Geometrically balanced model of cell growth

Alexei Vazquez[1] and Tomáš Gedeon[2]

[1]ORCID 0000-0003-2764-3244
Nodes & Links Ltd, Salisbury House, Station Road, Cambridge CB1 2LA, UK
[2]ORCID 0000-0001-5555-6741
Department of Mathematical Sciences, Montana State University, Bozeman, MT USA
Correspondence: avazque1@protonmail.com, tgedeon@montana.edu

Abstract

The proteome balance constraint in metabolic flux balance analysis asserts that the proteome is constructed by ribosomes, which themselves contain many proteins. This leads to a fundamental question of optimal allocation of limited proteome among different pools of enzymes, which include ribosomes themselves. However, recent work points to additional constraints imposed by the cell geometry. In this paper we deduce the *proteogeometric constraint* $\bar{\pi}_A = \pi_A + \theta\pi_L/\pi_P$**, where** π_A**,** π_P **and** π_L **are the proteomic fractions allocated to the cell surface area, protein synthesis and cell membrane phospholipids synthesis and** $\bar{\pi}_A$ **and** θ **are constants imposed by geometry of the cell. We illustrate the relevance of this constraint using a reduced model of cell metabolism, illuminating the interplay between cell metabolism and cell geometry.**

1 Introduction

In the last 20 years we have advanced our understanding of physical and chemical constraints shaping the metabolic capabilities of cells. It all started with mass balance, underpinning the creation of flux balance models [1, 2, 3, 4, 5, 6, 7, 8]. It was followed by molecular crowding of the cell volume [9, 10, 11], the proteome balance constraint [12] and molecular crowding of the cell membrane [13]. Today the proteomic balance constraint is well stablished and used in most studies [14, 15, 16, 17, 18, 19, 20, 21]. However, recent publications have highlighted the need for a deeper understanding of the mode of operation of the crowding constraints of the cell volume and the cell membrane area [22, 23, 24].

At the intuitive level, making more cell membrane implies a metabolic cost and a natural selection for increased molecular crowding [23]. At the same time, nutrient transporters and other metabolic enzymes are located in the cell membrane, which would favor larger cellular membrane, or at least imply a requirement of a minimal area of membrane that can support required growth rate [13, 25, 24]. The surface area that holds the cell membrane components also encloses the volume which contains all cell components. Finally, these allocation processes need to happen under the self-replicative nature of protein synthesis: ribosomes make new ribosomes and synthesis of any other protein takes away resources from ribosome synthesis, which is the basics of the proteomic constraint [15].

Here we derive the basic equation for the volume to area balance that is consistent with the metabolic requirements of a cell. We also analyze a reduced model focusing on interpretation. In spite of its simplicity, this reduced model predicts proteomic fractions that are close to those observed for the bacterium *E. coli* and brings new insights into the interplay between molecular crowding and the cell membrane area to volume ratio under different conditions. Our work demonstrates that the cell volume-area constraint is independent of the proteome allocation constraint and should be included in flux balance models of cell metabolism.

2 Volume to area balance

To derive a volume to area balance constraint we focus on major macromolecular synthesis processes. That includes protein synthesis driven by ribosomes, RNA nucleotide synthesis catalyzed by RNA polymerase and membrane synthesis by some membrane synthesis pathway. We denote by $\mathcal{P}$, $\mathcal{N}$ and $\mathcal{L}$ the set of all metabolic enzymes required for the biosynthesis of proteins, RNA nucleotides and cell surface area lipids. We further denote by $\mathcal{M} = \{\mathcal{P}, \mathcal{N}, \mathcal{L}, ...\}$ the full set of enzymes, including any additional enzymes that will be used in our cell metabolism model.

Our core flux balance equations reflect protein, RNA and membrane lipid balance

$$\lambda \sum_{i \in \mathcal{M}} n_i E_i = k_P E_P, \tag{1}$$

$$\lambda N = k_N E_N, \tag{2}$$

$$\lambda L = k_L E_L \tag{3}$$

where λ is the growth rate, E_i is the moles of i-pathway enzymes per cell, n_i is the moles of amino acids per mole of enzymes of pathway i, N the number of RNA nucleotides per cell, L are the moles of lipid bilayer units per cell, and k_i is the i-th pathway turnover number.

These flux balance equations are complemented by the geometric balance equations for cell volume V and cellular surface area A

$$V = V_W + \sum_{i \in \mathcal{M}} v_i E_i + v_n N, \tag{4}$$

$$A = a_L L + \sum_{i \in \mathcal{M}} a_i E_i, \tag{5}$$

where V_W is the volume of water, v_i the molar volume of enzyme i, v_n the RNA volume per RNA nucleotide and a_i the molar cell membrane area occupied by enzyme i, in units of area per mole of enzyme. Note that $a_i = 0$ for enzymes that do not localize to the cell membrane.

We now introduce important quantities that will simplify our model. We measure the total protein content in units of number of amino acids in proteins and denote its magnitude by

$$P = \sum_{i \in \mathcal{M}} n_i E_i,$$

the macromolecular volume fraction by

$$\phi = \frac{V - V_W}{V},$$

the RNA to protein ratio by

$$N = \nu P,$$

the proteomic fractions by

$$\pi_i = \frac{n_i E_i}{P},$$

and the proteomic fraction located at the membrane by

$$\pi_A = \sum_{i \in \mathcal{M}} \sigma_i \pi_i, \tag{6}$$

where the the subscript A stands for the cell area and σ_i is the fraction of the i-th protein pool that is allocated to the membrane.

We assume that the molar volumes and areas of enzymes scale with their size quantified in number of amino acids

$$v_i \approx v_a n_i, \tag{7}$$

$$a_i \approx a_a n_i, \tag{8}$$

where v_a and a_a are the typical i-th pool protein volume and area per amino acid. Deviations from this scaling are negligible compared to the uncertainty about the enzyme turnover numbers. The scaling approximations (7) and (8) simplify the analysis significantly. For example

$$\sum_{i \in \mathcal{M}} v_i E_i \approx v_a \sum_{i \in M} n_i E_i = v_a P$$

$$\sum_{i \in \mathcal{M}} a_i E_i \approx a_a \sum_{i \in A} n_i E_i = a_a \pi_A P.$$

We should bear in mind these are approximations. A more precise model should take into account the variability of the specific protein volume and area across proteins. These numbers are not available in general, especially for the protein area requirements. In that case we would need to work directly with the constraints imposed by equations (4) and (5).

