiration. This distinction between respirofermentative and purely respiratory growth is one of the most important characteristics of microbial growth.

From the point of view of coarse-grained modelling, there are two primary differences between respirofermentative and purely respiratory growth. On the one hand, the fermentation pathway consists of few different enzymes. Although the individual enzymes are highly abundant, there are so few that the total expression burden is markedly smaller than that of the respiratory pathway that would generate an equal energy flux. On the other hand, fermentation requires more nutrients from the environment: carbon is consumed rapidly and converted into ethanol or acetate. These two differences point towards the existence of a tradeoff between the efficient use of *external* consumables and that of *internal* macromolecular resources. In abundant conditions, the balance tips towards internal efficiency, whereas the cell uses nutrients more parsimoniously in more restrictive conditions.

4.1.3 Biomass comprises nitrogen as well as carbon

Earlier coarse-grained modelling efforts chiefly considered carbon modulations as representing the effect of the nutrient quality in general, because carbon modulation was more commonly used in *E. coli*. In such models, metabolism was typically considered as a linear pathway from nutrient to protein production. This included the models proposed by Weiße et al. (2015) and by Bertaux et al. (2020). Molenaar et al. (2009) did incorporate the uptake of free nitrogen, which was directly consumed by ribosomes rather than first combined into amino acid precursors. Contrasting with this earlier work, the growth rate modulation strategy explored in chapter 2 of this thesis made use of different nitrogen sources, rather than carbon sources. We aimed to better understand the effect of nitrogen modulation on resource allocation in a coarse-grained modelling context as well. Chiefly, we saw that many enzymes involved in carbon metabolism were expressed differently across the nitrogen sources, although none of the nitrogen sources appeared to repress the fermentation pathway. We therefore wondered if we could reproduce these findings in a coarse-grained model, or exclude cross-talk between the two pathways in certain parametrisations.

As a first approximation, the presence of nitrogen in biomass can simply be modelled by a nitrogen assimilation pathway that runs in parallel to the carbon uptake pathway and culminates in an amino acid biosynthesis that incorporates the two elements. Indeed, I believe this is a good representation of the metabolism that occurs when free nitrogen is provided to the cells in the form of ammonium ions. However, the amino acids used as nitrogen sources also contain carbon atoms. In order to account for this, a general understanding is required of the general mechanism of amino acid uptake. The exact pathways involved are surprisingly poorly determined in *S. pombe*, but there is no reason to believe that the mechanisms from *S. cerevisiae* and other organisms are not preserved. The key step in nitrogen assimilation from most amino acids consists of a de- or transamination reaction. In deamination reactions, an amino group is removed from the amino acid, in the process leaving free ammonium and the amino acid backbone. For all proteinogenic amino acids bar proline, this remaining backbone is a ketoacid. Transamination reactions can occur with either alpha-ketoglutarate or glutamate, forming glutamate or glutamine, respectively. Glutamate, glutamine, and alpha-ketoglutarate are easily interconverted and particularly glutamate is very abundant. One might as well consider glutamate a storage of free nitrogen in addition to ammonium, which

is toxic in large cytosolic concentrations. De- or transamination may take the original amino acid as a substrate or they might occur after some modifications, but regardless, the remaining ketoacid needs to be recycled or excreted. In summary, there are two ways in which nitrogen sources differ in their nutrient quality: in their uptake, and in their recycling and/or excretion.

4.2 Model formulation

In this section, I describe my construction of a base metabolic model, that includes pathways for both nitrogen and carbon uptake, and the synthesis of proteins from amino acids, themselves constructed from both nitrogen and carbon. I will explain how parametrisations of this model can represent different pathways such as fermentative and respiratory energy generation. To represent amino acids as nitrogen sources, I introduced recycling and carbon export pathways. I will describe the full model first, before giving more details about smaller, simplified, submodels used to generate part of the results. For most of this chapter, the allocation was assumed to be optimal as in section §3.5, although the simplest submodels permitted the choice of explicitly regulated ribosomal allocation, which I have indicated in the text.

4.2.1 Full metabolic model

The model from Bertaux et al. (2020), explained in detail in section §3.4, served as my starting point. I aimed to describe the steady-state behaviour of unperturbed wild-type cultures. Therefore, I disregarded the cell-cycle protein X and the useless protein U, as well as the inactivation of ribosomes by the translational inhibitor. Furthermore, I introduced additional metabolic pathways representing (i) the uptake of carbon and nitrogen, (ii) fermentative and respiratory energy generation, represented by different parametrisations of a similar enzyme, and (iii) the recycling and excretion of carbon from complex nitrogen sources containing both nitrogen and carbon.

In this subsection I will describe the formulation of the full model without much comment on the interpretation of variables and parameters included. More extensive explanations are provided in the following subsections, alongside the definitions of simpler subsystems of this full model.

The reactions included in the full model are pictured in Figure 4.1, and the model follows the formalism from equation (3.20), here repeated:

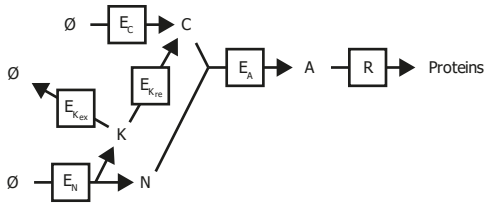


Figure 4.1: Illustration of the full metabolic model. It incorporates the uptake of two types of nutrient from the environment, the recycling of ketoacid backbones, as well as the production of proteins from amino acid precursors. The transporters E_C import carbon precursors C, and the transporters E_N represent the uptake of amino acids as nitrogen sources, which nets free nitrogen N as well as ketoacid backbones K. The latter are recycled into carbon precursors by enzymes E_{Kre} or excreted from the cell by enzymes E_{Kex}. Both carbon precursors and free nitrogen are required for the synthesis of amino acid precursors A by the enzyme E_A. Two parametrisations exist for this enzyme, representing fermentative and respiratory growth. The ribosomes R produce all the proteins in the system, which includes housekeeping proteins Q.

$$\dot{x} = \begin{pmatrix} \dot{a} \\ \dot{m} \\ \dot{p} \end{pmatrix} = \begin{pmatrix} S^A \cdot \hat{j}(x) - j_R(x) - \mu a \\ \hat{S}\hat{j}(x) - \mu m \\ j_R(x) f - \mu p \end{pmatrix}. \tag{4.1}$$

The quantities described by this equation are defined as the amino acids a, metabolite vector $m = (k, c, n)$, protein vector $p = (e_{K_{re}}, e_{K_{ex}}, e_C, e_{A_F}, e_{A_R}, e_N, r, q)$, flux vector

$$j = \begin{pmatrix} j_{K_{re}} \\ j_{K_{ex}} \\ j_C \\ j_{A_F} \\ j_{A_R} \\ j_N \\ j_R \end{pmatrix} = \begin{pmatrix} k_{K_{re}} e_{K_{re}} \frac{k}{k+k_{sat}} \\ k_{K_{ex}} e_{K_{ex}} \frac{k}{k+k_{sat}} \\ k_C e_C \\ k_{E_F} e_{A_F} \frac{cn}{(c+c_{sat})(n+n_{sat})+\Delta_{E_F}} \\ k_{E_R} e_{A_R} \frac{cn}{(c+c_{sat})(n+n_{sat})+\Delta_{E_R}} \\ k_N e_N \\ k_R r \frac{a}{a+a_{sat}} \end{pmatrix}, \tag{4.2}$$

allocation vector $\mathbf{f} = (f_{K_{re}}, f_{K_{ex}}, f_C, f_{E_F}, f_{E_R}, f_N, f_R, f_Q)$, and stoichiometry

$$\begin{pmatrix} (\mathbf{S}^A)^T \\ \hat{S} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & +1 & +1 & 0 \\ -1 & -1 & 0 & 0 & 0 & \gamma_K \\ 1 & 0 & 1 & -\alpha_{C_F} & -\alpha_{C_R} & 0 \\ 0 & 0 & 0 & -\alpha_{N_F} & -\alpha_{N_R} & \gamma_N \end{pmatrix}. \tag{4.3}$$

To avoid the need to introduce regulatory feedback for the expression of each protein class in the model, I chose allocations $\mathbf{f}$ that maximised the growth rate (section §3.5). Exceptions to this were the less complex models described in 4.2.2 and 4.2.3, for which regulation of the ribosomal sector was implemented as before (section §3.4, Bertaux et al. 2020).

Mass balance is maintained in all internal reactions, but the carbon and nitrogen transporters take up nutrients from the environment, and the ketoacid excretion enzyme can move ketoacid back into the environment. The mass balance is reflected by the following constraints:

$$\alpha_{C_F} + \alpha_{N_F} = 1, \tag{4.4}$$

$$\alpha_{C_R} + \alpha_{N_R} = 1, \tag{4.5}$$

$$\gamma_K + \gamma_N = 1. \tag{4.6}$$

In amino acid nitrogen sources, the ketoacid stoichiometry parameter γ_K can be calculated from n_C and n_N, the numbers of carbon and nitrogen atoms in a nutrient molecule, as

$$\gamma_K = \frac{n_C \times 12}{n_C \times 12 + n_N \times 14} = \frac{12\frac{n_C}{n_N}}{12\frac{n_C}{n_N} + 14}. \tag{4.7}$$

In some plots, the carbon-to-nitrogen ratio $\frac{n_C}{n_N}$ is used instead of γ_K to visualise the model behaviour.

When entered into equation (4.1), these definitions yield the full ODE system:

$$\dot{a} = j_{A_F} + j_{A_R} - j_R - \mu a, \tag{4.8}$$

$$\dot{k} = \gamma_K j_N - j_{K_{re}} - j_{K_{ex}} - \mu k, \tag{4.9}$$

$$\dot{c} = j_C + j_K - \alpha_{C_F} j_{A_F} - \alpha_{C_R} j_{A_R} - \mu c, \tag{4.10}$$

$$\dot{n} = \gamma_N j_N - \alpha_{N_F} j_{A_F} - \alpha_{N_R} j_{A_R} - \mu n, \tag{4.11}$$

$$\dot{e}_K = f_K j_R - \mu e_K, \tag{4.12}$$

$$\dot{e}_C = f_C j_R - \mu e_C, \tag{4.13}$$

$$\dot{e}_{A_F} = f_{E_F} j_R - \mu e_{A_F}, \tag{4.14}$$

$$\dot{e}_{A_R} = f_{E_R} j_R - \mu e_{A_R}, \tag{4.15}$$

$$\dot{e}_N = f_N j_R - \mu e_N, \tag{4.16}$$

$$\dot{r} = f_R j_R - \mu r, \tag{4.17}$$

$$\dot{q} = f_Q j_R - \mu q. \tag{4.18}$$

The growth rate μ in these equations is found by taking the sum of these equations. Using the allocation constraint $\sum_i f_i = 1$ and the concentration constraint $\sum_i x_i = 1$, such that $\sum_i \dot{x}_i = 0$, gives

$$\mu = j_C + j_N - j_{K_{ex}} = k_C e_C + k_N e_N - k_{K_{ex}} e_{K_{ex}} \frac{k}{k + k_{sat}}. \tag{4.19}$$

This is equal to the net uptake of nutrient from the environment.

4.2.2 Submodel I: one-enzyme metabolism

The simplest model considered in this thesis is a simplification from the model described by Bertaux et al. (2020), where the aforementioned ribosome inactivation reactions, cell cycle proteins, and unused proteins were discarded, but no additional metabolism was yet introduced. A cartoon representation of the system, referred to as submodel I, is given in Figure 4.2. The only protein species comprising the model are therefore an enzyme E representing nutrient uptake and metabolism, a housekeeping fraction Q, and a Michaelis–Menten ribosome produ-

cing these two proteins and itself. There is a single internal metabolite A. The ODEs that describe the system are:

$$\dot{a} = ke - \sigma\frac{a}{a + a_{sat}}r - \mu a, \tag{4.20}$$

$$\dot{e} = f_E\sigma\frac{a}{a + a_{sat}}r - \mu e, \tag{4.21}$$

$$\dot{r} = f_R\sigma\frac{a}{a + a_{sat}}r - \mu r, \tag{4.22}$$

$$\dot{q} = f_Q\sigma\frac{a}{a + a_{sat}}r - \mu q. \tag{4.23}$$

In this model, the allocation vector f was chosen as follows. First, f_Q was held fixed,

$$f_R \quad = \quad \delta a \tag{4.24}$$

was dynamically regulated, and

$$f_E = \max(0, 1 - f_Q - f_R) \tag{4.25}$$

followed the ribosomal allocation constraint.

The growth rate μ is calculated from the sum of the four equations using the allocation constraint $f_E+f_R+f_Q = 1$ and the concentration constraint $a+e+r+q = 1$ (so $\dot{a} + \dot{e} + \dot{r} + \dot{q} = 0$), resulting in

$$\mu = ke. \tag{4.26}$$

Note that this relation is true dynamically, because the concentration constraint is universally valid. We can therefore substitute this latter relation for μ in the set of dynamic equations posed above. It follows that the parameter k can be used to modulate the growth rate.

4.2.3 Submodel IIa: Two-enzyme linear metabolism

The first modification we made to the minimal model was to separate the enzyme into two parts, similar to the model presented by Weiße et al. (2015). This is illustrated in Figure 4.3 and referred to as submodel II. We termed the two enzymes the transporter E_T and the metabolic enzyme E_M for ease of reference. An additional metabolite S represented internal sugar molecules. As before, all were part of the

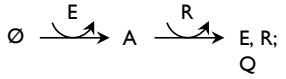


Figure 4.2: Illustration of submodel I, the most simplified one. It is based on the model by Bertaux et al. (2020) but includes only the metabolic enzymes E, active ribosomes R, and housekeeping proteins Q, with one metabolic intermediate A that cells import directly from the environment.

total volume, i.e. $1 = s + a + e_T + e_M + r + q$. We further assumed Michaelis–Menten kinetics for the metabolic enzyme, which converted the internal sugar into the substrate for the ribosomes. The dynamic equations from above were thus modified to:

$$\dot{s} = k_T e_T - k_M e_M \frac{s}{s + s_{sat}} - \mu s, \tag{4.27}$$

$$\dot{a} = k_M e_M \frac{s}{s + s_{sat}} - \sigma r \frac{a}{a + a_{sat}} - \mu a, \tag{4.28}$$

$$\dot{e}_T = f_T \sigma r \frac{a}{a + a_{sat}} - \mu e_T, \tag{4.29}$$

$$\dot{e}_M = f_M \sigma r \frac{a}{a + a_{sat}} - \mu e_M, \tag{4.30}$$

$$\dot{r} = f_R \sigma r \frac{a}{a + a_{sat}} - \mu r, \tag{4.31}$$

$$\dot{q} = f_Q \sigma r \frac{a}{a + a_{sat}} - \mu q, \tag{4.32}$$

with $f_T + f_M + f_R + f_Q = 1$. The housekeeping fraction was still represented by f_Q, fixed across all model instances, and

$$f_R = \delta a \tag{4.33}$$

was dynamically regulated as before. However, the extra enzyme introduced freedom in the choice of allocation vector $\mathbf{f}$ such that a trade-off existed between f_T and f_M while satisfying

$$f_T + f_M = 1 - f_Q - f_R. \tag{4.34}$$

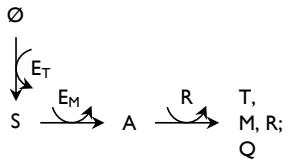


Figure 4.3: Illustration of submodel II, with two-step linear metabolism. Besides the protein production module illustrated in Figure 4.2, it includes two enzymes E_T and E_M as well as two metabolic intermediates S and A.

In practice, we explored the behaviour of the model under perturbations of f_M.

As before, the growth rate μ in the above equations was determined dynamically as

$$\mu = k_T e_T. \tag{4.35}$$

4.2.4 Submodel IIb: Assuming optimal allocation

As described in more detail in section §3.5, I simplified the analysis of the two-step linear pathway and its further extensions by choosing the allocation vector in such a way that the growth rate was maximised. That is, I removed the requirement that $f_R = \delta a$ and instead calculated the allocation (f_T, f_M, f_R) that optimised the growth rate μ for a given set of kinetic parameters.

4.2.5 Submodel III: separate nitrogen and carbon pathways feeding amino acid synthesis

Next, I modelled a system that contained both carbon and nitrogen metabolism, illustrated in Figure 4.4. For this model, I considered only the steady-state allocation that maximised the growth rate. Compared to the previous model, the transporter E_T is replaced by two parallel ones, named T_C and T_N, with the former importing a carbon precursor with relative abundance c and the latter producing a nitrogen precursor metabolite with relative abundance n.

The metabolic enzyme (now termed simply E) functioned anabolically in combining the two precursors into the amino acid. Mass balance was observed by this reaction: as much mass in the form of carbon c and nitrogen n was consumed as

was produced in the form of amino acid a. The relative requirements of the two substrates was modelled by the parameters α_C and α_N, with

$$\alpha_C + \alpha_N = 1. \tag{4.36}$$

The enzyme kinetics were modelled as the two-step enzyme described in section 3.3.2. As in section 3.3.2, the parameter Δ_E was set to zero.

The system of ODEs for the concentrations in this model was given by:

$$\dot{n} = k_N t_N - \alpha_N k_E e \frac{cn}{(c + c_{sat})(n + n_{sat}) + \Delta_E} - \mu n, \tag{4.37}$$

$$\dot{c} = k_C t_C - \alpha_C k_E e \frac{cn}{(c + c_{sat})(n + n_{sat}) + \Delta_E} - \mu c, \tag{4.38}$$

$$\dot{a} = k_E e \frac{cn}{(c + c_{sat})(n + n_{sat}) + \Delta_E} - \sigma r \frac{a}{a + a_{sat}} - \mu a, \tag{4.39}$$

$$\dot{t}_C = f_C \sigma r \frac{a}{a + a_{sat}} - \mu t_C, \tag{4.40}$$

$$\dot{t}_N = f_N \sigma r \frac{a}{a + a_{sat}} - \mu t_N, \tag{4.41}$$

$$\dot{e} = f_E \sigma r \frac{a}{a + a_{sat}} - \mu e, \tag{4.42}$$

$$\dot{r} = f_R \sigma r \frac{a}{a + a_{sat}} - \mu r, \tag{4.43}$$

$$\dot{q} = f_Q \sigma r \frac{a}{a + a_{sat}} - \mu q. \tag{4.44}$$

Using the concentration constraint to compute the dynamical growth rate μ works as before, with the growth rate still determined by the total uptake of mass from the system:

$$\mu = k_C t_C + k_N t_N. \tag{4.45}$$

4.2.6 Submodel IV: Parametrisations representing respirofermentative and respiratory growth

As mentioned before, there are two principal growth states of the cell: respirofermentative growth, and entirely respiratory growth. The utilisation of the fermentative pathway amounts to a higher carbon usage, because overflow metabolism is utilised to maintain energy levels in the form of ATP, and redox (NAD+/NADH) balance. The overflow metabolites are excreted—in yeast, typically in the form of ethanol, but in other organisms acetate or lactate fermentation are also present. I

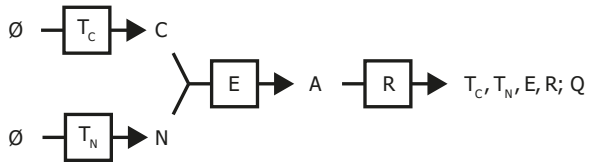


Figure 4.4: Illustration of submodel III, with simple branched metabolism. The model contains nitrogen and carbon precursors N and C, with associated transporters T_N and T_C, in addition to the base model illustrated in Figure 4.2.

did not model this pathway explicitly. Rather, I encoded the higher carbon usage in the parameter α_C. In other words, with respirofermentative growth represented by the subscript F and purely respiratory growth by the subscript R, parameter choices for these two iterations of the metabolic pathway were chosen that satisfied both $k_{E_F} > k_{E_R}$ and $\alpha_{C_F} > \alpha_{C_R}$.

For a rough estimate of the stoichiometry parameters α_C, I used that proteinogenic amino acids contain approximately 4 carbon atoms per nitrogen atom. This converted to the mass-action parameters gave $\alpha_{C_R} = \frac{4\times12}{4\times12+14} = {}^{24}/_{31}$ using molar masses of 12 g/mol and 14 g/mol for carbon and nitrogen, respectively. Another rough estimate that the excess carbon excreted in fermentation is approximately equal in amount to the carbon converted to biomass gave $\alpha_{C_F} = \frac{8\times12}{8\times12+14} = {}^{48}/_{55}$.

The two parametrisations were optimised simultaneously, by replacing the single enzyme E with two enzymes E_F and $E_R$that represented the respirofermentative

and respiratory pathways, respectively. The ODE equations for this model already closely resembled those of the full model:

$$\dot{c} = j_C - \alpha_{C_F} j_{A_F} - \alpha_{C_R} j_{A_R} - \mu c, \tag{4.46}$$

$$\dot{a} = j_{A_F} + j_{A_R} - j_R - \mu a, \tag{4.47}$$

$$\dot{n} = \gamma_N j_N - \alpha_{N_F} j_{A_F} - \alpha_{N_R} j_{A_R} - \mu n, \tag{4.48}$$

$$\dot{t}_C = f_C j_R - \mu t_C, \tag{4.49}$$

$$\dot{e}_{A_F} = f_{E_F} j_R - \mu e_{A_F}, \tag{4.50}$$

$$\dot{e}_{A_R} = f_{E_R} j_R - \mu e_{A_R}, \tag{4.51}$$

$$\dot{t}_N = f_N j_R - \mu t_N, \tag{4.52}$$

$$\dot{r} = f_R j_R - \mu r, \tag{4.53}$$

$$\dot{q} = f_Q j_R - \mu q. \tag{4.54}$$

The fluxes followed the functional forms described in equation (4.2), and the α-parameters the stoichiometric constraints from equations (4.4) and (4.5).

4.2.7 Amino acids as sources of nitrogen

The final component of the full model was the uptake of more complex nitrogen sources. Freely usable nitrogen comes in the form of ammonium ions (NH_4^+), which are used in the biosynthesis of nitrogen-containing compounds including amino acids. However, cells can assimilate nitrogen from many other sources. In the experiments described in chapter 2, the cells were presented with various amino acids as their sole nitrogen source. In budding yeast, these pathways typically involve deamination of the amino acid into free ammonium ions and the leftover ketoacid, either by a single enzyme or as the net result of a longer pathway (Godard et al. 2007). The ketoacid, a carbohydrate, is typically recycled into biomass or excreted.

To account for this process in the model, I modified the nitrogen transporter to produce both free nitrogen and a carbon-containing ketoacid metabolite with their relative proportions determined by parameters γ_N and γ_K. Additionally, I introduced two Michaelis–Menten enzymes, E_{Kre} and E_{Kex} that, respectively, recycled the ketoacid into usable carbon precursor or excreted it from the cell. This completed the model described in 4.2.1.

4.3 Results

I will now describe some key behaviours of the model constructed in the previous section. The results are presented in increasing order of the complexity of the model used to generate them. The submodels are referred to by the numerals used in the above. For submodels I and IIa, explicit ribosomal allocation regulation was included, whereas for the more involved submodels (IIb and following) all allocation fractions were assumed to maximise growth rate.

4.3.1 Submodel I: The ribosomal growth law followed from model assumptions

A central point in the definition of the regulated models I and IIa is the linear dependence of the ribosome allocation on the metabolite concentration:

$$f_R = \delta a, \tag{4.55}$$

with δ a fixed parameter. Furthermore, the requirement that all biomass is produced by stable ribosomes gives the ribosome ODE

$$\dot{r} = f_R \sigma r \frac{a}{a + a_{sat}} - \mu r, \tag{4.56}$$

which means that

$$\mu = f_R \sigma \frac{a}{a + a_{sat}} \tag{4.57}$$

in steady-state. Note that these assumptions are also met in the original model.

For $a \gg a_{sat}$ (so $f_R \gg \delta a_{sat}$) it now follows that

$$\mu = \sigma \left(\frac{f_R^2}{\delta a_{sat} + f_R} \right) \tag{4.58}$$

$$= \sigma \left(f_R - \delta a_{sat} + \frac{\delta^2 a_{sat}^2}{f_R + \delta a_{sat}} \right) \tag{4.59}$$

$$\approx \sigma \left(f_R - \delta a_{sat} \right). \tag{4.60}$$

In balanced growth, the allocation fraction f_R is equal to the more familiar proteome mass fraction ϕ_R, so

$$\phi_R \approx \phi_{R0} + \sigma^{-1} \mu \tag{4.61}$$

with

$$\phi_{R0} = \delta a_{sat}. \tag{4.62}$$

This is the familiar ribosomal growth law.

4.3.2 Submodel IIa: Enzymatic cost parameters modulated growth rate

In the two-step model, the regulation of f_R by a, the imposed housekeeping fraction f_Q, and the allocation constraint $f_T + f_M + f_R + f_Q = 1$ restricted only three out of four allocation fractions. Therefore, we were able to modulate the growth rate μ by changing the allocation f_M towards E_M. A curve of the resulting growth rates is plotted in Figure 4.5A. The growth rate was maximised at some $f_M = f_M^*$, separating two regimes. For $f_M < f_M^*$, the metabolic enzyme was limiting and increasing its allocation lead to an increase in growth. Here, the sugar s accumulated as the metabolic enzyme could not cope with the load, as shown in Figure 4.5B. For $f_M > f_M^*$, increasing allocation towards the metabolic enzyme constituted an extra burden, and thereby decreased the growth rate. In this second regime, both metabolites were low in abundance.

The first regime is unlikely to represent biological reality, because metabolites are known to constitute less biomass than macromolecules (Milo et al. 2016, pp. 128–132). However, the second regime can be used to represent the situation where the burden of metabolism depends on the composition of the growth environment. A higher burden results in slower growth, as resources are reallocated from the ribosome and transporter towards the metabolic enzyme. This is illustrated in Figure 4.5C, where we show the four proteome fractions as a function of the growth rate. Note how the ribosome fraction ϕ_R followed the growth law for medium to fast growth, as predicted by equation (4.61).

4.3.3 Submodel IIa: Transporter and metabolic enzyme expression were anticorrelated

In this model with this parametrisation, the transporter T was part of the "R"-fraction: f_T correlates positively with growth rate (see Figure 4.5C). In fact, this fact can be seen directly from equation (4.35), since k was kept constant. Some evidence for this was found in our proteomics data as well: for example, the enzyme catalysing the first step of glycolysis, Hxk2, was found to be strongly positively correlated with the growth rate as modulated by the nitrogen source, whereas

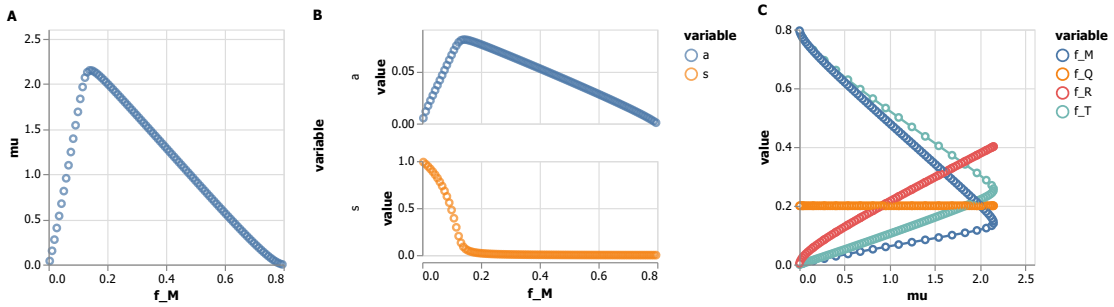


Figure 4.5: Analysis of the two-step model upon variation of the parameter f_M. **A.** Growth rate μ, **B.** concentration of internal metabolites s and a as a function of f_M. **C.** Mass fractions of each of the four proteins, parametrised by growth rate. Note the ordering according to the choice of f_M, so f_T is large when f_M is small and *vice versa*. Parametrization: $k_T = 10.0\ h^{-1}$, $k_M = 20.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $s_{sat} = 0.0167$, $a_{sat} = 0.0167$, $f_Q = 0.2$, $\delta = 5.0$.

the total burden of metabolism was found to be negatively correlated with the growth rate. This behaviour was also noted by Molenaar et al. (2009), but it is notably different from the two-step model described by Weiße et al. (2015), who assumed that the transporter and metabolic enzyme were both regulated identically. They imposed that both ϕ_T and ϕ_M correlated negatively with growth rate—for us, while ϕ_M and $\phi_M + \phi_T$ do,

$$\phi_T = \frac{e_T}{1 - a - s} = \frac{\mu}{k_T (1 - a - s)} \tag{4.63}$$

does not.

