What makes up a cell

Pranas Grigaitis[1] and Diana Széliová[2]

[1] Systems Biology Lab, Amsterdam Institute of Molecular and Life Sciences (AIMMS), Vrije Universiteit Amsterdam, Amsterdam, the Netherlands; [2] Department of Analytical Chemistry, University of Vienna, Vienna, Austria

Abstract

In this chapter, we describe the building blocks of cells, their varied composition in different environments and cell types, and the processes involved in creating these components. We also discuss how cells allocate resources and coordinate their cellular processes. The chapter begins by describing the basic cellular components and their quantities. While cells are mainly composed of a few chemical elements (CHNOPS), they contain a wide array of biological molecules such as proteins, nucleic acids, lipids, carbohydrates, and small molecules. Some of these molecules assemble into complex structures such as membranes, organelles, or highly specialized molecular machines such as polymerases and ribosomes. Although there are significant differences in the size and structure of different cell types (bacteria, yeast, mammalian cells), the composition of their biomass is similar, with proteins being the primary component. However, this composition is not staticit varies in different environments, at different growth rates, and even within a population. We also explore the large variability of cell size across different species and, conversely, the low variability of the cell buoyant density. Furthermore, we describe physical constraints that shape cellular composition, such as temperature, pH, diffusion limit, and macromolecular crowding. In the final section, we look at the synthesis of the biological molecules and their time scales. Metabolic reactions provide building blocks for macromolecule synthesis. In turn, macromolecules catalyze the metabolic reactions (enzymes) and their own synthesis (ribosomes). These processes happen on different time scales and have to be coordinated so that cells can efficiently replicate themselves. Furthermore, the allocation to the different processes is limited by the available proteome space. Finally, we discuss the remaining gaps in our knowledge of cell biology and explore how cell composition is represented in various mathematical models.

Keywords: Biomass composition, Cell size, Cell density, Biosynthesis, Macromolecule synthesis, Proteome space, Physical constraints

Contributions: This chapter was planned and outlined by P. Grigaitis and D. Széliová, with the help of W. Liebermeister and E. Noor. Feedback was provided by S. van den Bogaard.

To cite this chapter: P. Grigaitis and D. Széliová. What makes up a cell. Version April 2024. The Economic Cell Collective. Economic Principles in Cell Biology (2024). No commercial publisher | Authors open access book.

The authors are listed in alphabetical order.

This is a chapter from the free textbook "Economic Principles in Cell Biology".
Free download from principlescellphysiology.org/book-economic-principles/.
Lecture slides for this chapter are available on the website.

© 2024 The Economic Cell Collective.
Licensed under Creative Commons License CC-BY-SA 4.0.
An authors open access book. No publisher has been paid.

Chapter overview

- Cells use the same building blocks to give rise to a high number of molecular species
- There are many parallel processes happening in cells, with similar precursors
- Composition of cells is environment-dependent
- Different time-scales give rise to coordination of processes

2.1. Describing and counting cellular components

Cells contain a diverse spectrum of molecules, needed to create two cells out of one (as Rudolf Virchow proposed, *omnis cellula e cellula*, all cells come from cells). These molecules come in different sizes and properties, and therefore create a demand for a cell to keep these components in different places (*spatial* organization) with different patterns of use (*temporal* organization), and book-keep their quantities. Cell composition directly influences the function of the cell: thus we observe different cellular make-up in different organisms or even in different cells of the same organism. Both historical research and the latest advancements in instrumentation allow us to characterize the constituents of cells in more and more depth. Thus in this chapter, we will do a census of cellular components: we will discuss what molecules make up a cell, what they are derived from, how to measure these components in the lab and we will briefly consider allocation of resources, directed to synthesize individual cellular components.

2.2. The components of a cell

2.2.1. Cell composition and structures

Although living matter comes in different shapes and sizes, over 99% of the cellular mass can be described by only a handful of chemical elements. 6 most abundant elements form the famous CHNOPS notation: carbon (C), hydrogen (H), nitrogen (N), oxygen (O), phosphorus (P), and sulfur (S). Taken together, these 6 elements encompass the vast majority of the mass, namely, ca. 97.5% in budding yeast *Saccharomyces cerevisiae* [1]. Living cells also contain minute amounts of different metal ions, such as sodium (Na), potassium (K), iron (Fe), molybdenum (Mo) and others – usually facilitating signal transduction or supporting enzymatic catalysis.

In an extremely simplified way, cells can be looked at as bags of fluid-like material, kept together by a membrane. These “bags of things” can also contain other membrane structures inside them, forming so-called organelles. In cell biology, we call cells *prokaryotic* if they do not possess these membrane structures, and *eukaryotic* if they do. The divide between prokaryotes and eukaryotes can be illustrated by comparing two organisms: the prokaryotic bacterium *Escherichia coli* and the eukaryotic yeast *Saccharomyces cerevisiae*. They both are organisms, composed of a single cell (thus called *unicellular*), and they both are very small, compared to a typical human cell. However, *E. coli* does not contain any additional membrane structures except from the plasma membrane (which encompasses the cellular contents). Meanwhile, a handful of different organelles can be observed in *S. cerevisiae*. The cellular organization of these cells is shown in Figure 2.1.

Most biological membranes and membrane-based structures, including the plasma membrane itself, have multiple functions (not only separating space), and are highly dynamic. Some membranes can fold into very compact structures with extremely high surface area (endoplasmic reticulum, Golgi apparatus), occupy different volumes - from small vesicles to large vacuoles, occupying a major fraction of the cell volume. Moreover, some molecules can form very large structures, which might be transient (short-lived), thus capturing and defining them remains a major challenge. For these reasons, the fine structure of cells is unclear - some findings (e.g. organelle contact sites, see [2] for a recent review) hint into some functional organization of organelles, yet the canonical way to look at the cellular structure remains as to a “bag of things”.

A notable example of a highly specialized organelle is the mitochondrion. The mitochondrion is separated from the rest of the cell by two (outer and inner) membranes; this feature is essential for their function. In eukaryotes, mitochondria are a major hub of metabolism: they house essential biochemical pathways, such as tricarboxylic acid cycle (also known as citric acid-, or Krebs cycle), as well as the so-called *respiratory chain*, the machinery for generating energy with the use of oxygen (see Chapter 3 in [3] for more details). While the most biochemical interconversions happen inside the mitochondria (in mitochondrial *matrix*), the respiratory chain proteins are located in the inner mitochondrial membrane: these proteins create an *electrochemical gradient* across this membrane, and use it to drive the conversion of energy, stored in nutrients, into the energy the cell can use (in a form of ATP). What makes mitochondria even more interesting is that they also contain mitochondria-specific genetic information (mitochondrial DNA), which is essential for mitochondria to function inside the cell. In many organisms, the loss of mitochondrial DNA results in impaired growth (in yeasts, that is called the *petite* phenotype) [4], and some organisms cannot grow unless mitochondrial DNA is present (*petite-negative* yeasts).

2.2.2. Biological molecules

Although cells contain many different molecular *species* (“molecular identities”), we can crudely categorize them into *small molecules* and *macromolecules* based on their molecular weight and complexity. Small molecules, as the name suggests, are small chemical compounds, up to 1000 Daltons in mass (1 Dalton $=$ 1 atomic mass unit, 1 amu), and are usually composed of a non-repeating single chemical unit (called *monomer*). Macromolecules, on the contrary, are up to several megadaltons (MDa $= 10^6$ Da) in weight, and are frequently composed of multiple monomers (forming so-called *polymers*). Compounds in the cells, both macro- and small molecules, based on their chemical nature, fall into 5 big groups: proteins, nucleic acids (both macromolecules), carbohydrates (exist as both small molecules and polymers), lipids (small molecules), and cofactors/other small molecules.

Proteins are polymers, composed of amino acids. Proteins are an exceptionally diverse class of molecules: in Nature, 20 amino acids can be incorporated into proteins (so-called *proteogenic* amino acids), which, combinatorially provides 20 options for each position in the protein chain. Therefore, there is an enormous amount of possible combinations to make a protein of a length of 100 amino acids (20^{100}, to be precise), even for a amino acid chain way shorter than the average in *E. coli*, around 325 amino acids (BioNumbers ID (BNID) [5] 108986). This diversity gives rise to the spectrum of functions proteins can do, for instance, catalysis (catalytic proteins are also called *enzymes*), transport of molecules, keeping structural integrity of membranes, and others. Also two notable properties of proteins are that they (1) need to acquire a specific three-dimensional structure (“to fold”) in order to become functionally active, and (2) sometimes, they also need to form complexes of the same or other proteins (called *multimers*). Protein production is a major consumer of energy and biosynthetic intermediates in the cell, therefore, in this book we will frequently consider proteins as central players in implementing economic principles in cell physiology.

Nucleic acids are another category of macromolecules; their monomers are called nucleotides. There are two major classes of nucleic acids, RNA (ribonucleic acid) and DNA (deoxyribonucleic acid). RNA and DNA chemically have a slight, yet critical difference: the sugar, which is a part of the nucleotides, differs between RNA (ribose) and DNA (deoxyribose). The two sugars are almost the same but for one chemical

Box 2.A : Macromolecular machines

An important consideration about both proteins and nucleic acids is that they are polymerized by very specialized protein- and protein-nucleic acid complexes. These molecular motors use energy (in terms of ATP equivalents) to form chains of the respective monomers. In the case of proteins, the individual amino acids are combined into a so-called *peptide chain* by a ribosome, a macromolecular complex made from proteins and RNA. The nucleic acids are synthesized by a class of enzymatic complexes, called *nucleic acid polymerases*. There are two major classes of them, specific to the nucleic acid: RNA and DNA polymerases, respectively.

group: one of the carbon atoms in ribose is connected to two another carbon atoms, a hydrogen atom, and a chemical group, called hydroxy- ($-OH$). In deoxyribose, the hydroxy-group is substituted with another hydrogen atom, hence the prefix "deoxy-" ("minus oxygen"). RNA and DNA have different functions in the cell: the primary function of DNA is to store genetic information, while RNA can work both as an intermediate agent to transfer that genetic information to protein production (messenger RNA, mRNA) or to participate in catalysis and protein production in general (e.g. transfer and ribosomal RNA, tRNA and rRNA, respectively). Outside the polymers, nucleotides can also act as energy-accumulating compounds (e.g. ATP, adenosine triphosphate) or signaling molecules (e.g. cyclic adenosine monophosphate, cAMP). In this text, we will mostly refer to the energy-storing function of the nucleotides, although other functions, such as signaling, also are essential aspects of describing cell physiology.

Carbohydrates are another major class of biological molecules, and are important both as monomers and high molecular-weight polymers. Monomeric carbohydrates (sometimes also referred to as sugars) are mainly used as carbon and energy sources for organisms, e.g. glucose or fructose. In oligomeric form (up to 10 monomers), carbohydrate chains are essential for cellular sensing systems, to be specific, receptor-ligand binding. Finally, polymers of carbohydrates usually serve as structural components (part of peptidoglycan, major part of bacterial cell walls) or energy/carbon storage (glycogen in, e.g. yeasts and animal cells, or starch in plants).

Lipids are a vaguely-described class of compounds, which have an overarching similarity, being water-insoluble. The major function of lipids in biological cells is structural: a very abundant subclass of lipids, phospholipids, is an essential constitutent of biological membranes. As discussed in Section 2.2.1, membranes themselves have a variety of functions, which are mostly carried out by lipids (structural) or proteins (transport, sensing, signaling etc.). Some lipids can also undertake other functions, such as signaling (various sterols), or energy storage (tryglycerides, or fats).