Finally, taking into account that

$$k_i E_i / P = (k_i / n_i) n_i E_i / P = (k_i / n_i) \pi_i,$$

we define the specific turnover rates

$$\lambda_i = \frac{k_i}{n_i}.$$

Note that we define the specific rates per unit of amino acid, because we count protein content in units of number of amino acids. That is in contrast to the more common use of specific rates per protein mass. We argue that the specific rates per unit of amino acid have better measurement units for interpretation (1/time). In fact, as illustrated below, λ_i can be interpreted as the cell growth rate if the pathway i would be limiting.

The equation (1) can be then rewritten as

$$\lambda = \lambda_P \pi_P, \tag{9}$$

and the equation (2) as

$$\pi_N = \nu \frac{\lambda_P}{\lambda_N} \pi_P. \tag{10}$$

Using (9) the membrane balance equation (3) becomes

$$\frac{L}{P} = \frac{\lambda_L}{\lambda_P} \frac{\pi_L}{\pi_P}, \tag{11}$$

Finally, the volume balance and area balance equations (4)-(5) become

$$V = \frac{1}{\phi} v_a \left(1 + \nu \frac{v_n}{v_a}\right) P, \tag{12}$$

$$A = \left(a_L \frac{L}{P} + a_a \pi_A\right) P. \tag{13}$$

Our aim is now to obtain a membrane-volume balance equation from the equation (13). Dividing by P and using (11) we get

$$\begin{aligned} \frac{A}{P} &= a_L \frac{L}{P} + a_a \pi_A \\ \frac{1}{a_a}\frac{A}{P} &= \frac{a_L}{a_a}\frac{L}{P} + \pi_A \\ \frac{1}{a_a}\frac{A}{P} &= \theta \frac{\pi_L}{\pi_R} + \pi_A \end{aligned}$$

where in the last equation we introduced

$$\theta := \frac{a_L}{a_a}\frac{\lambda_L}{\lambda_P}.$$

At this point we introduce a surface to volume scaling. Our cell population level model operates on time scales longer than the cell cycle. Because of this longer time scale we do not take into account geometrical changes during the cell cycle and assume that the cell's geometry does not change. Given that the cell volume to area ratio is a key variable in bacterial morphogenesis [26], we parametrize the cell geometry by the cell volume to area ratio

$$l = V/A.$$

With this scaling and using (12) we obtain the final equation

$$\bar{\pi}_A = \theta \frac{\pi_L}{\pi_P} + \pi_A, \tag{14}$$

where we again introduced a new constant on the left hand side

$$\bar{\pi}_A := \frac{1}{\phi}\frac{l_P}{l}\left(1 + \nu \frac{v_n}{v_a}\right) \tag{15}$$

where

$$l_P := \frac{v_a}{a_a},$$

is the typical volume to area ratio of proteins.

We arrive at our final model in which the membrane-volume balance equation (14) is complemented by the growth rate equation, from (9),

$$\lambda = \lambda_P \pi_P, \tag{16}$$

RNA polymerase proteome fraction, from equation (10),

$$\pi_N = \nu \frac{\lambda_P}{\lambda_N} \pi_P,$$

and the proteome balance

$$1 = \sum_{i \in \mathcal{M}} \pi_i. \tag{17}$$

The last three equations are part of a standard proteome constraint approach. The membrane-volume balance equation (14) is the new constraint.

3 General observations

We call equation (14) the *proteogeometric constraint*, as it relates proteomic fractions and the cell geometry. It also serves as a self consistency check between geometry and macromolecular composition. We can reach important conclusions by inspecting this equation. First, it is obvious that the proteomic fraction at the cell membrane is bounded by the cell geometry: $\pi_A \leq \bar{\pi}_A$. If we specify the linear dimension of proteins l_P and the cell l and the macromolecular volume fraction ϕ, then $\bar{\pi}_A$ gives us the maximum proteomic fraction that can be located in the cell surface area. For example, based on that for the PTS glucose transporter of *E. coli*, $l_P \sim 0.009$ μm, cell volume/area measurements for *E. coli* report $l \sim 0.15$ and the macromolecular volume fraction is about $\phi \sim 0.4$ (Supplementary Information), resulting in $\bar{\pi}_A \approx 0.2$. That is, only about 20% of the cell proteome of *E. coli* can be located at the cell membrane. Experimental values are scattered around 20% [27, 28], indicating that the upper bound $\pi_A \leq \bar{\pi}_A$ suggested by our simple model is correct.

The second observation is the role of the proteomic fraction π_L of the membrane lipids synthesis pathway, which enters the equation divided by the proteomic fraction of the protein synthesis pathway π_P. The numerator π_L accounts for membrane lipids that are required to have a closed surface area, and the denominator π_P describes the bulk of proteins contained within the cell membrane area. This introduces a non-linearity, in contrast with previously used assumptions of linear constraints in the proteomic fractions. In fact, finding the optimal metabolic rates under the proteogeometric constraint becomes a quadratic optimization problem. This quadratic programming problem is convex because it contains only one non-linear term ($\pi_L \pi_P$). Therefore, its numerical solution scales linearly with the number of proteomic fractions in the model. That is relevant when applying the proteogeometric constraint to genome scale models.

Finally, the proteogeometric constraint (14), which emerged from the area constraint (5), that describes what fraction of the cell membrane is occupied by proteins or by lipids. In the right hand side, π_A is a scaled version of the cell membrane fraction occupied by proteins, while the remaining $\bar{\pi}_A - \pi_A$ is is a scaled version of the area occupied by membrane lipids. Therefore, we can use the ratio

$$A_P = \frac{\pi_A}{\bar{\pi}_A}, \tag{18}$$

to quantify the cell membrane fraction occupied by proteins.

4 Simple model with energy balance

To illustrate the relevance of the volume-area balance constraint, we consider a reduced model with four enzymes/pathways: $\mathcal{M}^* = \{\mathcal{R}, \mathcal{N}, \mathcal{M}, \mathcal{E}\}$, where $\mathcal{E}$ stands for the energy generating pathway. We further assume that amino acids and glucose are available in the extracellular media. In this way our model takes into account a trade-off between energy cost of the biosynthesis of the membrane and biosynthesis of enzymes of central metabolism and those involved in RNA synthesis. However, a further extension of our model would also acknowledge that external precursors usually enter cell through membrane based transporters and that cell needs to match the number of these transporters to the number of internal enzymes and ribosomes in a way that maximize their flux.