4.3.4 Submodel IIa: Housekeeping proteins were less abundant than widely assumed

Proteins associated with translation are estimated to constitute approximately 45% of the total protein mass in the fastest-growing *E. coli* cultures (Scott et al. 2010; Hui et al. 2015; A. Schmidt et al. 2016). A minimal partitioning of the proteome based on growth rate correlations suggested that around half of the proteome mass does not change with the growth rate (Scott et al. 2010, 2014). The model by Weiße et al. (2015) reflected this, partitioning a two-step model similar to ours into a metabolic P-sector (corresponding to our T and M), a translational

R-sector, and a constitutive Q-sector. In the parametrisation that best represented the ribosomal growth laws, the Q-proteins took up around 70% of the total protein mass, and were negatively correlated with the growth rate.

For the parametrisation underlying Figure 4.5, the allocation towards housekeeping proteins was small: $f_Q = 0.2$. However, this parametrisation was chosen such that the ribosomal growth law agreed reasonably well with the experimental data. The discrepancy in the estimated housekeeping allocation is explained by the considerable cost of the two enzymes even at maximal growth, at around 25% for T and 15% for M in the current parametrisation. Because proteins whose expressions is constant across all experimental conditions are indeed quite rare and most proteins are either correlated with growth rate or induced in a medium-specific manner (chapter 2, as well as Hui et al. 2015; A. Schmidt et al. 2016), I think this smaller allocation towards Q-proteins is an important feature of the f_M-modulated model.

4.3.5 Submodel IIb: Growth-optimised allocation was a good approximation for the regulated model

The results in the previous sections were obtained using a regulated model, where $f_R = \delta a$ with fixed $\delta = 5.0$ as taken from Bertaux et al. (2020). However, for the complete model I considered only model parametrisations where the allocation vector maximised growth rate. In the simpler Model II with linear metabolism, I investigated the similarities and discrepancies between the regulated model (IIa) and the optimised model (IIb) under modulations of k_C (Figure 4.6) and k_E (figure 4.7).

As expected, for both parametrisations, the dependency of the growth rate on the modulation parameter approximated a Monod curve (Figures 4.6A and 4.7A), and the components of the allocation vector f were approximately linearly correlated with growth rate (Figures 4.6B and 4.7B).

Next, we considered the relationship between f_R and a, as plotted in Figures 4.6C and 4.7C. In the optimal model IIb, the allocation to ribosomes was approximately proportional to the amino acid concentration, although a small nonlinearity was present. However, the best-fit proportionality factor δ was smaller than the value of 5.0 that was derived from the relation between elongation rate and ribosome abundance (Dai et al. 2016). As derived in the previous section, such a decreased value of δ would decrease the offset in the ribosomal growth law in the regulated

model. Indeed, the deviation from proportionality in $f_R(\mu)$ was larger in the regulated model than in the optimised model (compare Figures 4.6B and 4.7B to Figure 4.5C).

Last, the mass fractions occupied by the metabolites, c and α, were appreciable (Figures 4.6D and 4.7D). Particularly for the k_C-modulated model, the relative abundance of the carbon metabolite increased to more than half of the total biomass. In experimental cultures, metabolites comprise only a small fraction of the biomass, so the growth-optimised model was a poor approximation for conditions where this occurred.

4.3.6 Submodel III: The two-substrate metabolism was similar to the linear pathway

From the above, I concluded that growth optimisation was a reasonable approximation to make for our coarse-grained models. In the remainder of this section, I discuss such optimised models where explicit nitrogen consumption was introduced. First, I discuss results for submodel III, where nitrogen uptake was introduced but the metabolic pathway was fixed. Then, I introduce the two parallel enzymatic pathways representing different levels of respiration (submodel IV). Last, I include the results on the recycling and excretion of the backbones resulting from the metabolism of complex nitrogen sources.

Modulated transporter rates k_N and k_C

Under perturbations of transporter uptake rates, the two-precursor model showed the same behaviours of the linear two-step model, as plotted in Figures 4.8 and 4.9 for modulation of k_N and k_C, respectively. The growth rate depended on both transporter rates according to Monod law, and the allocation fractions mostly followed the expected phenomenological trade-offs, where all freely optimised allocation fractions were approximately linearly correlated with the growth rate, with the allocation of the modulated transporter correlated negatively and all other proteins positively. One notable exception was the growth-rate dependence of f_E, the allocation towards the metabolic enzyme, which was clearly nonlinear under perturbations of k_N, the nitrogen transporter uptake rate (Figure 4.8B). For both modulations, the optimal f_R and α were approximately linearly correlated, further supporting that a simple regulation of the ribosomal allocation fraction can approach growth optimisation. The modulation of one transporter rate (e.g. k_N)

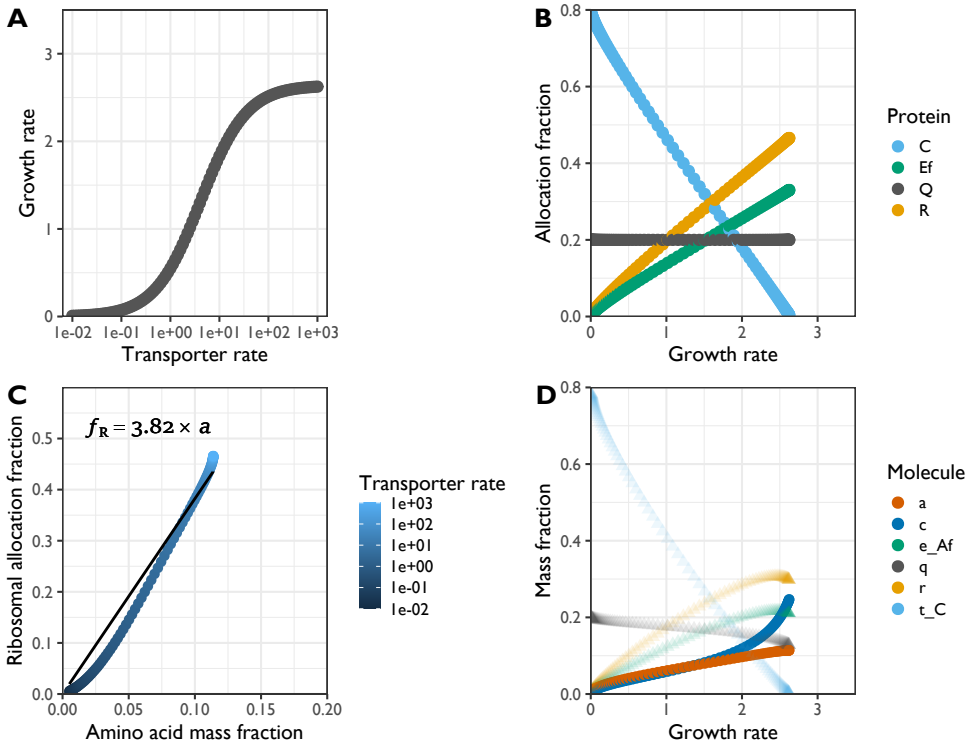


Figure 4.6: Analysis of two-step allocation-optimised model under modulation of the transporter rate. The uptake rate parameter k_C of the transporter was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen k_C. Other parameters were fixed to the same values as for Figure 4.5. **A.** Growth rate μ as a function of uptake rate k_C. **B.** Components of the protein allocation vector (f_T, f_E, f_R, f_Q) as a function of growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying k_C (blue points) as well as an ordinary least squares linear fit to the data, with forced zero intercept. **D.** Mass fractions of constituent molecules as a function of growth rate μ. Metabolites were plotted as solid circles and protein components as opaque triangles.

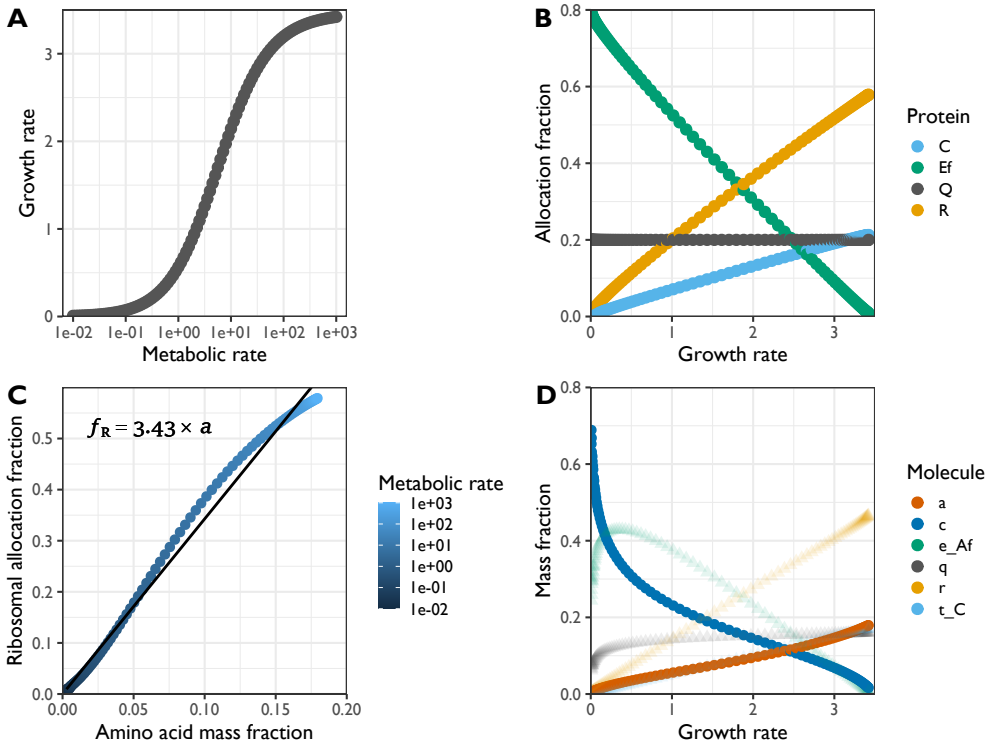


Figure 4.7: Analysis of two-step allocation-optimised model under modulation of the metabolic rate. The turnover rate parameter k_E of the metabolic enzyme was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen k_E. Other parameters were fixed to the same values as for Figure 4.5. **A.** Growth rate μ as a function of turnover rate k_E. **B.** Components of the protein allocation vector (f_T, f_E, f_R, f_Q) as a function of growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying k_E (blue points) as well as an ordinary least squares linear fit to the data, with forced zero intercept. **D.** Mass fractions of constituent molecules a function of growth rate μ. Metabolites were plotted as solid circles and protein components as opaque triangles.

barely affected homeostasis of the metabolite in the other pathway (e.g. c). As in the two-step model, expression of metabolites was considerable but only became greater than approximately 1/3 the biomass at large rates k_C and k_N, corresponding to small allocations f_C and f_N.

Modulated metabolic rate k_E

Furthermore, I modulated the growth rate through k_E, again in analogy to the situation in the simpler model, and again with very similar results. Like the two-step model, metabolites c and n were very highly abundant at slow growth (Figure 4.10D).

4.3.7 Submodel III: The ribosomal growth law was robust to random parameter sampling

Besides studying the model behaviour under modulations of single rate parameters, I explored the model under concurrent variations of the three rate parameters. I drew 100 triplets of (k_C, k_N, k_E) from independent uniform distributions with support $[0, 20]$. The ribosomal growth law was nearly exactly satisfied (Figure 4.11A), even though expression of the two transporters and the metabolic enzyme was highly variable between parameter choices. Furthermore, the concentration of the amino acid a was tightly related to growth rate, and it was almost always less abundant than c and n (Figure 4.11B). The tight growth rate correlations for both f_R and a again followed from tight covariation between the two (Figure 4.11C).

Together with the previous sections, these results suggest that a single regulatory f_R–a relationship is an appropriate simplification if a dynamic model of ribosome regulation is required. For the regime explored here, with small saturation constants, the simple proportionality proposed by Bertaux et al. (2020) works reasonably well, although $f_R \propto (a - a_0)$ could be an appropriate improvement.

4.3.8 Submodel IV: Respiration was induced by limiting the rate of carbon metabolism, but not of nitrogen

Next, I wondered under what circumstances respiration would be triggered in the model. I modulated the transporter rate parameters in model IV with two parallel metabolic pathways with different parametrisations ($k_{E_f} = 10.0\ h^{-1}$ and $\alpha_{C_f} = 48/55$ versus $k_{E_r} = 5.0\ h^{-1}$ and $\alpha_{C_r} = 24/31$, see 4.2.6). The allocation fractions

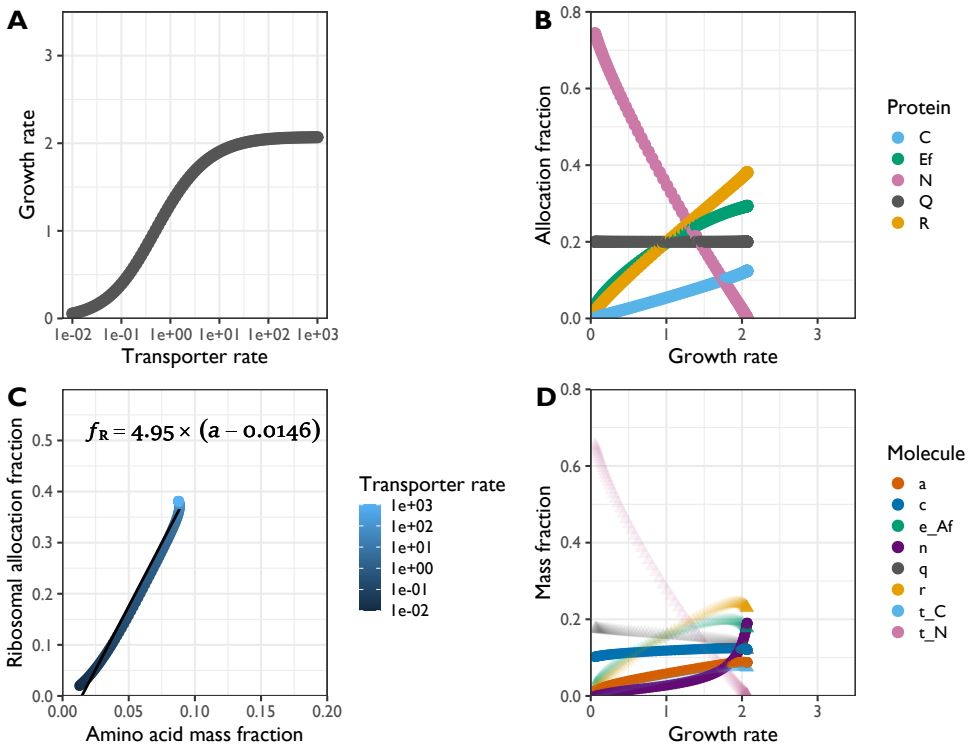


Figure 4.8: Analysis of allocation-optimised model with nitrogen and carbon uptake under modulation of the nitrogen transporter rate. The uptake rate parameter k_N of the nitrogen transporter was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen k_N. **A.** Growth rate μ as a function of uptake rate k_N. **B.** Components of the protein allocation vector $(f_C, f_E, f_N, f_R, f_Q)$ as a function of growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying k_N (blue points) as well as an ordinary least squares linear fit to the data. **D.** Mass fractions of constituent molecules a function of growth rate μ. Metabolites were plotted as solid circles and protein components as opaque triangles. Fixed parameters: $k_C = 20.0\ h^{-1}$, $k_E = 10.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_E = 0.0$, $\alpha_C = 48/55$, $f_Q = 0.2$.

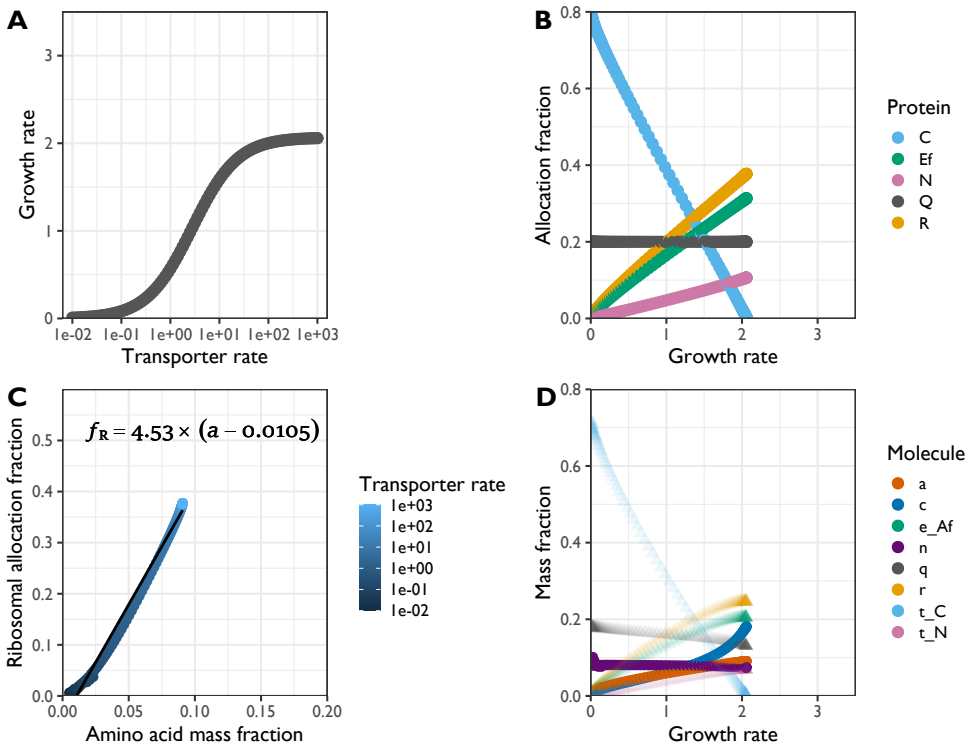


Figure 4.9: Analysis of allocation-optimised model with nitrogen and carbon uptake under modulation of the carbon transporter rate. The uptake rate parameter k_C of the nitrogen transporter was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen k_C. **A.** Growth rate μ as a function of uptake rate k_C. **B.** Components of the protein allocation vector $(f_C, f_E, f_N, f_R, f_Q)$ as a function of growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying k_C (blue points) as well as an ordinary least squares linear fit to the data. **D.** Mass fractions of constituent molecules a function of growth rate μ. Metabolites were plotted as solid circles and protein components as opaque triangles. Fixed parameters: $k_E = 10.0\ h^{-1}$, $k_N = 5.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_E = 0.0$, $\alpha_C = 48/55$, $f_Q = 0.2$.

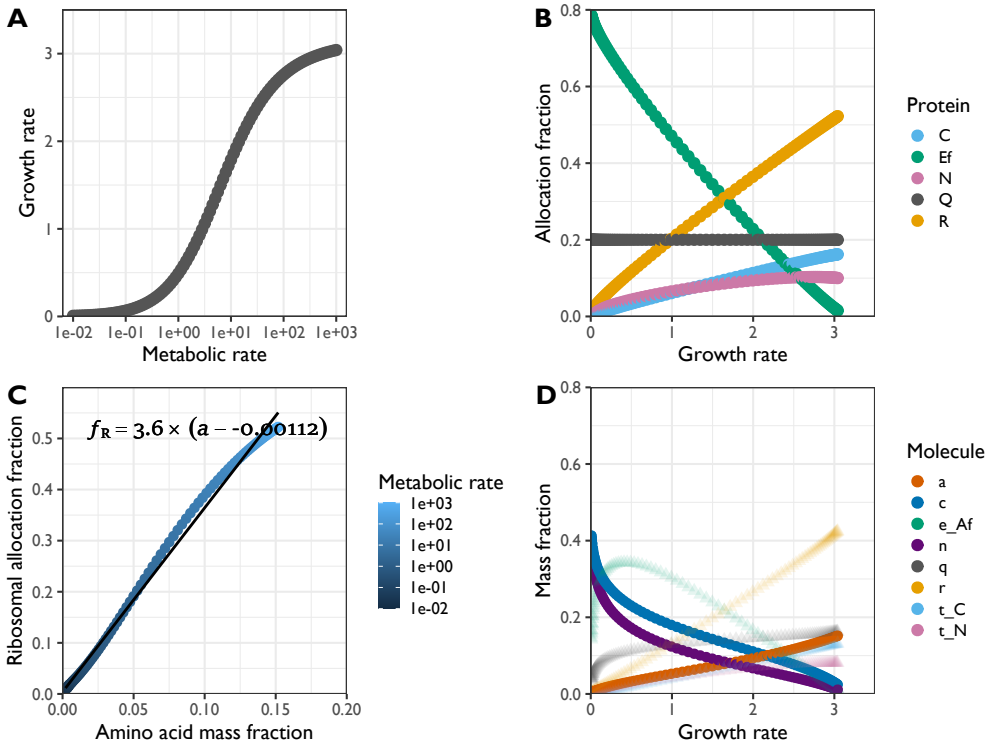


Figure 4.10: Analysis of allocation-optimised model with nitrogen and carbon uptake under modulation of the metabolic enzyme turnover rate. The turnover rate parameter k_E of the metabolic enzyme was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen k_E. **A.** Components of the protein allocation vector $(f_C, f_E, f_N, f_R, f_Q)$ as a function of growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying k_E (blue points) as well as an ordinary least squares linear fit to the data. **D.** Mass fractions of constituent molecules a function of growth rate μ. Metabolites were plotted as solid circles and protein components as opaque triangles. Fixed parameters: $k_C = 20.0\ h^{-1}$, $k_N = 5.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_E = 0.0$, $\alpha_C = 48/55$, $f_Q = 0.2$.

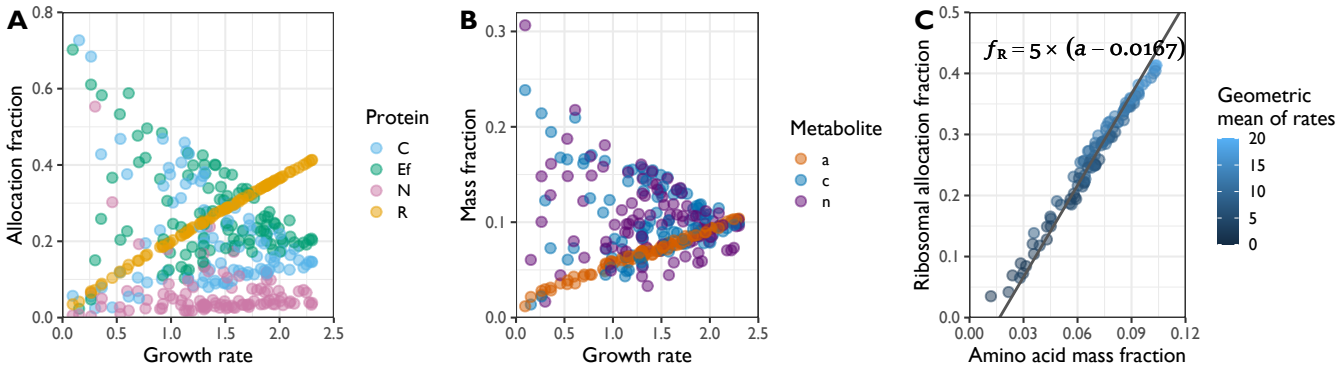


Figure 4.11: Analysis of allocation-optimised model with nitrogen and carbon uptake and randomly chosen rate parameters. The rate parameters k_{E_f}, k_C, and k_N were randomly drawn from a uniform distribution between 0 and 20.0 h^{-1}. The allocation parameters were chosen to maximise the growth rate for each set of (k_C, k_E, k_N). **A.** Components of the protein allocation vector (f_C, f_E, f_N, f_R) as a function of growth rate μ. f_Q was omitted to improve clarity. **B.** Steady-state biomass fractions of metabolites plotted against growth rate μ. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model (blue points, coloured according to $\sqrt[3]{k_C k_E k_N}$) as well as a line to guide comparison with previous figures. Fixed parameters: $\sigma = 6.46\ h^{-1}$, $c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_E = 0.0$, $\alpha_C = 48/55$, $f_Q = 0.2$.

to both enzymes were included in the growth rate optimisation. Because the two pathways are parallel, the usage of one of the two was preferred over the other, and in each parametrization with optimized allocation, exactly one of f_{E_f} and f_{E_r} was equal to zero. As shown in figure 4.12, the respiratory enzyme was only induced by low values of the carbon uptake rate k_C. The fermentative enzyme was preferred in all conditions where k_N was varied from its default.

A complication arose for slow growth rates induced by either transporter rates. With an extra allocation parameter to optimise, the optimisation problem was more difficult to solve and more prone to numerical error. The metabolite concentrations did not smoothly vary with the growth rate (Figure 4.12B), suggesting that the optimisation did not always reach the global minimum.

4.3.9 Full model: Recycling of ketoacids disturbed carbon metabolism

Now I discuss the full model including the introduction of ketoacid molecules as part of nitrogen uptake. At first, I explored the model behaviour when ketoacid was not being excreted, i.e. $f_{K_{ex}} = 0$; this restriction was lifted later. I perturbed the parameter γ_K, which represents the relative mass of recycled ketoacid with respect to the total mass taken up by the nitrogen (amino acid) transporter. This parametrisation was meant to represent different amino acids. For example, glycine molecules contain two carbon atoms for each nitrogen atom, which would be represented by $\gamma_K = \frac{2\times12}{2\times12+14} = 12/19 \approx 0.63$, whereas each molecule of isoleucine contains nine carbon atoms, for $\gamma_K = \frac{9\times12}{9\times12+14} = 54/61 \approx 0.89$. As seen in Figure 4.13A, the optimal growth rate decreased ever steeper with increasing γ_K, dropping to zero when $\gamma_K = 1$, which represents no nitrogen uptake at all.

The allocation fractions leading to the optimal growth rates are plotted in Figure 4.13B. Respiratory enzyme was not expressed, i.e. $f_{E_r} = 0.0$ was optimal across all simulations. Notably, the optimal cultures expressed considerable amounts of ribosomes even at vanishing growth rates.

