Optimal density of biological cells

Tin Yau Pang[1], Martin J. Lercher[1,*]

[1] Institute for Computer Science & Department of Biology, Heinrich Heine University, 40225 Düsseldorf, Germany

* To whom correspondence should be addressed

Research reports

Keywords: macromolecular crowding, Michaelis Menten kinetics, metabolic reaction, ribosome, translation

Abstract

A substantial fraction of the bacterial cytosol is occupied by catalysts and their substrates. While a higher volume density of catalysts and substrates might boost biochemical fluxes, the resulting molecular crowding can slow down diffusion, perturb the reactions' Gibbs free energies, and reduce the catalytic efficiency of proteins. Due to these tradeoffs, dry mass density likely possesses an optimum that facilitates maximal cellular growth and that is interdependent on the cytosolic molecule size distribution. Here, we analyse the balanced growth of a model cell with metabolic and ribosomal reactions, accounting systematically for crowding effects on reaction kinetics. We find that changes in cytosolic density affect biochemical efficiency more strongly for ribosomal reactions than for metabolic reactions, which involve much smaller catalysts and reactants. Accordingly, optimal cytosolic density depends on cellular resource allocation into ribosomal vs. metabolic reactions. A shift in the relative contributions of these sectors to the cellular economy explains the 10% difference in the cytosolic density between *E. coli* bacteria growing in nutrient-rich and -poor environments. We conclude that cytosolic density variation in *E. coli* is consistent with an optimality principle of cellular efficiency.

Significance statement

The cellular cytosol harbours diverse molecules, whose crowding slows down diffusion and perturbs the chemical equilibrium of biochemical reactions. Reaction rates thus depend not only on the reactants themselves, but also on the background density of other molecules; consequently, maximal cell growth requires an optimal density. Here, we simulate a model cell with crowding-adjusted metabolic reaction kinetics. Its cytosol accommodates two types of reactions: metabolic reactions involving small molecules, and protein production reactions involving much larger molecules. These two cellular subsystems have distinct optimal densities, and a shift in their relative contribution to the cellular biomass explains the 10% difference in the cytosolic density between *E. coli* bacteria growing in nutrient-rich and -poor environments.

Introduction

The dry mass solved in the major compartment of bacterial cells, the cytosol, comprises hundreds of species of molecules, including proteins, metabolites, polysaccharides, and nucleic acids. These molecules can be roughly classified into two sectors: the ribosomal sector, dominated by ribosomes and tRNA; and the non-ribosomal sector, comprising mostly metabolites, enzymes, and other proteins (1). The molecules in these two sectors have very different size distributions: each ribosome is 65 times larger than the median enzyme size (2,600kDa (2) vs. 40kDa (3)), and tRNAs are about 300 times larger than typical metabolites (26kDa (4) vs. 89Da, the mass of alanine). The allocation of dry mass between the two sectors of the cellular economy can be summarized by a single parameter, the growth rate μ, with the dry mass fraction of the ribosomal sector increasing almost linearly with μ (1, 5). Accordingly, the ribosome-rich cytosol at fast growth in nutrient-rich environments and the

ribosome-meager cytosol at slow growth in nutritionally poor environments exhibit very different distributions of molecule sizes.

Experiments have found an approximately constant dry mass density of the cytosol at 300g/L across minimal nutrient conditions with low growth rates, but a roughly 10% less dense cytosol in a rich medium supporting high growth rates (see Fig. 2B of Ref. (6)). We hypothesized that the observed difference in cytosolic density between slow and fast growth may be an evolutionary consequence of the growth-related differences in molecular composition. Cellular physiology has evolved under natural selection; thus, if the level of molecular crowding—the dry mass density of the cytosol—affects cellular efficiency and hence fitness, we expect density regulation to have evolved to a near-optimal, possibly condition-dependent state. While previous work has analyzed the influence of (macro-)molecular crowding on bacterial physiology and growth rates, these analyses assumed a constant, hard limit on the total cytosolic protein (7, 8) or dry mass (9, 10) concentration. These works do not justify the existence and magnitude of the density limit based on physicochemistry, and they cannot explain differences in the level of crowding (dry mass density) across conditions.