Here we assume an existence of a pathway $\mathcal{E}$ for energy generation with associated index $i = E$. First, from equation (17) we obtain the reduced protegeometric constraint

$$1 = \sum_{i \in \mathcal{M}^*} \pi_i. \tag{19}$$

Second, from the volume-area balance equation (14) and the definition of proteomic fraction contributing to the cell membrane area (6) we obtain the reduced volume-area balance equation

$$\bar{\pi}_A = \theta \frac{\pi_L}{\pi_P} + \sum_{i \in \mathcal{M}^*} \sigma_i \pi_i. \tag{20}$$

Finally, recalling that $k_i E_i = \lambda_i \pi_E$, we add the energy balance equation

$$\sum_{i=P,N,L} e_i \lambda_i \pi_i = \lambda_E \pi_E, \tag{21}$$

where e_i is the energy cost of pathway i in units of dephosphorylations per ligated building block.

We use equation (21) to solve for the proteomic fraction of energy generation

$$\pi_E = \sum_{i=P,N,L} (\alpha_i - 1)\pi_i,$$

where

$$\alpha_i = 1 + \frac{e_i \lambda_i}{\lambda_E},$$

Inserting this result into (19) yields

$$1 = \sum_{i=P,N,L} \alpha_i \pi_i, \tag{22}$$

Similarly, starting with the surface balance equation (20) we get

$$\bar{\pi}_A = \theta \frac{\pi_L}{\pi_P} + \sum_{i=P,N,L} \beta_i \pi_i, \tag{23}$$

where

$$\beta_i = \sigma_i + \sigma_E \frac{e_i \lambda_i}{\lambda_E}.$$

The RNA polymerase proteomic fraction is related to the the ribosome proteomic fraction via equation (10). Substituting this result into equations (22) and (23) we obtain

$$1 = \alpha_{P,N} \pi_P + \alpha_L \pi_L, \tag{24}$$

$$\bar{\pi}_A = \theta \frac{\pi_L}{\pi_P} + \beta_{P,N} \pi_P + \beta_L \pi_L, \tag{25}$$

where

$$\alpha_{P,N} = \alpha_P + \nu \frac{\lambda_P}{\lambda_N} \alpha_N$$

$$\beta_{P,N} = \beta_P + \nu \frac{\lambda_P}{\lambda_N} \beta_N.$$

To summarize our development to this point the energy production is subject to two constraints: (a) proteome balance constraint, and (b) the proteogeometric constraint which directly led to equations (24) and (25). In our final step we solve the system of equations (24) and (25) by expressing π_L in terms of π_P from (24)

$$\pi_L = \frac{1}{\alpha_L} \left(1 - \alpha_{P,N} \pi_P\right).$$

and then using (25) to arrive at our final equation

$$\pi_A^* = \frac{\theta}{\alpha_L}\frac{1}{\pi_P} + \kappa\pi_P \tag{26}$$

where we define new constants

$$\pi_A^* := \bar{\pi}_A + \frac{\theta\alpha_{P,N}}{\alpha_L} - \frac{\beta_L}{\alpha_L},$$

and

$$\kappa := \beta_{P,N} - \frac{\alpha_{P,N}\beta_L}{\alpha_L}.$$

Importantly, equation (26) is quadratic equation in π_P and therefore has two solutions

$$\pi_{P,\pm} = \frac{\pi_A^* \pm \sqrt{\Delta}}{2\kappa}, \tag{27}$$

where the discriminant is defined by

$$\Delta = {\pi_A^*}^2 - \frac{4\kappa\theta}{\alpha_L} \tag{28}$$

In the following we investigate whether both of these central metabolisms allocations π_P represent biologically meaningful solutions.

4.1 Parameter estimation

The key kinetic parameters λ_i quantify the enzyme (resp. pathway) rate per unit of amino acid contained in the enzyme (resp. pathway enzymes). Biologically, they quantify the cell growth rate when the corresponding enzyme or pathway is limiting. The smaller the λ_i value the higher is its importance in setting the maximum growth rate. To estimate the λ_i of a pathway we use the following reasoning.

Suppose the pathway labelled by index i has different enzymes $j = 1, 2, \ldots$. At steady state the rates of individual enzymes r_{ij} are related to the overall pathway rate r_i by $r_{ij} = \gamma_{ij} r_i$, where γ_{ij} is a yield coefficient. For example, if $i =$ "*glycolysis*" and we measure the glycolysis rate as the rate of pyruvate formation, then all reactions above the Glyceraldehyde 3-phosphate dehydrogenase (Gapdh) step will have $\gamma_{i,j} = 1/2$, while those below including Gapdh will have $\gamma_{i,j} = 1$. Then, assuming the kinetic model $r_{ij} = k_{ij}E_{ij}$ and the aforementioned relation $r_{ij} = \gamma_{ij} r_i$ we obtain

$$\lambda_i = \frac{r_i}{n_i} = \frac{r_i}{\sum_j n_{ij}E_{ij}} = \frac{r_i}{\sum_j n_{ij}r_{ij}/k_{ij}} = \frac{1}{\sum_j n_{ij}\gamma_{ij}/k_{ij}} = \frac{1}{\sum_j \frac{\gamma_{ij}}{\lambda_{ij}}}\ . \tag{29}$$

This shows that the kinetic rate λ_i of a pathway i is the geometric mean of the kinetic rates λ_{ij} for the pathway biochemical reactions. Using this approximation and known enzyme parameters (Supplementary Information) we have estimated λ_E, λ_L and λ_P for *E. coli* (Table 1).

The specific rates λ_i are convenient to make calculations involving proteomic fractions. For example, given the pathway i, the proteomic sub-fraction located to the cell membrane can be calculated as

$$\sigma_i = \frac{\sum_j \sigma_{ij}\pi_{ij}}{\sum_j \pi_{ij}} = \frac{\sum_j \sigma_{ij}n_{ij}E_{ij}}{\sum_j n_{ij}E_{ij}} = \frac{\sum_j \sigma_{ij}n_{ij}\gamma_{ij}r_i/k_{ij}}{\sum_j n_{ij}\gamma_{ij}r_i/k_{ij}} = \frac{\sum_j \sigma_{ij}\frac{\gamma_{ij}}{\lambda_{ij}}}{\sum_j \frac{\gamma_{ij}}{\lambda_{ij}}}.$$

The aggregation of enzyme quantities is weighted by their inverse λ. Using this equation and and known enzyme parameters (Supplementary Information) we have estimated σ_i for the reduced model pathways (Table 1).