There were three growth rate regimes, characterised by the types of proteins expressed. In the fastest growing regime 1, ketoacid built up but no recycling enzyme was expressed, i.e. $f_K = 0.0$. In this regime, the growth rate decreased because the metabolite k did not contribute to growth in any way, but the cost of expressing its recycling enzyme would have exceeded that of the wasted metabolite. In the intermediate-growth regime 2, both ketoacid recycling enzyme and carbon transporters were expressed, and intracellular carbon originated from both ketoacid and direct uptake. Only in this regime, the ribosome allocation fraction approx-

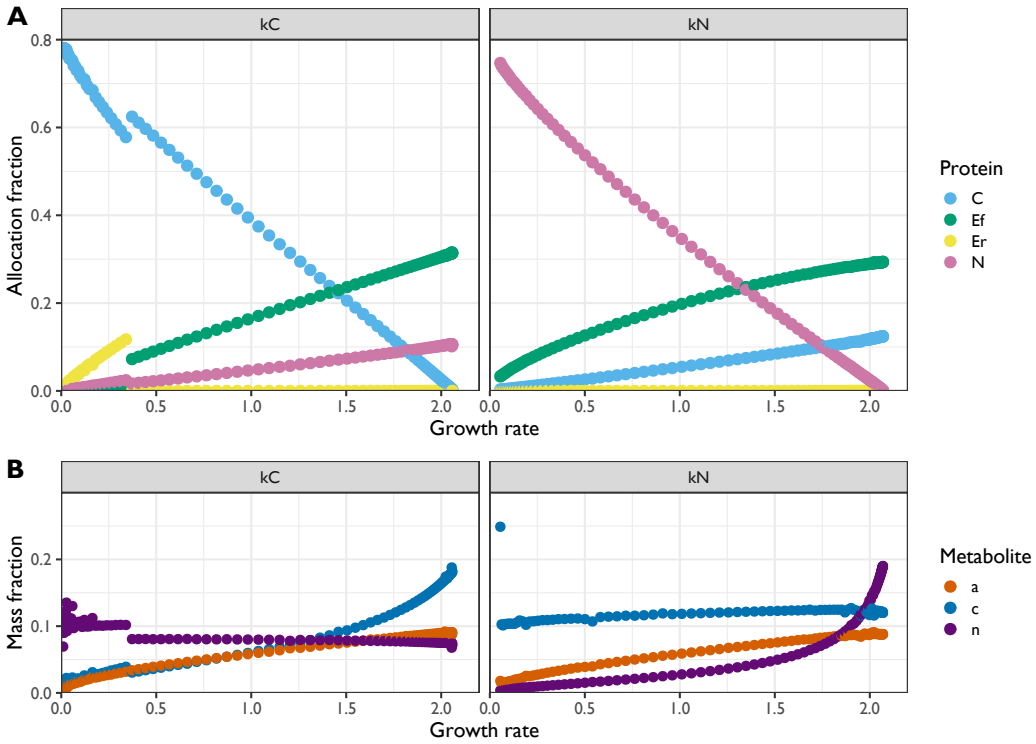


Figure 4.12: Analysis of allocation-optimised model with parallel metabolic enzymes under modulation of the transporter uptake rates. Allocation parameters were chosen to maximise the growth rate for each chosen k_C (left panels) and k_N (right panels). Both uptake rates were varied separately between 1.0×10^{-2} h^{-1} and 1.0×10^{3} h^{-1}. **A.** Components of the protein allocation vector $(f_C, f_E, f_N, f_R, f_Q)$ as a function of growth rate μ. The f_R- and f_Q-components are omitted from the figure to improve clarity. **B.** Steady-state biomass fractions of metabolites a function of growth rate μ. Fixed parameters: $k_{E_f} = 10.0$ h^{-1}, $k_{E_r} = 5.0$ h^{-1}, $\sigma = 6.46$ h^{-1}, $c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_{E_f} = \Delta_{E_r} = 0.0$, $\alpha_{C_f} = {}^{48}/_{55}$, $\alpha_{C_r} = {}^{24}/_{31}$, $f_Q = 0.2$.

imated the linear dependence on amino acid concentrations shown in the previous sections (Figure 4.13C). Lastly, in the slowest regime 3, all carbon originated from ketoacid, and $f_C = 0.0$. This was the case when $\gamma_K > \alpha_{C_f} \approx 0.87$, such that with each nitrogen, more carbon was being taken up by E_N than could be anabolised by E_A. In agreement with this, ketoacid built up considerably in Regime 3 (Figure 4.13D). It is expected that regime 3 would be negated by the possible excretion of superfluous ketoacid, which is what we explored next.

4.3.10 Full model: Different parametrisations induced complex trade-offs between carbon uptake, ketoacid recycling, and excretion

Finally, we explored the behaviour of the full model with no restrictions on ketoacid metabolism. We considered the optimal allocation when the ratio of carbon to nitrogen atoms in the nutrient was varied—by way of the ketoacid stoichiometry γ_K—alongside the maximal turnover rates (efficiencies) k_N, $k_{K_{ex}}$, and $k_{K_{re}}$ of the enzymes involved in ketoacid metabolism. The optimally allocated cell expressed only one or two out of the ketoacid recycling, ketoacid excretion, and carbon uptake enzymes depending on the parametrisation (Figure 4.14). Neither recycling nor excretion was expressed in nitrogen sources not also containing carbon, and additionally this was optimal even for carbon-containing nitrogen sources when the ketoacid recycling and excretion enzymes were inefficient relative to the nitrogen uptake enzyme (Figure 4.14ADE). The value of the ketoacid recycling rate $k_{K_{re}}$ below which ketoacid recycling was suboptimal depended weakly on the ketoacid stoichiometry (Figure 4.14A). When recycling enzymes were expressed, the trade-off between the excretion and ketoacid uptake was heavily influenced by the nitrogen source's carbon content and all three enzyme rates (Figure 4.14ABC). Low-carbon nutrient sources required additional carbon uptake whereas high-carbon nutrient sources generally required ketoacid excretion, with pure recycling being favoured in regimes with intermediate carbon content or inefficient excretion.

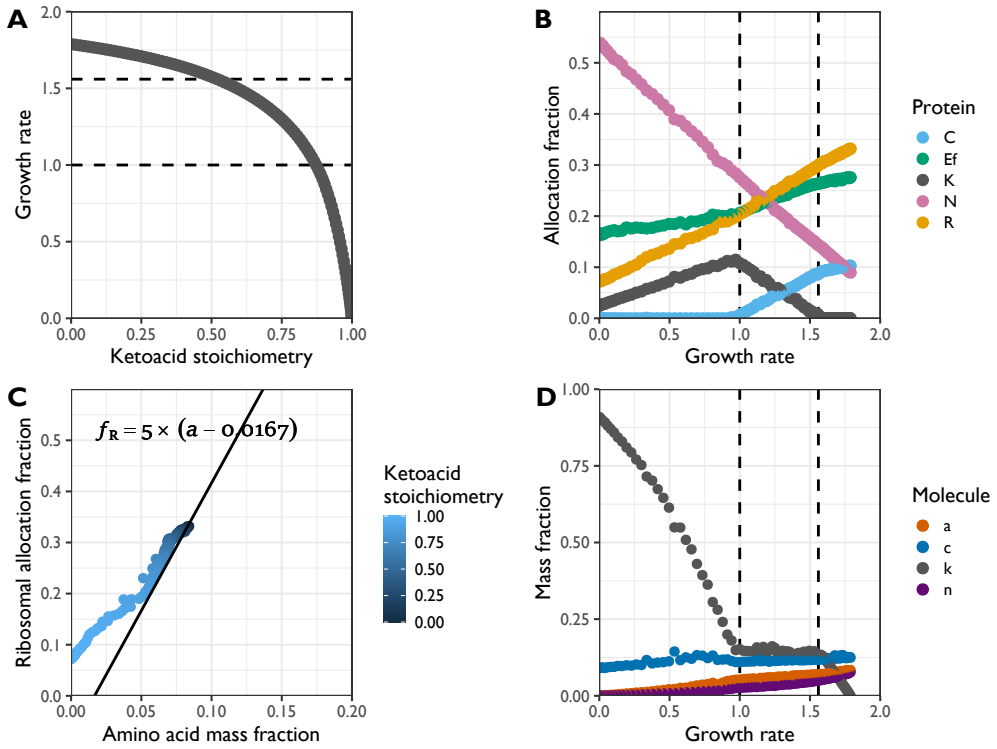


Figure 4.13: Analysis of allocation-optimised model with recycling of ketoacid under modulation of the ketoacid stoichiometry. The relative mass of ketoacid with respect to external amino acid γ_K was modulated to yield different growth rates. The allocation parameters were chosen to maximise the growth rate for each chosen γ_K. Dashed lines indicate regimes with qualitatively different behaviour. **A.** Growth rate μ as a function of ketoacid stoichiometry γ_K. **B.** Components of the protein allocation vector $(f_C, f_{E_f}, f_{K_{re}}, f_N, f_R)$ as a function of growth rate μ. $f_Q = 0.2$ and $f_{E_r} = 0.0$ were omitted for clarity. **C.** Relationship between amino acid concentration a and ribosomal allocation f_R in the model for varying γ_K (blue points) as well as line to guide comparison with previous figures. **D.** Mass fractions of metabolites as a function of growth rate μ. Fixed parameters: $k_{K_{re}} = 10.0\ h^{-1}$, $k_C = 20.0\ h^{-1}$, $k_{E_f} = 10.0\ h^{-1}$, $k_{E_r} = 5.0\ h^{-1}$, $k_N = 5.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $k_{sat} = c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_{E_f} = \Delta_{E_r} = 0.0$, $\alpha_{C_f} = 48/55$, $\alpha_{C_r} = 24/31$, $f_Q = 0.2$, $f_{K_{ex}} = 0$.

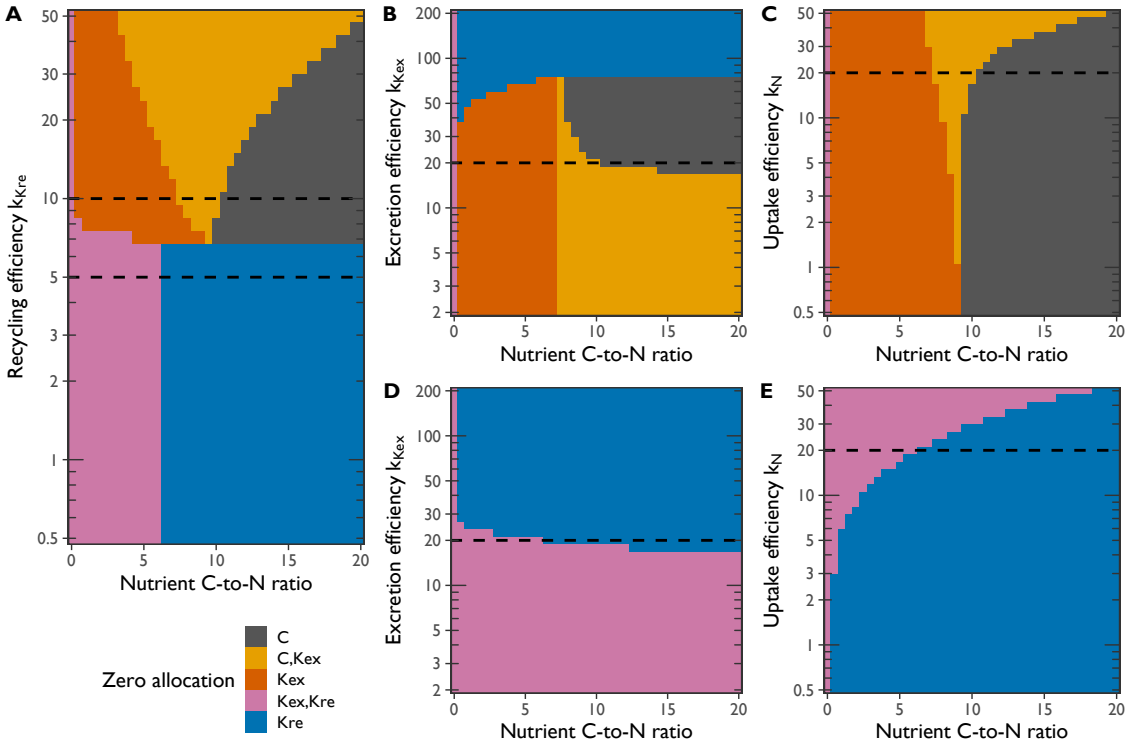


Figure 4.14: Phase diagrams of ketoacid recycling, excretion, and carbon uptake under perturbations of enzyme efficiencies and nutrient carbon-to-nitrogen ratio. For each figure, the optimal allocation was determined for 41^2 combinations of the ketoacid stoichiometry γ_K and one of the enzyme efficiencies. The ketoacid stoichiometry was expressed in terms of the ratio of carbon and nitrogen atoms in nutrient molecules as per equation (4.7). The enzyme efficiencies were log-transformed; parameters were chosen to be equidistant in this representation. Colours indicate which of the enzymes were not expressed in the optimal allocation. Dashed lines indicate parameter values fixed in the other panels. **A.** Ketoacid recycling rate $k_{K_{re}}$ varied, ketoacid excretion rate $k_{K_{ex}} = 20.0\ h^{-1}$ and nutrient uptake rate $k_N = 20.0\ h^{-1}$ held fixed. **B.** $k_{K_{ex}}$ varied, $k_{K_{re}} = 10.0\ h^{-1}$ and $k_N = 20.0\ h^{-1}$ fixed. **C.** k_N varied, $k_{K_{re}} = 10.0\ h^{-1}$ and $k_{K_{ex}} = 20.0\ h^{-1}$ fixed. **D.** and **E.** as **B.** and **C.** but with $k_{K_{re}} = 5.0\ h^{-1}$. Fixed parameters for all panels: $k_C = 10.0\ h^{-1}$, $k_{E_f} = 15.0\ h^{-1}$, $\sigma = 6.46\ h^{-1}$, $k_{sat} = c_{sat} = a_{sat} = n_{sat} = 0.0167$, $\Delta_{E_f} = \Delta_{E_r} = 0.0$, $\alpha_{C_f} = 48/55$, $\alpha_{C_r} = 24/31$, $f_Q = 0.2$, $f_{E_r} = 0$.

4.4 Discussion

4.4.1 Simple feedback of free amino acids setting ribosome allocation

Throughout this chapter it appeared that the proportional regulation

$$f_R = \delta a \tag{4.64}$$

was a good approximation to the growth-rate maximising allocation. The approximation was robust to many variations of the parameters describing the nutrient quality. This relation is remarkable: in principle, optimal allocation could depend on all internal metabolite concentrations and be highly nonlinear, but optimality is approximated by this linear dependence on only a single metabolite. Such approximately proportional regulation also emerged in the earlier model by Weiße et al. (2015), where this was implemented at the level of transcription. This result must be seen in the context of the evolutionary pressure on unicellular organisms to maximise their growth rate, at least when external resources are relatively abundant. The model formulated in this chapter suggests that this pressure is translated into a concrete requirement to monitor the abundance of the self-replication machinery and its substrate, and express both in proportion to each other.

It will be interesting to see if similar regulations exist for other allocation fractions in the model. I did not find another protein whose allocation was linearly correlated with a single metabolite, but the ribosomes are the only protein that is the sole consumer of a single internal metabolite. A systematic exploration of the allocation fractions of all enzymes and the concentrations of their products and substrates might yield suitable candidate regulation functions.

4.4.2 The ribosomal growth law offset

The origin of the factor δ in the preceding paragraph is still somewhat mysterious. In particular, it is an open question whether the optimal value of $\delta \approx 5.0$ can be derived from the other parameters using suitable assumptions. However, when simply assuming that ribosomal allocation is regulated as $f_R = \delta a$, its value is related to any observed offset ϕ_{R0} in the ribosomal growth law $\phi_R = \phi_{R0} + \sigma^{-1}\mu$ through

$$\phi_{R0} \approx \delta a_{sat}, \tag{4.65}$$

as derived in section 4.3.1. This offset has been explained by the presence of a pool of ribosomes not actively involved in translation (Dai et al. 2016; Metzl-Raz et al. 2017), but such a mechanism is absent in our model and further investigations are therefore required. I expect that a quantitative explanation for the origin of the offset in the ribosomal growth law will have profound implications on the understanding of the interplay between ribosome synthesis, excess translational capacity, and cell growth (Dai and Zhu 2020).

4.4.3 Starvation response and internal metabolite levels

In section 4.3.6, it was seen that modulating the efficiency of given enzymes repressed the abundance of its substrate, while the abundance of other metabolites remained at levels seen during fast growth. Such a mismatch may have implications for the starvation response, seen at very low growth rates. When nutrient conditions deteriorate such that growth only proceeds very slowly, cells often exit the cell cycle instead of continuing replication. An intriguing possibility is that the starvation response is induced precisely by mismatched metabolite concentrations rather than falling internal concentrations of the limiting nutrient. In fission yeast it is known that carbon starvation induces cell death, whereas nitrogen starvation induces the transition into a quiescent state wherein cells can survive for very long periods without proliferating. In this light, the observation that carbon starvation induces respiratory behaviour whereas nitrogen starvation does not is particularly intriguing.

On the other hand, metabolite build-up is often toxic in addition to just wasteful, and many pathways contain end-product inhibition to avoid this situation. This behaviour was not modelled but if implemented may keep metabolite levels more coordinated. Unfortunately, introducing metabolite toxicity and/or end-product inhibition would have required the addition of many extra parameters (at least one per metabolite) and it was therefore not explored further.

4.4.4 The fate of the carbon backbone

In the full model, modulations of the ketoacid stoichiometry γ_K gave rise to a wide range of growth rates in a monotonically decreasing manner. However, as observed in our fission yeast experiments described in chapter 2, the carbon-to-nitrogen ratio cannot be the sole determinant of a nitrogen source's quality. For example, glycine and tryptophan media gave rise to very similar growth rates in *S.*

Table 4.1: Carbon-to-nitrogen ratios of selected amino acids. The amino acids are ordered according to the growth rate that they support in *S. pombe* minimal media (Figure 2.1). Note that tryptophan molecule contains two nitrogen atoms; the other selected amino acid molecules contain no nitrogen apart from the backbone amino group. An estimate for the ketoacid stoichiometry parameter γ_K from the model is given by $\gamma_K = \frac{n_C \times 12}{n_C \times 12 + n_N \times 14}$.

Amino acid	**C:N ratio**	n_C	n_N	γ_K
Tryptophan	5.5	11	2	$61/80 \approx 0.83$
Glycine	2	2	1	$12/19 \approx 0.63$
Phenylalanine	9	9	1	$54/61 \approx 0.89$
Serine	3	3	1	$36/50 = 0.72$
Isoleucine	6	6	1	$36/43 \approx 0.84$
Proline	5	5	1	$30/37 \approx 0.81$
Glutamate	5	5	1	$30/37 \approx 0.81$

pombe, but carbon is present in a 2:1 ratio in glycine and in a 5.5:1 ratio in tryptophan (Table 4.1). This suggests that each growth medium is not only associated with γ_K, but that at least one out of the enzymatic rates k_N, $k_{K_{ex}}$,and $k_{K_{re}}$ must also be modulated by the choice of nitrogen source. As shown in section 4.3.10, the model behaviour of carbon-to-nitrogen ratio is highly complex but the coarse-grained model can still be used to obtain qualitative intuition.

For a quantitative mapping between the coarse-grained proteome sectors and proteome data, a good understanding of the nitrogen uptake pathways is necessary. Whether recycling or excretion is preferred would depend on the exact reactions that are available to the organism, and the implications this has for the effective recycling and excretion rates $k_{K_{re}}$ and $k_{K_{ex}}$. If ketoacid recycling effectively feeds into other synthesis pathways, this would correspond to large $k_{K_{re}}$. Similarly, the availability of excretion pathways would influence the value of $k_{K_{ex}}$. The metabolism of indigestible ketoacids has been well studied in *S. cerevisiae*, whose excretions of such "fusel oils" can spoil industrial applications (Godard et al. 2007). Unfortunately, no high-quality metabolic model exists as yet for *S. pombe*, so I was unable to calibrate the coarse-grained model from this chapter with our obtained expression data.

4.4.5 Explicit overflow metabolism and energy generation

The model accounted for the distinction between respiratory and respirofermentative growth by adjusting a single parameter, α_C, which represents the stoichiometry of carbon in biomass production relative to nitrogen. We saw that this minimal model accounted for the different behaviours under nitrogen and carbon limitation. Carbon limitation induced a switch to fully respiratory growth upon decreasing growth rates but nitrogen limitation did not. However, one important feature of fermentative growth, namely overflow metabolism, was not explicitly modelled. Another caveat to the results described in this chapter is the lack of energy metabolism in the model. These two potential model constituents are related, as cellular energy in the form of ATP equivalents is generated in different amounts by fermentation and respiration pathways. Adding an explicit fermentative pathway would therefore also be an opportunity to account for the energy balance.

4.4.6 Non-protein biomass

While the parameter α_C represented the biomass carbon-to-nitrogen ratio in the model, it cannot be directly equated to observed dry mass compositions in real cells (see Table 4.2). This is partially due to the inclusion of extra carbon that real cells excrete during fermentation (see previous paragraph), but also because the biomass of real cells consists of more macromolecules besides proteins. One important pathway in this regard is the biosynthesis of nucleotides, especially (ribosomal) RNA. Importantly, RNA biosynthesis requires the pentose phosphate pathway (PPP). However, the PPP shares with glycolysis and amino acid metabolism multiple intermediate metabolites. A coarse-grained model that includes both protein and nucleotide synthesis must therefore account for two parallel (and possibly interlocking) carbon metabolic pathways. Each interlocking metabolite must be represented by a metabolite in the model. This is not compatible with the minimal nature of the coarse-grained models, and I therefore think that the interplay between nucleotide and protein biosynthesis is better explored with models incorporating genome-wide metabolic maps.

Another important contribution to the biomass of *S. pombe* comes from its cell wall, which mostly consists of carbon. The relative importance of the cell wall

Table 4.2: Approximate relative dry mass composition of selected organisms. Most of these numbers were obtained via the Bionumbers database and corresponding identifiers (BNIDs) are indicated (Milo et al. 2010). The protein content of *S. pombe* was approximated from the following observations: (i) the cell measures approximately 14 by 3.5 microns (Fantes and Nurse 1977), resulting in a volume of 0.123 pl if its shape is assumed to be a spherocylinder; (ii) the average dry mass density is 282 pg/pl (Odermatt et al. 2021), such that the dry mass per cell is approximately 35 pg; (iii) the protein content (of haploid wild type cells in glucose minimal medium) is 15 pg/cell (Fantes and Nurse 1977).

Species	Chemical formula	Protein content (g/g)
E. coli	$CH_{1.77}O_{0.49}N_{0.24}$ (BNID 101800; von Stockar and Liu 1999)	0.55 (BNID 101436; Neidhardt et al. 1990)
S. cerevisiae	$CH_{1.613}O_{0.557}N_{0.158}$ (BNID 101801; von Stockar and Liu 1999)	0.53(BNID 102328; Ertugay and Hamamci 1997)
S. pombe	$CH_{1.9}O_{0.61}N_{0.155}$(de Queiroz et al. 1993)	~0.43

changes with the size and shape of the cell, both of which depend on growth conditions and fluctuate during the cell cycle. The interplay of resource allocation, cell size, cell shape, and growth is explored in the next chapter.

5 A C-GRAM of *S. pombe* metabolism, growth, size, and geometry

5.1 Introduction

In this final chapter I will discuss a coarse-grained resource allocation model of the fission yeast *Schizosaccharomyces pombe* that describes its size and geometry in addition to its metabolism. Two important observations about the yeast cell that were not modelled in the previous chapter are (i) the existence of the cell wall, and (ii) the cell cycle. Importantly, the relative contribution of the cell wall to the biomass depends both on the shape and size of the cell and on the cell cycle. This is because the ratio of the cell surface area to its volume is necessarily larger for smaller cells. The model described in this chapter will expand upon and quantify the intuition behind these statements.

The internal metabolic model is based on the C-GRAM described in the previous chapter 4. However, the parametrisation has been adjusted to better describe the physiology of *S. pombe*. In addition to metabolism and growth, the model discussed in this chapter describes the geometry of the fission yeast cell. This necessitated the introduction of lipid metabolism to describe the cell wall. Furthermore, the system was rewritten in terms of macromolecular amounts instead of mass fractions, so as to calculate cell volume, surface area, and total dry mass separately. As we will see, a further complication introduced was a basic model of ribosomal efficiency as a function of cytoplasmic density.

In the first section of this chapter I will introduce the experimental observations that informed the choices I made in the model construction. I will then describe the model itself in detail, followed by my efforts to choose suitable parametrisations. I will further discuss the core behaviours that the model represents.

5.1.1 The cell wall is mainly composed of polysaccharides

From a metabolic standpoint, the biomass sequestered in the (fission) yeast cell wall should be expected to make an important contribution to resource allocation, because it comprises an appreciable fraction of the total dry mass. One study

found that the cell wall amounted to around 30% of the total dry mass across nine different yeast species (Nguyen et al. 1998), with around 85% of the cell wall mass consisting of polysaccharides, the remaining being proteins. For *S. cerevisiae*, the cell wall dry mass was shown to be dependent on growth conditions; cell wall dry mass amounted to 18.3 ± 2.0 percent of the total dry mass in glucose minimal media (Aguilar-Uscanga and François 2003). This is in line with an older estimate of the *S. pombe* cell wall mass in an obsolete rich medium, which amounted to 19.3 ± 1.9 percent of dry mass (mean and standard deviation of four reported values from Kogane and Yanagita (1962)). Unfortunately I found no *S. pombe* data for defined media. In summary, the cell wall comprises around 20% of the yeast biomass, which is predominantly composed of carbon.

5.1.2 *S. pombe* cell cycle progression consists of elongation and self-replication

The standard eukaryotic cell cycle progresses through four phases: G1, S, G2, and M phase (Alberts et al. 2015, ch. 17). Self-replication occurs in S and M phase, with DNA synthesis and chromosome duplication defining S phase, and nuclear and cell division comprising M phase. G1 is the gap phase between cell division and the start of S phase, and G2 the gap phase between S phase and entry into M phase. Many cell types, including unicellular organisms such as fission yeast, grow in size and/or mass during S phase and both gap phases.

The entry into S phase is called the Start transition; it generally requires the cells to reach a minimum size. Wild-type *S. pombe* cells grown in standard (Edinburgh) minimal media (EMM) exceed the Start size threshold at birth and therefore grow little in G1/S phase (Nasmyth et al. 1979). However, mutant strains, such as *wee1Δ*, do have to pass Start and their characterisation has been instrumental in understanding the control of cell cycle progression (Russell and Nurse 1987). Furthermore, wild-type *S. pombe* can be induced to spend more time in G1 by a suitable choice of growth media (Carlson et al. 1999). In particular, cells grown in EMMTrp media (see chapter 2) are mostly in G1.

When biomass accumulates, the *S. pombe* cell envelope grows in size accordingly. The envelope expands outwards from the cell tips, with the net result that the cell elongates with a constant width (Mitchison and Nurse 1985). Initially elongation occurs only from the so-called old end (i.e. the cell tip which was also part of the mother cell), but elongation starts from the new end at the "new end take

off" or NETO point in the cell cycle. The cell tips are rounded and approximately hemispherical, though slightly pointier than hemispheres when elongation occurs (Abenza et al. 2015).

The final phase of the cell cycle is called M phase, and it consists of mitosis and cytokinesis (Alberts et al. 2015, ch. 17). During mitosis the duplicate chromosomes are separated and assembled into two daughter nuclei on opposite ends of the cell. Yeasts such as *S. pombe* are able to perform this process without dissolving the nucleus into the cytoplasm, in so-called 'closed' mitosis (Mori and Oliferenko 2020). When the nuclei are physically separated, a septum is formed that splits the cell into two compartments. Importantly, *S. pombe* elongation halts during M phase (Mitchison and Nurse 1985). When both chromosome separation and septation are complete, cytokinesis triggers as regulated by the septation initiation network (Simanis 2015). The septum constricts and separates the cell into two halves, and the turgor pressure quickly shapes the new ends into approximate hemispheres (Atilgan et al. 2015). Each daughter cell contains a new nucleus containing a single copy of the genetic code and the cell cycle restarts.

5.1.3 Dry mass density fluctuates with *S. pombe* cell cycle progression

A key question that arises when describing the interplay between translational resource allocation and cellular geometry is how biomass production and volume expansion are linked. A recent study has probed exactly this question. Using a novel method of analysing bright-field microscopy images, Odermatt et al. (2021) measured the dry mass density along the cell cycle in a spatially resolved manner. The average dry mass density was 282 ± 16 mg/ml; it fell by approximately 5% during G2 phase, increased by ~10% during M phase, and rapidly decreased by ~5% directly after cytokinesis. Meanwhile, cells grew steadily in length during G2 phase, but elongation was almost stopped from the G2–M transition onwards. The ratio of surface area (including the septal surface) to total dry mass remained approximately constant along the cell cycle, as previously seen in *E. coli* (Oldewurtel et al. 2019). These observations, and follow-up experiments with cells halted in G2 and M phase, were consistent with a model comprising (i) a constant, exponential, increase in cell mass throughout all cell cycle phases; (ii) a constant, exponential, increase in cell volume during G2 phase, and (iii) halted elongation during M phase (Odermatt et al. 2021).

5.1.4 Simple models of molecular crowding optimise ribosomal volume occupancy

An important biophysical process to account for when considering varying densities is molecular crowding. The rate of biochemical reactions is influenced by effective enzyme concentrations, and therefore the total protein density (Dill et al. 2011). When the total protein density is very low, reaction propensities will be diffusion-limited, whereas for high density, molecular crowding impedes reaction propensities instead. At some optimal protein density, the effects are similarly small and reactions occur at their maximal rate. The balance between crowding and diffusion therefore influences the optimal allocation between envelope and biomass production: if the cell does not produce enough envelope, the cytoplasm will be too crowded, but a large allocation towards the cell envelope will dilute the macromolecules too much in addition to being wasteful of cellular resources.