Biochemical reaction fluxes typically increase with increasing encounter rates of the molecule species involved; ignoring crowding effects on diffusion, encounter rates increase with increasing density of the respective molecules. At the same time, the molecular crowding caused by other molecule species in the background (volume-excluding co-solutes) affects fluxes in at least three ways. (i) Crowding slows down the diffusion of a catalyst and its substrates, thereby reducing their encounter rates (11, 12). (ii) Crowding limits the available volume and thus reduces a solute's entropy (the "excluded volume phenomenon"), thereby changing the free energies of the molecules involved in the reaction and consequently shifting the equilibrium concentrations of substrates and products (13). (iii) Crowding can affect the structure of the protein catalyst, the process of its folding, and its conformational stability (14); these structural changes may disturb the reaction flux if they affect the active site (14–18). Due to these opposing effects, there may be an optimal cytosolic density where cellular efficiency and hence fitness are maximal.

The effect of crowding depends on the size of the catalyst and its substrate: in the presence of other volume-excluding cosolutes, the larger the size of a solute, the stronger the reduction of its diffusion (19) and the perturbation of its free energy (20, 21). Hence, when the size distribution of the molecules changes, the cell needs to adjust its cytosolic density in order to optimize its physiological efficiency.

In a pioneering theoretical study, Vazquez (22) considered how the objective flux of a metabolic network is affected by crowding, assuming that the enzymes are also the crowders in their own right. This study shows that, at low cytosolic density, although further expression of enzymes increases the level of crowding, it nonetheless can increase the biochemical reaction fluxes contributing to biomass formation; the converse is true at high cytosolic density, because the benefit of reducing the level of crowding compensates the reduction of enzymes available to catalyse reactions. Hence, there exists an optimal cytosolic density, which maximizes reaction fluxes and depends on details of the network, such as the ratio between diffusion-limited and transition-state-limited reactions. While this

study demonstrated the existence of a flux-optimizing cytosolic density, natural selection maximizes fitness, not metabolic reaction fluxes. For non-interacting cells in a uniform environment, fitness is equivalent to growth rate (23).

The cellular physiology at maximal growth rate—and hence maximal fitness—can be described mathematically through growth balance analysis (GBA) (10).This modeling framework simulates the balanced growth of a self-replicating bacterial cell while accounting for the major physicochemical constraints on cellular growth: mass balance of metabolism and protein production, non-linear reaction kinetics that depend on the concentrations of catalysts and their substrates, and the effects of molecular crowding. Standard GBA assumes a given, hard limit on dry mass density (9, 10) or on total protein concentration (8). Here, to assess the effects of molecular crowding on fitness, we apply a generalization of GBA that instead describes the kinetics of metabolic reactions and protein translation through crowding-adjusted Michaelis-Menten kinetics (21, 24). We maximize the balanced growth rate while varying the concentrations of transporters, catalytic proteins, and metabolites; these molecular species also form the volume-excluding co-solutes in the background of each reaction, affecting diffusion, free energies, and hence reaction kinetics through molecular crowding. Consistent with experimental observations, we find that the cytosolic density of optimal growth strategies depends on the cellular growth rate.

Results

Crowding-adjusted reaction kinetics

The effects of molecular crowding on biochemical reaction kinetics are due to volume exclusion effects. Thus, the relevant parameter for their quantification is not the total density (dry mass per volume) but the total occupancy ρ, i.e., the fraction of cytosolic volume occupied by dry mass. However, as molecular mass and volume are approximately proportional (25) and as experimental measurements for the cytosolic occupancy are not available, we treat density and occupancy as interchangeable, subject to a scaling coefficient that quantifies mass/volume of cytosolic dry mass.

To model biochemical reaction kinetics as a function of molecular crowding, we use a modified description of irreversible Michaelis-Menten kinetics (21, 24) (see Methods for details),

$$v = k_{\mathrm{cat}} \frac{[s][E]}{K_{\mathrm{M}}^{*}+[s]} \ , \qquad (1)$$

where [S] and [E] are the concentrations of the substrate and catalyst, respectively; k_{cat} is the catalytic rate constant (turnover number); and K_{M}^{*} is the crowding-adjusted Michaelis parameter (21, 24), which corresponds to the enzyme-substrate dissociation constant under certain conditions. Following previous authors (12, 21, 24), we derive K_{M}^{*} by considering two limiting cases of catalyst-substrate interaction (Fig. 1): (i) the transition state-limited case and (ii) the diffusion-limited case. In a typical catalytic reaction, a substrate molecule S encounters a catalyst molecule E to form a catalyst-substrate complex ES, which either proceeds forward to convert the substrate into the product, or reverts back to release the substrate. When the encounter rate is much higher than the conversion rate, the reaction is

transition-state limited, in which case the complex exists in near equilibrium with the substrate and catalyst (Fig. 1A); in the opposite limit, when the conversion rate is much higher than the encounter rate, the reaction is diffusion limited (Fig. 1B) (21). When the concentration of volume-excluding co-solutes is low, the catalyst-substrate complex of most enzymatic reactions is in equilibrium with the free substrate and enzyme—the assumption behind Michaelis-Menten kinetics—and therefore the reactions fall largely into the transition-state limited case (21). A generalized treatment describes the intermediate case through a weighted combination of the two limiting cases (21, 24):

$$K_{\mathrm{M}}^{*} = K_{\mathrm{M}}^{0} \frac{\Gamma + \theta e^{-g\rho}}{(1+\theta) \times \Gamma \times e^{-g\rho}} . \qquad (2)$$