Pathway	λ (1/s)	σ	e (ATP)
Protein synthesis†	0.0028	0.0	4.2
RNA synthesis‡	0.026	0.0	18
Glycolysis, ATP production	0.02	0.007	-2
Glycolysis + OxPhos$^\diamond$	0.02	1.0	-26
Lipids synthesis*	0.03	1.0	4

Table 1: Effective parameters estimates for the reduced model pathways. †Estimated from the ribosome translation rate, but neglecting the contribution of amino acids synthesis. ‡The λ values were estimated from the RNA polymerase transcription rate, while the e value takes into account the energy cost of nucleotide synthesis. $^\diamond$ This estimate gives about the same value as glycolysis based on estimates for eukaryote cells [29]. The actual value is currently under debate [30, 31]. *Estimated using cardiolipin synthesis as a reference.

Finally, we emphasize that the kinetic parameter estimates are inaccurate and different experimental reports often yield very different values. In addition, there likely does not exist a "true value of a kinetic parameter" for a given reaction. There are isoenzymes, changes in enzyme activity due to molecular crowding or other modifying conditions, as well the dependency of the reaction rate on substrate availability. We will account for all that by utilizing random variations of the λ_i reported in Table 1. To this end we generate random numbers extracted from a triangular distribution with min, mode and max at $\lambda_i/2$, λ_i, $2\lambda_i$. Then we calculate the median and the confidence interval between quantiles 0.2 and 0.8. The goal of this simulation is to illustrate the effect of parameter uncertainty, rather than to make quantitative statements. Therefore there is no specific biological reason for the choice of distribution nor the magnitude of the variation.

4.2 Results

Based on our parameter estimates $\kappa < 0$ and the only valid solution in equation (27) is $\pi_{P,-}$. Figure 1 shows the distribution of $\pi_{P,-}$ solutions of equation (26), assuming that energy is generated by glucose fermentation and introducing variability in the λ_i values reported in Table 1. The mode is around 0.5, slightly higher than the experimentally observed maximum values of the ribosome proteome fraction (~ 0.4) [32]). Yet, given the simplicity of the reduced model the close fit to the experimental data is a remarkable achievement.

4.3 Impact of molecular crowding

Besides calculating proteomic fractions for typical values of the enzyme parameters, we use our model to investigate how the proteomic fractions depend on key model parameters. The first question we address is if the macromolecular crowding in the cell restricts proteomic fraction in the cell. To do this we investigate the dependency of the model predictions on the macromolecular volume fraction ϕ (Fig. 2A-F, ϕ on the lower x-axis). First we confirmed that the parameter κ stays negative for all tested parameter values of ϕ (Fig. 2A). Second, we noted significant changes in the dilute limit $\phi \to 0$. When $\phi \to 0$ the ribosome and RNA polymerase proteomic fractions get close to zero (Fig. 2B, C). In contrast, the proteomic fractions associated with membrane synthesis and energy generation increase significantly when $\phi \to 0$. Therefore a cell with a low macromolecular fraction ϕ has a clear growth disadvantage. On the other hand, the protein and nucleotide synthesis proteome fractions saturate to their maximum predicted value when $\phi >$

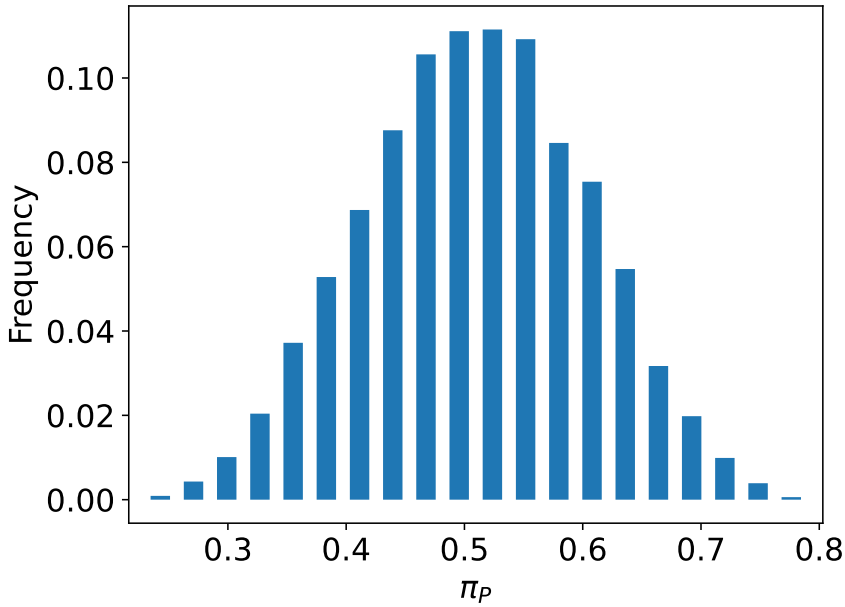

Figure 1: The distribution of the protein synthesis proteomic fraction π_P for random samplings the of λ_i parameters. Note the variability in π_P is commensurate with the input variability in the kinetics parameters, between 1/2 and 2 times the typical value.

0.05 (Fig. 2B,C), indicating there is no further selection beyond beyond macromolecular volume fractions of about 0.05.

Next we focus on the cell membrane area fraction A_P occupied by proteins, as defined in equation (18). The behavior of this quantity with the macromolecular fraction ϕ is shown in Fig. 2F. In this context A_P is small. The membrane synthesis pathway is the only component allocating a significant proteome fraction to the cell membrane and π_L is small for $\phi > 0.05$ (Fig. 2D). However, as noted below this behavior changes when energy is generated by oxidative phosphorylation and the oxidative phosphorylation localizes to the cell membrane.

4.4 Impact of cell size

Since the volume-to-area ratio l is changing with the size of the cell and is different for cells of different shapes, we also investigate the dependency of the reduced model predictions with the volume to area ratio l. From equation (15) it follows that the model dependency on the parameter l is exactly the same as on the parameter ϕ (Fig. 2A-F, top -axis). Therefore $l \to 0$ corresponds with the dilute limit described above. Indeed, as $l \to 0$ the volume to area ratio is getting smaller and the cell is shifting most of its resources to membrane synthesis.