A quantitative intuition can be obtained from a simple biophysical model. Dill et al. (2011) considered a representative chemical reaction rate r_d between a stationary reactant and moving particles with concentration c diffusing with diffusion constant D as $r_d \propto cD$. Approximating the cytosol as a colloidal glass, $D \propto \exp(-\alpha\phi)$ for some parameter α, where ϕ is the volume fraction occupancy (Zhou et al. 2009; Vazquez 2012); an estimate of $\alpha = 5.8$ was used by Vazquez (2012). With the diffusion length set by the concentration c or, equivalently, by the volume fraction ϕ (Dill et al. 2012), the reaction rate is then proportional to

$$r_d \propto \phi \exp(-\alpha\phi). \tag{5.1}$$

It therefore has an optimum at $1/\alpha = 0.17$—approximately a third of the hard-sphere packing fraction. Although the exact value heavily depends on the exact assumptions and values used, this analysis nonetheless highlights that the remarkable amount of molecular crowding found in the cell can be consistent with models that maximise enzymatic fluxes.

Direct measurements of molecular crowding have been obtained for particles with similar sizes as RNA polymerases and ribosomes (Delarue et al. 2018). The intracellular diffusion of purposely designed nanoparticles with diameters of ~20 nm and ~40 nm was shown to depend on the ribosomal abundance. This suggests that ribosomes obstruct their own diffusion as well as other roughly similar-sized

particles such as RNA polymerases. Therefore, in a cytoplasm crowded with ribosomes, the effective biomass production flux catalysed by the ribosomes will be reduced.

A more detailed treatment of crowding-affected enzymes and ribosomes was given by Pang and Lercher (2020). They considered Michaelis–Menten enzyme kinetics with an effective Michaelis constant K_M^* that accounted for molecular crowding and depended on several parameters, including the total volume occupancy ρ and the sizes of the enzymes and substrates. Importantly, ribosomes (with radius ~13 nm) and their substrates (tRNA, radius ~2.4 nm) are substantially bigger than metabolic enzymes (~2.4 nm) and their substrates (~0.34 nm). Using suitable parametrisations, Pang and Lercher (2020) numerically calculated the flux through representative pathways at various volume occupancies. Parametrisations representing ribosomal fluxes were peaked around an occupancy of ~0.12, with the relative flux dropping more steeply than expected from equation (5.1). However, metabolic pathway representations did not vary more than around 10% for the entire range of permitted occupancies.

5.1.5 Carbon and nitrogen limitation have different effects on surface-to-volume ratio

Data on the shape and size of *S. pombe* in conditions with varying growth rates were generated in our lab by François Bertaux, as summarised in Figures 5.1 and 5.2. Cells were cultured in turbidostats and the growth rate was varied by the concentration of ammonium or glucose in otherwise standard minimal medium (EMM). In slow-growing cultures, the cell size was generally decreased in both types of culture (Figure 5.1), and the correlation of volume and growth rate appeared identical in both culture types. However, the average length of cells was actually larger in cells only moderately limited by nitrogen availability. Therefore, the surface-to-volume (S/V) ratio differed between glucose- and ammonium-limited cultures (Figure 5.2), with the S/V ratio slightly elevated during strong glucose limitation, but significantly more so during ammonium limitation. Cells became shorter and narrower under increasing limitation, with nitrogen-limited cultures decreasing more in width and carbon-limited cultures decreasing more in length.

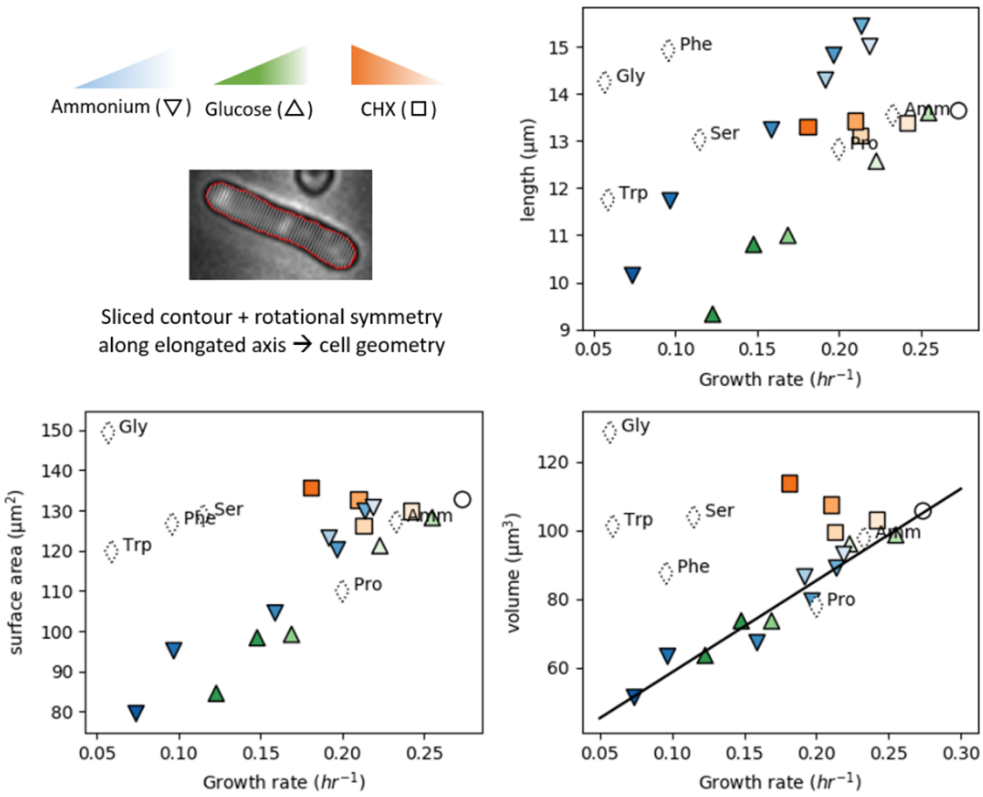


Figure 5.1: Length, surface area, and volume of fission yeast cells grown in environments supporting a range of growth rates. Blue triangles denote nitrogen modulation by the concentration of ammonium in the medium, and green triangles denote carbon modulation by the concentration of glucose. Orange squares denote translation inhibition by cycloheximide (CHX) and open diamonds denote nutrient quality modulation by nitrogen sources. Only nitrogen and carbon concentration modulations were investigated in this section. The black line is a joint linear fit to the nitrogen- and carbon-limited conditions. Image kindly provided by François Bertaux.

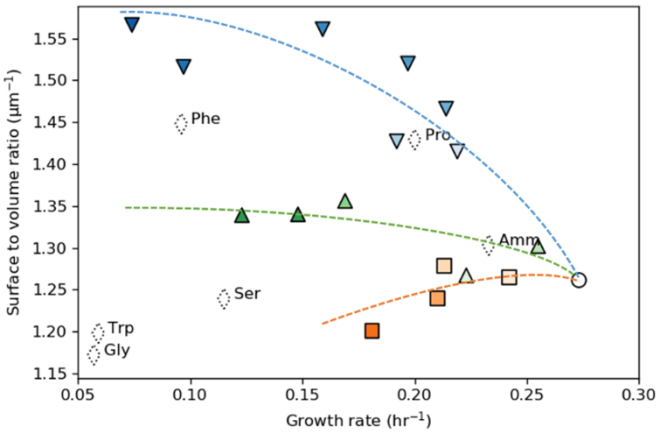


Figure 5.2: Surface-to-volume ratio of fission yeast cells grown in environments supporting a range of growth rates. Colours and shapes are as in Figure 5.1. Dashed lines were drawn to guide the eye. Image kindly provided by François Bertaux.

5.2 Model formulation

5.2.1 The core metabolic pathway was restated in terms of absolute abundances

For the metabolism of the yeast model described in this chapter, I adapted the model described in chapter 4. However, as mentioned in section 3.3.1, the non-extensivity of the Michaelis–Menten rate laws means that the parametrisation had to be reinterpreted. In the following, I explain this for the ribosomal rate law, but the argument and derivation is identical for all the other enzymes.

The ribosomal flux, normalised to the system size, was previously assumed to take the form

$$j_R(a;r) = k_R r \frac{a}{a + a_{sat}}, \tag{5.2}$$

with a and r the amino acid and ribosome concentrations, k_R and a_{sat} were fixed parameters. With an explicit cell volume V and writing A and R for the amino acid and ribosome abundances per cell, the Michaelis–Menten assumption gives a flux per cell of

$$v_R(A;R) = V \cdot k_R \frac{R}{V} \frac{A/V}{A/V + K_A} = k_R R \frac{A}{A + VK_A}. \tag{5.3}$$

If Ω is the total mass of the cell, note that

$$\frac{v_R}{\Omega} = k_R \frac{R}{\Omega} \frac{A/\Omega}{A/\Omega + \frac{V}{\Omega} K_A} = k_R r \frac{a}{a + \frac{V}{\Omega} K_A} \neq j_R, \tag{5.4}$$

because the model is constructed such that $\frac{V}{\Omega}$, the inverse of the dry mass density, fluctuates during the cell cycle. However, as these fluctuations are minor (approximately 25%, see section 5.3.2), approximating the system-size normalised flux by 5.2 is still appropriate absent an explicit interest in the cell cycle.

5.2.2 Cells were assumed to have constant width throughout all simulations

To ensure that the cell width remained constant during the cell cycle, I assumed that the width w was a constant parameter for each instantiation of the model. This simplified the modelling considerably, because it enabled me to convert between length, surface area, and volume according to the equations given in section 5.2.3. I discarded alternative assumptions because they led to variations in cell width along each cell cycle, which is not consistent with *S. pombe* physiology. However, this assumption means that the model is unable to represent changes between environments, as such a transition would require the width to be variable, rather than a fixed parameter. Some alternative assumptions are discussed further in section 5.4.4.

5.2.3 Useful identities for spherocylinders

Consider a spherocylindrical cell with length λ and width w. The surface area Σ and volume V are related through the following equalities:

$$\Sigma = \pi w \lambda; \tag{5.5}$$

$$V = \frac{\pi}{4} w^2 \lambda - \frac{\pi}{12} w^3. \tag{5.6}$$

Note that two out of these four geometric quantities are sufficient to specify the other two. I assumed that the width w was constant, because *S. pombe* cell width remains constant during balanced growth (Knapp et al. 2019). The identities that relate the surface Σ, volume V, and length λ to one another and the width w are therefore particularly useful:

$$\lambda = \frac{\Sigma}{\pi w} \tag{5.7}$$

$$= \frac{4}{\pi}\frac{V}{w^2} + \frac{w}{3}; \tag{5.8}$$

$$\Sigma = 4\frac{V}{w} + \frac{\pi}{3}w^2; \tag{5.9}$$

$$V = \frac{w\Sigma}{4} - \frac{\pi}{12}w^3. \tag{5.10}$$

Now consider the surface-to-volume ratio $\frac{\Sigma}{V}$, which can be expressed as a function of width and length, surface, or volume accordingly:

$$\frac{\Sigma}{V} = \frac{4}{w}\frac{1}{1-\frac{w}{3\lambda}} \tag{5.11}$$

$$= \frac{4}{w}\frac{1}{1-\frac{\pi}{3}\frac{w^2}{\Sigma}} \tag{5.12}$$

$$= \frac{4}{w} + \frac{\pi}{3}\frac{w^2}{V}. \tag{5.13}$$

Over the course of a cell cycle, the length, surface, and volume all increase. It is quickly seen, then, that the assumption of constant width corresponds to a surface-to-volume ratio that decreases during the elongation phase of the cell cycle, when λ, Σ,and V all increase.

5.2.4 Cell surface was synthesised proportionally to biomass generation

The surface of the cell separates the bulk from the environment. The fission yeast cell surface consists of a cell wall and a cell membrane. In the model, both were jointly represented by a metabolite L (the abbreviation was chosen for lipids, the main component of the membrane). The species L was assumed to be present only in the cell surface, and therefore the surface area Σ of the cell was directly proportional to the amount L of lipid present in the cell:

$$\Sigma = \frac{L}{\rho_S}, \tag{5.14}$$

with ρ_S the concentration of molecule L per surface area (in molecule equivalents per square micron). Together with the fixed width w, this surface area Σ uniquely determined the size and shape of the cell: the volume V and length λ are calculated via equations (5.7) and equation (5.10).

Because both the cell wall and the membrane consist mostly of carbon compounds (polysaccharides in the case of the wall and lipids for the membrane), the metabolite L was assumed to be produced directly from the carbon precursor C. I assumed that this reaction was not limited by enzyme abundances, but rather by some process that keeps the internal density in line. In other words, even though many processes have been implicated here, I assumed that expression of these enzymes was cheap and did not significantly affect ribosomal allocation. Alternatively, one can think of these proteins as part of the Q-sector of proteins with fixed allocation. In practice, all this means that the lipid production flux was proportional to the ribosomal flux producing (protein) biomass:

$$v_L = \sigma v_R, \tag{5.15}$$

with σ a parameter that governs the surface stoichiometry.

5.2.5 The cytosolic occupancy of ribosomal proteins determined an effective crowding factor

As noted in 5.1.4, the molecular crowding affects enzymatic fluxes. The effect is far stronger for larger complexes, such as ribosomes, than for metabolic enzymes (Pang and Lercher 2020). With crowding of metabolic enzymes only affecting the flux by at most ~10%, I opted to neglect this contribution and focus on the ribosomes only. Consider, therefore, the cytosolic occupancy of ribosomal proteins, given by

$$\omega_R = \frac{V_{rib} R}{V}. \tag{5.16}$$

Here, V_{rib} is a parameter that reflects the volume occupied by one unit of R. Note that one unit of R in the model is not equal to one ribosome, so V_{rib} is an effective parameter and not equal to the volume occupied by a single ribosome as experimentally determined. To model the effect of crowding, I used the following two-parameter crowding function g_c, which takes as its argument an occupancy $0 \leqslant \omega \leqslant \omega_{max}$.

$$g_c(\omega) = \exp\left(-\zeta \frac{\omega}{\omega_{max} - \omega}\right). \tag{5.17}$$

Here, ζ and ω_{max} are parameters. This equation was previously used by Delarue et al. (2018); it is based on Doolittle's model of molecular crowding (Doolittle 1952). ω_{max} is the maximal cytoplasmic volume fraction occupied by ribosomes, and ζ is a phenomenological parameter that represents the strength of the crowding interaction. Delarue et al. (2018) estimated $\omega_{max} = 0.5$ and $\zeta = 0.6 \pm 0.2$ in *S. cerevisiae* for 40nm-diameter nanoparticles of their own design, which they showed to mostly interact with the ribosomes, I used these parameters unchanged to model the crowding of ribosomes themselves.

Based on the above, molecular crowding of the ribosomes was modelled as a slowdown of the non-crowded protein production reaction by an effective crowding factor $g_c(\omega_R)$ given by equation (5.17). The protein production flux after this correction was defined as

$$v_R = g_c(\omega_R)\, k_R R \frac{A}{A + VK_A}. \tag{5.18}$$

All other fluxes were left unmodified from the model proposed in 4.2.1, but only respirofermentative growth on free nitrogen was explored in this chapter. Because $g_c(\omega_R) = g_c(V_{rib}\frac{R}{V})$ decreases exponentially with the ribosome concentration whereas the remaining term increases linearly, the volume-normalised ribosomal flux $j_R = \frac{v_R}{V}$ will be maximised by some intermediate ribosome concentration.

5.2.6 During cytokinesis, septum was produced instead of cell wall

I modelled the cell cycle in three different phases. From cell birth, the cell was initiated in an elongation phase, in which the lipid production flux was incorporated directly as surface growth, i.e.

$$\frac{d\Sigma}{dt} = \frac{v_L}{\rho_S}. \tag{5.19}$$

This most closely aligned with the G2 phase in wild-type *S. pombe* in standard medium. In order to account for the time required for chromosome segregation and nuclear division, I explicitly modelled a separate phase of the cell cycle, where the lipid flux was halted and, consequently, free carbon built up. The entry into nuclear division was triggered by cell cycle proteins, as explained in 5.2.7, and this second phase describing nuclear division lasted for a fixed amount of time $\delta t_{ND} = 0.25\ h^{-1}$. In the third cell cycle phase, the lipid flux—still given by equation (5.15)—

was redirected into septum construction, such that surface area Σ (still) remained constant. This septation phase ended when the septum was fully constructed, i.e. when enough septum had built up to construct the two hemispherical end caps, with total surface area

$$\Sigma_{caps} = \pi w^2, \tag{5.20}$$

which corresponds to a "lipid" copy number of

$$L_{caps} = \pi \rho_S w^2. \tag{5.21}$$

Together, the nuclear division phase and the septation phase model the M/G1/S phases of standard *S. pombe* growth. However, the important feature of the model phases is whether elongation occurs or not, rather than the DNA copy number. I therefore expect the model to remain valid in conditions that affect the cell cycle phases.

5.2.7 The build-up of cell cycle proteins signalled progression from growth to mitosis

The progression from elongation into cell division was initiated by the copy number of a signalling protein Y passing a pre-specified threshold. Protein Y assembled a ring structure in the cylindrical part of the cell, and cell cycle progression was triggered when the ring was fully assembled. The allocation f_Y towards the ring-forming protein Y did not in itself scale with the cell size. Rather, it was simply proportional to the aforementioned metabolic allocation. Therefore, f_Y was a small constant. Cell cycle progression was triggered when enough Y built up to span the circumference of the cell at a fixed linear concentration ρ_Y. I chose $\rho_Y = 0.1$ molecule equivalents/μm, which resulted in very small allocations f_Y such that the division protein did not meaningfully alter the optimal ribosomal allocation. The circumference of the cell is given by πw, and as already noted the cell width w was assumed to be constant. The division threshold for Y was therefore also constant across conditions:

$$Y_{div} = \pi \rho_Y w. \tag{5.22}$$

The mechanism by which Y triggered cell cycle progression resembled how the fission yeast protein Cdr2 is thought to measure cell size (Facchetti et al. 2019). Cdr2 does not construct a simple ring, but rather assembles in the nodal region,

which is a wide region in the centre of the cell. Crucially, the width of this region was found to be constant across conditions (Facchetti et al. 2019). It is therefore equivalent to a narrow ring structure, where the amount of Cdr2 present in the nodal region needs to be enough to circumvent the cell entirely at a fixed nodal density.

I ensured that the allocation to these cell cycle proteins was small, so that it did not interfere with the cell's metabolism. However, the ribosomal allocation towards the signalling proteins had to still satisfy the allocation constraint, even though the allocation f_Y to protein Y could in principle change. To satisfy the allocation constraint, the expression of these proteins was subtracted from the non-metabolic Q-proteins. Therefore, the allocation parameter f_Z only indirectly determined the expression of the Q-proteins, and

$$f_Q = f_Z - f_Y. \tag{5.23}$$

5.2.8 Optimising for maximum growth rate in the copy number model

Unlike in the concentration model, the full model including the cell cycle does not have a defined steady state. Rather, balanced growth is defined as a state where the cell changes vanishingly little between successive cell division events. The dynamic model included discrete events that represented cell divisions and changes between elongation and mitotic phases. These events I implemented within the DifferentialEquations.jl framework by defining suitable callback functions and using the DiffEqCallbacks.jl package (Rackauckas and Nie 2017). The Rosenbrock-type solver, which I used to good effect to solve the concentration-based ODEs to steady-state, was not well suited to the more complex model including these callback events. This was due to interpolation errors, which often multiplied leading to unstable solutions. I found that the OwrenZen3 method better handled the division and cell cycle progression events, and I used it for the evolution of the full model.

The full model evolution was rather more costly than that of the concentration-only model. To save computing time, I was able to start the evolution of the full model at initial conditions based on the steady-state solution of the concentration-only model. I scaled up the steady-state solution to the concentration model by ensuring that the lipid copy number was exactly equal to $2\pi\rho_S w^2$, which is the surface area of a cell whose aspect ratio is exactly 2. I further initiated the division protein Y as present at half threshold levels.

To optimise cell width or surface stoichiometry in addition to the allocation vector, I used a two-step approach. First, I maximised the growth rate of a simplified model of allocation only, with saturation constants approximated by their values when $V = \frac{w\Sigma}{4}$, negligible crowding slowdown, and negligible metabolite concentrations, such that the dry mass density is approximately

$$\rho_{approx} \approx 4\rho_S \frac{1+\sigma}{w\sigma}. \tag{5.24}$$

For the ribosomal flux equation (5.18) for example, this approximation results in a saturation constant

$$a_{sat} = \frac{K_A}{\rho_{approx}} \tag{5.25}$$

such that

$$j_R = k_R r \frac{a}{a + a_{sat}} \tag{5.26}$$

as in the allocation-only models from earlier. I then used this optimum to define the concentrations in the initial condition of the full model describing the copy number evolution. I solved this model for a pre-specified time span (20 doubling times of the simplified model unless otherwise noted) and ensured that the solution was approximately equal between the final two division events.

5.2.9 Global parameter estimation based on observed cell physiology and geometry

To obtain a parametrization that approximated observed cell geometries in three different conditions, I followed the following procedure. First, I defined the following cost function for each condition:

$$C = \left(\frac{\mu_{obs} - \mu_{fit}}{\delta\mu}\right)^2 + \left(\frac{V_{obs} - V_{fit}}{\delta V}\right)^2 + \left(\frac{\phi^M_{obs} - \phi^M_{fit}}{\delta\phi^M}\right)^2 + \left(\frac{\phi^S_{obs} - \phi^S_{fit}}{\delta\phi^S}\right)^2. \tag{5.27}$$

Here, μ is the growth rate, V the cell volume, ϕ^M the metabolite mass fraction, and ϕ^S the total surface biomass fraction. The subscripts denote the observed and fit values, and the δ-values correspond to a weighting of the four types of observations. This cost was based on the fit obtained for any parametrisation after the joint optimization of the surface-to-mass stoichiometry σ and the allocation fractions f_C, f_{E_f}, f_N, and f_R.

For any given parametrisation, I evaluated this cost function for three conditions, one corresponding to fast growth in EMM ($\mu_{obs} = 0.28$ h^{-1}; $V_{obs} = 100$ μm^3), and one each corresponding to carbon- and nitrogen-limited conditions supporting half the growth rate in EMM ($\mu_{obs} = 0.14$ h^{-1}; $V_{obs} = 70$ μm^3). The total cost of a parametrisation was the sum of the cost for each of the three conditions. I manually chose the width parameter to reflect the different surface-to-volume ratios of the three conditions. I set $w = 3.4$ μm for the EMM culture, $w = 3.6$ μm for low-carbon, and $w = 2.9$ μm to represent the low-nitrogen culture. This corresponded to approximate surface areas of $\Sigma = 130$ μm^2 for EMM, $\Sigma = 91$ μm^2 for low-carbon, and $\Sigma = 105$ μm^2 for low-nitrogen, such that the surface-to-volume ratio equalled (to three significant digits) 1.30 for both EMM and low-carbon cultures and 1.50 for the low-nitrogen culture. Furthermore, I aimed for a parametrisation that kept the metabolite mass fraction low, but had an appreciable fraction of biomass sequestered in the surface molecules (the septum and the cell wall equivalents). I therefore chose $\phi^M_{obs} = 0.05$ and $\phi^S_{obs} = 0.20$. Since these were manual estimates not based on direct experimental observation, I weighted the growth rate and volume fit more heavily: $\delta\mu = 0.01$ h^{-1}, $\delta V = 1$ μm^3, $\delta\phi^M = 0.05$, and $\delta\phi^S = 0.20$.

Using a generating-set-based global optimisation algorithm implemented in the Julia package BlackBoxOptim.jl (Kolda et al. 2003; Feldt and Stukalov 2019), I minimised this cost function to obtain the parametrisation summarized in Table 5.1 (for the parameters fit by this method) and Table 5.2 (for the parameters held at a constant, manually chosen, value). A comparison between the fit and observed cell physiologies is given in Table 5.3.

5.3 Results

5.3.1 Dynamics of cell physiology in the full model

Using the parametrisation outlined in Tables 5.1 and 5.2, I evolved the model to steady-state for the three parametrisations corresponding to (i) fast growth in EMM, (ii) carbon-limited growth for glucose-restricted cultures, and (iii) nitrogen-limited growth for ammonium-restricted cultures. The growth rates supported by the nutrient parametrisations in (ii) and (iii) were approximately equal to each other and half of the growth rate in the conditions mimicking EMM. The dynamics of the three model parametrisations over the course of a cell cycle are plotted in Figures 5.3, 5.4, and 5.5. These dynamics followed the expected behaviour, as described in the following paragraphs.

Table 5.1: Model parametrisation that best fit the observations outlined in Table 5.3. Values in italics denote parameters that were obtained via growth-rate optimisation in the inside optimisation loop. Values indicated with asterisks were set manually.

Description	Symbol	Unit	Fast	low-C	low-N
Surface lipid density	ρ_S	μm^{-2}	2.411×10^{4}	2.411×10^{4}	2.411×10^{4}
Ribosome volume	V_r	μm^3	6.827×10^{-7}	6.827×10^{-7}	6.827×10^{-7}
Ribosome efficiency	k_R	h^{-1}	0.6996	0.6996	0.6996
Carbon uptake efficiency	k_C	h^{-1}	2.0*	0.3078	2.0*
Nitrogen uptake efficiency	k_N	h^{-1}	4.0*	4.0*	0.02889
Division protein allocation	f_Y		5.606×10^{-8}	3.526×10^{-8}	2.954×10^{-8}
Width	w	μm	3.4*	3.6*	2.9*
Carbon enzyme allocation	f_C		*0.1696*	*0.4765*	*0.0805*
Fermentation enzyme allocation	f_{E_f}		*0.1623*	*0.0799*	*0.0977*
Nitrogen enzyme allocation	f_N		*0.0102*	*0.0041*	*0.3751*
Ribosome allocation	f_R		*0.4579*	*0.2394*	*0.2467*
Surface-to-mass stoichiometry	σ		*0.1771*	*0.0713*	*0.0853*

Table 5.2: Manually fixed parameters of the copy number model. In this chapter, if parameter values are not explicitly given, the values were as in this table. Note that the chosen value for the fermentative stoichiometry $\alpha_{C_f} = {}^{96}/_{103} \approx 0.932$ corresponds to a carbon-to-nitrogen atomic ratio of exactly 8:1 (4.2.6).

Description	Symbol	Unit	Value
Michaelis constant (all)	$K_C = K_A = K_N$	μm^{-3}	500.0
Fermentation enzyme efficiency	k_{E_f}	h^{-1}	2.0
Enzyme nonlinearity	Δ_{CN}		0.0
Non-metabolic allocation	f_Z		0.2
Fermentative stoichiometry	α_{C_f}		96/103
Ring-bound cyclin threshold	ρ_Y	μm^{-1}	0.1
Crowding interaction	ζ		0.6
Maximal ribosome occupancy	ω_m		0.5
Nuclear division time	t_{ND}	h	0.25

Table 5.3: Selected observations of fission yeast size, shape, and non-protein biomass fractions under different types of limitation, and best-fit approximation by the model parametrised as in Table 5.1. The cost according to equation (5.27) of this parameterization was approximately 22.1. The observed volumes were taken from Figure 5.1 and its underlying raw data. The surface-to-volume ratios were calculated using equation (5.13) from the fit volumes and manually set widths of 3.4, 3.6, and 2.9 µm, thereby approximating the observations from Figure 5.2. A single value for each of the surface lipid density ρ_S, ribosome volume V_r, and ribosome efficiency k_R was fit jointly to each of the three conditions.