Here, K_M^0 is the Michaelis parameter in the low-crowding limit; Γ is the correction term for the shift in Gibbs free energy ΔG due to molecular crowding; the exponential term exp($-g\rho$) accounts for the slow-down of diffusion, where the scaling factor g can be estimated from the size of the substrate, i.e., $g = g(r_S)$ for a spherical substrate with radius r_S; and θ is the relative weight between the two effects. The best available estimate for θ is 2.3, obtained for the ERK MAP kinase phosphorylation reaction (24)**;** we employ this value throughout our simulations.

Protein translation favours lower occupancy compared to metabolic pathways

To understand the effect of crowding on catalytic reactions, we first consider a simple, linear biochemical pathway model consisting of N=20 consecutive enzyme-catalyzed reactions at steady state, i.e., with identical flux through each reaction (Fig. 2A). The kinetics of each step in the pathway depend not only on the concentrations of the catalyzing enzyme and the substrate of the reaction, but—through the crowding effects on K_M^*—also on the concentrations of the molecules involved in the remaining N-1 reactions. We identified the combination of enzyme and metabolite concentrations that maximize the pathway output per dry mass, calculated from crowding-aware kinetics; adding these concentrations, weighted by the respective molecular volumes, resulted in an estimate of the optimal cytosolic occupancy ρ_{opt} for this system. To restrict the number of model parameters, we assumed identical enzyme and substrate sizes as well as identical crowding-aware kinetics (Eq. (1)) for all reactions in the pathway. We fixed the physicochemical parameter K_M^0=130μM, which is the median value for metabolic enzymes and their substrates (26) and is also close to the value estimated for the binding of ternary complexes to the ribosome (27), 120μM. The second physicochemical parameter in this model, k_{cat}, appears merely as a scaling factor of the pathway flux, and for simplicity we set k_{cat}=1s^{-1} without loss of generality.

We varied the sizes of the catalyst and its substrate, which we assumed to be both spherical. We considered (i) a metabolic pathway, where substrates have sizes typical for metabolites (r=0.34nm) and catalyst have sizes typical for globular proteins (r=2.4nm); and (ii) a ribosomal system with sizes resembling those of tRNAs (r=2.4nm) and ribosomes (r=13nm). We also assumed the catalyst-substrate complex to be spherical, with a volume equal to that of the catalyst plus the substrate. Note that while cells do not contain pathways of consecutive ribosomes, the model for the 20-step linear pathway depicted in Fig. 2A is mathematically identical to the model for 20 parallel reactions shown in Fig. 2B if all metabolite concentrations are assumed to be identical; this latter model can be interpreted

as a molecular snapshot showing 20 types of ribosomes, each presenting an anticodon for a different tRNA species.

Fig. 3 shows that the optimal occupancy is ρ_{opt}=0.30 for the metabolic system and ρ_{opt}=0.12 for the ribosomal system. The effect of the occupancy on the pathway output is substantial in the ribosomal system, where a 50% drop in occupancy from ρ_{opt} decreases pathway flux per dry mass by 21%, while it is small for the metabolic system, where a 50% drop in occupancy incurs a flux decrease per dry mass of only 1.3%. Increasing the number of steps in the pathway from 20 to 100 has almost no effects on the optimal occupancies ρ_{opt} for both systems; however, the reduction of flux per dry mass when ρ deviates from ρ_{opt} is substantially increased for longer pathways or more parallel reactions (Fig. 3, dashed lines).

We conclude that the results from the pathway model are consistent with the experimentally observed trend that the cytosol of fast-growing bacteria, which is dominated by the larger molecules of the ribosomal sector, favors a lower occupancy than the cytosol of a slowly growing cell dominated by smaller metabolites and enzymes. Moreover, our results suggest that compared to cells dominated by the ribosomal sector, cells dominated by the smaller molecules of the metabolic sector may suffer a smaller reduction in biochemical efficiency when the cytosol shifts away from ρ_{opt}.