As we see, the metabolism of biological molecules is tightly interlinked, although they exibit major differences in their abundance, size and chemical properties. Macromolecules are present in very low concentrations, and their biosynthesis usually takes minutes. Meanwhile, the time scale of small molecule reactions is usually seconds (or fraction of), and the concentrations of small molecules are usually several magnitudes higher than these of macromolecules. Yet, despite acting at different rates and concentrations, these two types of biological molecules work in an orchestrated manner. To begin with, a number of different small molecules are required to produce both other small molecules and the macromolecules. In return, the macromolecules ensure cell integrity and growth by, among other functions, operating the reaction networks of small molecule interconversions (which we usually refer to as metabolism). Additionally, presence of some small molecules can influence the function of macromolecules, both directly (e.g. essential cofactors, needed for enzymatic reactions; enzyme activation or inhibition), and indirectly (e.g. modulation of gene expression, signaling). Therefore, a lot of different processes have to happen in parallel to ensure the operation of the cells. Having

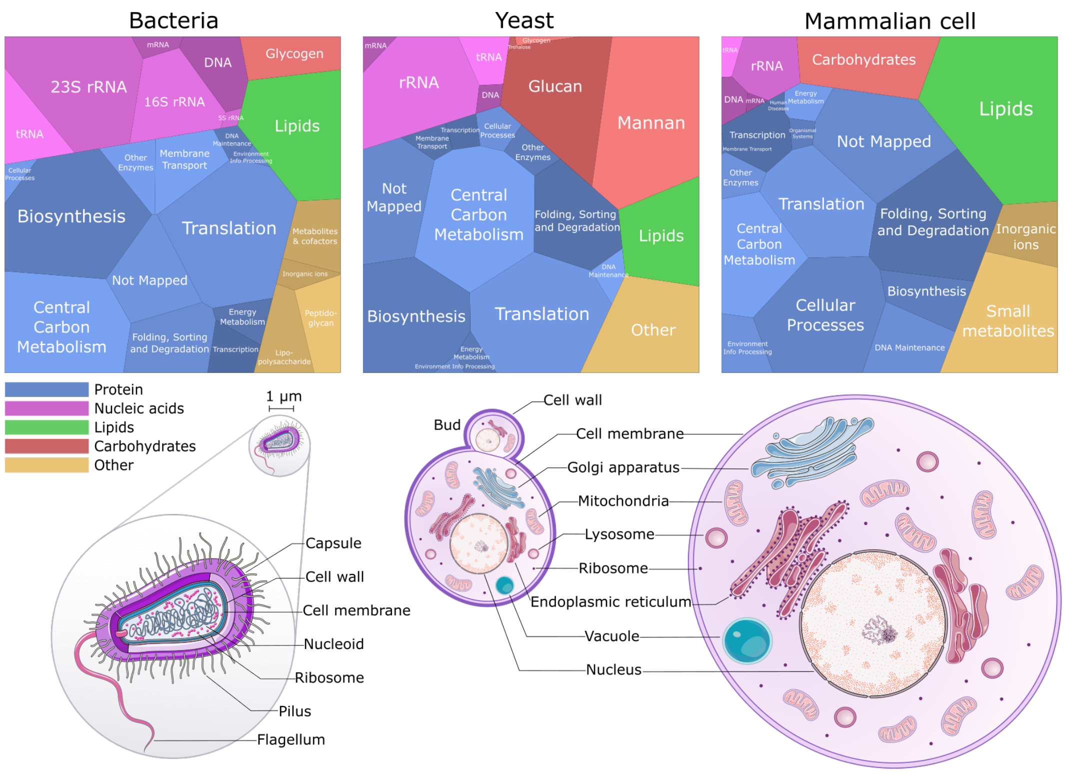

Figure 2.1: Biomass composition and cell structure of a typical bacterial, yeast, and a mammalian cell. The area of each polygon corresponds to a mass fraction of a component per cell. While the average composition is quite similar in the three groups, there are major differences in size and internal organization (especially when comparing prokaryotes with eukaryotes). Data for proteome groups (length-weighted protein abundances) was obtained from Proteomaps. Sources of composition data: bacteria [6], yeast (BNID 108200, 108196, 107234, 100261, [7]), mammalian cells (BNID 107131, 107235, 107234). Pictures of cells were created using Bioicons[1].

defined the major types of molecules we find in living cells, next we will discuss how abundant are different components of the cells.

2.3. Cell composition in numbers

2.3.1. Biomass composition

Cells are composed of around 70% water and 30% dry mass. As mentioned in the previous section, we can describe the composition of the dry mass with the most abundant chemical elements. For example, the elemental formula for *E. coli* is $CH_{1.77}O_{0.49}N_{0.24}$ (BNID 101800) and for *S. cerevisiae* $CH_{1.61}O_{0.56}N_{0.16}$ (BNID 101801).

However, more often, we are interested in biomass composition in terms of the main macromolecules (proteins, nucleic acids, lipids, and carbohydrates) and small molecules (metabolites, cofactors, and ions). Table 2.1 summarizes an average composition of *E. coli* and *S. cerevisiae* during exponential growth, the typical molecular masses and copy numbers of the components. The most abundant component is protein, which forms around half of the cell's dry mass. When we divide the proteome into functional groups, we find that

[1]The icons bacterium-interior, golgi-3d-1, mitochondrium-3, endoplasmatic-reticulum-3d-medium, endoplasmatic-reticulum-rough-3d-2, endoplasmatic-reticulum-rough-3d, and nucleus by Servier are licensed under CC-BY 3.0 Unported.

the biggest fractions belong to translation, central carbon metabolism, folding, sorting and degradation, and biosynthesis. A substantial fraction belongs to proteins that are not mapped (especially in mammalian cells), illustrating that we still lack knowledge about the function of many proteins (Figure 2.1).

RNA forms 20% of dry cell mass in *E. coli*, but this number is lower in eukaryotes, such as yeast (11%) or mammalian cells (4%). While the total amount of RNA is variable in different organisms, its relative composition is similar – most of the RNA mass is formed by rRNA (80%), followed by tRNA (15%) and mRNA (5%) (BNID 100258, 100261, 106154). Lipid content is the highest in mammalian cells (13%) compared to yeast and bacteria (4-10%, BNID 111209, Table 2.1). Remarkably, there are cases where engineered yeast cells accumulated up to 80 % of lipids per cell dry mass [8]. The content of storage carbohydrates varies from around 30% in yeast to 3% in bacteria (Table 2.1). In bacteria, carbohydrates are stored as the polysaccharide glycogen, while yeast cells use glycogen and the disaccharide trehalose. Yeast cells also contain structural polysaccharides, such as mannan and glucan [7]. Bacteria contain the structural molecule peptidoglycan (3% of dry mass) – a polymer of sugars and amino acids, which forms bacterial cell walls. In addition, some bacteria (e.g. *E. coli*) also have lipopolysaccharides on their cell wall (3% of dry mass).

A small fraction of the cell mass (2- 3%) is formed by small molecules ($<$ 1000 Da) such as metabolites and ions. This group contains thousands of different molecules with vastly different functions and concentrations. For illustration, the concentrations of the most abundant metabolites in *E. coli* range from 10^{-1} to 10^{-7} moles per cell, corresponding to a range of 10^8 to only 100 copies per cell [6]. Possibly, there are metabolites with even lower concentrations, but these are much more difficult to quantify. Similarly, the concentrations of the most common inorganic ions (K^+, Na^+, Mg^+, Ca^+, Cl^- span several orders of magnitude [6].

The quantities of biomass components are usually expressed in relation to other quantities. The most common units are copy numbers, moles, grams, or fractions which can be expressed per cell, per gram dry mass, or per cell volume. Membrane components can also be expressed per surface area. Often, experimental data for these quantities is not readily available, so we need to extract it from literature. Useful sources for average or "rule of thumb" values include BioNumbers database [5] and the book Cell Biology by the numbers [6]. Some useful quantities are summarized in Table 2.2 They are organized in increasing order with respect to the dimensions (1 – mass, size, thickness; 2 – area; 3 – volume, density). Notice how the dimensions influence the numerical values. For example, while the cell size differs only about 3-fold between bacteria and yeast, the surface area differs by more than tenfold and the volume by about 60-fold. Because volume grows faster than area, the ratio of cell surface area to volume (SA/V) gets smaller and smaller as cells get bigger (see more in Section 2.4).

2.3.2. Variability of biomass composition

Table 2.1 shows biomass composition of a typical *E. coli* and *S. cerevisiae* cell – these are average values in certain environmental conditions. However, cell size, mass, and composition vary with growth rate and environmental conditions. One of the most extensively studied relationships in the literature is the correlation of growth rate with cell size. The increase of cell mass and volume with growth rate has been observed in bacteria (Figure 2.2), yeast, and mammalian cells [10, 11, 12, 13] (BNID 107948, 110191, 105103). For example, the cell mass of *E. coli* can vary fivefold – 150 to 870 fg per cell for generation times between 100 and 24 minutes [6]. Larger cell mass goes hand in hand with larger amounts of individual biomass components. The absolute amounts of protein, RNA, and DNA increase with cell size. However, the ratios of the components do not stay the same and the relative composition changes with growth rate [10, 14].

One of the most consistent observations is that the relative amount of RNA per cell increases with a higher

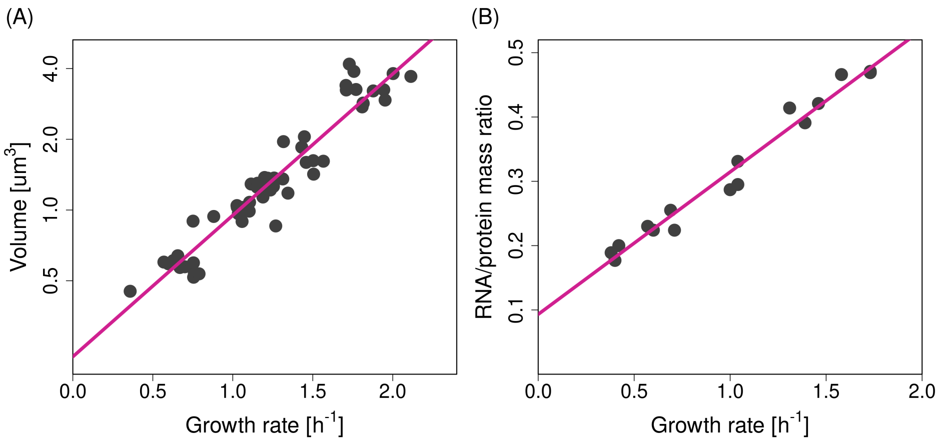

Figure 2.2: Growth laws in *E. coli*. (A) Cell volume grows exponentially with growth rate (data from [15]). (B) RNA/protein ratio grows linearly with growth rate (data from [16]). In both cases, growth rate was varied by changing medium composition.

growth rate [10, 14, 17], (BNID 111460, 111755, 108200). On the other hand, the data for relative protein content is more variable. For example, in bacteria, protein content decreases with growth rate in some studies [14, 17] but goes up and down in another (BNID 111460); in yeast, it increases (BNID 108200, 111755). Nevertheless, when looking at RNA:protein *ratio* we consistently find a positive correlation with growth rate across various species of bacteria (see Figure 2.2) and yeast [16, 18]. RNA:protein ratio is a measure of protein production capacity since most RNA is dedicated to protein synthesis. 80% is rRNA, which forms 2/3 of the mass of a bacterial ribosome – the molecular machine that makes proteins, and 15% is tRNA which brings new amino acids to the ribosome (for more details about ribosomes, see Section 2.7). Indeed, we also observe a correlation between ribosome content and growth rate. The increase of RNA:protein ratio and ribosome content with increasing growth rate reflect higher biosynthetic needs of faster-growing cells. To support higher growth rate, cells need to reallocate resources according to the growth demands (for example, make more ribosomes which can then make more proteins) [19, 18, 16, 20]. For more details about resource allocation and how it is modeled see Chapters 8 in [3] and 9 in [3].