	Growth rate (h^{-1})		Volume (μm^3)		*Surface/ volume (μm^{-1})*
	Target	Fit	Target	Fit	
Error	0.01		1		
Fast	0.28	0.2818	100	100.08	*1.2974*
low-C	0.14	0.1420	70	70.06	*1.3048*
low-N	0.14	0.1474	70	69.81	*1.5055*

	Metabolite mass fraction		Surface mass fraction	
	Target	Fit	Target	Fit
Error	0.05		0.05	
Fast	0.05	0.1418	0.20	0.1170
low-C	0.05	0.0735	0.20	0.0588
low-N	0.05	0.0742	0.20	0.0689

The cell cycle consisted of three phases. Directly after cell division, the cells were in the elongation phase until the ring-bound cyclin protein Y reached its division threshold (subfigures A). During the elongation phase, the lipid molecule comprising the surface built up in approximately exponential fashion (subfigures B), which halted at the progression into the subsequent nuclear division phase. During the nuclear division phase, internal carbon storage increased rapidly, because carbon consumption due to lipid and septum construction was halted (subfigures C). After the (constant) nuclear division time had passed, the cells entered the septation phase, in which the lipid flux was rerouted to septum construction (subfigures B). The cell divided into two daughter cells when the septum was fully constructed.

During the entire cell cycle, the total dry mass of the cells steadily increased, but the relative protein content appeared to decreased somewhat during the latter part of the cell cycle (subfigures D). The internal cell volume increased during the elongation phase and remained constant afterwards (subfigures E). Note that the volume did not halve at cell division, due to the instantaneous addition of the new-end hemisphere. The combined dynamics of the dry mass and volume resulted in typical fluctuations in the cell dry mass density (subfigures F), which decreased gradually during the elongation phase, increased strongly during the nuclear division and septation phases, and decreased in a discontinuous manner at the point of cell division. These fluctuations were approximately 25% of the dry mass density, which contrasts with the observed fluctuations of only approximately 10%. A possible explanation for this discrepancy will be explored in 5.3.2.

When comparing the three different conditions, note that the septation phase has the longest duration in the carbon-limited parametrisation. This is because the width, and therefore the relative size of the septum to the rest of the surface area, is largest in the carbon-limited cultures. In other words, in the carbon-limited parametrisation, the cell is short and a large proportion of the cell surface area is taken up by the hemispherical end caps. In contrast, in the nitrogen-limited parametrisation, the cell is long and thin and the hemispherical end caps are less important relative to the cylindrical cell mantle.

A further difference between the carbon- and nitrogen-limited cells is that the former has low internal carbon concentrations, whereas the latter have low internal nitrogen concentrations. This agrees with the behaviour observed in the allocation model of the previous chapter.

Furthermore, the absolute values of the dry mass density in the model representing fast growth in EMM are only approximately half those in the nutrient-limited cells, whose dry mass densities are of comparable magnitude. This is probably due to crowding only affecting the ribosomes. Related to this, note that in the two nutrient-limited parametrisations, the ribosome allocation is approximately half that in EMM Table 5.1. This of course follows the ribosomal growth law.

5.3.2 Dynamics of dry mass density in a simplified model

For a simplified model of a spherocylindrical cell with constant width, the dry mass densities at cell birth, cell division, and at the transition from elongation to septation are related to each other, parametrised only by the geometry of the cell at division. Importantly, the timing of the transition is also determined by the cell geometry. Consider a simplified model of the cell cycle composed of two phases: the elongation phase and the septation phase. The biomass M grows exponentially throughout the cell cycle. During the elongation phase, the surface area of the cell envelope Σ grows exponentially, whereas during the septation phase, the envelope remains at a constant size. The width w of the cell is assumed to be constant, so the volume V and length λ are linearly related to the surface of the cell according to equations (5.7) and (5.10).

Let

$$\mu_L = \frac{\dot{\lambda}}{\lambda} = \frac{\dot{\Sigma}}{\Sigma} \tag{5.28}$$

the elongation rate during the elongation phase. Note that $\Sigma = \pi w \lambda$ means that the elongation rate and the surface production rate are equal to each other. During the elongation phase, the length, surface area, and volume are therefore given by the following functions of time:

$$\begin{aligned}
\lambda(t) &= \lambda(0)\exp(\mu_L t), && (5.29)\\
\Sigma(t) &= \pi w \lambda(t) = && (5.30)\\
&= \Sigma(0)\exp(\mu_L t), \text{ and} && (5.31)\\
V(t) &= \frac{\pi\Sigma(t)}{4} - \frac{\pi}{12}w^3 = && (5.32)\\
&= V(0)\exp(\mu_L t) + \frac{\pi}{12}w^3\left[\exp(\mu_L t) - 1\right]. && (5.33)
\end{aligned}$$

Furthermore, let

$$\mu_M = \frac{\dot{M}}{M} \tag{5.34}$$

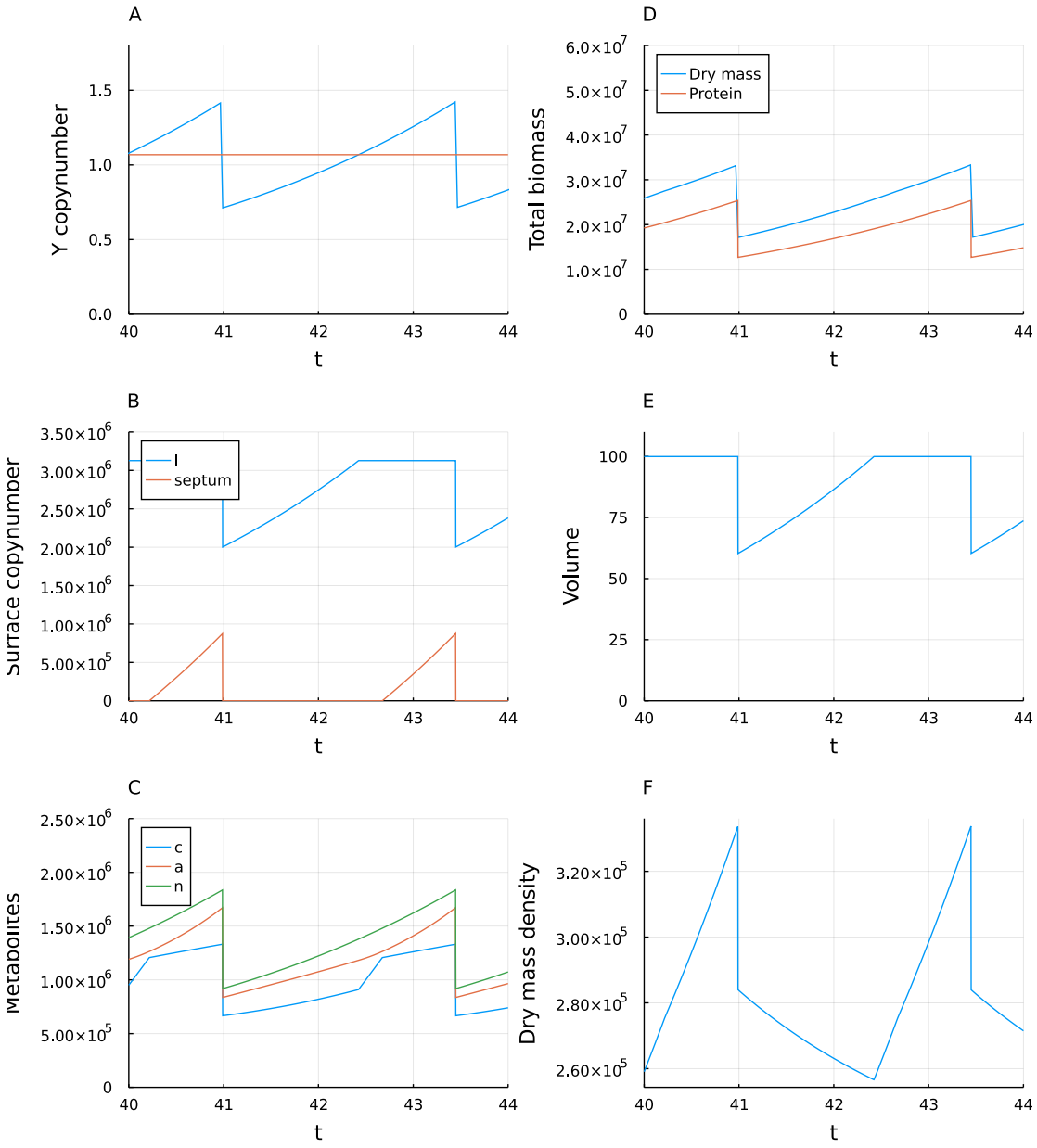


Figure 5.3: Dynamics of cell physiology in model parametrised as EMM culture. All plots are taken during the same time interval, corresponding to approximately 1.5 cell cycle times in balanced growth. **A.** Division protein copy number (blue) and cell cycle progression threshold (orange). **B.** Copynumbers of surface-like proteins: lipids (blue) and septum (orange). **C.** Copynumbers of internal metabolites representing carbon (blue), amino acids (orange), and nitrogen (green). **D.** Total dry mass (blue) and total protein biomass (orange), in units of copynumbers. **E.** Cell volume (in μm^3). **F.** Dry mass density (total dry mass divided by cell volume, in μm^{-3}).

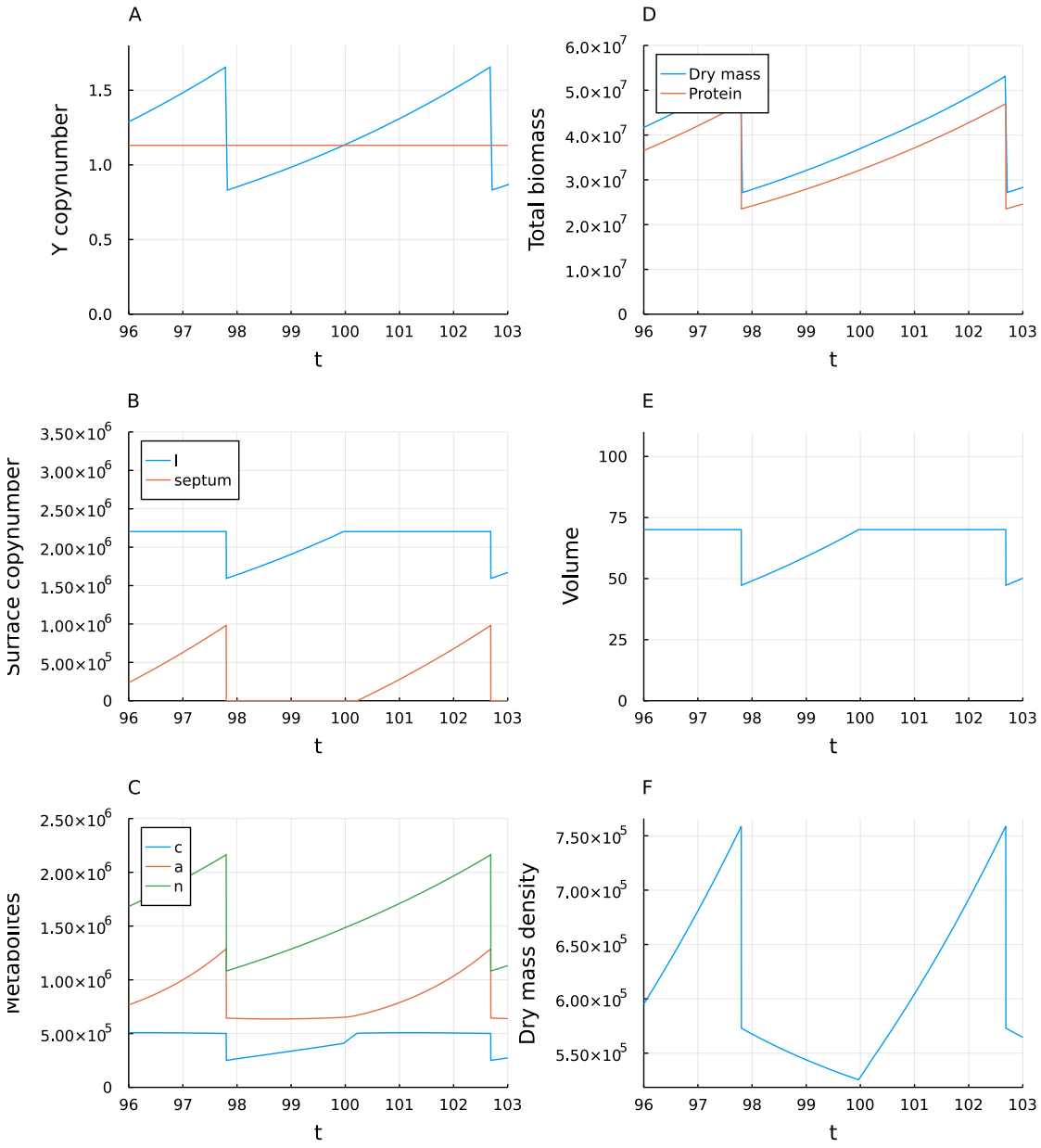


Figure 5.4: Dynamics of cell physiology in model parametrised for low carbon concentrations. All plots are taken during the same time interval, corresponding to approximately 1.5 cell cycle times in balanced growth. **A.** Division protein copy number (blue) and cell cycle progression threshold (orange). **B.** Copynumbers of surface-like proteins: lipids (blue) and septum (orange). **C.** Copynumbers of internal metabolites representing carbon (blue), amino acids (orange), and nitrogen (green). **D.** Total dry mass (blue) and total protein biomass (orange), in units of copynumbers. **E.** Cell volume (in μm^3). **F.** Dry mass density (total dry mass divided by cell volume, in μm^{-3}).

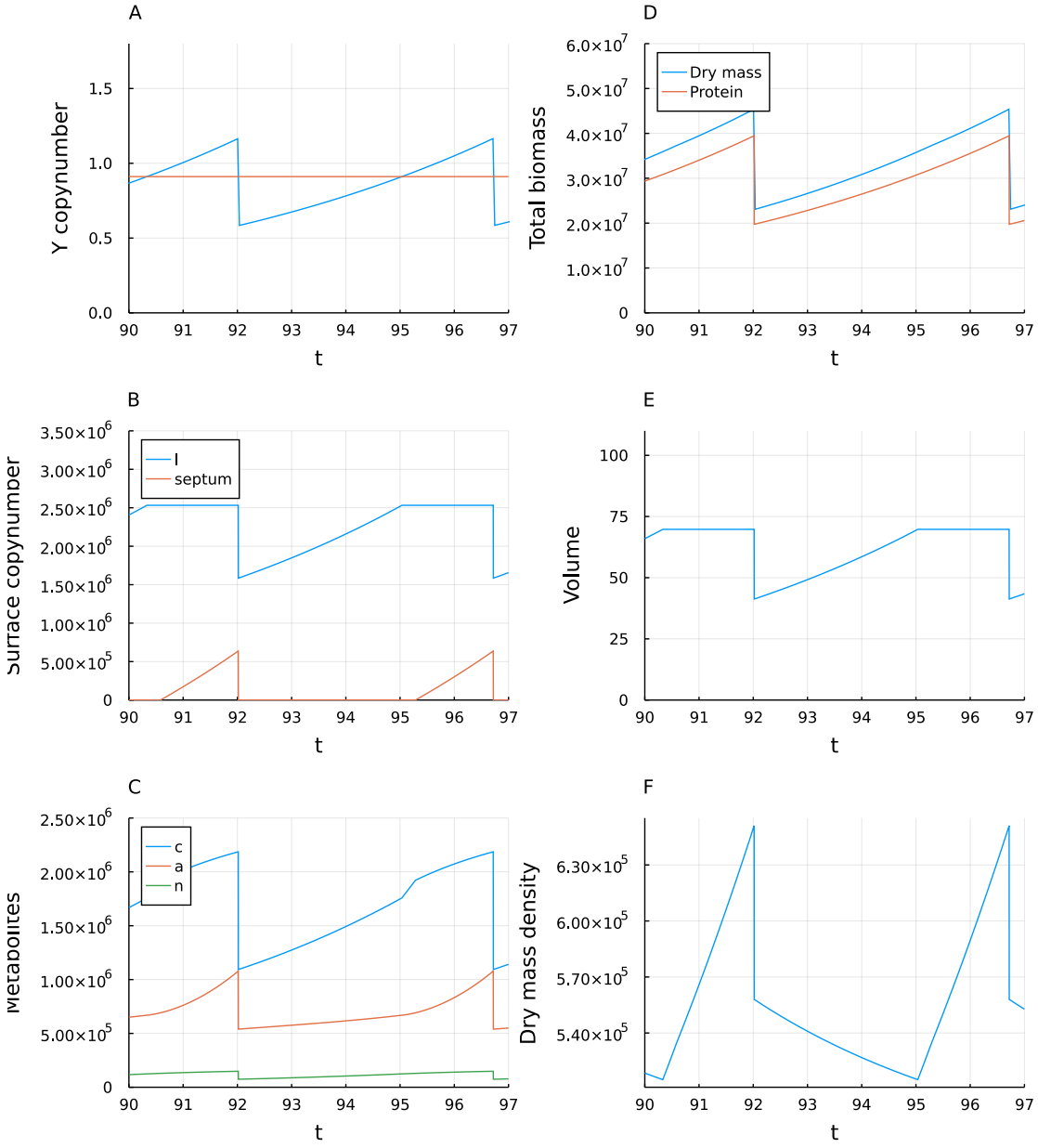


Figure 5.5: Dynamics of cell physiology in model parametrised for low nitrogen concentrations. All plots are taken during the same time interval, corresponding to approximately 1.5 cell cycle times in balanced growth. **A.** Division protein copy number (blue) and cell cycle progression threshold (orange). **B.** Copynumbers of surface-like proteins: lipids (blue) and septum (orange). **C.** Copynumbers of internal metabolites representing carbon (blue), amino acids (orange), and nitrogen (green). **D.** Total dry mass (blue) and total protein biomass (orange), in units of copynumbers. **E.** Cell volume (in μm^3). **F.** Dry mass density (total dry mass divided by cell volume, in μm^{-3}).

be the growth rate of biomass increase and, assume that the cell divides in half at the end of a cell cycle, such that the mass at birth $M_b = M(0)$ is exactly half of the mass at division $M_d = M(T_d) = 2M_b$. Therefore, the biomass growth rate is equal to

$$\mu_M = \frac{\ln 2}{T_d} \tag{5.35}$$

with T_d the doubling time, and the total biomass is given as a function of time as

$$M(t) = M_b \exp(\mu_M t). \tag{5.36}$$

Now, it may be further assumed that the total amount of "surface" molecules, i.e. the cell envelope plus the septum, also increases exponentially during the cell cycle and exactly halves at cell division. This is equivalent to assuming that surface production is exactly proportional to biomass production as then $\mu_L = \frac{\ln 2}{T_d} = \mu_M$. With this assumption, a septum accumulates until it is large enough to be converted instantaneously into two new hemispherical end caps, thereby producing two equal daughter cells. Many of the following derivations do not require this assumption and I will not equate the two rates μ_M and μ_L until necessary to make progress, in the interest of easy generalisation to a model where cell division is not limited by septum construction.

At division, the mother cell envelope is shared equally between the two daughter cells, and the envelopes of both daughter cells are expanded instantaneously by the construction of hemispherical end caps, one for each cell. The surface area of a hemispherical cap is $\frac{1}{2}4\pi\left(\frac{w}{2}\right)^2 = \frac{\pi}{2}w^2$. Therefore, the surface area at birth $\Sigma_b = \Sigma(t = 0)$ is related to the surface area at division $\Sigma_d = \Sigma(t = T_d)$ as

$$\Sigma_b = \frac{\Sigma_d}{2} + \frac{\pi}{2}w^2. \tag{5.37}$$

Likewise, the volume of one hemispherical cap is $\frac{1}{2}\frac{4\pi}{3}\left(\frac{w}{2}\right)^3 = \frac{\pi}{12}w^3$ and therefore the volumes at birth and division are related as

$$V_b = \frac{V_d}{2} + \frac{\pi}{12}w^3. \tag{5.38}$$

From this, the dry mass density at division is related to the density $D_b = D(0)$ at birth:

$$\begin{aligned} D_b &= \frac{M_b}{V_b} = \\ &= \frac{\frac{1}{2}M_d}{\frac{V_d}{2} + \frac{\pi}{12}w^3} = \\ &= D_d \frac{1}{1 + \frac{\pi}{6}\frac{w^3}{V_d}}, \end{aligned} \tag{5.39}$$

with $D_d = D(T_d)$ the dry mass density at division. This can be further rewritten in terms of the aspect ratio $\eta_d = \frac{\lambda_d}{w} = \frac{\Sigma_d}{\pi w^2}$ at division:

$$\begin{aligned} \frac{V_d}{\pi w^3} &= \frac{\frac{\pi}{4}w^2\lambda_d - \frac{\pi}{12}w^3}{\pi w^3} \\ &= \frac{1}{4}\left(\eta_d - \frac{1}{3}\right), \end{aligned} \tag{5.40}$$

so

$$D_b = \frac{3\eta_d - 1}{3\eta_d + 1} D_d. \tag{5.41}$$

A typical *S. pombe* aspect ratio at division is $\eta_d \approx 4$, giving $D_b = {}^{11}/_{13} D_d \approx 0.85 D_d$.

Let the transition between elongation and septation phases occur at time $0 < \tau < T_d$. At the septation time, the dry mass density is given by

$$\begin{aligned} D(\tau) &= \frac{M(\tau)}{V(\tau)} = \frac{M_b e^{\mu_M \tau}}{V_d} = \\ &= \frac{M_b}{V_d} \exp\left(\frac{\ln 2}{T_d}\tau\right) = \\ &= \frac{\frac{1}{2}M_d}{V_d} 2^{\frac{\tau}{T_d}} = \\ &= D_d 2^{\frac{\tau}{T_d} - 1}. \end{aligned} \tag{5.42}$$

The transition time τ can be written in terms of the aspect ratio if it is furthermore assumed that surface and biomass production are exactly proportional. In the following, let $\mu_L = \mu_M = \mu$, and septum construction continuously follows the

construction of the cell envelope, i.e. the amount of septum is given as a function of time as

$$\begin{aligned} S(t) &= \Sigma_b \exp(\mu t) - \Sigma_d = && (5.43)\\ &= \Sigma_d \left(e^{\mu(t-\tau)} - 1\right) && (5.44) \end{aligned}$$

during the division phase (and $S = 0$ during the elongation phase). If division is initiated when the septum can be converted into two hemispherical end caps,

$$S(T_d) = \pi w^2. \tag{5.45}$$

Then,

$$\begin{aligned} \pi w^2 &= \Sigma_d \left(e^{\mu(T_d-\tau)} - 1\right) \\ \frac{\pi w^2}{\Sigma_d} &= e^{\mu T_d} e^{-\mu\tau} - 1 \\ \frac{1}{\eta_d} + 1 &= 2\exp\left(-\frac{\ln 2}{T_d}\tau\right) \\ \frac{\eta_d + 1}{\eta_d} &= 2^{1-\frac{\tau}{T_d}} \\ 2^{\frac{\tau}{T_d}-1} &= \frac{\eta_d}{\eta_d + 1}, && (5.46) \end{aligned}$$

and therefore

$$D(\tau) = \frac{\eta_d}{\eta_d + 1} D_d. \tag{5.47}$$

For $\eta_d \approx 4$, $D(\tau) = 0.75 D_d$.

5.3.3 The full geometry-aware model recovered the main results from the allocation-only model

Next, I explored the behaviour of the model under modulations of the carbon uptake efficiency k_C and of the nitrogen uptake efficiency k_N. For this, I extrapolated from the parametrisation for EMM, keeping f_Y constant. Initially, I kept the width constant to $w = 3.4$ µm and left the surface stoichiometry parameter σ part of the growth-maximisation. The behaviour of this model is plotted as a function of the growth rate in Figure 5.6. The main behaviours of the allocation-only model from chapter 4 were reproduced, in particular the Monod laws (Figure 5.6AB) and the ribosomal growth law (Figure 5.6C). The surface stoichiometry decreased even more

strongly with decreasing growth rate than in the model parametrised by the best fit to the data (Figure 5.6D). The metabolite mass fractions remained small, with the largest metabolite storage happening in the fastest conditions (Figure 5.6E).

5.3.4 The observed cell geometry did not maximise the growth rate in the model

However, the main difference in cell geometry between the carbon- and nitrogen-limited cultures was the difference in width. I therefore explored an alternative strategy for maximising growth, namely optimising the width w jointly with the ribosomal allocation fraction $\vec{f}$ and keeping the surface stoichiometry fixed at its EM value of $\sigma = 0.1771$. This is shown in Figure 5.7. Overall, the results were similar to the parametrisation with constant width, with one important difference: at carbon and nitrogen uptake efficiencies that resulted in similar optimal growth rates, the optimal cell width was considerably smaller in the nitrogen-limited cultures than in the carbon-limited ones (Figure 5.7D). The direction of this difference was in agreement with the observed difference in cell width in the nitrogen- and carbon-limited cultures. However, the growth-maximising width was considerably smaller in the model than in the experiments. This suggests that the cell width is influenced by factors other than growth maximisation, but that the cell may adjust its width in the direction of the pressure to maximise the growth in the nitrogen-limited cultures, when it is most beneficial to do so.

Further observations as a function of the nitrogen- and carbon-limited growth rates of both growth-maximisation strategies are plotted in Figure 5.8. The effect of ribosomal crowding was small but appreciable, reducing the ribosomal flux between approximately 7% and 20% for most parametrisations (Figure 5.8A.) An important difference between the fixed-width and fixed-surface-stoichiometry strategies is shown in Figure 5.8B. For the former, the fraction of time spend in mitosis (nuclear division and septation phase, as opposed to the "G2" or elongation phase) approximated 1 at slow growth. In other words, for the fixed-width strategy, small cells were spherical. This was also apparent from the cell size and shape variables (Figure 5.8C–F). In contrast, for fixed stoichiometry and variable width, the division length of the cell was almost constant across carbon- and nitrogen perturbations (Figure 5.8C), but as the width and growth rate decreased together with increasing limitation (Figure 5.7D), the fraction of time spent in mitosis decreased to approximately zero (Figure 5.8B), and the surface-to-volume ra-

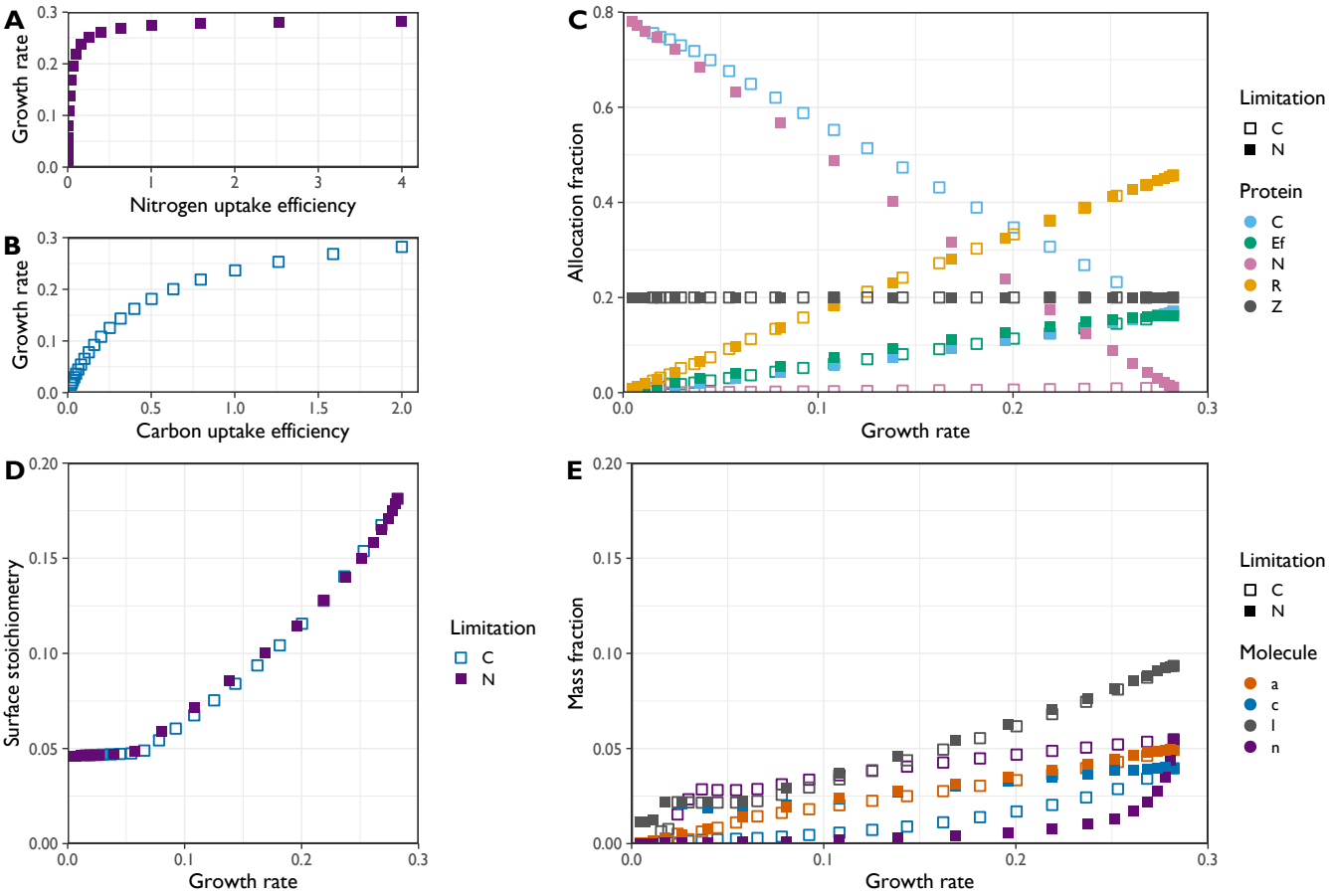


Figure 5.6: Behaviour of fixed-width parametrisation as a function of growth rate under perturbation of the nitrogen or carbon uptake efficiency. **A.–B.** Maximal growth rate as a function of the modulated nitrogen uptake efficiencies k_N while the latter was modulated. **B.** As A. for the carbon uptake efficiency k_C. **C.** Ribosomal allocation fractions $\vec{f}$, **D.** surface stoichiometry σ that together maximised the growth rate. **E.** Abundances of metabolites and lipids in the growth-rate-maximising parameterization as a fraction of total biomass directly preceding cell division.