Our results suggest that there is a minimum volume to area size requirement to reach high π_P values, but above ~ 0.01 μm the maximal protein synthesis proteomic fraction is attained. That may sound counterintuitive but it is what may have been expected. If the volume to area is too small, then there will too few ribosomes to sustain the synthesis of enzymes in the membrane synthesis pathway. For the reference, the reported volume to area ratio for *E. coli* are in the range of $0.1 - 0.3\mu$m [33].

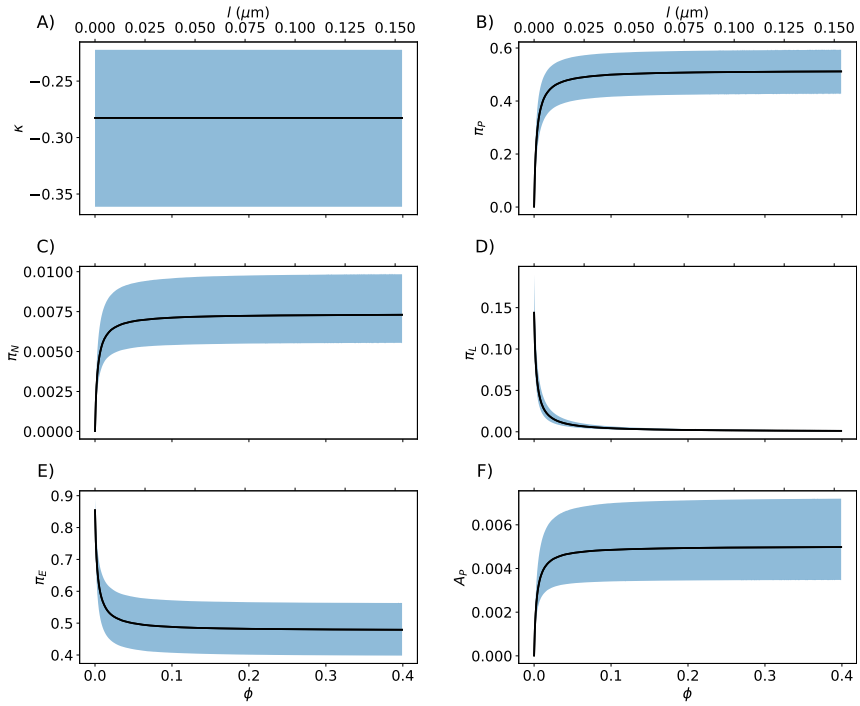

Figure 2: Functional dependency with the macromolecular volume fraction ϕ (lower X-axis) or volume to area ratio l (top X-axis) when assuming energy generation via glucose fermentation to acetate. The solid line is the median and the shaded are the confidence interval between the 0.2 and 0.8 quantiles. The dependent variables are previously defined protein fractions, κ and the fraction of the membrane area that is occupied by proteins A_P, defined in (18).

4.5 Energy generation via oxidative phosphorylation

We can repeat the simulations assuming instead that energy is generated via full oxidation of glucose, via glycolysis, the TCA cycle and oxidative phosphorylation. Unfortunately, we don't have any estimate of the effective lambda for the complete oxidation of glucose. Here we will assume the same value of glucose fermentation to acetate (Table 1). However, we do know the ATP yield per molecule of glucose. We assume that most of the pathway localizes to the cell membrane, as it is the case in *E. coli*. This is clearly an approximation since during glycolysis some TCA steps take place in the cytoplasm. Here again $\kappa < 0$ (Fig. 3A) and the functional dependency of the proteomic fractions is similar to what observed with glucose fermentation as the energy source (Fig. 3B-E vs 2B-E). In contrast, when using oxidative phosphorylation there is a significant increase in the cell membrane area ratio occupied by proteins (Fig. 2F). This is the consequence of utilizing oxidative phosphorylation that localizes to the cell membrane. In this scenario volume and membrane crowding increase together.

4.6 Impact of nutrient availability

So far we have assumed that the nutrients of the reduced model, glucose and amino acids, are in excess. Now we investigate the low glucose limit of *E. coli* cells growing on glucose as the only carbon source. To simulate glucose availability we set the λ of the glucose uptake step of glycolysis

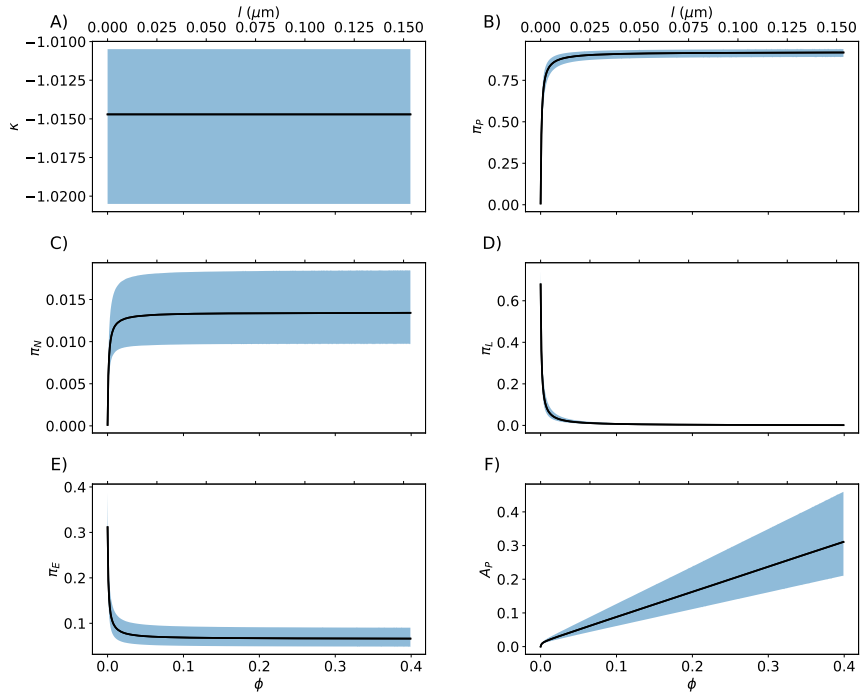

Figure 3: Functional dependency with the macromolecular volume fraction ϕ (lower X-axis) or volume to area ratio l (top X-axis), when assuming energy generation via complete oxidation of glucose (glycolysis + TCA cycle + oxidative phosphorylation). The solid line is the median and the shaded are the confidence interval between the 0.2 and 0.8 quantiles.