Similarly to protein content, there is no clear correlation between the relative DNA and lipid content with growth rate across studies [14] (BNID 111460, 111755, 108196). The content of storage carbohydrates decreases at higher growth rates in yeast and bacteria [17] (BNID 111755, 111460).

As we have seen, biomass composition changes with growth rate, and for some components, we can describe this relationship with simple mathematical equations [19, 17, 14]. However, the growth rate is a result of environmental conditions (the amount or the quality of a carbon source, temperature, oxygen concentration, presence of inhibitors, and so on), and the same growth rate can be achieved in different ways. However, it may not lead to the same changes in cell physiology [11]. For example, modulation of growth rate by temperature rather than medium composition does not significantly alter cell size, and composition [10, 14]. The inhibition of ribosomes with an antibiotic decreases growth rate but increases the ribosome content (as opposed to reduced ribosome content in a medium with a "worse" carbon source) [16].

Conversely, environmental factors can influence cell composition without affecting growth rate. This shows that cell metabolism is flexible – cells can reach the same growth rate in different ways, depending on the conditions. For example, in yeast, changes of O_2 concentration lead to changes in biomass composition while keeping the growth constant using a chemostat [21]. In mammalian cells, a change of a cultivation

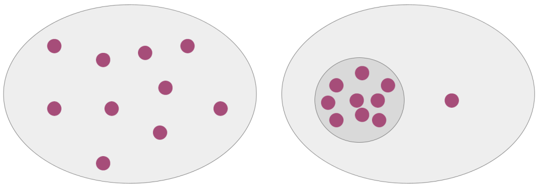

Figure 2.3: Two cells with the same number of molecules per cell but with different concentrations.

medium leads to significant changes in lipid composition without having a considerable effect on growth rate [13]. Genetic background (mutations or a presence of a transgene) can also affect cell characteristics without changing the growth rate [11, 21].

Interestingly, even though the total protein content is variable, the amino acid composition is roughly constant at different growth rates/conditions in bacteria, yeast, and mammalian cells and can even be predicted from a genome sequence with reasonable accuracy [21, 22, 13].

2.3.3. Biomass composition is not uniform

In the previous paragraphs, we considered average cells with a homogeneous composition across the cell. However, we need to keep in mind that cells have an internal structure, and the biomass components are not uniformly distributed throughout the cell (as illustrated in Figure 2.3). Even though prokaryotic cells do not have compartments separated by membranes, they have some internal organization. For example, DNA is not spread across the cytoplasm, but wrapped around proteins and packed in a compact structure called a nucleoid. Another example is the preferential localization of certain proteins on the poles in rod-shaped bacteria. Eukaryotes have compartments with distinct compositions, pH, and membrane potential. DNA is localized only in the nucleus and mitochondria, and many proteins localize only in a particular compartment. Small molecules and ions also have different concentrations in the different compartments. Often they cannot freely diffuse through membranes, but the transport is regulated and requires energy.

These differences in concentrations have implications for cellular functions. Some processes are restricted only to a particular compartment/area. For example, transcription only happens in the nucleus and mitochondria (nucleoid), and some metabolic pathways occur only in a specific compartment (e.g. tricarboxylic acid cycle in the mitochondria). Even if the same enzyme is present in several compartments, it might work at a different rate or in the opposite direction because of the different concentrations of substrates or products. In eukaryotes, certain digestive enzymes only work at low pH present in lysosomes (thus preventing a cell from digesting itself). Sometimes, consecutive enzymes in a metabolic pathway are not freely floating in a cell but form an assembly or bind to a scaffold, allowing intermediates to be channeled directly from one enzyme to another. This accelerates metabolic reactions because intermediates do not diffuse away into the bulk solution and are not consumed by competing reactions.

Finally, we need to zoom out from a single-cell (or average) view of a cell and consider the heterogeneity at the population level. This heterogeneity is often neglected, and we use a single number to describe a concentration of a molecule in a cell/compartment – an average value of the population. However, biological processes are *stochastic* (noisy), and the actual molecule numbers follow a certain distribution (Figure 2.4), which can be characterized by mean and variance. The effect of the heterogeneity becomes especially important at low copy numbers.

The heterogeneity in molecule copy numbers leads to a heterogeneity in cell phenotypes such as generation time, cell size, stress tolerance and others. Population heterogeneity can impact fitness in a positive or

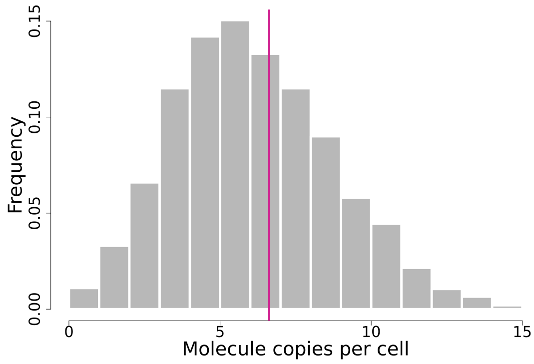

Figure 2.4: Number of molecules per cell in a population (for example protein or mRNA). The red line is the population mean, which is often the value we use (for modeling). However, the values at a single-cell level can differ several fold.

negative way, depending on conditions. For example, when a cell population encounters an unexpected environment, a certain subpopulation might be better suited to survive. In a different environment, another subpopulation might thrive. We can view this as a microbial "bet-hedging" which increases the chances that at least some part of a population will survive the new conditions. However, when cells try to maximize growth rate, the variability in the population can decrease fitness because it decreases the average population growth rate [23]. This topic is discussed in detail in the Chapter 13 in [3].

2.4. Cell size

There is a remarkable variability of cell sizes in nature (Figure 2.5). Figure 2.1 shows the typical sizes of bacterial, yeast and mammalian cells, which range from 1 to 15 µm. However, we can easily find more extreme values. For example, human egg cell has 100 µm (BNID 111184). The smallest known bacteria *Mycoplasma* has only 0.2 µm in diameter (BNID 104717) while the largest bacteria *Thiomargarita magnifica* can reach up to 2 cm [34] which is even more than most mammalian cells. However, this giant bacteria looks very different from typical bacteria like *E. coli* – it has hundreds of thousands of genome copies in organelle-like structures. There are exceptional cases where cells can reach even bigger sizes. The largest known single-celled organism is the alga *Caulerpa taxifolia*. It has many nuclei that are not separated by a membrane, and it reaches up to one meter [35]. Another special case is a neuron – its body has a small diameter (100 µm), but its axons can extend to more than a meter (BNID 109548).

For many organisms, cell size changes with environmental conditions. As already mentioned in Section 2.3.2, cell size varies with growth rate, and it depends on how a particular growth rate is reached. More than 60 years ago, Schaechter et al. discovered the nutrient growth law – cell volume increases exponentially with growth rate (as a result of the nutrient availability in the medium) [10]. Since then, the correlation between

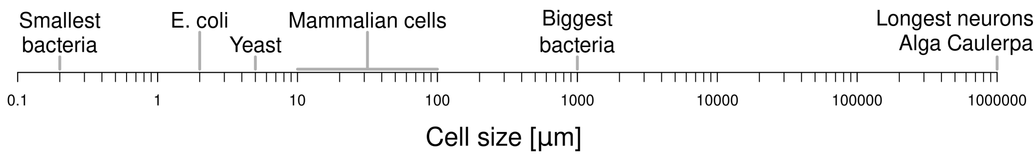

Figure 2.5: Variability of cell size across organisms

Experimental methods 2.B : Experimental quantification of biomass composition

We can measure biomass composition at different levels of detail – from a coarse elemental or macromolecular composition of an average cell to the quantities of individual molecules in each cellular compartment.
To quantify the main chemical elements (CHNOPS), we can use devices called elemental analyzers. The main macromolecular components – the total protein, lipid, carbohydrate, DNA, and RNA content – can be quantified with simple assays such as detection with fluorescent dyes, chemical reactions that lead to color change or extraction and weighing of a component. Going into more detail typically requires more sophisticated methods such as liquid or gas chromatography (LC, GC), mass spectrometry (MS) or nuclear magnetic resonance (NMR). For example, for proteins, we can measure an average amino acid composition, and for lipids, the main lipid classes (glycerophospholipids, sphingolipids, sterols, etc.). For comprehensive reviews and protocols, see for example [24, 25, 26].
If we go down to the level of individual molecules, we enter fields of study collectively termed as *omics*, which aim to characterize and quantify certain pools of biomolecules. Omics methods typically involve high-throughput measurements of hundreds or thousands of different molecules and require a lot of resources (specialized equipment, computational resources) and expertise. The classic omics fields include genomics [27], transcriptomics [28] and proteomics [29] and study the components of the central dogma of molecular biology – DNA, RNA and proteins. Other examples include metabolomics which focuses on small metabolites [30] or fluxomics which measures metabolic fluxes (for example ^{13}C metabolic flux analysis [31]).
Combinations of different omics can help us obtain other parameters that are difficult to measure. For example, turnover numbers of enzymes (k_{cat}) are notoriously difficult to quantify because the measurements are error-prone and low-throughput. With proteomics and fluxomics data we can calculate apparent turnover numbers (k_{app}) at various conditions (see Figure 2.8) and use the maximum value ($k_{\mathrm{app}}^{\mathrm{max}}$) as an estimate of *in vivo* k_{cat} [32, 33].

cell size and growth rate was also observed for other organisms [11, 12, 13] (BNID 107948, 110191, 105103). However, when the growth rate is changed by other means, for example by temperature, this relationship is not observed [10, 14]. In some cases, even the opposite is observed. For example, for a mammalian culture, it was observed that cell division stops at the end of the exponential phase, but cell volume continues to increase threefold [36].

The relationships above refer to an average volume in the population. However, size changes throughout the cell cycle at the level of single cells. Before cells divide, they need to increase their cell size. Otherwise, they would get smaller and smaller with each division. However, they also cannot grow too much, or the average cell size would get bigger and bigger. There are various mechanisms of how cells maintain a cell size homeostasis, and they are discussed in detail in Chapter 11 in [3].

Finally, we need to discuss the importance of cellular shape. Different cell types come in different shapes, such as spheres, ovals, rods, or spirals. Differently shaped cells may have the same volume but very different surface area and surface area to volume ratio (SA/V). Spheres have the lowest possible SA/V while more complicated shapes have higher SA/V (e.g. endoplasmic reticulum). What happens to the shape when a cell changes its volume (for example, in response to environmental conditions)? For many cells, the shape remains roughly the same – for example *E. coli* always looks like a rod. As a result, SA/V decreases when cells get bigger. We see a decreasing linear relationship if we plot SA/V against growth rate. On the other hand, some cells vary their size and shape but maintain a constant condition-specific SA/V. [37]

Experimental methods 2.C : Examples of biomass quantification methods

Component/parameter	Examples of quantification methods
Cell size	microscopically
Dry cell mass	weighing of a defined amount of dry cells
Buoyant density	Percoll gradient
Protein	colorimetric (Bradford assay; Lowry assay)
Lipid	weighing of extracted and dried lipids
Carbohydrates	colorimetric (anthrone assay; phenol-sulphuric acid assay)
RNA	fluorimetric (RiboGreen), spectrophotometric
DNA	fluorimetric (PicoGreen, Hoechst), spectrophotometric
Amino acids/lipid classes	LC/MS, GC/MS
Genomics	next-generation sequencing (NGS) - Illumina, PacBio, Nanopore
Transcriptomics	NGS (RNA-seq), DNA microarrays
Proteomics/metabolomics	LC/MS, GC/MS, NMR

To visualize composition data, consider using Voronoi diagrams instead of the traditional pie charts or bar plots. An online tool is available at bionic-vis.biologie.uni-greifswald.de for proteomics data, but there is also a tool that works with any type of input data (GitLab repository on the book website).