Optimal occupancy for a self-replicating cell at balanced growth

Can the single-pathway results be generalized to more realistic models of cellular growth, which combine metabolic and ribosomal activities, and where we can directly assess the effect of concentration changes on cellular growth rates? How does the cytosolic occupancy optimal for growth, ρ_{opt}, change when the number of active metabolic reactions changes, e.g., when switching from a minimal medium, where all biomass components have to be synthesized from a single carbon source through multi-step biochemical pathways, to a rich medium that provides many cellular building blocks through simple transport processes?

To answer these questions, we considered a schematic GBA model of a bacterial cell (10), the cytosol of which comprises interdependent ribosomal and metabolic sectors (**Fig. 2C**). In this model cell, a nutrient s_{ext} is imported into the cell by transport protein T; the nutrient is then converted into precursor p by an N-steps metabolic pathway; finally, the ribosome R uses p to synthesize all catalytic proteins, including T, R, and the metabolic enzymes M_1 to M_N. For simplicity, all molecules are again assumed to be spherical in shape. All macromolecules, except for the transporter protein T, are located in the cytosol and are thus crowders in their own right; T is assumed to be fully integrated into the membrane and does not contribute to crowding. As before, the molecules of the metabolic sector are small (metabolites s_i with r_s=0.34nm and metabolic enzymes M_i with r_M=2.4nm), while the constituents of the ribosomal sector are much larger (precursor p with r_p=2.4nm and ribosome R with r_R=13nm).

To estimate the appropriate number of active enzymatic reactions in the metabolic sector, N, we used flux balance analysis constrained by enzyme concentration (28); simulations were performed with an improved implementation parameterized for a genome-scale model of *Escherichia coli* metabolism (29). We found 259 active metabolic enzymes for growth in a

minimal medium with glucose as the sole carbon source; 206 active enzymes for the same medium supplemented with amino acids; and 174 active enzymes for growth in a rich medium (Methods).

To facilitate the numerical determination of the state with maximal growth rate, we approximate the metabolic pathway through a single, lumped reaction with catalyst *M* and substrate *s*, scaling kinetic parameters and molecular masses to account for the pathway length *N*; this approximation only neglects the dilution of intermediate metabolite concentrations through volume growth (Methods). The solution space of the model with *N* metabolic enzyme-catalyzed reactions is spanned by 5 concentration variables (**Fig. 2C**), describing the transporter protein with concentration [*T*], the cytosolic substrates with identical concentrations $[s_1]=...=[s_N]=:[s]$, the protein precursor ("ternary complex") with concentration [*p*], the metabolic enzymes with identical concentrations $[M_1]=...=[M_N]=:[M]$, and the ribosome with concentration [*R*].

The metabolic and ribosomal reactions are described by crowding-adjusted irreversible Michaelis Menten kinetics (Eq. (1)), whereas the transporter reaction is described by conventional irreversible Michaelis-Menten kinetics with constant K_M^T. We investigated how the cytosolic occupancy that allows the fastest growth varies with two parameters, the number of enzyme species in the metabolic pathway, *N*, inversely related to the richness of the nutrient composition and entering the model through scaling the lumped metabolic reaction; and the concentration of the external nutrient, $[s_{ext}]$, which parameterizes the degree to which the available nutrients are limiting growth.

In the whole-cell model, the optimal occupancy ρ_{opt} generally falls between 0.1 and 0.3 (Fig. 4). ρ_{opt} increases when the number of simultaneously active metabolic reactions *N* increases or when the external nutrient concentration s_{ext} (and consequently the growth rate) decreases (Fig. 5). These effects are due to shifts in the relative dry mass fractions of the metabolic sector (favoring higher occupancy) and the ribosomal sector (favoring lower occupancy). At increasing *N* and constant s_{ext} , the fraction of the cytosolic volume occupied by the metabolic sector expands at the expense of the ribosomal sector (Fig. S2A), and the saturation of the ribosome with its substrate drops, whereas the saturation of the metabolic enzymes with their substrates varies within a small range (Fig. S2B). At constant *N*=250 and improving nutrient conditions (increasing s_{ext}), the fraction of the cytosol occupied by the metabolic sector also expands slightly at the expense of the ribosomal sector (Fig. S4A), while the saturation of both metabolic enzymes and ribosomes increases (Fig. S4B). Note that while the cytosolic concentration of ribosomes decreases with increasing growth rate (Fig. S3B), the ribosomal proteome fraction increases (Fig. S3C), mirroring experimental observations (1, 30).