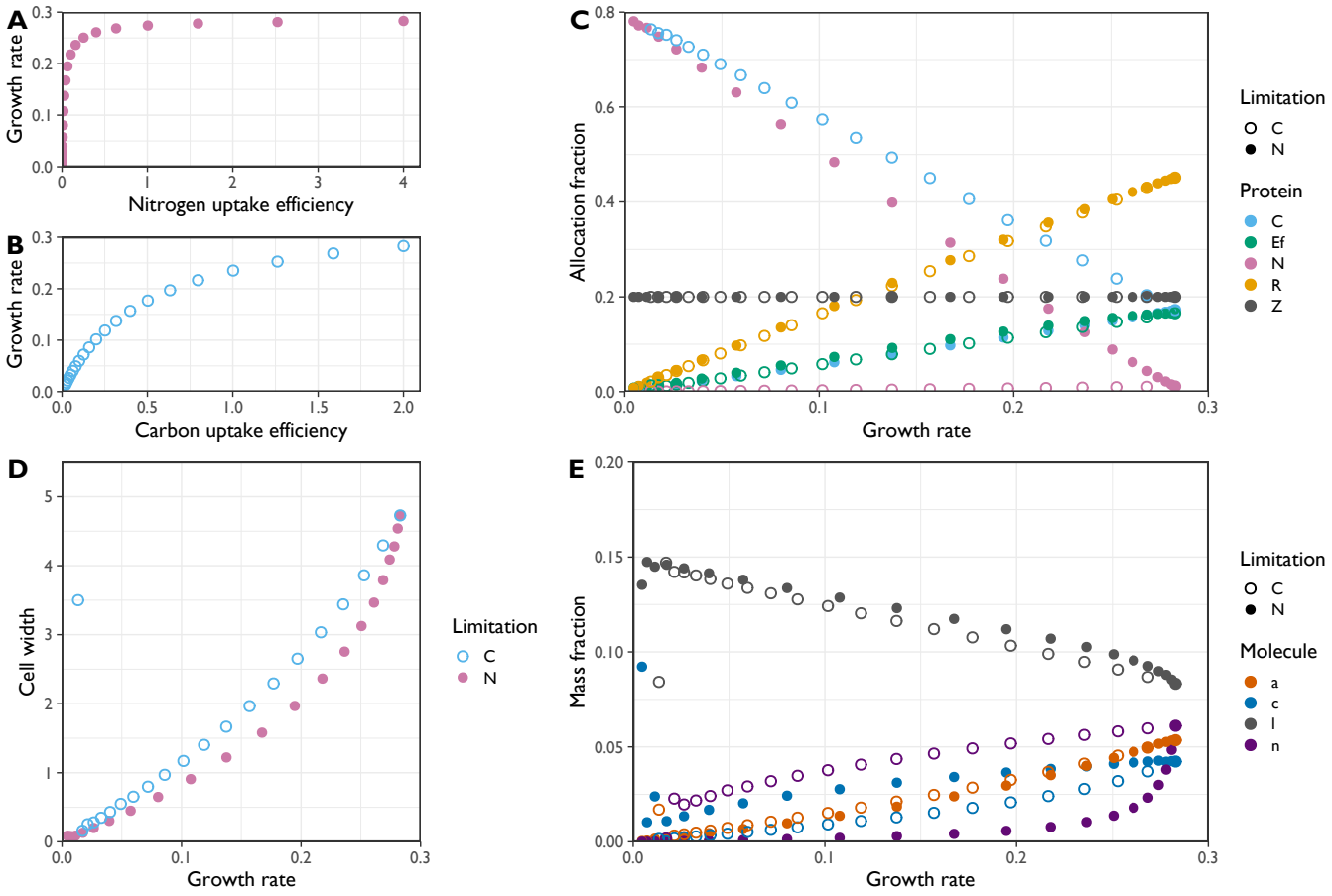


Figure 5.7: Behaviour of fixed-surface-stoichiometry parametrisation as a function of growth rate under perturbation of the nitrogen or carbon uptake efficiency. **A.–B.** Maximal growth rate as a function of the modulated nitrogen uptake efficiencies k_N while the latter was modulated. **B.** As A. for the carbon uptake efficiency k_C. **C.** Ribosomal allocation fractions $\vec{f}$, **D.** cell width w that together maximised the growth rate. **E.** Abundances of metabolites and lipids in the growth-rate-maximising parameterization as a fraction of total biomass directly preceding cell division.

tio increased several-fold (Figure 5.8F). This again suggests that neither is a good approximation for the experimentally observed behaviour, where these variables stayed in more moderate ranges.

5.4 Discussion

While I made good progress in constructing a model that approximated key observations about *S. pombe* physiology, the results described in this chapter should be seen as a work-in-progress. Here I outline three main shortcomings of the model, as well as possible modifications to the model that may provide improvements. I end this section by discussing alternative assumptions for coupling cell shape, surface production, and biosynthesis that were made in other models but that I discarded.

5.4.1 Varying allocation towards cell cycle proteins

In Table 5.1, it was seen that the division protein allocation parameter f_Y depended on the environment. However, in 5.3.3 and 5.3.4 it was held constant when exploring the models with optimal width or surface stoichiometry. The effect of increases in f_Y would be a more rapid build-up of the division protein Y, such that division would occur sooner. The parameter thereby affects the cell size at division. In the bacterial model from which I started my efforts, Bertaux et al. (2020) showed that a scaling of f_Y with enzymatic abundances appropriately explained cell sizes across a range of limitations, including translational inhibition and the expression of metabolically useless proteins in addition to nutrient limitation. However, in their model, only a single enzyme was present, which I replaced by three different ones. It is not immediately obvious from Table 5.1 how to proceed: there is no single enzyme in my current model whose allocation is proportional to that of the division protein across all three modelled conditions. Some progress might be made by including the effect of translational inhibition and/or the expression of metabolically useless proteins in the model. Like nutrient modulation, both of these processes reduce the growth rate, but each have different effects on the cell size and shape.

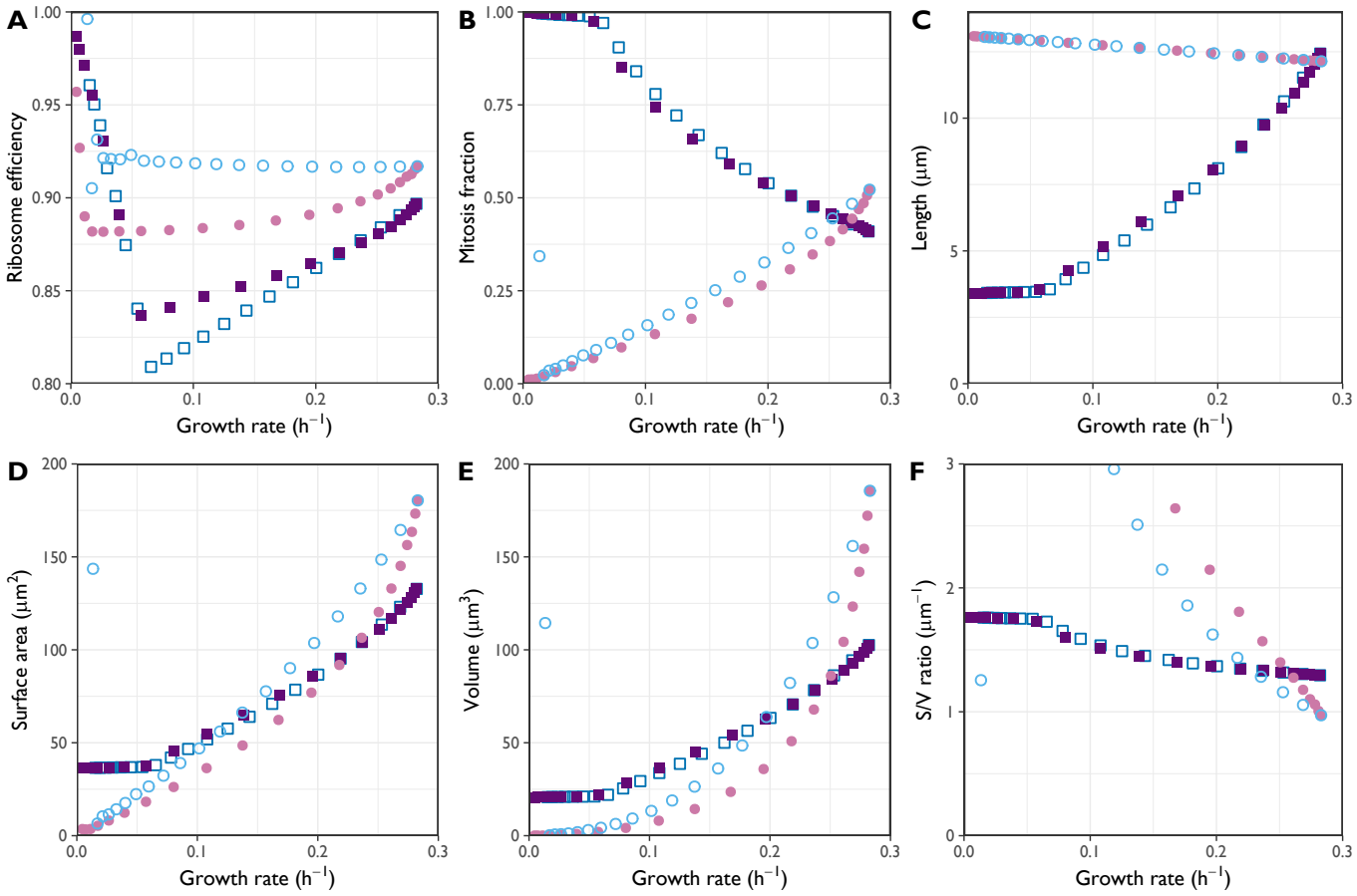


Figure 5.8: Further behaviour of fixed-width and fixed-surface-stoichiometry growth-maximisation strategies as a function of growth rate. Colours and shapes are as in Figures 5.6 and 5.7: purple and dark blue squares denote the fixed-width strategy and pink and light blue circles denote the fixed-surface-stoichiometry strategy; open symbols denote carbon limitation and filled symbols denote nitrogen limitation. Plotted are **A.** the relative efficiency of the ribosomes compared to a non-crowded model (see $g_c(\omega)$ in equation (5.18)), **B.** the fraction of time each cell spent in the nuclear division and septation phases, **C.** the cell length, **D.** the surface area, **E.** the volume of the cell, and **F.** the surface-to-volume ratio. For A. and C.–F. the value was taken at the moment immediately preceding cell division.

5.4.2 Crowding of non-ribosomes

A major shortcoming of the model are the observed large difference in dry mass density between EMM and nutrient-limited cultures (compare the axes on Figures 5.3F, 5.4F, and 5.5F). It is generally thought that the dry mass density does not vary much between conditions (van den Berg et al. 2017). This discrepancy is likely due to the crowding mechanism: in the model, the ribosomal crowding does not affect the total dry mass density, only the ribosomal density itself. In conditions supporting slow growth, the ribosomes are less abundant following the ribosomal growth law. However, it is not obvious how the crowding of small metabolites should be modelled. Analogous to how one can fit an entire bag of rice in a pot already filled with potatoes, small metabolites are likely to find accessible volume even when macromolecules such as ribosomes are strongly affected by crowding.

5.4.3 Septum–new end transition and changes in width

In 5.3.1 it was noted that the dry mass density fluctuations in the model were considerably larger than observed experimentally by Odermatt et al. (2021). In 5.3.2 I showed that these large dry mass density fluctuations were reproduced in a simpler model too. Related to this, Odermatt et al. (2021) also showed that the septum was constructed within approximately 30–40 minutes, whereas this took longer in our model, especially in carbon-limited parametrisations. Here a key assumption is the constant surface density of the surface area, including the new-end hemispheres, as well as the assumption that all the mass of the septum is transformed into the cell wall at the new end. In contrast, a mechanistic model of the cell wall, such as presented by Abenza et al. (2015), could implement interactions between the cell wall expansion and its biophysical properties. Locally, cell wall expansion decreases the cell wall thickness and a thicker wall inhibits growth (Davì et al. 2018). Following cytokinesis, there may therefore be a need for the new end to be reshaped and strengthened prior to new-end take off. This would shorten the time required to construct the septum, because the cell wall would be thinner at the new end.

Such a mechanistic model of the septum–new end transition may provide a further benefit, as then the assumption that the cell width is constant may be relaxed. Differences in balanced-growth width have been observed across conditions in both *S. pombe* and the related fission yeast *Schizosaccharomyces japonicus* (Gu

and Oliferenko 2019). The transient response of fission yeast cell width to changing environments is an open question, and a C-GRAM incorporating fluctuations in width may provide valuable insights. However, these biophysical models are a large step up in terms of model complexity and this was therefore not pursued further.

5.4.4 Alternative assumptions coupling cell shape, surface production, and biosynthesis

I stress here that the assumption of constant width is specific to *S. pombe* in steady-state nutrient environments. For *E. coli*, the aspect ratio is approximately constant across nutrient conditions, and a model implementing aspect ratio homeostasis was proposed by Ojkic et al. (2019). However, this model is incompatible with the current data on *S. pombe* which is consistent with a constant width during the cell cycle in balanced growth. In contrast, *E. coli* does indeed vary its cell width during the cell cycle in balanced growth (Harris and Theriot 2016). Another possible assumption is that the surface-to-volume ratio $\frac{\Sigma}{V}$ is held constant during the cell cycle. However, consider equation (5.13). Because the volume V increases during the cell cycle, the width w has to decrease accordingly if $\frac{\Sigma}{V}$ is to be constant. Again, this is not appropriate for a model of *S. pombe*. Because neither constant aspect ratio nor constant surface-to-volume ratio describe *S. pombe* well, I did not explore these assumptions further.

A different assumption was made to model *E. coli* growth by Oldewurtel et al. (2019), who let surface area be produced in direct proportion to the production of total dry mass. This behaviour is mirrored in our model, because the total biomass production, including that of the lipids and septum and hence surface area, is proportional to v_R. However, the model described in this chapter additionally considers that elongation halts during M phase in yeast, a complication not present in the bacterial model.

6 Discussion and conclusion

In this final chapter I provide a general summary of the main results presented in this thesis. I furthermore discuss points that connect the four chapters, as well as describe possible further developments of the broad ideas presented in this thesis.

6.1 General summary

In chapter 2, I described the extent of growth-rate-correlated and medium-specific gene expression in fission yeast grown in conditions with different nitrogen sources. Gene expression positively correlated with the growth rate was found across all classes of proteins involved in the production of proteins. This was counteracted to some degree by negative growth-rate-correlations of proteins associated with the environmental stress response. The expression of metabolic enzymes was mostly medium-specific, although the aggregate burden of metabolism on translational resource allocation was negatively correlated with the growth rate, thereby balancing the growth-rate-correlated effect on protein production.

A holistic understanding of growth, gene expression, and resource allocation can be described in coarse-grained resource allocation models (C-GRAMs), where a fundamental understanding is distilled into each component of the model. In chapter 3, I formulated a general methodology for this type of modelling, which I applied to two models of microbial growth. In chapter 4, I constructed a minimal C-GRAM of microbial metabolism that accounted for the metabolism of both carbon and nitrogen, simplifying the parametrisation by considering resource allocation that maximised the growth rate. This induced complex trade-offs in metabolism when describing growth on complex nitrogen sources. Furthermore, in chapter 5, I proposed a C-GRAM of *S. pombe* that additionally accounted for the biomass sequestered in the cell wall, as well as for the effects of molecular crowding and the existence of multiple stages in the cell cycle. This model showed how the shape and size of fission yeast adjusted in the direction of growth-rate-maximising evolutionary pressures but were not themselves optimal for growth rate. Across both

models, the ribosomal growth law was very robust, and the optimization of ribosomal resource allocation for maximal growth rate was in good agreement with our understanding of the data.

6.2 Stress–growth trade-off and optimal ribosome allocation

In chapter 2, I quantified the ribosomal growth law in fission yeast and showed it to be part of a broader trade-off between stress- and growth-related gene expression programmes. As I noted in the introductory chapter 1, this stress–growth trade-off is thought to be under the control of the TORC1 complex in *S. pombe*, whose activity is modulated by nitrogen starvation mediated by AMPK (Davie et al. 2015; Weisman 2016; González and Hall 2017; Ling et al. 2020; Morozumi and Shiozaki 2021). In the allocation model of chapter 4, regulation of the ribosomes directly proportional to the availability of free amino acids was a good approximation to growth-optimal behaviour. This raises the intriguing possibility that the TORC1-AMPK axis of regulation functions to regulate the expression of the protein production programme, and ribosomes specifically, to approximate growth rate maximisation. This would be analogous to the role of the stringent response mediated by ppGpp in bacteria (M. Zhu et al. 2019; Irving et al. 2020).

6.3 Non-protein biomass involved in translation: ribosomes and other factors

In the data analysis of chapter 2, ribosomal proteins, proteins involved in ribosome biogenesis, and other translation-related factors were all positively correlated with the growth rate, but with different offsets when extrapolated to zero growth. In other words, the normalised slope of the growth law differed between these three categories. However, in the C-GRAMs I did not account for this difference to keep the model minimal. Instead, I accounted for all proteins involved in translation as the single R-sector and referred to them as ribosomes. Here I provide more detail and compare the observed proteome mass fractions to the allocation fractions in the *S. pombe* C-GRAM of chapter 5.

An initial estimate of the total R-sector allocation fraction comes directly from the observed proteome mass fractions: $\phi_{\mathrm{RP}} \approx 0.17$, $\phi_{\mathrm{RiBi}} \approx 0.02$, and $\phi_{\mathrm{IET}} \approx 0.07$ in EMM, for the non-overlapping ribosomal protein (RP), ribosome biogenesis (RiBi), and translation initiation, elongation, and termination (IET) functional

classes plotted in Figure 2.18A. However, ribosomes consist of several large RNA molecules in addition to many small ribosomal proteins—in a fashion highly optimised for autocatalysis (Reuveni et al. 2017). Without introducing RNA to the models (which has a different C-to-N stoichiometry), the rRNA can be accounted for as additional ribosomal proteins, which in eukaryotes account for approximately 45% of the total ribosome mass (Reuveni et al. 2017). Approximating slightly, $m_{rRNA} \approx m_{RP}$, which importantly must also be accounted for in the total "protein" biomass, which now includes rRNA and all proteins. Therefore, a better estimate for the R-fraction in the model based on the observed proteome mass fractions for ϕ_{RP}, ϕ_{RiBi}, and ϕ_{IET} is then

$$\begin{aligned}\hat{\phi}_R &= \frac{m_{rRNA} + m_{RP} + m_{RiBi} + m_{IET}}{m_{all\,protein} + m_{rRNA}} \approx \\ &\approx \frac{2m_{RP} + m_{RiBi} + m_{IET}}{m_{all\,protein} + m_{RP}} = \\ &= \frac{2\phi_{RP} + \phi_{RiBi} + \phi_{IET}}{1 + \phi_{RP}}. \end{aligned} \tag{6.1}$$

At fast growth with rate $\mu \approx 0.28\,h^{-1}$, corresponding to EMM in the experiments, this equals $\hat{\phi}_R \approx \frac{2\times0.17+0.07+0.02}{1+0.17} \approx 0.37$; at intermediate growth with $\mu \approx 0.14\,h^{-1}$, $\hat{\phi}_R \approx \frac{2\times0.11+0.055+0.015}{1+0.11} \approx 0.26$; extrapolating to zero growth, it gives an intercept of about $\hat{\phi}_{R,0} \approx \frac{2\times0.06+0.04+0.01}{1+0.06} \approx 0.15$ and a fold-change of around $\frac{0.37-0.15}{0.26} \approx 0.85$.

In the model (Table 5.1), the ribosomal protein sector accounts for approximately $f_R \approx 0.46$ in EMM, which is perhaps a slight overestimate. Assuming the ribosomal growth law is quantitatively the same between the nitrogen-source modulations of chapter 2 and the carbon- and nitrogen titrations modelled in chapter 5, the estimate at intermediate growth is closer at approximately $f_R \approx 0.24$. I expect that the allocation parameters could be tweaked somewhat by tweaking the baseline rates k_C and k_N for the carbon and nitrogen uptake enzymes in the model. However, I did not explicitly include the ribosome fraction in the parameter fitting used to obtain Table 5.1, because the estimate obtained in the above is somewhat uncertain due to the necessary rRNA assumptions, as well as the generally underestimated intercept already in the allocation model.

6.4 Protein reserves and underutilised enzymes

The assumption that expression of all proteins is optimised for growth in any given condition considerably simplifies the modelling. It is not obvious how the alternative of explicitly regulated allocation would be implemented and in any case it would require many extra parameters. However, recent evidence has challenged the view that all allocation is growth-optimal. It is thought that significant fractions of the *E. coli* (Valgepea et al. 2013; Peebo et al. 2015; Mori et al. 2017) and budding yeast (Metzl-Raz et al. 2017; R. Yu et al. 2020) proteome are not immediately required for sustaining the growth rate and are instead held in reserve. This reserve pool of protein could support cell adaption to sudden environmental changes. It has furthermore been suggested that central carbon metabolism has a large reserve capacity, suggesting that many enzymes may also not be utilised solely to maximise metabolic fluxes (O'Brien et al. 2016; Christodoulou et al. 2018; R. Yu et al. 2021).

These considerations are particularly important for the translational proteins. A reserve pool of ribosomes has been proposed to account for the y-intercept in the ribosomal growth law (Metzl-Raz et al. 2017; Mori et al. 2017). However, the mechanism by which the reserve pool is sequestered from the actively translating ribosome fraction is unclear and in the C-GRAMs proposed in this thesis, this effect was therefore not explicitly accounted for. Intriguingly, as described in section 4.3.1, part of the nonzero y-intercept could be explained by a combination of the nonlinearity in the dependence of ribosomal flux on its amino acid substrate and proportional regulation of the ribosomal allocation with the substrate. However, the observed intercept is larger than the model offset explained this way. Therefore, while coarse-grained modelling points to a partial explanation of the effect, further experiments are required to clarify the mechanism behind underutilised ribosomes. In light of the quantitatively different growth laws for ribosomal proteins, ribosome biogenesis proteins, and other translation factors (section 2.3.7), broadening such a search to include translational proteins beyond ribosomes could be fruitful.

Considering the degree of utilisation is also important for metabolic enzymes. Some nutrient-specific regulatory programmes were detected in both the transcriptome and the proteome of chapter 2, in particular responses specific to serine and tryptophan media. Most of these enzymes were present in all conditions to some degree, suggesting that they were under- or unutilised in most. Unlike the

Ser- and Trp-responses, the broader WFSP pattern was only present in the proteome, not in the transcriptome. A better understanding of post-transcriptional regulation in fission yeast is therefore required to fully understand the relationship between the high translational burden of metabolism and the extent to which that burden is required to sustain fluxes.

6.5 Medium-specific expression and the Q-sector

One feature of the allocation model in chapter 2 was a smaller Q-sector of housekeeping proteins than in previous models. This was due to basal expression of other enzymes. The mechanism that ensures the constant expression of the Q-sector is unclear; in the models their allocation f_Q was held fixed as including it in any growth-maximisation strategy would quickly generate $f_Q = 0$ as the optimal solution, since the Q-proteins definitionally do not contribute to metabolism and growth.

Likewise I found that there were few proteins in the *S. pombe* data set of chapter 2 whose expression was really constant across all conditions. Instead, many proteins without a clear growth-rate-correlated component to their expression exhibited medium-specific expression. This was unexpected as previous experiments pointed to growth rate correlations as the main driving force between resource allocation trade-offs. Still, relatively minimal changes in the growth medium (swapping out one saturating nitrogen source for another) resulted in large shifts in the resource allocation burden of metabolism. I repeat here that the aggregate burden of all metabolic proteins was negatively correlated with the growth rate. Importantly, this was also the case for the net burden of proteins not showing significant individual growth rate correlations.

Together, these points supports the hypothesis that the true Q-sector is indeed smaller than previously thought. Rather, many proteins can be expressed in a medium-dependent manner, thereby occupying a significant fraction of the proteome at the expense of other proteins more directly required for biomass production. In this view, illustrated in Figure 6.1, the medium-specific proteins function complementary to the original P-sector and I propose the name M-sector for them. Most of proteome is then a trade-off between the R-sector (biomass production), M-sector (metabolism), and P-sector (stress), with only a small ~20% allocation to a Q-sector that is somehow regulated to be expressed at a constant proportion of the proteome.

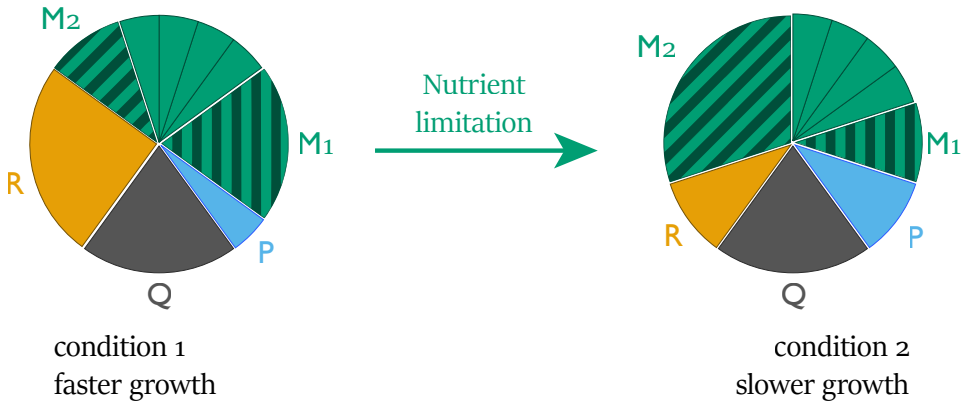


Figure 6.1: Proteome pie chart incorporating the trade-off between growth-rate-correlated and medium-specific expression. The green slices of the pie represent the parts of the proteome that are induced in a medium-specific manner; the orange slice is the R-sector (positively correlated with growth rate) and the blue slice the P-sector (negatively correlated with growth rate). The growth condition that induces the vertically hatched slice M1 supports faster growth than the condition inducing the diagonally hatched M2.