2.5. Cell density

Most cellular parameters we discussed so far – cell size, mass, and composition – vary greatly with the cell type, growth rate, or conditions. However, one quantity does not show such variability – buoyant cell density. Buoyant density is the ratio of cell mass to volume, usually expressed as $\mathrm{g\,mL^{-1}}$. For most organisms, prokaryotic or eukaryotic, the buoyant cell density is around 1.05-1.15 $\mathrm{g\,mL^{-1}}$ [38, 6]. This range results from the fact that cells are 70% water which has a density of $1\,\mathrm{g\,mL^{-1}}$ and that most dry mass is formed by proteins, which have a density of 1.2-1.4 relative to water (BNID 111208, 104272, 101502). Other components range from 1 for lipids (BNID 108142) to 1.4-2 for nucleic acids (BNID 111208). To try the calculation of bacterial density, see Problem 2.4.

For many organisms (*E. coli*, the yeast *Schizosaccharomyces pombe*, Chinese hamster ovary cells, mouse cells), cell density is constant throughout the cell cycle and at different growth rates when growing exponentially. However, it was observed to increase in stationary phase for *E. coli* and *S. pombe* [38, 39]. On the other hand, the density of *S. cerevisiae* fluctuates during the cell cycle, which might be related to a different division mode. The organisms mentioned earlier divide by binary fission – cells divide in the middle and produce two (roughly) identical daughter cells. In contrast, *S.cerevisiae* divides asymmetrically - it grows a bud that breaks away and becomes a smaller daughter cell.

Nevertheless, despite the variability, the range of the observed values is relatively small and similar for most organisms, from bacteria to mammalian cells. There are special cases where cell density deviates from the characteristic values – for example, cells with very high fat content or gas bubbles have lower densities. However, assuming the density of $1.1\,\mathrm{g\,mL^{-1}}$ is probably a good guess unless you work with a particularly fatty or gassy cell type.

The invariability of cell density suggests that this property is highly regulated and brings us to the next question – is there an optimal density? And what are the constraints that (possibly) determine this optimum? These questions (among others) are discussed in the next section.

2.6. The physical constraints of cell growth

The living cells are constantly subject to a handful of so-called *physical constraints*, which are directly linked to the physics and the chemistry of life. Cells cannot override (evolve to bypass) these limits – only try to cope with them. Thus, sometimes these constraints are also called "hard" constraints. Notice that we consider the "hardness" of these constraints only in the space where conditions can still sustain life: some of these limitations could be relaxed by changing abiotic conditions, but would result in breakdown of biological systems.

One of the abiotic factors would be temperature; however, increased temperatures cause proteins to *denature* (lose their 3D-folded structure, thus functionality) and destabilize biological membranes. Although there are organisms, which live in extremely high temperatures (so-called *thermophiles*), as a rule of thumb, we usually consider the temperature above 393 K (120 °C) to be close to the limit of life. There is an organism known as *Strain 121* (*Geogemma barossii*) which can grow at 121 °C (hence the name), currently the highest temperature known [40]. Next, the suboptimal concentration of inorganic salts (osmolarity) or pH could also drive similar changes, disfavoring life. Here we will consider two prominent physical limits in life: the diffusion and density limits. These two limits describe two aspects of how molecules move in aqueous environments, in our case – living cells.

The diffusion limit describes the state where enzymatic catalysis is so specific and so fast that the reaction speed is determined only by the collisions of substrate molecules to the enzymes, which all result in conversions (i.e. no futile collisions) [41]. Usually, the number of futile collisions vary between 1 and 10^4 per successful conversion, and thus having as little futile collisions as possible greatly enhances the overall rate of the reaction. Enzymes approaching (operating at) the diffusion limit are also called *perfect* enzymes. Currently there are no enzymes reported which are considerably "above" diffusion limit (see [41] for an in-depth discussion), suggesting the universality of the underlying constraint. Nonetheless, cells do have a strategy to counter the diffusion limit. Consecutive enzymes from a pathway can be placed on a scaffold, which allows the product of one reaction to be channeled directly into the next reaction without diffusing away.

Another aspect to consider is the density, or sum concentration of molecules, of the fluid. As described in previous sections, cell cytosol contains a spectrum of different molecules at different sizes and concentrations. We normally assume that some sort of optimal cell density that maximizes fitness exists, however, the density is known to fluctuate substantially in time and across conditions [42]. One of the most prevalent properties, linked to cytosolic density, is *macromolecular crowding*. As the name suggests, it describes the concentration of biological macromolecules, mainly proteins, in cytosol (thus in bacteria, the genomic DNA also contributes to molecular crowding). For example, the macromolecular crowding is suggested to impose a limit on the protein translation [43], therefore, increased crowding would result in a growth rate decrease. The state of macromolecular crowding is relevant for the cellular function, and is proposed to be in homeostasis (reviewed in [44]): optimal macromolecular crowding corresponds to a state where crowding reduces the path proteins have to diffuse, yet does not substantially decrease the speed of diffusion. In such a way, maintaining high macromolecular crowding is suggested to maximize reaction rates in the cytosol [45].

2.7. Macromolecule synthesis and the resources needed

Now that we have explored the diversity of nature and abundance of biological molecules, in this and the next section we will consider the coordination of cell components in the biosynthesis of macromolecules. The overall cell growth can be called *self-replication*: a cell makes a copy of itself by synthesizing macromolecules by using molecules it either produces or takes up from the environment, all at right amounts and propor-

Economic analogy 2.D : A bakery

The diversity of metabolic intermediates/end products, stemming from small number of nutrients (e.g. minimal mineral media for yeast growth, containing glucose, ammonium, phosphate and sulphate salts), can be imagined as a bakery. Every pastry starts with a small array of ingredients (flour, water, salt, sugar, ...) and using some machinery (e.g. ovens), one ends up baking bread, pretzels, cookies, muffins etc., which are way diverse in their features, compared to the starting mixture. Likewise, by taking only a handful of compounds, cells, especially microorganisms, can synthesize most of the molecules they need to eventually replicate.

tions. Three essential types of resources are needed for synthesizing the macromolecules: (1) precursors, (2) catalysts, and (3) physical space/volume for the process to happen.

As discussed earlier in the chapter, macromolecules, primarily proteins, are essential for operating metabolic networks. As synthesis of different macromolecule species competes for the same classes of resources, macromolecule synthesis can be altered to change the operational metabolic network - to switch between metabolic strategies. In different conditions, different strategies are superior in the growth they support and the best manner to allocate the limited resources will be preferred. We thus will discuss how these resources are primed and used for macromolecule synthesis, together with different considerations surrounding each type of these resources.

2.7.1. Precursors of macromolecules

Biosynthesis of the macromolecule precursors (e.g. amino acids, nucleotides, energy equivalents) is a major part of every metabolic network. Many microorganisms can grow on a very limited number of nutrients (in the lab context, so-called minimal media), which usually consist of a single source for carbon, nitrogen, phosphorus, and sulfur. For instance, a minimal growth medium with glucose as the sole carbon source can fully support growth: glucose enters glycolysis as the main energy harvesting route, however, some of the glycolytic intermediates serve as substrates for, e.g. amino acid, lipid, or nucleotide biosynthesis.

A particularly interesting fact is that metabolic networks can be described as *bow-tie* structures [46]: a large variety of nutrients can be converted into a very small number (usually counted up to 12) essential metabolic intermediates, which give rise to, again, a diverse set of molecules (for a detailed discussion, see Chapter 3 in [3]). This gives two important insights into metabolic networks. First, this plasticity of the metabolic networks, allows organisms to grow in various environments, where different nutrients are available. Second, due to this organization, the biosynthesis of macromolecule precursors competes for the same starting molecules, independently from the initial nutrients.

2.7.2. Catalysts needed for macromolecule synthesis

Many steps of the biosynthesis of macromolecules, as discussed previously, need catalysis to proceed. Therefore, another kind of investment into macromolecule synthesis is expression of necessary proteins and RNAs (in the latter case - ribosomal RNA). Expression of proteins, starting from transcription of messenger RNAs, their translation into proteins, folding, and degradation, involve many steps with energy investment (ATP hydrolysis) and consume large amounts of precursors (nucleotides, amino acids). Talking in energetic terms alone, protein expression accounts for ca. 40% of energy investments in yeast *S. cerevisiae* [47], and the investments of energy for every stage of protein expression is illustrated in Table 2.3 for typical bacterial and eukaryotic cells. This concerted action of several systems, as described above, with substantial investments at every intermediate step, means that these investments thus happen on two levels: investments in the

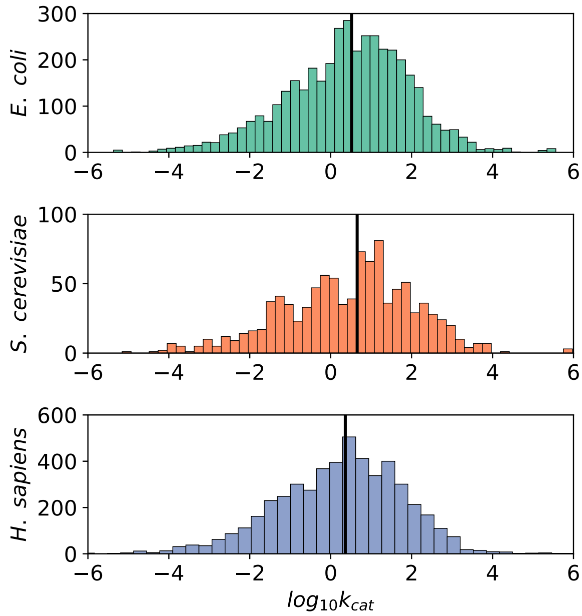

Figure 2.6: Distributions of the $k_{\rm cat}$ and $K_{\rm M}$ values (in s^{-1} and mM, respectively), collected for *E. coli*, yeast and human enzymes. The vertical solid line depicts the median of each distribution. Values were collected from the BRENDA database, release 2022.1 [50].

metabolic machinery and in the machinery, producing proteins themselves. We will consider these two levels in the following.

Metabolic enzymes. First, metabolic enzymes need to be expressed to convert nutrients into biosynthesis precursors. Some enzymes are active only in a form of complexes, which also creates a demand to express proteins at defined ratios. Enzymes and their complexes come in different sizes and flavors, and their activity can be described (in very coarse-grained way, for more details see Chapter 3 in [3]) by two kinetic aspects: the efficacy (represented by the turnover number $k_{\rm cat}$) and substrate specificity (Michaelis constant $K_{\rm M}$) of an enzyme. Importantly, these two parameters are intertwined: high substrate specificity usually comes at the cost of efficacy and vice versa. Therefore, although some enzymes tend towards extremes in terms of their specificity or efficacy, most of the enzymes land close to the average/median values of these parameters, when considering the distribution of enzyme parameters among different organisms [49] (Figure 2.6).