to $\lambda_G([\text{Glucose}]) = \lambda_G[\text{Glucose}]/(K+[\text{Glucose}])$, where [Glucose] denotes the glucose concentration and K the half-saturation constant for glucose uptake. We need to propagate this change to all pathways $i = \mathcal{P}, \mathcal{N}, \mathcal{L}, \mathcal{E}$. Since the λ of a pathway is the geometric mean of the λ of the individual pathway steps, the updated λ is given by

$$\lambda_i([\text{Glucose}]) = \frac{\gamma_i}{\frac{1}{\lambda_i} - \frac{\gamma_{i,G}}{\lambda_G} + \frac{\gamma_{i,G}}{\lambda_G([\text{Glucose}])}} = \frac{\gamma_i}{\frac{\gamma_i}{\lambda_i} + \frac{\gamma_{i,G} K}{\lambda_G[\text{Glucose}]}}$$

for $i = \mathcal{P}, \mathcal{N}, \mathcal{L}, \mathcal{E}$, where $\gamma_{i,G} = 1/2$ for $i = \mathcal{P}, \mathcal{N}, \mathcal{L}$ and $\gamma_{E,G} = 1$.

We simulated the reduced model metabolism as a function of the glucose concentration, denoted by [Glucose], using as input the λ_G of the *E. coli* PTS system for glucose uptake and the extracellular glucose half-saturation constant $K_M = 20\mu M$ [34]. We find that $\kappa < 0$ for all glucose concentrations tested (Fig. 4A), and therefore $\pi_P = \pi_{P,-}$ is the only biologically relevant solution as reported above. With increasing [Glucose], the model predicts a small increase of the protein synthesis proteomic fraction π_P, a decline of the proteomic fraction of the energy generating pathway (Fig. 4E) and an overall decrease of the proteome to lipid fraction making the cell membrane area (Fig. 4F). The decline of π_E with increasing [Glucose] can be explained by a reduction in the amount of glucose transporter required to sustain the glucose uptake with increasing the glucose concentration in the extracellular media. The increase in π_L with increasing [Glucose] is explained by the reduction in the glucose transporter proteomic fraction, which localizes to the cell membrane. The latter needs to be replaced by membrane lipids thus increasing the proteomic fraction of the membrane synthesis pathway.

For the smallest glucose concentrations the reduced model yields solutions with $\pi_L \approx 0$ and $A_P \approx 1$ (Fig. 4D). Based on packing constraints, A_P should not exceed the densest packing of disks (≈ 0.9 [35]) (Fig. 4F). In fact, we should not expect the perfect hexagonal configuration of the densest packing. At most random dense packings, which achieve a packing fraction of 0.77 [36]. Biophysical factors associated with membrane crowding are relevant as well. The cell membrane bends when the protein occupancy exceeds 20% [37], affecting cell viability in ways not accounted by our metabolic model.

Recall that in this simulation we have assumed a fixed cell geometry, *i.e.* fixed value of $\bar{\pi}_A$. However, the cell could adjust its geometry or macromolecular volume fraction to increase $\bar{\pi}_A = l_P/(l\phi)$, resulting in a feasible growth solution with $\pi_L > 0$ and $A_P < A_{P,\max}$, where $A_{P,\max}$ is some biologically relevant upper bound of the maximum membrane area fraction occupied by proteins. The cell can achieve that by reducing the volume to area ratio l and concomitantly decreasing the macromolecular content per cell to keep the macromolecular fraction ϕ constant. That observation is in agreement with empirical evidence. Bacteria and eukaryote cells have signaling pathways to reduce the cell size when nutrients are scarce. The proteogeometric constraint thus provides a hypothesis for the need of such adaptation.

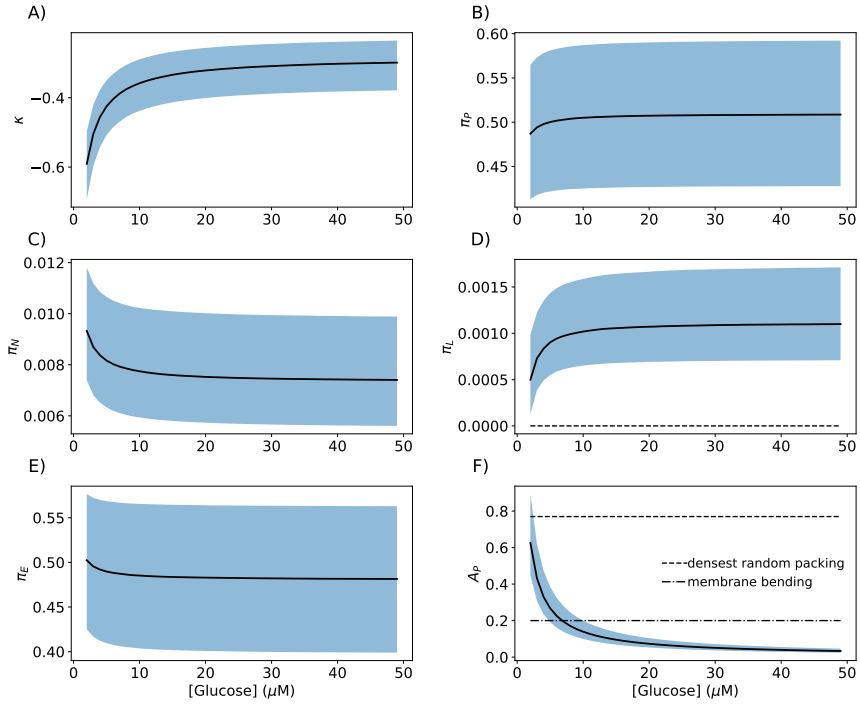

Figure 4: Functional dependency with the glucose concentration. The solid line is the median and the shaded are the confidence interval between the 0.2 and 0.8 quantiles.