At optimal occupancy ρ_{opt} for different *N* and s_{ext} , the crowding-adjusted Michaelis parameter K_M^* of the ribosomal reaction shows a marked increase with ρ_{opt} (Fig. 6; two-sided Spearman rank correlation coefficient r=0.995, $P<10^{-15}$), consistent with our observations of a strong dependence of the ribosomal flux on ρ in the simple pathway model (Fig. 3). In contrast, the K_M^* of the metabolic reactions is almost invariant when plotted against ρ_{opt} (r=0.034, P=0.85). These observations support the intuitive notion that the optimal occupancy in the whole-cell model reflects a tradeoff between the saturation of metabolic enzymes, favoring larger

occupancies with higher encounter rates, and the inhibition of the ribosomes, favoring lower occupancies with unhindered diffusion of tRNAs. In agreement with our findings for the simple pathway model, the dependence of the growth rate on ρ appears to be moderate: e.g., at N=150 and s_{ext}=1μM, a 50% reduction of ρ from ρ_{opt} reduces μ by only 1.1% (Fig. 4).

The slow-down of diffusion and the perturbation of Gibbs free energies have opposing effects on reaction efficiencies. To consider these two effects separately, we define the transition state-perturbation only Michaelis parameter as given by Eq. (2) when setting the diffusion scaling exponent g=0, and we define the diffusion-perturbation only Michaelis parameter as given by Eq. (2) when setting the Gibbs perturbation term Γ=1. We used these hypothetical Michaelis parameters in renewed simulations of the pathway models of the metabolic and ribosomal systems (Fig. 2A,B). In the diffusion-perturbation only model, crowding increases K_M^* and reduces the reaction fluxes (Fig. 7, dotted lines); in contrast, crowding in the transition state-perturbation only model decreases K_M^* and boosts the fluxes (Fig. 7, dashed lines). At high occupancies ($\rho \gtrsim 0.5$), the transition state-perturbation effects reach a plateau, whereas the diffusion-perturbation effects continue to increase. Thus, at larger occupancies, the flux-reducing slow-down of diffusion always dominates over the flux-enhancing perturbation of Gibbs free energy when considering their joint effect (Fig. 7, solid lines). Comparison of Fig. 7A and B shows that these trends are largely independent of N, the number of reactions in the system. While the trend for the slow-down of diffusion depends on the sizes of the reacting molecules, the trend for the perturbation of Gibbs free energies is largely independent of molecule sizes.

Discussion

The linear pathway model shows that reactions with larger catalysts and substrates favour a lower occupancy than reactions with smaller molecules (Fig. 3). This effect explains the observation of lower optimal occupancies for decreasing pathway length N in the GBA model cell, which simulates the switch from minimal media, requiring on the order of 260 metabolic enzymes to convert a small number of nutrients to the full range of cellular building blocks, to increasingly richer media, where progressively more biomass components can be taken up directly from the environment. With decreasing numbers of active metabolic reactions N, the ribosomal sector expands at the expense of the metabolic sector, pushing ρ_{opt} to lower values (Fig. S2A).

The whole-cell model predicts that ρ_{opt} also decreases with increasing nutrient concentration at a constant number of active metabolic reactions N, i.e., without switching between metabolic pathways. While no direct experimental observations of the occupancy ρ are available, occupancy is expected to scale approximately proportional to the cytosolic dry mass density ρ_{DM} (25). Empirical observations indicate that ρ_{DM} does not vary noticeably with μ across minimal media (see Fig. 2B of Ref (6)). This suggests either that (i) the cell does not fine-tune its cytosolic occupancy across minimal media of different quality, or that (ii) the uncertainty of the measurements is too large to capture subtle variations of ρ. Indeed, the uncertainty of ρ measurements is on the order of 50% of the change in ρ due to pathway switching (Fig. 2B of Ref (6)), the same order of magnitude as the change in the modelled ρ_{opt} in response to a 10-fold increase in external nutrient concentrations (Fig. 5).

Our simulations also show that even if the cell remains at the state of optimal occupancy, a large reduction in the nutrient level results in only a small decrease of the ribosome's saturation with its substrate (0.81 to 0.72, Fig. S4B); at the same time, the ribosome's Michaelis parameter K_M^* increases by 51% (Fig. S4C). Thus, as the cytosolic occupancy adjusts itself to support an increasing metabolic sector, the concentration of the ribosomal substrate increases to compensate for its catalyst's reduced affinity. These findings are consistent with experimental observations that changes in translation rate per ribosome in *E. coli*, a direct consequence of ribosomal substrate saturation, are much smaller than the simultaneous changes in growth rate (31). At increasing nutrient levels, despite the small decrease in ribosome concentrations, a higher growth rate μ is facilitated by relocating a substantial amount of the ribosome's synthesis capacity from transporter proteins to metabolic enzymes and additional ribosomes (Fig. S3C).