The metabolic networks need to work in a concerted manner, even though different enzymes need to perform different amounts of "work" (described as metabolite *flux* through these enzymes, v). Thus, even given the similarities in "average" (or "moderate") enzyme properties, the expression of proteins and the abundance of their substrates span several orders of magnitude. Based on the kinetic interpretation of enzyme kinetic parameters, we can link them to either expression level of the enzyme e ($e \propto \frac{v}{k_{\rm cat}}$) or substrate concentration s (usually, $0.1K_{\rm M} \leq s \leq 10K_{\rm M}$). Note that for substrate concentrations, the suggested range (order-of-magnitude difference from the $K_{\rm M}$ to each side) is arbitrary, yet supported by empirical observations. On the higher end, the benefit from high substrate concentration becomes negligible (saturation kinetics) as the concentration moves from the order of magnitude of $K_{\rm M}$ (see Exercises for an example). The lower bound of concentrations is defined through high demand of enzymes: in order to sustain flux, a lot of enzyme would have to be produced. As cells have a finite volume to accommodate proteins, such a strategy works only for a very small number of enzymes. Taken together the limitations on the both sides of the spectrum, enzyme

(A) Methionine dropout (B) Complete medium

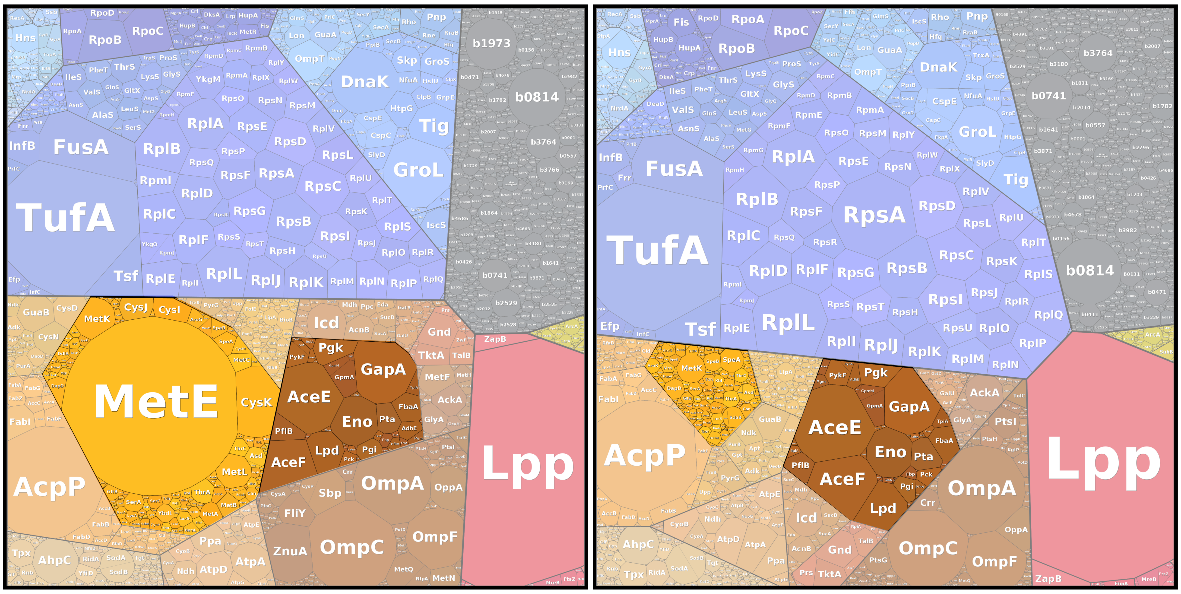

Figure 2.7: Proteome composition of *E. coli*, grown on the growth medium with full amino acid supplement (right) or its version without amino acid methionine (left). Proteome composition data from [51].

kinetics set the bounds for the concentrations of metabolites in the cells.

To illustrate the diversity of enzyme turnover values k_{cat} and the condition-dependent expression of enzymes (dictated by the flux v these enzymes have to sustain), we can consider the proteome composition of *E. coli* under two conditions: growth medium with the complete supplement of amino acids (all 20 proteogenic amino acids present in medium), in contrast to the supplement with a single amino acid missing (a "dropout" medium) (Figure 2.7). The growth of *E. coli* in a nutrient-rich medium (glucose + amino acid supplement) is indeed a very fast one (with doubling time of $\tau_{d,rich} = 21.5 \pm 0.4$ vs. $\tau_{d,minimal} = 56.3 \pm 0.5$ minutes). The omission of methionine from the amino acid supplement does increase the doubling time ($\tau_{d,-Met} = 26.5 \pm 1.1$ minutes), yet the growth rate remains high, and so is the methionine biosynthesis demand in these conditions.

Methionine is an amino acid that is energetically the most expensive to make [52], and the final enzymatic reaction in the methionine synthesis pathway is so-called *rate-limiting*, or the reaction which dictates the flux through the whole pathway. Moreover, the enzyme methionine synthase (*MetE*) is a very slow enzyme (Figure 2.7, table on the bottom), thus required at large quantities to provide enough methionine for protein synthesis at high growth. Consequently, it was observed that *MetE* alone could occupy up to ca. 7.5% of the total proteome (by mass) in medium lacking methionine, and growth on a medium, containing methionine, would reduce the proteome fraction by ca. 800-fold, to 0.009% [51]. To contrast this highly condition-dependent expression of *MetE*, we considered a protein in the lower glycolysis, called enolase *Eno* (Table 2.4). The expression of glycolytic proteins, including *Eno*, was determined to be similar, as both the complete- and the methionine-free media contained glucose as the main carbon source. A noticeable contrast of *Eno* *vs.* *MetE* is also a ca. 3 orders-of-magnitude higher k_{cat} value compared to the one of *MetE*: having to invest less (per mass) into this enzyme contributes to the ability to sustain a very high flux through enolase when cells grow fast on glucose [51] (see Chapter 6 in [3] for a more detailed discussion).

The variable concentrations of metabolic substrates, and their relation to the enzyme parameters (K_{M} in this case), also bring additional kinetic considerations. The above-introduced turnover value k_{cat} represents

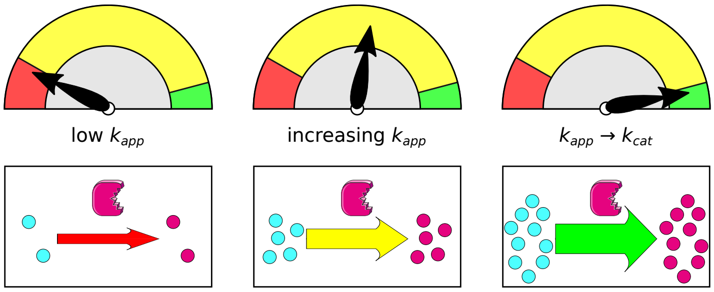

Figure 2.8: The relation between the apparent and measured turnover value(k_{app} and k_{cat}, respectively). Factors, leading to low net rate of reaction per unit of protein (e.g. low substrate concentration) lead to k_{app} being significantly lower than the measured k_{cat} value, latter of which corresponds to the maximal rate of the reaction.

the highest possible efficacy of the enzyme, where all substrates are accessible in concentration needed to sustain this efficacy (also called *saturating* concentrations). Turnover values are usually measured *in vitro*, with all the substrates highly in excess, thus deliberately minimizing many kinetic effects (enzyme saturation, reversibility of reactions etc.) that are prevalent in more physiological conditions (see Chapter 3 in [3] for details). Therefore, what we usually observe in living cells is not the enzyme efficacy in terms of the k_{cat}, but rather their *apparent turnover value* k_{app} (Figure 2.8). The ratio of these values ($\frac{k_{\mathrm{app}}}{k_{\mathrm{cat}}}$) is then called the *enzyme efficiency* and can be used to infer how far away the enzyme is from its optimal working conditions. The k_{app} value of an enzyme *in vivo* can be computed as follows: knowing the k_{cat} value, the flux through the reaction, one can compute the minimal demand (in moles) of the enzyme to sustain that flux. Then, the k_{app} value can be computed by taking the ratio of predicted minimal enzyme demand and the enzyme abundance in the cells.

Macromolecule polymerization. Moving from the metabolic enzymes to the macromolecular synthesis machinery, the polymerization of the macromolecules (DNA replication, RNA transcription and protein translation) are catalyzed by large enzyme (and RNA, in the case of ribosomes) complexes: DNA and RNA polymerases (DNAP, RNAP) and ribosomes. Resources, needed for expressing them also significantly contribute to the total costs of macromolecule biosynthesis. For instance, the molecular weight of an intact ribosome in *E. coli* is ca. 2.3 MDa (BNID 111560), and the *E. coli* ribosome consists of 62% RNA and 38% protein (in mass %, BNID 109047). Meanwhile, eukaryal ribosomes are even larger, ca. 3.3 MDa for *S. cerevisiae* and ca. 4.3 MDa for human *H. sapiens* (BNID 111560), and have higher protein content [53]. For a comparison, the average length of a protein in *E. coli* is ca. 300 amino acids (BNID 100017) and average amino acid weight is ca. 109 Da (BNID 104877). By multiplying these numbers, the molecular mass of an average protein is ca. 32.7 kDa, roughly $70\times$ lower than the ribosome that synthesizes this protein.

The nature of these large complexes requires an exceptional coordination of resources. The first consideration is the number of individual proteins that form these complexes: the RNA polymerases of *S. cerevisiae* contain up to 17 subunits (BNID 111568), and 79 ribosomal proteins form a fully functional ribosome [54]. Therefore, the assembly of these complexes must be fast and robust: thus cells contain a number of assembly factors for facilitating these processes. Next, the coordination also has to be temporal, especially for prokaryotes, where both messenger RNA transcription and protein translation can happen simultaneously. In *E. coli*, this is well illustrated by the 3-fold difference between elongation rates of mRNAs and proteins, ca. 62 $nt\ s^{-1}$ and 21 $aa\ s^{-1}$, respectively (BNID 103021, 107868). This coordination is essential for coordinated transcription and translation happening at the same time [55], as translation happens in steps, 3 nt each (so-called *triplets*). Even in eukaryote *S. cerevisiae* we observe a similar pattern: mRNA elongation rate of ca. 30 $nt\ s^{-1}$

(BNID 103016) and protein chain elongation rate of ca. 10.5 $aa\ s^{-1}$ [56], nearly a $3\times$ difference. Also, the polymerization of macromolecules is very tightly connected to the metabolism: different kinds of growth limitations (limiting amounts of nutrients) were shown to create bottlenecks at different stages of protein expression [57], and the optimal regulation of these processes were selected for by the evolution [58, 59].

2.7.3. Physical proteome space

A final type of asset required for macromolecule synthesis is the physical volume in the cell. As the cells are, again, "bags of things", they possess a finite volume, thus different processes compete for available proteome volume (also called "proteome space" interchangeably). A general trend across microorganisms is that ribosomes occupy larger proteome mass fraction (in the range of 10-40% total proteome) with increasing growth rate [16, 60], with an estimated maximum in *E. coli* of ca. 55% of total proteome mass [16]. Alongside ribosomes, biosynthetic pathways also occupy a substantial share of total proteome (e.g. enzymes, required for amino acid biosynthesis occupy up to 15% of the proteome space in *S. cerevisiae* [60]). Experimentally, the optimal allocation of proteome space can be challenged by, e.g. varying expression of an unneeded (gratuitous) protein. Both for *E. coli* and *S. cerevisiae* it was shown that increasing gratuitous protein expression directly affects the maximal growth rate on both minimal and rich media [61, 57], suggesting that the decrease in growth rate is not dependent on the nutrient status of the cell.