5 Eukaryote cells

The extension of the proteogeometric constraint to eukaryote cells is not straightforward. Eukaryote cells contain organelles with their own membranes and so the cell contains membranes encapsulating each set of organelles and the cell membrane separating all the cell content from the extracellular

media. In a first approximation, we can assume that the macromolecular density is homogeneous across compartments (cytosol, nucleus, mitochondria, etc). Under this approximation the volume balance in equation (4) can be used for eukaryote cells. The equation (5) for the cell membrane area balance remains valid as well, with the interpretation of L as the lipid content of the cell membrane and E_L as the amount of enzyme associated with synthesis of cell membrane lipids (not including the synthesis if intracellular membranes). In that context all equations (1)-(5) are valid and the proteogeometric constraint in equation (14) is derived.

Proteogeometric constraints are also valid for each organelle separately. The derivation for each organelle follows the same steps as we have done for the whole cell. To this end we denote by π_o and π_{Ao} the proteomic fractions localizing to the organelle o as a whole and to the organelle o membrane, respectively. We also denote by π_{Lo} the proteomic fraction associated with organelle membrane lipids synthesis that resides within the organelle. Finally, we replace P by $\pi_o P$ in the right hand side of the volume balance equation (4). With this definitions, and following the same steps as for the whole cell, we obtain the organelle specific proteogeometric constraints

$$\frac{\bar{\pi}_{Ao}}{\pi_o} = \pi_{Ao} + \theta_o \frac{\pi_{Lo}}{\pi_P}. \tag{30}$$

The investigation of these constraints in specific examples and, in particular, their coupling, will be the subject of future work.

6 Discussion

We have presented evidence demonstrating that the cell volume to area balance constrains cell metabolism beyond what is already accounted by proteomic balance constraints. Deviation of protein synthesis resources to the synthesis of cell membrane components implies a metabolic cost, which indirectly selects for compact cells with molecular crowding.

We may ask why flux balance models with a proteomic balance constraint alone, but no accounting of the volume to area constraint, have been successful in fitting experimental data. The answer is that usually such models use macromolecular composition of cells as an input and/or constraints. By doing so the models are restricting the analysis to the experimentally observed macromolecular proportions. In particular, the proportion between membrane lipids and enzyme proteins is usually assumed, rather than it emerges naturally from the model.

In a deeper sense, the challenge for systems biology is to make no assumptions about macromolecular composition. With composition of available nutrients as the only input, the challenge is to predict what is the fastest growth in that specified environment, and what macromolecular composition supports such growth. We present evidence in this paper that such models need to account for cost of membrane biosynthesis, and for relationship between surface area and volume of the cell.

In conclusion, the proteogeometric constraint $\bar{\pi}_A = \pi_A + \theta \pi_L / \pi_P$ relates the cell geometry with its macromolecular composition.

Supplementary Information

Tables with the reduced model parameter estimates.

Funding

Nodes & Links Ltd provided support in the form of salary for AV, but did not have any additional role in the conceptualization of the study, analysis, decision to publish, or preparation of the manuscript. TG has no funding to declare in relation to this work.

Data availability

All reported data is contained in the manuscript or the supplementary information. The reduced model and code to generate the figures are available at the github repository proteogeometric.

References

[1] R. A. Majewski and M. M. Domach. Simple constrained-optimization view of acetate overflow in E. coli. Biotechnology and Bioengineering, 35(7):732–738, 1990. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bit.260350711.

[2] Amit Varma and Bernhard O. Palsson. Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use. Nat Biotechnol, 12(10):994–998, October 1994. Publisher: Nature Publishing Group.

[3] A Varma and B O Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Applied and Environmental Microbiology, 60(10):3724–3731, October 1994. Publisher: American Society for Microbiology.

[4] Torben L. Nissen, Ulrik Schulze, Jens Nielsen, and John Villadsen. Flux Distributions in Anaerobic, Glucose-Limited Continuous Cultures of Saccharomyces Cerevisiae. Microbiology, 143(1):203–218, 1997. Publisher: Microbiology Society,.

[5] 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.

[6] Jeremy S. Edwards and Bernhard O. Palsson. Metabolic flux balance analysis and the in silico analysis of Escherichia coli K-12 gene deletions. BMC Bioinformatics, 1(1):1, July 2000.

[7] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. Proceedings of the National Academy of Sciences, 99(23):15112–15117, November 2002. Publisher: Proceedings of the National Academy of Sciences.

[8] E. Almaas, B. Kovács, T. Vicsek, Z. N. Oltvai, and A.-L. Barabási. Global organization of metabolic fluxes in the bacterium Escherichia coli. Nature, 427(6977):839–843, February 2004. Publisher: Nature Publishing Group.

[9] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. Proceedings of the National Academy of Sciences, 104(31):12663–12668, July 2007. Publisher: Proceedings of the National Academy of Sciences.

[10] Alexei Vazquez, Qasim K. Beg, Marcio A. deMenezes, Jason Ernst, Ziv Bar-Joseph, Albert-László Barabási, László G. Boros, and Zoltán N. Oltvai. Impact of the solvent capacity constraint on E. coli metabolism. BMC Systems Biology, 2(1):7, January 2008.

[11] Douwe Molenaar, Rogier van Berlo, Dick de Ridder, and Bas Teusink. Shifts in growth strategies reflect tradeoffs in cellular economics. Molecular Systems Biology, 5(1):323, January 2009. Publisher: John Wiley & Sons, Ltd.

[12] 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, November 2010. Publisher: American Association for the Advancement of Science.

[13] Kai Zhuang, Goutham N Vemuri, and Radhakrishnan Mahadevan. Economics of membrane occupancy and respiro-fermentation. Molecular Systems Biology, 7(1):500, January 2011. Publisher: John Wiley & Sons, Ltd.

[14] Tomer Shlomi, Tomer Benyamini, Eyal Gottlieb, Roded Sharan, and Eytan Ruppin. Genome-Scale Metabolic Modeling Elucidates the Role of Proliferative Adaptation in Causing the Warburg Effect. PLOS Computational Biology, 7(3):e1002018, March 2011. Publisher: Public Library of Science.

[15] Matthew Scott, Stefan Klumpp, Eduard M Mateescu, and Terence Hwa. Emergence of robust growth laws from optimal regulation of ribosome synthesis. Molecular Systems Biology, 10(8):747, August 2014. Publisher: John Wiley & Sons, Ltd.