The growth rate μ of the whole-cell model decreases only mildly when the occupancy ρ deviates from ρ_{opt} (Fig. 4). The difference of *E. coli* cytosolic dry mass density observed between low growth rates (minimal media, μ<0.8/h) and high growth rates (rich medium, μ>0.8/h) is (6) $\Delta\rho = (0.31g/mL - 0.28g/mL) \div 1.35g/mL = 0.02$ (where 1.35g/mL is the mass to volume ratio of dry mass components, estimated from the protein dry mass to volume ratio (32)). In the whole-cell model, a deviation of 0.02 from ρ_{opt} results in a 0.02% drop in μ. Using growth rate as a proxy for fitness, this corresponds to a selection coefficient s=0.0002. While this may appear low, the effective population size of most bacterial species is on the order of $N_e=10^8$ (33), and we thus have $s \gg 1/N_e = 10^{-8}$; accordingly, natural selection would be sufficiently strong to explain the difference in average ρ_{DM} observed between the two physiological states. According to the experimental observations, there is substantial between-cell variation in each nutritional condition (6): the observed dry mass densities show coefficients of variation (standard deviation / mean) around 5%, which translates to a variation in occupancy of approximately $\Delta\rho = 5\% \times 0.31g/mL \div 1.35g/mL = 0.01$. According to the whole-cell model, such a difference corresponds to a selection coefficient of $s = 10^{-4} \gg 1/N_e$, indicating that a large variation in ρ may persist despite deleterious selection.

Ref (6) reported a cytosolic mass density of around ρ_{DM}=0.31±0.02g/mL at slow growth in minimal media (μ<0.8/h), and ρ_{DM}=0.28±0.01g/mL at fast growth in a rich medium (μ>0.8/h). Assuming a mass-to-volume-ratio of 1.35g/mL for the dry mass (32), this is equivalent to volume occupancies of ρ=0.23±0.01 in minimal media and ρ=0.21±0.01 in rich media. Our schematic whole-cell model predicts that, in a minimal medium (N=250), ρ=0.280 when the nutrient concentration in the environment s_{ext}=0.01μM, or ρ=0.234 when s_{ext}=1μM; , and in a rich medium (N=150), ρ=0.204 when s_{ext}=10μM. Given the minimalistic nature of the whole-cell model, which assumes that all metabolic reactions follow identical kinetics and approximates protein production through a single Michaelis-Menten type reaction, it is striking that the model not only predicts differences across physiological states, but predicts experimentally observed values quantitatively with an error below 22%.

Questions about optimal allocation of the protein resource can be addressed by maximizing the growth rate (or flux) in computational models of cellular growth with fixed, crowding-unaware kinetic parameters. While such simulations produce meaningful

predictions for the relative amounts of the proteins, they cannot limit absolute protein concentrations. To solve this problem, existing models (9, 10, 34) implement hard, phenomenological constraints on the total concentration of protein or cellular dry mass, based on experimental observations that found these to be approximately constant across growth conditions (6, 30, 35, 36). Our results elucidate the biophysical origin of these observations, showing that the cellular dry mass density represents a compromise between the saturation of metabolic enzymes with their substrates and the effects of reduced diffusion on the effective affinity of the ribosome for its much larger substrate, the ternary complex.

Methods

Crowding-adjusted Michaelis-Menten kinetics

Macromolecular crowding affects the flux of a metabolic reaction in multiple ways. It can (i) slow down diffusion; (ii) affect the free energy of substrate, catalyst, and the substrate-catalyst complex and thereby change their relative equilibrium ratios; and (iii) disturb the folding of a protein and affect the shape of the active site. In our modelling framework, we followed the derivation proposed in Ref (21) that systematically accounts for the effects of crowding on metabolic fluxes caused by effects (i) and (ii).
In this section, let us consider the metabolic reaction carried out by an enzyme *E* that converts substrate *S* into product *P*, in the presence of other volume-excluding co-solutes that, collectively, constitute the dry mass of the solution. The metabolic reaction is described by the chemical equation following Michaelis-Menten kinetics:

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P}$$

It is described by two parameters: the enzyme-substrate dissociation (or Michaelis) parameter $K_M \sim k_{off}/k_{on}$, and the catalytic rate constant (or turnover number) k_{cat}. Note that while K_M is usually assumed to be invariable and is hence referred to as “Michaelis constant”, we here examine crowding-dependent changes in K_M und hence refer to it as the “Michaelis parameter”. We assume that all reactions follow effectively irreversible Michaelis Menten kinetics and thus free energy changes are irrelevant to the transition rate k_{cat}; moreover, we ignore crowding effects on the enzyme structure, and thus k_{cat} is not affected by crowding. To derive the effect of crowding on the Michaelis parameter K_M, we first consider two extreme cases: diffusion-limited, where effect (i) dominates the perturbation of the enzymatic reaction through crowding, and transition-state limited, where effect (ii) dominates.

Diffusion limited case

The diffusion-limited case is relevant when the rate of encounter of *E* and *S* is much smaller than the rate of conversion of the complex *ES* into the product *P*. At given concentrations of *E* and *S*, the rate of formation of *ES* is thus mostly determined by the rate of encounter, which is proportional to the sum of *E*’s and *S*’s diffusion coefficients. Slow-down of diffusion shifts the equilibrium of *ES* concentration; this can be accounted for in the following way (21):

$$\frac{[ES]}{[E][S]} = \frac{1}{K_{\mathrm{M}}^{\mathrm{diff}}} \simeq \frac{1}{K_{\mathrm{M}}^{0}} \exp(-g(r_{\mathrm{S}})\rho) \quad \text{(S1)}$$

Here, K_M^{diff} is the Michaelis parameter in the diffusion-limited case, while K_M^0 is the Michaelis parameter in the low-crowding limit; ρ is the volume occupancy of the volume-excluding co-solutes (dry mass) of the solution (with range $0 < \rho < 1$), and g is a function that depends on the shape of E, S, and other volume-excluding co-solutes. Since S is typically much smaller than E, the diffusion coefficient of S in a crowded solution is in general much higher than that of E. Hence, we estimate this scaling term $\exp(-\ g\rho)$ solely from the diffusion coefficient of S. Approximating S as a sphere of radius r_S, we can write it as $\exp(-\ g(r_{\mathrm{S}})\rho)$.

The bacterial cytosol is crowded, slowing down the diffusion of all molecular species. The extent of this slow-down, however, is non-uniform and depends largely on the size of the affected molecule: the larger the molecule, the more it is slowed down. The slow-down of diffusion in the *E. coli* cytosol is summarized by the following empirical scaling law, which was inferred from molecular dynamics simulations (19):

$$\ln\left(\frac{D_0(r_h)}{D_{cyto}(r_h)}\right) = \left(\frac{\xi^2}{R^2} + \frac{\xi^2}{r_h^{\ 2}}\right)^{-\frac{a}{2}} \quad \text{(S2)}$$

where $r_h = 1.3(r + 1.4\text{Å})$ is the hydrodynamic radius of a molecule with radius *r* (37), *i.e.*, its effective radius including the attached water molecules; $D_0(r_h)$ is the diffusion coefficient in the low crowding limit, while $D_{cyto}(r_h)$ is the diffusion coefficient in the crowded cytosol condition; *ξ*=0.51nm is the average distance between the surfaces of volume-excluding co-solutes in the cytosol of *E. coli*; *R*=42nm is the radius of the largest common crowders in the cytosol; and *a*=0.53 is an empirical scaling factor (19). Note that in general, the parameter *ξ* depends on *ρ;* however, as we are only interested in the relationship between reaction fluxes, growth rate, and *ρ* in a small range centered around the native *E. coli* cytosolic density, we can approximate *ξ* by a constant in our analysis. As the cytosolic dry mass density is ~0.3g/mL (6) and the mass-to-volume-ratio of protein is 1.35g/mL (32), the cytosolic volume occupancy *ρ* is approximately 0.22=0.3/1.35. If the reaction rate is proportional to the rate of encounter between *E* and *S*, and this rate of encounter is in turn approximately proportional to the diffusion coefficient of *S*, then Eq. (S2) can be used to calculate the scaling factor $g(r_{\mathrm{S}})$:

$$g(r_S) = \left(\frac{\xi^2}{R^2} + \frac{\xi^2}{(1.3(r_S+1.4\text{Å}))^2}\right)^{-\frac{a}{2}} / 0.22$$

and the reaction rate in this diffusion limited case is denoted as k_{diff}:

$$k_{\mathrm{diff}} = k_{\mathrm{cat}}[\mathrm{ES}] = \frac{k_{\mathrm{cat}}[\mathrm{E}][\mathrm{S}]}{K_{\mathrm{M}}^{\mathrm{diff}}}$$