Numbers provided above were measured for cells, grown in minimal medium, and some of the costs we discussed - not only proteome space, but also precursors and enzymes - could be alleviated by growth in rich medium. Uptake of biosynthetic precursors usually is less costly than biosynthesis, as expression of a single type of transporter can substitute the need of expressing a biosynthetic pathway with tens of enzymes associated. Indeed, transfer of *S. cerevisiae* cells to a amino acid-rich growth medium resulted in an increase of growth rate, caused by increased proteome allocation to ribosomes, in place of the proteins of *de novo* amino acid biosynthesis [62]. In conclusion, the physical space that proteins can occupy is also an asset that the proteins are competing for, and thus the optimal allocation of the available space is key for the cells to grow in the most favorable way under specific conditions.

2.8. Concluding remarks

In this chapter, we discussed the properties and the quantities of the main cellular components, how the composition changes in different environmental conditions, and what resources are needed for a cell to replicate itself. It may seem that we already have a vast amount of data, but a lot is still missing. Most available data comes from model organisms such as *E. coli*, *S. cerevisiae*, or humans, but the data for other organisms is still limited. Single-cell data (ideally with subcellular resolution) is also not widely available. Even though we can sequence a genome within a few hours or days, we still do not know the functions of many genes. Many experiments still need to be done, and new high-throughput experimental methods developed to fill the gaps in our knowledge.

Nevertheless, with the basic knowledge from this chapter, we can dive deeper into studying cellular economics and resource allocation with mathematical modeling. How is biomass represented in mathematical models? Often, models only focus on proteome as it is a cell's most abundant and expensive component. However, some models also include other major components (RNA, DNA, lipids, carbohydrates, cofactors, etc.). The components can be modeled at different levels of detail. For example, the cell proteome can be represented simply as a total proteome, divided into protein subgroups (e.g. metabolic, ribosomal, other), or modeled as individual proteins. Finally, there are two contrasting ways to include biomass in mathematical models. On the one hand, some models consider a fixed biomass composition based on measurements or literature (see

Chapters 4 in [3] and 5 in [3]). On the other hand, some models *predict* the biomass composition (i.e. they calculate optimal resource allocation or enumerate all possible compositions, see Chapter 9 in [3]).

Apart from biomass composition, we can include other cellular properties as constraints or parameters in the models, depending on the type of a model and how detailed it is. For example, we can constrain the transcription/translation rates, enzyme turnover rates, cell surface area or volume,

In conclusion, this chapter introduced the basic building blocks of a cell, processes that make them, how they are coordinated and how they depend on environmental conditions. In the next chapters you will learn how to translate this information into mathematical models and how to use them to gain deeper knowledge of cell biology.

Recommended readings

Cell biology

- Alberts, B., et al. (2022). Molecular Biology of the Cell. WW Norton & Co.

Numbers in cell biology

- Moran, U., Phillips, R., & Milo, R. (2010). SnapShot: key numbers in biology. Cell, 141(7), 1262-1262.
- Milo, R., & Phillips, R. (2015). Cell biology by the numbers. Garland Science.

Problems

Problem 2.1 Intuition for biological numbers. Try to answer the following questions, and only then look up the results:

- What is the volume of a cell?
- What is the size of a protein?
- What is bigger, a protein or the mRNA that encoded it?
- How many protein molecules are there in a cell?
- What is the number of genes in a genome?
- How long does it take to transcribe a gene?
- How long does it take to produce a protein molecule?
- What is the minimal doubling time of a cell?
- What other questions come to your mind?

Precise values do not matter here think about orders of magnitude.

Problem 2.2 Proteins per cell - estimate one. How many proteins are there in a bacterial/yeast/mammalian cell [6]? Use data from the following table:

Protein mass per volume	0.2 g mL^{-1}
Molecular mass of a protein	40000 g mol^{-1}
Avogadro's number	$6 \cdot 10^{23}$ 1/mol
E. coli volume	1 μm^3
S. cerevisiae volume	60 μm^3
Mammalian cell volume	3000 μm^3

Problem 2.3 Proteins/ribosomes per cell - estimate two. A typical protein has a volume of 25 nm^3

(BNID 101828) and a ribosome 3400 nm^3 (BNID 104919). Given that 70% of a cell volume is water, what is the maximum number of protein/ribosome molecules that fit into a typical *E. coli* cell (see Table 2.2)? Compare your answers with the previous problem/values in BioNumbers database.

Problem 2.4 Buoyant cell density. Calculate the buoyant density of a typical bacteria using the following data:

Component	Density ($g\,mL^{-1}$)	Mass fraction per cell
Water	1	0.7
Proteins	1.3	0.18
Nucleic acids	1.7	0.08
Lipids	1	0.03
Carbohydrates	1.5	0.01

Problem 2.5 Concentrations enzymes and substrates. Dourado *et al.* [63] suggested that there is a relationship between the concentrations of enzymes and their substrates in *E. coli*, which is a result of a constraint on the biomass density. They showed that the reaction flux is maximal when the dry mass of each substrate is equal to the dry mass of the unsaturated (free) enzyme. What is the concentration of one enzyme per cell for *E. coli* (in $mol\,L^{-1}$)? What would be the optimal concentration of its substrate? Use protein mass and cell volume from Problem 2.2 and the mass of glucose as substrate.

Problem 2.6 Cell size in different dimensions. Imagine a spherical cell that increases its diameter from 1 to 2 um. How much do the surface area, volume, and SA/V change? Think about how this could influence the import of nutrients and the diffusion across the cell.

Problem 2.7 Alien lifeforms. Imagine alien lifeforms. Would they be composed of cells? Why? What features of cells could be completely different? What features are so much dictated by physics that they could not be different in any type of alien cell?

Problem 2.8 Substrate demand to saturate an enzyme. Take the following rate law: $v = v_{max}\frac{S}{K_{\mathrm{M}}+S}$ (also known as irreversible Michaelis-Menten rate law, see Chapter 3 in [3]), where v_{max} is the maximal reaction velocity. Plug in the values for v and compare the substrate concentration needed for the reaction rate to increase from (i) 10% to (ii) 90% of the maximal rate v_{max}. Hint: express the S in terms of K_{M} and take the ratio.

	% of dry mass		Mass per cell [fg]		Molecular mass [Da]		Copy number	
	E. c.	*S. c.*	*E. c.*	*S. c.*	*E.c.*	*S. c.*	*E. c.*	*S. c.*
Proteins	55	51	165	7650	40000	55000	3×10^6	10^8
RNA	20	11	60	1650	10^4-10^6	10^4-10^6	3×10^5	4×10^6
DNA (chromosomal)	3	0.5	9	75	3×10^9	2.5×10^8	2	16
Lipids	9	6	27	900	800	800	2×10^7	10^9
Storage carbohydrates	3	0.5	9	75	10^6	variable	4000	–
Structural polymers	6	23	18	3450	variable	variable	–	–
Metabolites/cofactors	3	2	9	300	< 1000	< 1000	–	–
Other	1	6	3	900	–	–	–	–

Table 2.1: Amounts, characteristic molecular masses and copy numbers of the main biomass components for *Escherichia coli (E. c.)* and *Saccharomyces cerevisiae (S. c.)*. The composition data is shown for *E. c.* with a doubling time of 40 minutes (BNID 104954) and for *S. c.* with a doubling time of 110 minutes ([9], BNID 111755). The storage carbohydrates include glycogen for *E. c.* / glycogen and trehalose for *S. c.*. The structural carbohydrates include peptidoglycan and lipopolysaccharides for *E. c.* / mannan and glucan for *S. c.*. Sources for molecular masses (BNID 105861, 115091, 101838, 104886, 107678, 109645, 102502, 100459); molecule copy numbers (BNID 108248, 108197, 114950).

Name	Unit	*E. coli*	*S. cerevisiae*	BNID/Reference
Surface area/volume (SA/V)	μm^{-1}	6	1.2	calculated here
Dry cell mass	pg	0.3	15	104954, 108315
Total cell mass (with water)	pg	1	60	104954, 108315
Bilayer membrane thickness	nm	4	4	[6]
Cell size	µm	$1-2$	5	[6], 101796
Cell surface area	μm^2	6	70	101792, 113854
Cell volume	μm^3	1	60	101788, 101794
Cell density	$g\,mL^{-1}$	1.1	1.1	103875, 103876

Table 2.2: Useful quantities for unit conversions. Note that these are average or "rule of thumb" values. In reality, these values typically cover a broad range and depend on environmental conditions.

Expression stage	Bacteria	Eukaryotes
DNA synthesis	$101\ L_g$	$263\ L_g$ ($\times 2$ for diploids)
RNA transcription	$2\ N_r\ L_g(23 + \delta_r t)$	$N_r(46 \times L_{r,mat} + 2.17 \times \delta_r t L_{r,pre})$
Protein synthesis	$N_p L_p[(\bar{c}_{AA} - 1) + 5\delta_p t]$	

Table 2.3: The estimated energetic costs (units of ATP hydrolysis) of biosynthesis of a gene, as computed by [48]. The estimates are represented as functions of the following parameters: L_g, gene length; N_r, the steady-state number of mRNAs; $L_{r,pre}$ and $L_{r,mat}$, the length of precursor and mature mRNA, respectively; δ_r, the degradation rate of mRNA; t, division time of a cell; N_p, the steady-state number of protein molecules; L_p, length of the protein chain; $\bar{c}_{AA}$, average cost of an amino acid; δ_p, the degradation rate of proteins.

Pathway	Enzyme	Proteome mass fraction (%)		k_{cat} (s^{-1})
		Met dropout	Complete	
Glycolysis	Enolase (Eno)	0.53	0.53	192.95
Amino acid biosynthesis	Methionine synthase (MetE)	7.45	0.009	0.12

Table 2.4: Abundance and k_{cat} values of two selected proteins from Figure 2.7: enolase (independent on amino acid supplementation) and methionine synthase (dependent on amino acid supplementation).

Bibliography

[1] H. C. Lange and J. J. Heijnen. Statistical reconciliation of the elemental and molecular biomass composition of Saccharomyces cerevisiae. *Biotechnology and Bioengineering*, 75(3):334–344, 2001. doi: 10.1002/bit.10054.

[2] Sarah Cohen, Alex M Valm, and Jennifer Lippincott-Schwartz. Interacting organelles. *Current Opinion in Cell Biology*, 53:84–91, aug 2018. doi: 10.1016/j.ceb.2018.06.003.

[3] The Economic Cell Collective, editor. *Economic Principles in Cell Biology*. Free online book, 2023. doi: 10.5281/zenodo.8156386.

[4] Marian Carlson, Barbara C Osmond, and David Botstein. Mutants of yeast defective in sucrose utilization. *Genetics*, 98(1):25–40, 1981. doi: 10.1093/genetics/98.1.25.

[5] Ron Milo, Paul Jorgensen, Uri Moran, Griffin Weber, and Michael Springer. BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Research*, 38(suppl_1):D750–D753, oct 2010. doi: 10.1093/nar/gkp889.

[6] Ron Milo and Rob Phillips. *Cell biology by the numbers*. Garland Science, 2015. doi: 10.1201/9780429258770.

[7] MT Küenzi and A Fiechter. Regulation of carbohydrate composition of Saccharomyces cerevisiae under growth limitation. *Archiv für Mikrobiologie*, 84(3):254–265, 1972. doi: 10.1007/BF00425203.