6.6 The difficulty of mapping real proteins to C-GRAM sectors

As touched upon in the discussion about protein reserves, it can be difficult to match observed proteome fractions to the burden of proteome sectors in the coarse-grained models. An initial thought-provoking example is provided by the possibility of inactive translational proteins. Intriguingly, sequestered ribosomes may be thought of as a component of the "housekeeping" sector if they do not influence metabolism in any way. If this thought is followed, the y-intercept of the ribosomal growth law can never be directly explained from the properties of the model ribosomes, as it should be implemented as a proportion of the housekeeping proteins.

The complications for assigning metabolic enzymes to C-GRAM sectors are even larger, because of the complexity inherent in the metabolic network. For example, TCA cycle intermediates may be thought of as prime candidates representing the carbon metabolite C, since these molecules are required for energy generation, amino acid production, and the production of lipids. This view is consistent with including the glycolytic pathway as the carbon-generating enzyme E_C. However,

several glycolytic intermediates are required as part of amino acid synthesis as well, so enzymes constructing them may equally be included in the amino-acid metabolism enzymes E_A. The largest problem comes with the uptake of nitrogen in the form of amino acids: some of those pathways are thought to be the reverse of the amino acid synthesis pathways, and regardless of the nitrogen assimilation pathway, the recycling of the carbon backbones can be part of glycolysis, the TCA cycle, or the pentose phosphate pathway.

Because of these difficulties mapping proteins to coarse-grained sectors, I did not directly use the observed proteome fractions from chapter 2 to inform the allocation model of chapter 4 and beyond. However, I do not think that this is necessarily a weakness. The coarse-grainedness of the modelling approach presented in this thesis implies that only few model parameters directly correspond to biological variables. Rather, the components of the model represent broad concepts, such as "nutrient quality", that are implicitly defined by the resulting model behaviours. The analysis of coarse-grained models provides a way to generate hypotheses and quantify the interplay between the model components even when a direct match between model variable and observations may be lacking. However, it can sometimes be tempting to draw unwarranted conclusions and the onus is on the modeller to keep this in mind.

6.7 Outlook

In spite of the pithy assertion that all models are wrong but some are useful, I believe that the models presented in this thesis may be adapted to become more 'illuminating and useful' (Box 1979) by reducing their wrongness. To start with, the model from chapter 5 could be improved by a better understanding of molecular crowding beyond the effect on ribosomes only. A further improvement may be to explicitly implement mechanisms that are responsible for the shape of the cell wall, particularly at the growth regions in the cell tips. Additionally, the coarse-grained metabolism could be expanded to account for a cellular pool of chemical energy in the form of ATP. A further expansion to metabolism would be accounting differentially for biomass in nucleotides and in proteins, considering that ribosomes are primarily composed of RNA, which has a higher nitrogen content than protein.

A broader use for C-GRAMs such as I presented in this thesis may be to form the basis of a stochastic or multi-scale model that provides a more detailed descrip-

tion of a particular biological problem. For example, the coarse-grained model of bacterial physiology by Weiße et al. (2015) was successfully adapted to provide a stochastic view of growth dynamics in single cells (Thomas et al. 2018). Consistent with the view of biotic systems as processors of information (Hogeweg 2011), biological processes are interlinked across spatio-temporal scales and multi-scale models require efficient linking of the models describing each level of detail and a deep understanding of each individual level (Cilfone et al. 2015). As a specific example, consider that microbes heavily share resources in their extracellular metabolome (J. S. L. Yu et al. 2022). A suitably adapted C-GRAM based on the model from chapter 5 may be incorporated into a broader ecological model to aid our understanding of the interplay between resource allocation, the optimality of growth, and the sharing of metabolites.

6.8 Conclusion

In this thesis I have presented how global resource allocation in fission yeast can be primarily divided into growth-rate-correlated and medium-specific expression. I have further shown how coarse-grained mathematical modelling provides an improved understanding of the interplay between cellular growth and the physiological and translational burden of gene expression. Further study of these topics across diverse biological and mathematical systems will influence a wide range of research areas such as microbiology, synthetic biology, and cancer research. Beyond its contribution to our understanding of gene regulation, the results and further hypotheses presented in this thesis will support future experimental and modelling efforts aimed at defining the nature of the trade-offs involved in growth, stress resistance, and metabolism across the tree of life.

Bibliography

Abenza, Juan F., Etienne Couturier, James Dodgson, Johanna Dickmann, Anatole Chessel et al. (2015). "Wall Mechanics and Exocytosis Define the Shape of Growth Domains in Fission Yeast". In: *Nature Communications* 6.1 (1), p. 8400. DOI: 10.1038/ncomms9400.

Aguilar-Uscanga, B. and J. M. François (2003). "A Study of the Yeast Cell Wall Composition and Structure in Response to Growth Conditions and Mode of Cultivation". In: *Letters in Applied Microbiology* 37.3, pp. 268–274. DOI: 10.1046/j.1472-765X.2003.01394.x.

Airoldi, Edoardo M., Curtis Huttenhower, David Gresham, Charles Lu, Amy A. Caudy et al. (2009). "Predicting Cellular Growth from Gene Expression Signatures". In: *PLoS Computational Biology* 5.1. DOI: 10.1371/journal.pcbi.1000257.

Airoldi, Edoardo M., Darach Miller, Rodoniki Athanasiadou, Nathan Brandt, Farah Abdul-Rahman et al. (2016). "Steady-State and Dynamic Gene Expression Programs in Saccharomyces Cerevisiae in Response to Variation in Environmental Nitrogen". In: *Molecular Biology of the Cell* 27.8, pp. 1383–1396. DOI: 10.1091/mbc.E14-05-1013.

Alam, Mohammad Tauqeer, Aleksej Zelezniak, Michael Mülleder, Pavel Shliaha, Roland Schwarz et al. (2016). "The Metabolic Background Is a Global Player in Saccharomyces Gene Expression Epistasis". In: *Nature Microbiology* 1, p. 15030. DOI: 10.1038/nmicrobiol.2015.30.

Alberts, Bruce, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff et al. (2015). *Molecular Biology of the Cell*. 6th ed. New York, NY: Garland Science. ISBN: 978-0-8153-4464-3.

Atilgan, Erdinc, Valentin Magidson, Alexey Khodjakov and Fred Chang (2015). "Morphogenesis of the Fission Yeast Cell through Cell Wall Expansion". In: *Current Biology* 25.16, pp. 2150–2157. DOI: 10.1016/j.cub.2015.06.059.

Atkinson, Sophie R., Samuel Marguerat, Danny A. Bitton, Maria Rodríguez-López, Charalampos Rallis et al. (2018). "Long Noncoding RNA Repertoire and Targeting by Nuclear Exosome, Cytoplasmic Exonuclease, and RNAi in Fission Yeast". In: *RNA* 24.9, pp. 1195–1213. DOI: 10.1261/rna.065524.118.

Aylett, Christopher H. S. and Nenad Ban (2017). “Eukaryotic Aspects of Translation Initiation Brought into Focus”. In: *Philosophical Transactions of the Royal Society B* 372.1716, p. 20160186. DOI: 10.1098/rstb.2016.0186.

Balakrishnan, Rohan, Matteo Mori, Igor Segota, Zhongge Zhang, Ruedi Aebersold et al. (2021). “Principles of Gene Regulation Quantitatively Connect DNA to RNA and Proteins in Bacteria”. In: *bioRxiv*, p. 2021.05.24.445329. DOI: 10.1101/2021.05.24.445329.

Barenholz, Uri, Leeat Keren, Eran Segal and Ron Milo (2016). “A Minimalistic Resource Allocation Model to Explain Ubiquitous Increase in Protein Expression with Growth Rate”. In: *PLOS ONE* 11.4, e0153344. DOI: 10.1371/journal.pone.0153344.

Basan, Markus, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen et al. (2015). “Overflow Metabolism in Escherichia Coli Results from Efficient Proteome Allocation”. In: *Nature* 528.7580, pp. 99–104. DOI: 10.1038/nature15765.

Berkhout, Jan, Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar, Dick de Ridder et al. (2013). “How Biochemical Constraints of Cellular Growth Shape Evolutionary Adaptations in Metabolism”. In: *Genetics* 194.2, pp. 505–512. DOI: 10.1534/genetics.113.150631.

Bertaux, François, Julius von Kügelgen, Samuel Marguerat and Vahid Shahrezaei (2020). “A Bacterial Size Law Revealed by a Coarse-Grained Model of Cell Physiology”. In: *PLOS Computational Biology* 16.9, e1008245. DOI: 10.1371/journal.pcbi.1008245.

Björkeroth, Johan, Kate Campbell, Carl Malina, Rosemary Yu, Francesca Di Bartolomeo et al. (2020). “Proteome Reallocation from Amino Acid Biosynthesis to Ribosomes Enables Yeast to Grow Faster in Rich Media”. In: *Proceedings of the National Academy of Sciences* 117.35, pp. 21804–21812. DOI: 10.1073/pnas.1921890117.

Box, G. E. P. (1979). “Robustness in the Strategy of Scientific Model Building”. In: *Robustness in Statistics*. Ed. by ROBERT L. LAUNER and GRAHAM N. WILKINSON. Academic Press, pp. 201–236. ISBN: 978-0-12-438150-6. DOI: 10.1016/B978-0-12-438150-6.50018-2.

Brauer, Matthew J., Curtis Huttenhower, Edoardo M. Airoldi, Rachel Rosenstein, John C. Matese et al. (2008). “Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast”. In: *Molecular Biology of the Cell* 19.1, pp. 352–367. DOI: 10.1091/mbc.E07-08-0779.

Bremer, Hans and Patrick P. Dennis (2008). “Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates”. In: *EcoSal Plus* 3.1. DOI: 10.1128/ecosal.5.2.3.

Bruggeman, Frank J, Robert Planqué, Douwe Molenaar and Bas Teusink (2020). “Searching for Principles of Microbial Physiology”. In: *FEMS Microbiology Reviews* 44.6, pp. 821–844. DOI: 10.1093/femsre/fuaa034.

Campbell, Kate, Lucia Herrera-Dominguez, Clara Correia-Melo, Aleksej Zelezniak and Markus Ralser (2018). “Biochemical Principles Enabling Metabolic Cooperativity and Phenotypic Heterogeneity at the Single Cell Level”. In: *Current Opinion in Systems Biology* 8, pp. 97–108. DOI: 10.1016/j.coisb.2017.12.001.

Carlson, C. R., B. Grallert, T. Stokke and E. Boye (1999). “Regulation of the Start of DNA Replication in Schizosaccharomyces Pombe”. In: *Journal of Cell Science* 112.6, pp. 939–946.

Chávez, Sebastián, José García-Martínez, Lidia Delgado-Ramos and José E. Pérez-Ortín (2016). “The Importance of Controlling mRNA Turnover during Cell Proliferation”. In: *Current Genetics* 62.4, pp. 701–710. DOI: 10.1007/s00294-016-0594-2.

Chen, Dongrong, W. Mark Toone, Juan Mata, Rachel Lyne, Gavin Burns et al. (2003). “Global Transcriptional Responses of Fission Yeast to Environmental Stress”. In: *Molecular Biology of the Cell* 14.1, pp. 214–229. DOI: 10.1091/mbc.E02-08-0499.

Chen, Yu, Eunice van Pelt-KleinJan, Berdien van Olst, Sieze Douwenga, Sjef Boeren et al. (2021). “Proteome Constraints Reveal Targets for Improving Microbial Fitness in Nutrient-Rich Environments”. In: *Molecular systems biology* 17.4, e10093. DOI: 10.15252/msb.202010093.

Christodoulou, Dimitris, Hannes Link, Tobias Fuhrer, Karl Kochanowski, Luca Gerosa et al. (2018). “Reserve Flux Capacity in the Pentose Phosphate Pathway Enables Escherichia Coli’s Rapid Response to Oxidative Stress”. In: *Cell Systems* 6.5, 569–578.e7. DOI: 10.1016/j.cels.2018.04.009.

Chubukov, Victor, Luca Gerosa, Karl Kochanowski and Uwe Sauer (2014). “Coordination of Microbial Metabolism”. In: *Nature Reviews Microbiology* 12.5, pp. 327–340. DOI: 10.1038/nrmicro3238.

Cilfone, Nicholas A., Denise E. Kirschner and Jennifer J. Linderman (2015). “Strategies for Efficient Numerical Implementation of Hybrid Multi-scale Agent-Based Models to Describe Biological Systems”. In: *Cellular and Molecular Bioengineering* 8.1, pp. 119–136. DOI: 10.1007/s12195-014-0363-6.

Csárdi, Gábor, Alexander Franks, David S. Choi, Edoardo M. Airoldi and D. Allan Drummond (2015). “Accounting for Experimental Noise Reveals That mRNA

Levels, Amplified by Post-Transcriptional Processes, Largely Determine Steady-State Protein Levels in Yeast". In: *PLOS Genetics* 11.5, e1005206. DOI: 10.1371/journal.pgen.1005206.

Dahal, Sanjeev, Jiao Zhao and Laurence Yang (2020). "Genome-Scale Modeling of Metabolism and Macromolecular Expression and Their Applications". In: *Biotechnology and Bioprocess Engineering* 25.6, pp. 931–943. DOI: 10.1007/s12257-020-0061-2.

Dai, Xiongfeng and Manlu Zhu (2020). "Coupling of Ribosome Synthesis and Translational Capacity with Cell Growth". In: *Trends in Biochemical Sciences* 45.8, pp. 681–692. DOI: 10.1016/j.tibs.2020.04.010.

Dai, Xiongfeng, Manlu Zhu, Mya Warren, Rohan Balakrishnan, Vadim Patsalo et al. (2016). "Reduction of Translating Ribosomes Enables *Escherichia Coli* to Maintain Elongation Rates during Slow Growth". In: *Nature Microbiology* 2.2, p. 2016231. DOI: 10.1038/nmicrobiol.2016.231.

Davì, Valeria, Hirokazu Tanimoto, Dmitry Ershov, Armin Haupt, Henry De Belly et al. (2018). "Mechanosensation Dynamically Coordinates Polar Growth and Cell Wall Assembly to Promote Cell Survival". In: *Developmental Cell* 45.2, 170–182.e7. DOI: 10.1016/j.devcel.2018.03.022.

Davie, Elizabeth, Gabriella M. A. Forte and Janni Petersen (2015). "Nitrogen Regulates AMPK to Control TORC1 Signaling". In: *Current Biology* 25.4, pp. 445–454. DOI: 10.1016/j.cub.2014.12.034.

De Groot, Daan H., Julia Lischke, Riccardo Muolo, Robert Planqué, Frank J. Bruggeman et al. (2019). "The Common Message of Constraint-Based Optimization Approaches: Overflow Metabolism Is Caused by Two Growth-Limiting Constraints". In: *Cellular and Molecular Life Sciences* 77, pp. 441–453. DOI: 10.1007/s00018-019-03380-2.

De Jong, Hidde, Stefano Casagranda, Nils Giordano, Eugenio Cinquemani, Delphine Ropers et al. (2017). "Mathematical Modelling of Microbes: Metabolism, Gene Expression and Growth". In: *Journal of The Royal Society Interface* 14.136, p. 20170502. DOI: 10.1098/rsif.2017.0502.

Dekel, Erez and Uri Alon (2005). "Optimality and Evolutionary Tuning of the Expression Level of a Protein". In: *Nature* 436.7050, nature03842. DOI: 10.1038/nature03842.

Delarue, M., G. P. Brittingham, S. Pfeffer, I. V. Surovtsev, S. Pinglay et al. (2018). “mTORC1 Controls Phase Separation and the Biophysical Properties of the Cytoplasm by Tuning Crowding”. In: *Cell* 174.2, 338–349.e20. DOI: 10.1016/j.cell.2018.05.042.

De Menezes, D. Q. F., D. M. Prata, A. R. Secchi and J. C. Pinto (2021). “A Review on Robust M-estimators for Regression Analysis”. In: *Computers & Chemical Engineering* 147, p. 107254. DOI: 10.1016/j.compchemeng.2021.107254.

De Queiroz, José Humberto, Jean-Louis Uribelarrea and Alain Pareilleux (1993). “Estimation of the Energetic Biomass Yield and Efficiency of Oxidative Phosphorylation in Cell-Recycle Cultures of Schizosaccharomyces Pombe”. In: *Applied Microbiology and Biotechnology* 39.4, pp. 609–614. DOI: 10.1007/BF00205061.

Dill, Ken A., Kingshuk Ghosh and Jeremy D. Schmit (2011). “Physical Limits of Cells and Proteomes”. In: *Proceedings of the National Academy of Sciences* 108.44, pp. 17876–17882. DOI: 10.1073/pnas.1114477108.

– (2012). “Reply to Vazquez: Optimal Density Is Robust to Variations in Calculation”. In: *Proceedings of the National Academy of Sciences* 109.9, E534–E534. DOI: 10.1073/pnas.1121337109.

Doolittle, Arthur K. (1952). “Studies in Newtonian Flow. III. The Dependence of the Viscosity of Liquids on Molecular Weight and Free Space (in Homologous Series)”. In: *Journal of Applied Physics* 23.2, pp. 236–239. DOI: 10.1063/1.1702182.

Dourado, Hugo and Martin J. Lercher (2020). “An Analytical Theory of Balanced Cellular Growth”. In: *Nature Communications* 11.1 (1), pp. 1–14. DOI: 10.1038/s41467-020-14751-w.

Ertugay, N. and H. Hamamci (1997). “Continuous Cultivation of Bakers’ Yeast: Change in Cell Composition at Different Dilution Rates and Effect of Heat Stress on Trehalose Level”. In: *Folia Microbiologica* 42.5, pp. 463–467. DOI: 10.1007/BF02826554.

Facchetti, Giuseppe, Benjamin Knapp, Ignacio Flor-Parra, Fred Chang and Martin Howard (2019). “Reprogramming Cdr2-Dependent Geometry-Based Cell Size Control in Fission Yeast”. In: *Current Biology* 29.2, 350–358.e4. DOI: 10.1016/j.cub.2018.12.017.

Fantes, P. and P. Nurse (1977). “Control of Cell Size at Division in Fission Yeast by a Growth-Modulated Size Control over Nuclear Division”. In: *Experimental Cell Research* 107.2, pp. 377–386. DOI: 10.1016/0014-4827(77)90359-7.

Feldt, Robert and Alexey Stukalov (2019). *BlackBoxOptim.Jl*. Version 0.5.0.

Franks, Alexander, Edoardo Airoldi and Nikolai Slavov (2017). “Post-Transcriptional Regulation across Human Tissues”. In: *PLoS computational biology* 13.5, e1005535. DOI: 10.1371/journal.pcbi.1005535.

Friedmann, Herbert Claus (2004). “From Butyribacterium to E. Coli : An Essay on Unity in Biochemistry”. In: *Perspectives in Biology and Medicine* 47.1, pp. 47–66. DOI: 10.1353/pbm.2004.0007.

Gao, Fuchang and Lixing Han (2012). “Implementing the Nelder-Mead Simplex Algorithm with Adaptive Parameters”. In: *Computational Optimization and Applications* 51.1, pp. 259–277. DOI: 10.1007/s10589-010-9329-3.

Gene Ontology Consortium, The (2019). “The Gene Ontology Resource: 20 Years and Still GOing Strong”. In: *Nucleic Acids Research* 47.D1, pp. D330–D338. DOI: 10.1093/nar/gky1055.

Godard, Patrice, Antonio Urrestarazu, Stéphan Vissers, Kevin Kontos, Gianluca Bontempi et al. (2007). “Effect of 21 Different Nitrogen Sources on Global Gene Expression in the Yeast Saccharomyces Cerevisiae”. In: *Molecular and Cellular Biology* 27.8, pp. 3065–3086. DOI: 10.1128/MCB.01084-06.

Goelzer, Anne and Vincent Fromion (2017). “Resource Allocation in Living Organisms”. In: *Biochemical Society Transactions* 45.4, pp. 945–952. DOI: 10.1042/BST20160436.

Goelzer, Anne, Jan Muntel, Victor Chubukov, Matthieu Jules, Eric Prestel et al. (2015). “Quantitative Prediction of Genome-Wide Resource Allocation in Bacteria”. In: *Metabolic Engineering* 32 (Supplement C), pp. 232–243. DOI: 10.1016/j.ymben.2015.10.003.

González, Asier and Michael N. Hall (2017). “Nutrient Sensing and TOR Signaling in Yeast and Mammals”. In: *The EMBO Journal* 36.4, pp. 397–408. DOI: 10.15252/embj.201696010.

Gu, Ying and Snezhana Oliferenko (2019). “Cellular Geometry Scaling Ensures Robust Division Site Positioning”. In: *Nature Communications* 10.1 (1), p. 268. DOI: 10.1038/s41467-018-08218-2.

Harris, Leigh K. and Julie A. Theriot (2016). “Relative Rates of Surface and Volume Synthesis Set Bacterial Cell Size”. In: *Cell* 165.6, pp. 1479–1492. DOI: 10.1016/j.cell.2016.05.045.

Hartl, Johannes, Patrick Kiefer, Fabian Meyer and Julia A. Vorholt (2017). “Longevity of Major Coenzymes Allows Minimal de Novo Synthesis in Microorganisms”. In: *Nature Microbiology* 2.7 (7), pp. 1–9. DOI: 10.1038/nmicrobiol.2017.73.

Heinrich, Reinhart and Stefan Schuster (1996). *The Regulation of Cellular Systems*. New York, NY: Chapman & Hall. 372 pp.

Hogeweg, Paulien (2011). "The Roots of Bioinformatics in Theoretical Biology". In: *PLOS Computational Biology* 7.3, e1002021. DOI: 10.1371/journal.pcbi.1002021.

Hu, Xiao-Pan, Hugo Dourado, Peter Schubert and Martin J. Lercher (2020). "The Protein Translation Machinery Is Expressed for Maximal Efficiency in Escherichia Coli". In: *Nature Communications* 11.1 (1), p. 5260. DOI: 10.1038/s41467-020-18948-x.

Huber, W., V. J. Carey, R. Gentleman, S. Anders, M. Carlson et al. (2015). "Orchestrating High-Throughput Genomic Analysis with Bioconductor". In: *Nature Methods* 12.2, pp. 115–121.

Hui, Sheng, Josh M. Silverman, Stephen S. Chen, David W. Erickson, Markus Basan et al. (2015). "Quantitative Proteomic Analysis Reveals a Simple Strategy of Global Resource Allocation in Bacteria". In: *Molecular Systems Biology* 11, p. 784.

Irving, Sophie E., Naznin R. Choudhury and Rebecca M. Corrigan (2020). "The Stringent Response and Physiological Roles of (Pp)pGpp in Bacteria". In: *Nature Reviews. Microbiology* 19, pp. 256–271. DOI: 10.1038/s41579-020-00470-y.

Jeffares, Daniel C., Charalampos Rallis, Adrien Rieux, Doug Speed, Martin Převorovský et al. (2015). "The Genomic and Phenotypic Diversity of Schizosaccharomyces Pombe". In: *Nature Genetics* 47.3, pp. 235–241. DOI: 10.1038/ng.3215.

Jun, Suckjoon, Fangwei Si, Rami Pugatch and Matthew Scott (2018). "Fundamental Principles in Bacterial Physiology—History, Recent Progress, and the Future with Focus on Cell Size Control: A Review". In: *Reports on Progress in Physics* 81.5, p. 056601. DOI: 10.1088/1361-6633/aaa628.

Kamrad, Stephan, Jan Grossbach, Maria Rodríguez-López, Michael Mülleder, StJohn Townsend et al. (2020). "Pyruvate Kinase Variant of Fission Yeast Tunes Carbon Metabolism, Cell Regulation, Growth and Stress Resistance". In: *Molecular Systems Biology* 16.4, e9270. DOI: 10.15252/msb.20199270.

Kanehisa, Minoru, Miho Furumichi, Mao Tanabe, Yoko Sato and Kanae Morishima (2017). "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs". In: *Nucleic Acids Research* 45.D1, pp. D353–D361. DOI: 10.1093/nar/gkw1092.

Kanou, Akihiko, Shinichi Nishimura, Toshitsugu Tabuchi, Akihisa Matsuyama, Minoru Yoshida et al. (2020). "Serine Catabolism Produces ROS, Sensitizes Cells to Actin Dysfunction, and Suppresses Cell Growth in Fission Yeast". In: *The Journal of Antibiotics* 73.8 (8), pp. 574–580. DOI: 10.1038/s41429-020-0305-6.

Keren, L., O. Zackay, M. Lotan-Pompan, U. Barenholz, E. Dekel et al. (2013). "Promoters Maintain Their Relative Activity Levels under Different Growth Conditions". In: *Molecular Systems Biology* 9.1, pp. 701–701. DOI: 10.1038/msb.2013.59.

Keseler, Ingrid M., Socorro Gama-Castro, Amanda Mackie, Richard Billington, César Bonavides-Martínez et al. (2021). "The EcoCyc Database in 2021". In: *Frontiers in Microbiology* 12, p. 711077. DOI: 10.3389/fmicb.2021.711077.

Kleijn, Istvan T., Amalia Martínez-Segura, François Bertaux, Malika Saint, Holger Kramer et al. (2022). "Growth-Rate-Dependent and Nutrient-Specific Gene Expression Resource Allocation in Fission Yeast". In: *Life Science Alliance* 5.5. DOI: 10.26508/lsa.202101223.

Klumpp, Stefan and Terence Hwa (2014). "Bacterial Growth: Global Effects on Gene Expression, Growth Feedback and Proteome Partition". In: *Current Opinion in Biotechnology*. Nanobiotechnology • Systems Biology 28, pp. 96–102. DOI: 10.1016/j.copbio.2014.01.001.

Knapp, Benjamin D., Pascal Odermatt, Enrique R. Rojas, Wenpeng Cheng, Xiangwei He et al. (2019). "Decoupling of Rates of Protein Synthesis from Cell Expansion Leads to Supergrowth". In: *Cell Systems* 9.5, 434–445.e6. DOI: 10.1016/j.cels.2019.10.001.

Koch, Arthur L. (1988). "Why Can't a Cell Grow Infinitely Fast?" In: *Canadian Journal of Microbiology* 34.4, pp. 421–426. DOI: 10.1139/m88-074.

Kogane, Fusae and Tomomichi Yanagita (1962). "Qualitative and Quantitative Studies on the Cell Wall Carbohydrate in the Yeast, Schizosaccharomyces Pombe." In: *Chemical & Pharmaceutical Bulletin* 10.1, pp. 61–66. DOI: 10.1248/cpb.10.61.

Kohanim, Yael Korem, Dikla Levi, Ghil Jona, Benjamin D. Towbin, Anat Bren et al. (2018). "A Bacterial Growth Law out of Steady State". In: *Cell Reports* 23.10, pp. 2891–2900. DOI: 10.1016/j.celrep.2018.05.007.

Kolda, Tamara G., Robert Michael Lewis and Virginia Torczon (2003). "Optimization by Direct Search: New Perspectives on Some Classical and Modern Methods". In: *SIAM Review* 45.3, pp. 385–482. DOI: 10.1137/S003614450242889.

Komsta, Lukasz (2019). *Mblm: Median-Based Linear Models*. Version 0.12.1.

Korotkevich, Gennady, Vladimir Sukhov and Alexey Sergushichev (2019). "Fast Gene Set Enrichment Analysis". In: *bioRxiv : the preprint server for biology*. DOI: 10.1101/060012.

Laboucarié, Thomas, Dylane Detilleux, Ricard A. Rodriguez-Mias, Céline Faux, Yves Romeo et al. (2017). "TORC1 and TORC2 Converge to Regulate the SAGA Co-activator in Response to Nutrient Availability". In: *EMBO reports* 18.12, pp. 2197–2218. DOI: 10.15252/embr.201744942.

Liao, Chen, Andrew E. Blanchard and Ting Lu (2017). "An Integrative Circuit–Host Modelling Framework for Predicting Synthetic Gene Network Behaviours". In: *Nature Microbiology* 2, pp. 1658–1666. DOI: 10.1038/s41564-017-0022-5.