[16] Sheng Hui, Josh M Silverman, Stephen S Chen, David W Erickson, Markus Basan, Jilong Wang, Terence Hwa, and James R Williamson. Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. Molecular Systems Biology, 11(2):784, February 2015. Publisher: John Wiley & Sons, Ltd.

[17] Markus Basan, Sheng Hui, Hiroyuki Okano, Zhongge Zhang, Yang Shen, James R. Williamson, and Terence Hwa. Overflow metabolism in Escherichia coli results from efficient proteome allocation. Nature, 528(7580):99–104, December 2015. Publisher: Nature Publishing Group.

[18] Matteo Mori, Terence Hwa, Olivier C. Martin, Andrea De Martino, and Enzo Marinari. Constrained Allocation Flux Balance Analysis. PLOS Computational Biology, 12(6):e1004913, June 2016. Publisher: Public Library of Science.

[19] Benjamín J Sánchez, Cheng Zhang, Avlant Nilsson, Petri-Jaan Lahtvee, Eduard J Kerkhoven, and Jens Nielsen. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Molecular Systems Biology, 13(8):935, August 2017. Publisher: John Wiley & Sons, Ltd.

[20] Yu Chen and Jens Nielsen. Energy metabolism controls phenotypes by protein efficiency and allocation. Proceedings of the National Academy of Sciences, 116(35):17592–17597, August 2019. Publisher: Proceedings of the National Academy of Sciences.

[21] Pierre Salvy and Vassily Hatzimanikatis. Emergence of diauxie as an optimal growth strategy under resource allocation constraints in cellular metabolism. Proceedings of the National Academy of Sciences, 118(8):e2013836118, February 2021. Publisher: Proceedings of the National Academy of Sciences.

[22] Nathan M. Belliveau, Griffin Chure, Christina L. Hueschen, Hernan G. Garcia, Jane Kondev, Daniel S. Fisher, Julie A. Theriot, and Rob Phillips. Fundamental limits on the rate of bacterial growth and their influence on proteomic composition. Cell Syst, 12(9):924–944.e2, September 2021.

[23] Alexei Vazquez. Impact of life origins on metabolism. bioRxiv, 2024.06.07.597902, 2024.

[24] Ross P. Carlson, Ashley E. Beck, Mauricio Garcia Benitez, William R. Harcombe, Radhakrishnan Mahadevan, and Tomáš Gedeon. Cell geometry and membrane protein crowding constrain growth rate, overflow metabolism, respiration, and maintenance energy. bioRxiv, 2024.08.21.609071, 2024.

[25] Mariola Szenk, Ken A. Dill, and Adam M. R. de Graff. Why Do Fast-Growing Bacteria Enter Overflow Metabolism? Testing the Membrane Real Estate Hypothesis. Cell Syst, 5(2):95–104, August 2017. Publisher: Elsevier.

[26] 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, October 2018. Publisher: Elsevier.

[27] Xiao-Pan Hu, Stefan Schroeder, and Martin J. Lercher. Proteome efficiency of metabolic pathways in Escherichia coli increases along the nutrient flow. mSystems, 8(5):e00760–23, October 2023. Publisher: American Society for Microbiology.

[28] Malvina Papanastasiou, Georgia Orfanoudaki, Marina Koukaki, Nikos Kountourakis, Marios Frantzeskos Sardis, Michalis Aivaliotis, Spyridoula Karamanou, and Anastassios Economou. The Escherichia coli Peripheral Inner Membrane Proteome *. Molecular & Cellular Proteomics, 12(3):599–610, March 2013. Publisher: Elsevier.

[29] Jorge Fernandez-de Cossio-Diaz and Alexei Vazquez. Limits of aerobic metabolism in cancer cells. Sci Rep, 7(1):13488, October 2017. Publisher: Nature Publishing Group.

[30] Yihui Shen, Hoang V. Dinh, Edward R. Cruz, Zihong Chen, Caroline R. Bartman, Tianxia Xiao, Catherine M. Call, Rolf-Peter Ryseck, Jimmy Pratas, Daniel Weilandt, Heide Baron, Arjuna Subramanian, Zia Fatma, Zong-Yen Wu, Sudharsan Dwaraknath, John I. Hendry, Vinh G. Tran, Lifeng Yang, Yasuo Yoshikuni, Huimin Zhao, Costas D. Maranas, Martin Wühr, and Joshua D. Rabinowitz. Mitochondrial ATP generation is more proteome efficient than glycolysis. Nat Chem Biol, pages 1–10, March 2024. Publisher: Nature Publishing Group.

[31] Alexei Vazquez. Selective advantage of aerobic glycolysis over oxidative phosphorylation. arXiv, 2403.14741, 2024.

[32] Matteo Mori, Vadim Patsalo, Christian Euler, James R. Williamson, and Matthew Scott. Proteome partitioning constraints in long-term laboratory evolution. Nat Commun, 15(1):4087, May 2024. Publisher: Nature Publishing Group.

[33] Handuo Shi, Yan Hu, Pascal D. Odermatt, Carlos G. Gonzalez, Lichao Zhang, Joshua E. Elias, Fred Chang, and Kerwyn Casey Huang. Precise regulation of the relative rates of surface area and volume synthesis in bacterial cells growing in dynamic environments. Nat Commun, 12(1):1975, March 2021. Publisher: Nature Publishing Group.

[34] J.M. Rohwer, D.N. Meadow, S. Roseman, H.V. Westerhoff, and P.W. Postma. Understanding Glucose Transport by the Bacterial Phosphoenolpyruvate:Glycose Phosphotransferase System on the Basis of Kinetic Measurements in Vitro. Journal of Biological Chemistry, 275(45):34909–34921, November 2000. Publisher: Elsevier.

[35] Denis Weaire and Tomaso Aste. The Pursuit of Perfect Packing. CRC Press, Boca Raton, 2 edition, March 2008.

[36] Einar L. Hinrichsen, Jens Feder, and Torstein Jøssang. Random packing of disks in two dimensions. Phys. Rev. A, 41(8):4199–4209, April 1990. Publisher: American Physical Society.

[37] Jeanne C. Stachowiak, Eva M. Schmid, Christopher J. Ryan, Hyoung Sook Ann, Darryl Y. Sasaki, Michael B. Sherman, Phillip L. Geissler, Daniel A. Fletcher, and Carl C. Hayden. Membrane bending by protein-protein crowding. Nat Cell Biol, 14(9):944–949, September 2012.