[8] Jinpeng Wang, Rodrigo Ledesma-Amaro, Yongjun Wei, Boyang Ji, and Xiao-Jun Ji. Metabolic engineering for increased lipid accumulation in Yarrowia lipolytica - A Review. *Bioresource Technology*, 313: 123707, 2020. doi: 10.1016/j.biortech.2020.123707.

[9] André B. Canelas, Cor Ras, Angela ten Pierick, Walter M. van Gulik, and Joseph J. Heijnen. An in vivo data-driven framework for classification and quantification of enzyme kinetics and determination of apparent thermodynamic data. *Metabolic Engineering*, 13(3):294–306, 2011. doi: 10.1016/j.ymben.2011.02.005.

[10] M. Schaechter, O. Maaloe, and N. O. Kjeldgaard. Dependency on medium and temperature of cell size and chemical composition during balanced grown of Salmonella typhimurium. *Journal of General Microbiology*, 19(3):592–606, 1958. doi: 10.1099/00221287-19-3-592.

[11] Stephen Vadia and Petra Anne Levin. Growth rate and cell size: a re-examination of the growth law. *Current Opinion in Microbiology*, 24:96–103, 2015. doi: 10.1016/j.mib.2015.01.011.

[12] Martí Aldea, Kirsten Jenkins, and Attila Csikász-Nagy. Growth rate as a direct regulator of the start network to set cell size. *Frontiers in Cell and Developmental Biology*, 5:57, 2017. doi: 10.3389/fcell.2017.00057.

[13] Diana Széliová, David E. Ruckerbauer, Sarah N. Galleguillos, Lars B. Petersen, Klaus Natter, Michael Hanscho, Christina Troyer, Tim Causon, Harald Schoeny, Hanne B. Christensen, Dong-Yup Lee, Nathan E. Lewis, Gunda Koellensperger, Stephan Hann, Lars K. Nielsen, Nicole Borth, and Jürgen

Zanghellini. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*, 61:288–300, 2020. doi: 10.1016/j.ymben.2020.06.002.

[14] Hans Bremer and Patrick P. Dennis. Modulation of chemical composition and other parameters of the cell by growth rate. *EcoSal Plus*, 3(1), 2008. doi: 10.1128/ecosal.5.2.3.

[15] Fangwei Si, Dongyang Li, Sarah E. Cox, John T. Sauls, Omid Azizi, Cindy Sou, Amy B. Schwartz, Michael J. Erickstad, Yonggun Jun, Xintian Li, and Suckjoon Jun. Invariance of initiation mass and predictability of cell size in Escherichia coli. *Current Biology*, 27(9):1278–1287, 2017. doi: 10.1016/j.cub.2017.03.022.

[16] Matthew Scott, Carl W Gunderson, Eduard M Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of cell growth and gene expression: origins and consequences. *Science*, 330(6007):1099–1102, 2010. doi: 10.1126/science.1192588.

[17] J. Pramanik and J. D. Keasling. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering*, 56(4):398–421, 1997. doi: 10.1002/(SICI)1097-0290(19971120)56:4<398::AID-BIT6>3.0.CO;2-J.

[18] Tatiana V. Karpinets, Duncan J. Greenwood, Carl E. Sams, and John T. Ammons. RNA:protein ratio of the unicellular organism as a characteristic of phosphorous and nitrogen stoichiometry and of the cellular requirement of ribosomes for protein synthesis. *BMC Biology*, 4(1):30, 2006. doi: 10.1186/1741-7007-4-30.

[19] Allen G Marr. Growth rate of Escherichia coli. *Microbiological reviews*, 55(2):316–333, 1991. doi: 10.1128/mr.55.2.316-333.1991.

[20] Sophia Hsin-Jung Li, Zhiyuan Li, Junyoung O. Park, Christopher G. King, Joshua D. Rabinowitz, Ned S. Wingreen, and Zemer Gitai. Escherichia coli translation strategies differ across carbon, nitrogen and phosphorus limitation conditions. *Nature Microbiology*, 3(8):939–947, 2018. doi: 10.1038/s41564-018-0199-2.

[21] Marc Carnicer, Kristin Baumann, Isabelle Töplitz, Francesc Sánchez-Ferrando, Diethard Mattanovich, Pau Ferrer, and Joan Albiol. Macromolecular and elemental composition analysis and extracellular metabolite balances of Pichia pastoris growing at different oxygen levels. *Microbial Cell Factories*, 8(1):65, 2009. doi: 10.1186/1475-2859-8-65.

[22] J. Pramanik and J. D. Keasling. Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnology and Bioengineering*, 60(2):230–238, 1998. doi: 10.1002/(SICI)1097-0290(19981020)60:2<230::AID-BIT10>3.0.CO;2-Q.

[23] Jie Lin and Ariel Amir. The effects of stochasticity at the single-cell level and cell size control on the population growth. *Cell Systems*, 5(4):358–367.e4, oct 2017. doi: 10.1016/j.cels.2017.08.015.

[24] Ashley E. Beck, Kristopher A. Hunt, and Ross P. Carlson. Measuring cellular biomass composition for computational biology applications. *Processes*, 6(5):38, 2018. doi: 10.3390/pr6050038.

[25] Christopher P Long and Maciek R Antoniewicz. Quantifying biomass composition by gas chromatography/mass spectrometry. *Analytical chemistry*, 86(19):9423–9427, 2014. doi: 10.1021/ac502734e.

[26] Diana Széliová, Harald Schoeny, Špela Knez, Christina Troyer, Cristina Coman, Evelyn Rampler, Gunda Koellensperger, Robert Ahrends, Stephen Hann, Nicole Borth, Jürgen Zanghellini, and David E. Ruckerbauer. Robust analytical methods for the accurate quantification of the total biomass composition of mammalian cells. *Methods in Molecular Biology (Clifton, N.J.)*, 2088:119–160, 2020. doi: 10.1007/978-1-0716-0159-4_7.

[27] Jay Shendure, Shankar Balasubramanian, George M. Church, Walter Gilbert, Jane Rogers, Jeffery A. Schloss, and Robert H. Waterston. DNA sequencing at 40: past, present and future. *Nature*, 550 (7676):345–353, 2017. doi: 10.1038/nature24286.

[28] Stark R, Grzelak M, and Hadfield J. RNA sequencing: the teenage years. *Nature reviews. Genetics*, 20 (11), 2019. doi: 10.1038/s41576-019-0150-2.

[29] Winston Timp and Gregory Timp. Beyond mass spectrometry, the next step in proteomics. *Science Advances*, 6(2):eaax8978, 2020. doi: 10.1126/sciadv.aax8978.

[30] Saleh Alseekh, Asaph Aharoni, Yariv Brotman, Kévin Contrepois, John D'Auria, Jan Ewald, Jennifer C. Ewald, Paul D. Fraser, Patrick Giavalisco, Robert D. Hall, Matthias Heinemann, Hannes Link, Jie Luo, Steffen Neumann, Jens Nielsen, Leonardo Perez de Souza, Kazuki Saito, Uwe Sauer, Frank C. Schroeder, Stefan Schuster, Gary Siuzdak, Aleksandra Skirycz, Lloyd W. Sumner, Michael P. Snyder, Huiru Tang, Takayuki Tohge, Yulan Wang, Weiwei Wen, Si Wu, Guowang Xu, Nicola Zamboni, and Alisdair R. Fernie. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. *Nature Methods*, 18(7):747–756, 2021. doi: 10.1038/s41592-021-01197-1.

[31] Christopher P. Long and Maciek R. Antoniewicz. High-resolution 13C metabolic flux analysis. *Nature Protocols*, 14(10):2856–2877, 2019. doi: 10.1038/s41596-019-0204-0.

[32] Dan Davidi, Elad Noor, Wolfram Liebermeister, Arren Bar-Even, Avi Flamholz, Katja Tummler, Uri Barenholz, Miki Goldenfeld, Tomer Shlomi, and Ron Milo. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proceedings of the National Academy of Sciences*, 113(12):3401–3406, 2016. doi: 10.1073/pnas.1514240113.

[33] David Heckmann, Anaamika Campeau, Colton J. Lloyd, Patrick V. Phaneuf, Ying Hefner, Marvic Carrillo-Terrazas, Adam M. Feist, David J. Gonzalez, and Bernhard O. Palsson. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proceedings of the National Academy of Sciences of the United States of America*, 117(37):23182–23190, 2020. doi: 10.1073/pnas.2001562117.

[34] Jean-Marie Volland, Silvina Gonzalez-Rizzo, Olivier Gros, Tomáš Tyml, Natalia Ivanova, Frederik Schulz, Danielle Goudeau, Nathalie H. Elisabeth, Nandita Nath, Daniel Udwary, Rex R. Malmstrom, Chantal Guidi-Rontani, Susanne Bolte-Kluge, Karen M. Davies, Maïtena R. Jean, Jean-Louis Mansot, Nigel J. Mouncey, Esther R. Angert, Tanja Woyke, and Shailesh V. Date. A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. *Science*, 376(6600):1453–1458, 2022. doi: 10.1126/science.abb3634.

[35] William P Jacobs. Caulerpa. *Scientific American*, 271(6):100–105, 1994. URL http://www.jstor.org/stable/24942951.

[36] Xiao Pan, Ciska Dalm, René H. Wijffels, and Dirk E. Martens. Metabolic characterization of a CHO cell size increase phase in fed-batch cultures. *Applied Microbiology and Biotechnology*, 101(22):8101–8113, 2017. doi: 10.1007/s00253-017-8531-y.

[37] Leigh K. Harris and Julie A. Theriot. Surface area to volume ratio: a natural variable for bacterial morphogenesis. *Trends in Microbiology*, 26(10):815–832, 2018. doi: 10.1016/j.tim.2018.04.008.

[38] Herbert E. Kubitschek. Buoyant density variation during the cell cycle in microorganisms. *CRC Critical Reviews in Microbiology*, 14(1):73–97, 1987. doi: 10.3109/10408418709104436.

[39] H. E. Kubitschek and R. A. Ward. Buoyant density constancy of Schizosaccharomyces pombe cells. *Journal of Bacteriology*, 162(3):902–904, 1985. doi: 10.1128/jb.162.3.902-904.1985.

[40] Kazem Kashefi and Derek R. Lovley. Extending the upper temperature limit for life. *Science*, 301(5635): 934–934, 2003. doi: 10.1126/science.1086823.

[41] Dan Davidi, Liam M. Longo, Jagoda Jabłońska, Ron Milo, and Dan S. Tawfik. A bird's-eye view of enzyme evolution: Chemical, physicochemical, and physiological considerations. *Chemical Reviews*, 118 (18):8786–8797, 2018. doi: 10.1021/acs.chemrev.8b00039.

[42] Pascal D Odermatt, Teemu P Miettinen, Joël Lemière, Joon Ho Kang, Emrah Bostan, Scott R Manalis, Kerwyn Casey Huang, and Fred Chang. Variations of intracellular density during the cell cycle arise from tip-growth regulation in fission yeast. *eLife*, 10:e64901, 2021. doi: 10.7554/eLife.64901.

[43] Stefan Klumpp, Matthew Scott, Steen Pedersen, and Terence Hwa. Molecular crowding limits translation and cell growth. *Proceedings of the National Academy of Sciences*, 110(42):16754–16759, 2013. doi: 10.1073/pnas.1310377110.

[44] Jonas van den Berg, Arnold J. Boersma, and Bert Poolman. Microorganisms maintain crowding homeostasis. *Nature Reviews Microbiology*, 15(5):309–318, 2017. doi: 10.1038/nrmicro.2017.17.

[45] Ken A. Dill, Kingshuk Ghosh, and Jeremy D. Schmit. Physical limits of cells and proteomes. *Proceedings of the National Academy of Sciences*, 108(44):17876–17882, 2011. doi: 10.1073/pnas.1114477108.