Ling, Naomi X. Y., Adrian Kaczmarek, Ashfaqul Hoque, Elizabeth Davie, Kevin R. W. Ngoei et al. (2020). "mTORC1 Directly Inhibits AMPK to Promote Cell Proliferation under Nutrient Stress". In: *Nature Metabolism* 2.1, pp. 41–49. DOI: 10.1038/s42255-019-0157-1.

Lock, Antonia, Kim Rutherford, Midori A. Harris, Jacqueline Hayles, Stephen G. Oliver et al. (2019). "PomBase 2018: User-Driven Reimplementation of the Fission Yeast Database Provides Rapid and Intuitive Access to Diverse, Interconnected Information". In: *Nucleic Acids Research* 47.D1, pp. D821–D827. DOI: 10.1093/nar/gky961.

López-Maury, Luis, Samuel Marguerat and Jürg Bähler (2008). "Tuning Gene Expression to Changing Environments: From Rapid Responses to Evolutionary Adaptation". In: *Nature Reviews Genetics* 9.8, pp. 583–593. DOI: 10.1038/nrg2398.

Love, Michael I., Wolfgang Huber and Simon Anders (2014). "Moderated Estimation of Fold Change and Dispersion for RNA-seq Data with DESeq2". In: *Genome Biology* 15, p. 550. DOI: 10.1186/s13059-014-0550-8.

Lynch, Michael, Mark C. Field, Holly V. Goodson, Harmit S. Malik, José B. Pereira-Leal et al. (2014). "Evolutionary Cell Biology: Two Origins, One Objective". In: *Proceedings of the National Academy of Sciences* 111.48, pp. 16990–16994. DOI: 10.1073/pnas.1415861111.

Maitra, Arijit and Ken A. Dill (2015). "Bacterial Growth Laws Reflect the Evolutionary Importance of Energy Efficiency". In: *Proceedings of the National Academy of Sciences* 112.2, pp. 406–411. DOI: 10.1073/pnas.1421138111.

– (2016). "Modeling the Overproduction of Ribosomes When Antibacterial Drugs Act on Cells". In: *Biophysical Journal* 110.3, pp. 743–748. DOI: 10.1016/j.bpj.2015.12.016.

Malecki, Michal, Stephan Kamrad, Markus Ralser and Jürg Bähler (2020). "Mitochondrial Respiration Is Required to Provide Amino Acids during Fermentative Proliferation of Fission Yeast". In: *EMBO reports* 21.11, e50845. DOI: 10.15252/embr.202050845.

Marangoni, Alejandro G. (2003). *Enzyme Kinetics: A Modern Approach*. Hoboken, NJ: John Wiley & Sons, Inc. ISBN: 978-0-471-15985-8.

Marguerat, Samuel, Alexander Schmidt, Sandra Codlin, Wei Chen, Ruedi Aebersold et al. (2012). "Quantitative Analysis of Fission Yeast Transcriptomes and Proteomes in Proliferating and Quiescent Cells". In: *Cell* 151.3, pp. 671–683. DOI: 10.1016/j.cell.2012.09.019.

Mata, Juan, Rachel Lyne, Gavin Burns and Jürg Bähler (2002). "The Transcriptional Program of Meiosis and Sporulation in Fission Yeast". In: *Nature Genetics* 32.1 (1), pp. 143–147. DOI: 10.1038/ng951.

Metzl-Raz, Eyal, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich et al. (2017). "Principles of Cellular Resource Allocation Revealed by Condition-Dependent Proteome Profiling". In: *eLife* 6, e28034. DOI: 10.7554/eLife.28034.

Milo, Ron, Paul Jorgensen, Uri Moran, Griffin Weber and Michael Springer (2010). "BioNumbers—the Database of Key Numbers in Molecular and Cell Biology". In: *Nucleic Acids Research* 38 (suppl_1), pp. D750–D753. DOI: 10.1093/nar/gkp889.

Milo, Ron, Rob Phillips and Nigel Orme (2016). *Cell Biology by the Numbers*. Abingdon: Garland Science. ISBN: 978-0-8153-4537-4.

Mitchison, J. M. and K. G. Lark (1962). "Incorporation of 3H-adenine into RNA during the Cell Cycle of Schizosaccharomyces Pombe". In: *Experimental Cell Research* 28.2, pp. 452–455. DOI: 10.1016/0014-4827(62)90304-X.

Mitchison, J. M. and Paul Nurse (1985). "Growth in Cell Length in the Fission Yeast Schizosaccharomyces Pombe". In: *Journal of Cell Science* 75.1, pp. 357–376.

Mogensen, Patrick K. and Asbjørn N. Riseth (2018). "Optim: A Mathematical Optimization Package for Julia". In: *Journal of Open Source Software* 3.24, p. 615. DOI: 10.21105/joss.00615.

Molenaar, Douwe, Rogier van Berlo, Dick de Ridder and Bas Teusink (2009). "Shifts in Growth Strategies Reflect Tradeoffs in Cellular Economics". In: *Molecular Systems Biology* 5.1, p. 323. DOI: 10.1038/msb.2009.82.

Monod, Jacques (1949). "The Growth of Bacterial Cultures". In: *Annual Review of Microbiology* 3.1, pp. 371–394. DOI: 10.1146/annurev.mi.03.100149.002103.

Mori, Matteo, Severin Schink, David W. Erickson, Ulrich Gerland and Terence Hwa (2017). "Quantifying the Benefit of a Proteome Reserve in Fluctuating Environments". In: *Nature Communications* 8.1, p. 1225. DOI: 10.1038/s41467-017-01242-8.

Mori, Risa and Snezhana Oliferenko (2020). "Cell Biology: An Open Solution for Closed Mitosis". In: *Current Biology* 30.16, R942–R944. DOI: 10.1016/j.cub.2020.06.067.

Morozumi, Yuichi and Kazuhiro Shiozaki (2021). “Conserved and Divergent Mechanisms That Control TORC1 in Yeasts and Mammals”. In: *Genes* 12.1 (1), p. 88. DOI: 10.3390/genes12010088.

Muchinsky, Paul M. (1996). “The Correction for Attenuation”. In: *Educational and Psychological Measurement* 56.1, pp. 63–75. DOI: 10.1177/0013164496056001004.

Mülleder, Michael, Enrica Calvani, Mohammad Tauqeer Alam, Richard Kangda Wang, Florian Eckerstorfer et al. (2016). “Functional Metabolomics Describes the Yeast Biosynthetic Regulome”. In: *Cell* 167.2, 553–565.e12. DOI: 10.1016/j.cell.2016.09.007.

Nasmyth, Kim, Paul Nurse and R. S. S. Fraser (1979). “The Effect of Cell Mass on the Cell Cycle Timing and Duration of S-Phase in Fission Yeast”. In: *Journal of Cell Science* 39.1, pp. 215–233.

Neidhardt, Frederick C., John L. Ingraham and Moselio Schaechter (1990). *Physiology of the Bacterial Cell*. Sunderland, MA: Sinauer Associates.

Nelder, J. A. and R. Mead (1965). “A Simplex Method for Function Minimization”. In: *The Computer Journal* 7.4, pp. 308–313. DOI: 10.1093/comjnl/7.4.308.

Nguyen, T. H., G. H. Fleet and P. L. Rogers (1998). “Composition of the Cell Walls of Several Yeast Species”. In: *Applied Microbiology and Biotechnology* 50.2, pp. 206–212. DOI: 10.1007/s002530051278.

O’Brien, Edward J., Joshua A. Lerman, Roger L. Chang, Daniel R. Hyduke and Bernhard Ø Palsson (2013). “Genome-scale Models of Metabolism and Gene Expression Extend and Refine Growth Phenotype Prediction”. In: *Molecular Systems Biology* 9.1, p. 693. DOI: 10.1038/msb.2013.52.

O’Brien, Edward J., Jose Utrilla and Bernhard O. Palsson (2016). “Quantification and Classification of E. Coli Proteome Utilization and Unused Protein Costs across Environments”. In: *PLOS Computational Biology* 12.6, e1004998. DOI: 10.1371/journal.pcbi.1004998.

Odermatt, Pascal D, Teemu P Miettinen, Joël Lemière, Joon Ho Kang, Emrah Bostan et al. (2021). “Variations of Intracellular Density during the Cell Cycle Arise from Tip-Growth Regulation in Fission Yeast”. In: *eLife* 10. Ed. by Mohan K Balasubramanian, Naama Barkai and Matthieu Piel, e64901. DOI: 10.7554/eLife.64901.

Ojkic, Nikola, Diana Serbanescu and Shiladitya Banerjee (2019). “Surface-to-Volume Scaling and Aspect Ratio Preservation in Rod-Shaped Bacteria”. In: *eLife* 8. Ed. by Raymond E Goldstein, Naama Barkai and Charles W Wolgemuth, e47033. DOI: 10.7554/eLife.47033.

Oldewurtel, Enno R., Yuki Kitahara, Baptiste Cordier, Gizem Özbaykal and Sven van Teeffelen (2019). "Bacteria Control Cell Volume by Coupling Cell-Surface Expansion to Dry-Mass Growth". In: *bioRxiv*, p. 769786. DOI: 10.1101/769786.

Orth, Jeffrey D., Ines Thiele and Bernhard Ø Palsson (2010). "What Is Flux Balance Analysis?" In: *Nature Biotechnology* 28.3 (3), pp. 245–248. DOI: 10.1038/nbt.1614.

Padovan-Merhar, Olivia, Gautham P. Nair, Andrew G. Biaesch, Andreas Mayer, Steven Scarfone et al. (2015). "Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms". In: *Molecular Cell* 58.2, pp. 339–352. DOI: 10.1016/j.molcel.2015.03.005.

Pancaldi, Vera, Falk Schubert and Jürg Bähler (2010). "Meta-Analysis of Genome Regulation and Expression Variability across Hundreds of Environmental and Genetic Perturbations in Fission Yeast". In: *Molecular BioSystems* 6.3, pp. 543–552. DOI: 10.1039/B913876P.

Pandey, Parth Pratim and Sanjay Jain (2016). "Analytic Derivation of Bacterial Growth Laws from a Simple Model of Intracellular Chemical Dynamics". In: *Theory in Biosciences* 135.3, pp. 121–130. DOI: 10.1007/s12064-016-0227-9.

Pang, Tin Yau and Martin J. Lercher (2020). "Optimal Density of Biological Cells". In: *bioRxiv*, p. 2020.11.18.388744. DOI: 10.1101/2020.11.18.388744.

Paulo, Joao A., Jeremy D. O'Connell, Robert A. Everley, Jonathon O'Brien, Micah A. Gygi et al. (2016). "Quantitative Mass Spectrometry-Based Multiplexing Compares the Abundance of 5000 S. Cerevisiae Proteins across 10 Carbon Sources". In: *Journal of Proteomics* 148 (Supplement C), pp. 85–93. DOI: 10.1016/j.jprot.2016.07.005.

Paulo, Joao A., Jeremy D. O'Connell, Aleksandr Gaun and Steven P. Gygi (2015). "Proteome-Wide Quantitative Multiplexed Profiling of Protein Expression: Carbon-Source Dependency in Saccharomyces Cerevisiae". In: *Molecular Biology of the Cell* 26.22, pp. 4063–4074. DOI: 10.1091/mbc.E15-07-0499.

Peebo, Karl, Kaspar Valgepea, Andres Maser, Ranno Nahku, Kaarel Adamberg et al. (2015). "Proteome Reallocation in Escherichia Coli with Increasing Specific Growth Rate". In: *Molecular BioSystems* 11.4, pp. 1184–1193. DOI: 10.1039/C4MB00721B.

Petersen, Janni and Paul Russell (2016). "Growth and the Environment of Schizosaccharomyces Pombe". In: *Cold Spring Harbor Protocols* 2016.3, pdb.top079764. DOI: 10.1101/pdb.top079764.

Petibon, Cyrielle, Mustafa Malik Ghulam, Mathieu Catala and Sherif Abou Elela (2020). “Regulation of Ribosomal Protein Genes: An Ordered Anarchy”. In: *Wiley interdisciplinary reviews. RNA* 12.3, e1632. DOI: 10.1002/wrna.1632.

Planqué, Robert, Josephus Hulshof, Bas Teusink, Johannes C. Hendriks and Frank J. Bruggeman (2018). “Maintaining Maximal Metabolic Flux by Gene Expression Control”. In: *PLOS Computational Biology* 14.9, e1006412. DOI: 10.1371/journal.pcbi.1006412.

Rackauckas, Christopher and Qing Nie (2017). “DifferentialEquations.Jl – A Performant and Feature-Rich Ecosystem for Solving Differential Equations in Julia”. In: *Journal of Open Research Software* 5.1. DOI: 10.5334/jors.151.

Rallis, C., S. Codlin and J. Bähler (2013). “TORC1 Signaling Inhibition by Rapamycin and Caffeine Affect Lifespan, Global Gene Expression, and Cell Proliferation of Fission Yeast.” In: *Aging cell* 12.4, pp. 563–573. DOI: 10.1111/acel.12080.

Rallis, C., L. López-Maury, T. Georgescu, V. Pancaldi and J. Bähler (2014). “Systematic Screen for Mutants Resistant to TORC1 Inhibition in Fission Yeast Reveals Genes Involved in Cellular Ageing and Growth.” In: *Biology open* 3.2, pp. 161–171. DOI: 10.1242/bio.20147245.

Reuveni, Shlomi, Måns Ehrenberg and Johan Paulsson (2017). “Ribosomes Are Optimized for Autocatalytic Production”. In: *Nature* 547.7663, pp. 293–297. DOI: 10.1038/nature22998.

Russell, Paul and Paul Nurse (1987). “Negative Regulation of Mitosis by Wee1+, a Gene Encoding a Protein Kinase Homolog”. In: *Cell* 49.4, pp. 559–567. DOI: 10.1016/0092-8674(87)90458-2.

Rustici, Gabriella, Juan Mata, Katja Kivinen, Pietro Lió, Christopher J. Penkett et al. (2004). “Periodic Gene Expression Program of the Fission Yeast Cell Cycle”. In: *Nature Genetics* 36.8 (8), pp. 809–817. DOI: 10.1038/ng1377.

Saint, Malika, François Bertaux, Wenhao Tang, Xi-Ming Sun, Laurence Game et al. (2019). “Single-Cell Imaging and RNA Sequencing Reveal Patterns of Gene Expression Heterogeneity during Fission Yeast Growth and Adaptation”. In: *Nature Microbiology* 4, pp. 480–491. DOI: 10.1038/s41564-018-0330-4.

Sánchez, Benjamín J, Cheng Zhang, Avlant Nilsson, Petri-Jaan Lahtvee, Eduard J Kerkhoven et al. (2017). “Improving the Phenotype Predictions of a Yeast Genome-Scale Metabolic Model by Incorporating Enzymatic Constraints”. In: *Molecular Systems Biology* 13.8, p. 935. DOI: 10.15252/msb.20167411.

Sánchez, Benjamín J., Petri-Jaan Lahtvee, Kate Campbell, Sergo Kasvandik, Rosemary Yu et al. (2021). "Benchmarking Accuracy and Precision of Intensity-Based Absolute Quantification of Protein Abundances in Saccharomyces Cerevisiae". In: *PROTEOMICS* 21.6, p. 2000093. DOI: 10.1002/pmic.202000093.

Schaechter, M., O. Maaløe and N. O. Kjeldgaard (1958). "Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella Typhimurium". In: *Journal of General Microbiology* 19.3, pp. 592–606. DOI: 10.1099/00221287-19-3-592.

Schaechter, Moselio (2006). "From Growth Physiology to Systems Biology". In: *International Microbiology: The Official Journal of the Spanish Society for Microbiology* 9.3, pp. 157–161.

Schmidt, Alexander, Karl Kochanowski, Silke Vedelaar, Erik Ahrné, Benjamin Volkmer et al. (2016). "The Quantitative and Condition-Dependent Escherichia Coli Proteome". In: *Nature Biotechnology* 34.1, pp. 104–110. DOI: 10.1038/nbt.3418.

Schmidt, Michael W, Andres Houseman, Alexander R Ivanov and Dieter A Wolf (2007). "Comparative Proteomic and Transcriptomic Profiling of the Fission Yeast Schizosaccharomyces Pombe". In: *Molecular Systems Biology* 3.1, p. 79. DOI: 10.1038/msb4100117.

Scott, Matthew, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang and Terence Hwa (2010). "Interdependence of Cell Growth and Gene Expression: Origins and Consequences". In: *Science* 330, pp. 1099–1102. DOI: 10.1126/science.1192588.

Scott, Matthew, Stefan Klumpp, Eduard M. Mateescu and Terence Hwa (2014). "Emergence of Robust Growth Laws from Optimal Regulation of Ribosome Synthesis". In: *Molecular Systems Biology* 10.8, p. 747. DOI: 10.15252/msb.20145379.

Sen, Pranab Kumar (1968). "Estimates of the Regression Coefficient Based on Kendall's Tau". In: *Journal of the American Statistical Association* 63.324, pp. 1379–1389. DOI: 10.1080/01621459.1968.10480934.

Shah, Meera, Dan Su, Judith S. Scheliga, Tomáš Pluskal, Susanna Boronat et al. (2016). "A Transcript-Specific eIF3 Complex Mediates Global Translational Control of Energy Metabolism". In: *Cell Reports* 16.7, pp. 1891–1902. DOI: 10.1016/j.celrep.2016.07.006.

Shahrezaei, Vahid and Samuel Marguerat (2015). "Connecting Growth with Gene Expression: Of Noise and Numbers". In: *Current Opinion in Microbiology*. Environmental Microbiology • Extremophiles 25, pp. 127–135. DOI: 10.1016/j.mib.2015.05.012.

Shimizu, Kazuyuki and Yu Matsuoka (2018). “Regulation of Glycolytic Flux and Overflow Metabolism Depending on the Source of Energy Generation for Energy Demand”. In: *Biotechnology Advances* 37.2, pp. 284–305. DOI: 10.1016/j.biotechadv.2018.12.007.

Siegel, Andrew F. (1982). “Robust Regression Using Repeated Medians”. In: *Biometrika* 69.1, pp. 242–244. DOI: 10.1093/biomet/69.1.242.

Simanis, Viesturs (2015). “Pombe’s Thirteen – Control of Fission Yeast Cell Division by the Septation Initiation Network”. In: *Journal of Cell Science* 128.8, pp. 1465–1474. DOI: 10.1242/jcs.094821.

Spearman, C. (1904). “The Proof and Measurement of Association between Two Things”. In: *The American Journal of Psychology* 15.1, pp. 72–101. DOI: 10.2307/1412159.

Spinelli, Jessica B. and Marcia C. Haigis (2018). “The Multifaceted Contributions of Mitochondria to Cellular Metabolism”. In: *Nature Cell Biology* 20.7 (7), pp. 745–754. DOI: 10.1038/s41556-018-0124-1.

Strimmer, Korbinian (2008). “Fdrtool: A Versatile R Package for Estimating Local and Tail Area-Based False Discovery Rates”. In: *Bioinformatics* 24.12, pp. 1461–1462. DOI: 10.1093/bioinformatics/btn209.

Sun, Xi-Ming, Anthony Bowman, Miles Priestman, Francois Bertaux, Amalia Martinez-Segura et al. (2020). “Size-Dependent Increase in RNA Polymerase II Initiation Rates Mediates Gene Expression Scaling with Cell Size”. In: *Current Biology* 30.7, 1217–1230.e7. DOI: 10.1016/j.cub.2020.01.053.

Szenk, Mariola, Ken A. Dill and Adam M. R. de Graff (2017). “Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis”. In: *Cell Systems* 5.2, pp. 95–104. DOI: 10.1016/j.cels.2017.06.005.

Theil, Henri (1950). “A Rank-Invariant Method of Linear and Polynomial Regression Analysis: I, II, and III”. In: *Proceedings of the Koninklijke Nederlandse Academie van Wetenschappen Series A* 53 (3, 4, and 9), pp. 386–392, 521–525, 1397–1412.

Thomas, Philipp, Guillaume Terradot, Vincent Danos and Andrea Y. Weiße (2018). “Sources, Propagation and Consequences of Stochasticity in Cellular Growth”. In: *Nature Communications* 9.1, p. 4528. DOI: 10.1038/s41467-018-06912-9.

Towbin, Benjamin D., Yael Korem, Anat Bren, Shany Doron, Rotem Sorek et al. (2017). “Optimality and Sub-Optimality in a Bacterial Growth Law”. In: *Nature Communications* 8, p. 14123. DOI: 10.1038/ncomms14123.

Uprety, Bhawana, Rwik Sen and Sukesh R. Bhaumik (2015). “Eaf1p Is Required for Recruitment of NuA4 in Targeting TFIID to the Promoters of the Ribosomal Protein Genes for Transcriptional Initiation In Vivo”. In: *Molecular and Cellular Biology* 35.17, pp. 2947–2964. DOI: 10.1128/MCB.01524-14.

Vadia, Stephen and Petra Anne Levin (2015). “Growth Rate and Cell Size: A Re-Examination of the Growth Law”. In: *Current Opinion in Microbiology* 24, pp. 96–103. DOI: 10.1016/j.mib.2015.01.011.

Valbuena, Noelia, Ana Elisa Rozalén and Sergio Moreno (2012). “Fission Yeast TORC1 Prevents eIF2α Phosphorylation in Response to Nitrogen and Amino Acids via Gcn2 Kinase”. In: *Journal of Cell Science* 125.24, pp. 5955–5959. DOI: 10.1242/jcs.105395.

Valgepea, Kaspar, Kaarel Adamberg, Andrus Seiman and Raivo Vilu (2013). “Escherichia Coli Achieves Faster Growth by Increasing Catalytic and Translation Rates of Proteins”. In: *Molecular BioSystems* 9.9, pp. 2344–2358. DOI: 10.1039/C3MB70119K.

Van den Berg, Jonas, Arnold J. Boersma and Bert Poolman (2017). “Microorganisms Maintain Crowding Homeostasis”. In: *Nature Reviews Microbiology* 15.5 (5), pp. 309–318. DOI: 10.1038/nrmicro.2017.17.

Vander Heiden, Matthew G., Lewis C. Cantley and Craig B. Thompson (2009). “Understanding the Warburg Effect: The Metabolic Requirements of Cell Proliferation”. In: *Science* 324.5930, pp. 1029–1033. DOI: 10.1126/science.1160809.

Van Hoek, Milan JA and Roeland MH Merks (2012). “Redox Balance Is Key to Explaining Full vs. Partial Switching to Low-Yield Metabolism”. In: *BMC Systems Biology* 6.1, p. 22. DOI: 10.1186/1752-0509-6-22.

Vazquez, Alexei (2012). “Optimal Macromolecular Density in the Cell”. In: *Proceedings of the National Academy of Sciences* 109.9, E533–E533. DOI: 10.1073/pnas.1118500109.

Vazquez, Alexei, Qasim K. Beg, Marcio A. deMenezes, Jason Ernst, Ziv Bar-Joseph et al. (2008). “Impact of the Solvent Capacity Constraint on E. Coli Metabolism”. In: *BMC Systems Biology* 2, p. 7. DOI: 10.1186/1752-0509-2-7.

Vemuri, G. N., M. A. Eiteman, J. E. McEwen, L. Olsson and J. Nielsen (2007). “Increasing NADH Oxidation Reduces Overflow Metabolism in Saccharomyces Cerevisiae”. In: *Proceedings of the National Academy of Sciences* 104.7, pp. 2402–2407. DOI: 10.1073/pnas.0607469104.

Von Stockar, U. and J. -S. Liu (1999). “Does Microbial Life Always Feed on Negative Entropy? Thermodynamic Analysis of Microbial Growth”. In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1412.3, pp. 191–211. DOI: 10.1016/S0005-2728(99)00065-1.

Waldron, C. and F. Lacroute (1975). “Effect of Growth Rate on the Amounts of Ribosomal and Transfer Ribonucleic Acids in Yeast.” In: *Journal of Bacteriology* 122.3, pp. 855–865.

Warner, Jonathan R. and Kerri B. McIntosh (2009). “How Common Are Extraribosomal Functions of Ribosomal Proteins?” In: *Molecular Cell* 34.1, pp. 3–11. DOI: 10.1016/j.molcel.2009.03.006.

Watanabe, Daisuke, Takuma Kajihara, Yukiko Sugimoto, Kenichi Takagi, Megumi Mizuno et al. (2019). “Nutrient Signaling via the TORC1-Greatwall-PP2AB55δ Pathway Is Responsible for the High Initial Rates of Alcoholic Fermentation in Sake Yeast Strains of Saccharomyces Cerevisiae”. In: *Applied and Environmental Microbiology* 85.1, e02083–18. DOI: 10.1128/AEM.02083-18.

Weisman, Ronit (2016). “Target of Rapamycin (TOR) Regulates Growth in Response to Nutritional Signals”. In: *Microbiology Spectrum* 4.5, FUNK-0006–2016. DOI: 10.1128/microbiolspec.FUNK-0006-2016.

Weiße, Andrea Y., Diego A. Oyarzún, Vincent Danos and Peter S. Swain (2015). “Mechanistic Links between Cellular Trade-Offs, Gene Expression, and Growth”. In: *Proceedings of the National Academy of Sciences* 112.9, E1038–E1047. DOI: 10.1073/pnas.1416533112.

Yang, Laurence, James T Yurkovich, Zachary A King and Bernhard O Palsson (2018). “Modeling the Multi-Scale Mechanisms of Macromolecular Resource Allocation”. In: *Current Opinion in Microbiology*. Antimicrobials * Microbial Systems Biology 45, pp. 8–15. DOI: 10.1016/j.mib.2018.01.002.

You, Conghui, Hiroyuki Okano, Sheng Hui, Zhongge Zhang, Minsu Kim et al. (2013). “Coordination of Bacterial Proteome with Metabolism by Cyclic AMP Signalling.” In: *Nature* 500, pp. 301–306. DOI: 10.1038/nature12446.

Yu, Jason S. L., Clara Correia-Melo, Francisco Zorrilla, Lucia Herrera-Dominguez, Mary Y. Wu et al. (2022). “Microbial Communities Form Rich Extracellular Metabolomes That Foster Metabolic Interactions and Promote Drug Tolerance”. In: *Nature Microbiology*, pp. 1–14. DOI: 10.1038/s41564-022-01072-5.

Yu, Rosemary, Kate Campbell, Rui Pereira, Johan Björkeroth, Qi Qi et al. (2020). “Nitrogen Limitation Reveals Large Reserves in Metabolic and Translational Capacities of Yeast”. In: *Nature Communications* 11.1 (1), pp. 1–12. DOI: 10.1038/s41467-020-15749-0.

Yu, Rosemary, Egor Vorontsov, Carina Sihlbom and Jens Nielsen (2021). “Quantifying Absolute Gene Expression Profiles Reveals Distinct Regulation of Central Carbon Metabolism Genes in Yeast”. In: *eLife* 10. Ed. by Kevin J Verstrepen, e65722. DOI: 10.7554/eLife.65722.

Zavřel, Tomáš, Marjan Faizi, Cristina Loureiro, Gereon Poschmann, Kai Stühler et al. (2019). “Quantitative Insights into the Cyanobacterial Cell Economy”. In: *eLife* 8. Ed. by Severin Sasso, Ian T Baldwin, Severin Sasso and Jörg Toepel, e42508. DOI: 10.7554/eLife.42508.

Zhou, E. H., X. Trepat, C. Y. Park, G. Lenormand, M. N. Oliver et al. (2009). “Universal Behavior of the Osmotically Compressed Cell and Its Analogy to the Colloidal Glass Transition”. In: *Proceedings of the National Academy of Sciences* 106.26, pp. 10632–10637. DOI: 10.1073/pnas.0901462106.

Zhu, Anqi, Joseph G Ibrahim and Michael I Love (2019). “Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences”. In: *Bioinformatics* 35.12, pp. 2084–2092. DOI: 10.1093/bioinformatics/bty895.

Zhu, Manlu, Yige Pan and Xiongfeng Dai (2019). “(P)ppGpp: The Magic Governor of Bacterial Growth Economy”. In: *Current Genetics*. DOI: 10.1007/s00294-019-00973-z.