[46] Tamar Friedlander, Avraham E. Mayo, Tsvi Tlusty, and Uri Alon. Evolution of bow-tie architectures in biology. *PLOS Computational Biology*, 11(3):e1004055, mar 2015. doi: 10.1371/journal.pcbi.1004055.

[47] Petri-Jaan Lahtvee, Benjamín J. Sánchez, Agata Smialowska, Sergo Kasvandik, Ibrahim E. Elsemman, Francesco Gatto, and Jens Nielsen. Absolute quantification of protein and mRNA abundances demonstrate variability in gene-specific translation efficiency in yeast. *Cell Systems*, 4(5):495–504.e5, may 2017. doi: 10.1016/j.cels.2017.03.003.

[48] Michael Lynch and Georgi K. Marinov. The bioenergetic costs of a gene. *Proceedings of the National Academy of Sciences*, 112(51):15690–15695, nov 2015. doi: 10.1073/pnas.1514974112.

[49] Arren Bar-Even, Elad Noor, Yonatan Savir, Wolfram Liebermeister, Dan Davidi, Dan S. Tawfik, and Ron Milo. The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, may 2011. doi: 10.1021/bi2002289.

[50] Antje Chang, Lisa Jeske, Sandra Ulbrich, Julia Hofmann, Julia Koblitz, Ida Schomburg, Meina Neumann-Schaal, Dieter Jahn, and Dietmar Schomburg. Brenda, the elixir core data resource in 2021: new developments and updates. *Nucleic Acids Research*, 49(D1):D498–D508, 2021. doi: 10.1093/nar/gkaa1025.

[51] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, 2014. doi: 10.1016/j.cell.2014.02.033.

[52] Christoph Kaleta, Sascha Schäuble, Ursula Rinas, and Stefan Schuster. Metabolic costs of amino acid and protein production in Escherichia coli. *Biotechnology Journal*, 8(9):1105–1114, 2013. doi: 10.1002/biot.201200267.

[53] Daniel N Wilson and Jamie H Doudna Cate. The structure and function of the eukaryotic ribosome. *Cold Spring Harbor perspectives in biology*, 4(5):a011536, 2012. doi: 10.1101/cshperspect.a011536.

[54] John L Woolford and Susan J Baserga. Ribosome biogenesis in the yeast Saccharomyces cerevisiae. *Genetics*, 195(3):643–681, nov 2013. doi: 10.1534/genetics.113.153197.

[55] Sergey Proshkin, A. Rachid Rahmouni, Alexander Mironov, and Evgeny Nudler. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, 328(5977):504–508, apr 2010. doi: 10.1126/science.1184939.

[56] Eyal Metzl-Raz, Moshe Kafri, Gilad Yaakov, Ilya Soifer, Yonat Gurvich, and Naama Barkai. Principles of cellular resource allocation revealed by condition-dependent proteome profiling. *eLife*, 6, aug 2017. doi: 10.7554/eLife.28034.

[57] Moshe Kafri, Eyal Metzl-Raz, Ghil Jona, and Naama Barkai. The cost of protein production. *Cell Reports*, 14(1):22–31, jan 2016. doi: 10.1016/j.celrep.2015.12.015.

[58] Jan Berkhout, Evert Bosdriesz, Emrah Nikerel, Douwe Molenaar, Dick de Ridder, Bas Teusink, and Frank J Bruggeman. How biochemical constraints of cellular growth shape evolutionary adaptations in metabolism. *Genetics*, 194(2):505–512, 2013. doi: 10.1534/genetics.113.150631.

[59] Arijit Maitra and Ken A Dill. Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proceedings of the National Academy of Sciences*, 112(2):406–411, 2015. doi: 10.1073/pnas.1421138111.

[60] Ibrahim E Elsemman, Angelica Rodriguez Prado, Pranas Grigaitis, Manuel Garcia Albornoz, Victoria Harman, Stephen W Holman, Johan van Heerden, Frank J Bruggeman, Mark MM Bisschops, Nikolaus Sonnenschein, et al. Whole-cell modeling in yeast predicts compartment-specific proteome constraints that drive metabolic strategies. *Nature Communications*, 13(1):1–12, 2022. doi: 10.1038/s41467-022-28467-6.

[61] Erez Dekel and Uri Alon. Optimality and evolutionary tuning of the expression level of a protein. *Nature*, 436(7050):588–592, July 2005.

[62] Johan Björkeroth, Kate Campbell, Carl Malina, Rosemary Yu, Francesca Di Bartolomeo, and Jens Nielsen. Proteome reallocation from amino acid biosynthesis to ribosomes enables yeast to grow faster in rich media. *Proceedings of the National Academy of Sciences*, 117(35):21804–21812, 2020. doi: 10.1073/pnas.1921890117.

[63] Hugo Dourado, Matteo Mori, Terence Hwa, and Martin J Lercher. On the optimality of the enzyme–substrate relationship in bacteria. *PLoS biology*, 19(10):e3001416, 2021. doi: 10.1371/journal.pbio.3001416.

Solutions to problems

Problem 1 (Intuition for biological numbers)

Check the results at http://bionumbers.hms.harvard.edu, http://book.bionumbers.org/ or https://doi.org/10.1016/j.cell.2010.06.019.

Problem 2 (Proteins per cell - estimate one)

$$\frac{\text{Proteins}}{\mathsf{mL}} = 0.2\frac{\mathsf{g}}{\mathsf{mL}} \cdot 6 \cdot 10^{23}\frac{1}{\mathsf{mol}} \cdot \frac{1}{40000}\frac{\mathsf{mol}}{\mathsf{g}} = 3 \cdot 10^{18}\frac{1}{\mathsf{mL}}$$

$$\frac{\text{Proteins}}{\mu\mathsf{m}^3} = 3 \cdot 10^{18}\frac{1}{\mathsf{mL}} \cdot 10^{-12}\frac{\mathsf{mL}}{\mu\mathsf{m}^3} = 3 \cdot 10^{6}\frac{1}{\mu\mathsf{m}^3}$$

$$\frac{\text{Proteins}}{\text{cell}} \approx \begin{cases} 3 \cdot 10^6 & \textit{E. coli} \\ 2 \cdot 10^8 & \textit{S. cerevisiae} \\ 9 \cdot 10^9 & \text{mammalian cells} \end{cases}$$

Problem 3 (Proteins/ribosomes per cell - estimate two)

$10^9 \cdot 0.3/25 \approx 1.2 \cdot 10^7$ proteins
$10^9 \cdot 0.3/3400 \approx 88000$ ribosomes

Problem 4 (Buoyant cell density)

$\text{Density} = 1 \cdot 0.7 + 1.3 \cdot 0.18 + 1.7 \cdot 0.08 + 1 \cdot 0.03 + 1.5 \cdot 0.01 = 1.115$

Problem 5 (Concentrations enzymes and substrates)

Concentration of one molecule per *E. coli* cell:

$$\frac{1}{\mu\mathsf{m}^3} \cdot 10^{15}\frac{\mu\mathsf{m}^3}{\mathsf{L}} \cdot \frac{1}{6 \cdot 10^{23}}\frac{\mathsf{mol}}{1} = 1.7 \cdot 10^{-9}\frac{\mathsf{mol}}{\mathsf{L}}$$

Enzyme mass: 40 000 g mol^{-1} (BNID 105861)
Glucose mass: 180 g mol^{-1}

$40000/180 = 222$

Assuming that the mass of the free enzyme and the mass of the substrate should be equal for optimal flux, we need approximately 222 molecules of substrate (glucose) per one molecule of free enzyme. This corresponds to the following concentrations:

Enzyme: $\sim$ 1.7 nmol L^{-1}
Substrate: $\sim$ 370 nmol L^{-1}

Problem 6 (Cell size in different dimensions)

	Cell 1	Cell 2	Ratio
Diameter	1	2	2
Surface area	3.14	12.57	4
Volume	0.52	4.19	8
SA/V	6	3	0.5

As the cell diameter increases, the volume grows faster than the surface area, leading to a decrease in the surface area-to-volume ratio (SA/V). This decrease can impact nutrient uptake as there is less relative surface area available for membrane transporters. Therefore, larger cells with lower SA/V ratios may face challenges in obtaining enough nutrients to meet their metabolic demands.

Problem 7 (Alien lifeforms)
We do not know the answer to this question. These are ideas from ChatGPT (https://openai.com, GPT-3.5, version 2023) – take them with a grain of salt:

When imagining alien lifeforms, it is difficult to make definitive statements about their nature since we have no direct knowledge of extraterrestrial life. However, we can explore some possibilities and consider the principles of biology and physics as we currently understand them.

1. Composition of Cells. While cells are the fundamental building blocks of life on Earth, it is not certain that extraterrestrial life would be composed of cells. Cells are a product of evolution on our planet and are highly efficient structures for carrying out biological processes. However, alternative forms of life may exist with entirely different structures and organization. For example, hypothetical non-cellular life forms based on self-replicating molecules or other complex arrangements have been proposed.

2. Different Features of Alien Cells. (a) Membranes: While cells on Earth have lipid bilayer membranes, it is conceivable that alien life forms might have different membrane compositions. For example, they could have membranes composed of different molecules or organized in alternative ways. (b) Genetic Material: On Earth, DNA and RNA are the primary carriers of genetic information. However, it is possible that alternative genetic molecules or information storage systems could exist in alien life. These could have different base pairs or use different coding mechanisms. (c) Metabolism: Alien life could have different metabolic processes compared to Earth organisms. They might utilize different energy sources, employ unique enzymatic reactions, or even rely on completely novel biochemical pathways. (d) Size and Structure: Cells on Earth exhibit a wide range of sizes, from microscopic bacteria to the largest known cells in organisms like ostrich eggs. It is conceivable that alien cells could differ significantly in size and overall structure, depending on the specific conditions and evolutionary paths of their respective environments.

3. Features Dictated by Physics. Certain fundamental principles of physics are likely to impose constraints on the functioning and structure of any kind of cell, including potential alien cells. These features include: (a) Biochemistry: Regardless of the specific molecular composition, alien cells would need a biochemistry that allows for the storage and utilization of energy, the replication and expression of genetic information, and the maintenance of internal equilibrium. (b) Thermodynamics: The laws of thermodynamics, such as energy conservation and entropy increase, are universal physical principles. Any living system, including alien cells, would need to adhere to these principles to maintain their internal processes. (c) Water: Water is a highly abundant molecule and a fundamental solvent for life on Earth. It provides a medium for biochemical reactions and allows for efficient transport of molecules within cells. It is possible that water or another suitable liquid would be essential for alien life, but alternative solvents cannot be ruled out entirely.

While these considerations provide a starting point for thinking about alien life, the possibilities are vast, and it is challenging to predict the specific characteristics of extraterrestrial organisms. Future discoveries and investigations in astrobiology will help refine our understanding of life beyond Earth.

Problem 8 (Substrate demand to saturate an enzyme) Take the irreversible Michaelis-Menten law, and plug in $v = 0.1 \times v_{max}$ and $v = 0.9 \times v_{max}$. By rearranging the terms to express S in terms of K_M, the answers are $S_{0.1} = \frac{0.1}{0.9} K_\mathrm{M} \approx 0.11 K_\mathrm{M}$ and $S_{0.9} = \frac{0.9}{0.1} K_\mathrm{M} = 9 K_\mathrm{M}$. This is approx. 81-fold difference to go from $0.1 \times v_{max}$ to $0.9 \times v_{max}$!