Transition-state limited case

The transition-state limited case is relevant when the rate of encounter of E and S is much larger than the rate of conversion of the ES complex into the product P. The complex ES thus exists in near equilibrium with E and S, and macromolecular crowding affects the

reaction rate parameter mainly through shifting this equilibrium. To assess the magnitude of this effect, we consider the reversible reaction

$$\mathrm{E} + \mathrm{S} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{ES}$$

and denote the crowding-adjusted equilibrium Michaelis parameter in this transition-state limit as K_M^{ts} (21):

$$K_M^{ts} = \frac{\gamma_E \gamma_S [E][S]}{\gamma_{ES}[ES]} = \Gamma \frac{[E][S]}{[ES]} = \Gamma K_M^0,$$

where γ_i denotes the activity coefficient of molecular species i, $\Gamma = \frac{\gamma_E \gamma_S}{\gamma_{ES}}$, and $K_M^0 = k_{off}/k_{on}$ is the enzyme-substrate dissociation parameter in the low-crowding limit. The activity coefficients are defined as

$$\gamma_i = \exp \frac{\mu_i - \mu_i^{ideal}}{RT}$$

with

$$\mu_i = \frac{\partial G}{\partial [i]};$$

here, G is the Gibbs free energy of the system; [i] is the concentration of molecular species i; μ_i is the chemical potential of i , and μ_i^{ideal} is the chemical potential in an idealized situation, i.e., without intermolecular interactions and in the absence of other volume-excluding co-solutes. In other words, the equilibrium of the reaction will be identical to the dissociation parameter K_M^0 if the system is ideal. The Γ term, therefore, accounts for the deviation of the Gibbs free energy from the idealized situation.

The γ_i of each molecular species i can be written as an expansion in terms of the concentrations of all molecular species (38):

$$\ln \gamma_i = \sum_j B_{ij}[j] + \sum_{j,k} B_{ijcok}[j][k] + \ldots$$

The coefficients B_{ij} (B_{ijk}, ...) reflect the interaction between 2 (3, ...) molecular species. For example, the coefficient B_{ij} is given as (39):

$$B_{ij} = 4\pi N_A \int_0^{\inf} \{1 - \exp[-\frac{U_{ij}(r)}{kT}]\} r^2 dr$$

where N_A is Avogadro's number, r is the distance between the center of mass of molecular species i and j, and $U_{ij}(r)$ is the potential of average force acting between the two molecular species. While the interaction potential among multiple molecular species is complex, it has been found that the colligative properties of solutions of globular proteins can be well accounted for over a wide range of concentrations by using a simple hard sphere potential, where two molecules cannot overlap but do not interact otherwise (see Ref (21) for a review):

$$U_{ij}(r) = \begin{cases} \infty & r \le r_i + r_j \\ 0 & r > r_i + r_j \end{cases}$$

The scaled particle theory applies this rigid sphere potential to calculate the activity coefficient γ_i of molecular species i (40):

$$\ln \gamma_i = -\ln(1 - \langle\langle V \rangle\rangle) + \frac{r_i \langle\langle S \rangle\rangle + S_i \langle\langle r \rangle\rangle + V_i \langle\langle 1 \rangle\rangle}{1 - \langle\langle V \rangle\rangle} + \frac{r_i^2 \langle\langle S \rangle\rangle^2 + 2V_i \langle\langle r \rangle\rangle \langle\langle S \rangle\rangle}{2(1 - \langle\langle V \rangle\rangle)^2} + \frac{V_i \langle\langle r^2 \rangle\rangle \langle\langle S \rangle\rangle^2}{3(1 - \langle\langle V \rangle\rangle)^3} \quad \text{(S3)}$$

Here, w_i is the concentration (number density) of molecular species i ; $<< X >>= \sum_i w_i X_i$ is a weighted sum of property X, with $<< 1 >>= \sum_i w_i$; $S_i = 4\pi r_i^2$ is the surface area and $V_i = \frac{4}{3}\pi r_i^3$ is the volume of molecular species i; in addition, $<< r >>= \sum_i w_i r_i$ and $<< r^2 >>= \sum_i w_i r_i^2$. As in the transition-state limited case, the reaction rate parameter is proportional to the concentration of the enzyme-substrate complex, and a shift of its equilibrium constant by Γ leads to a corresponding shift of the reaction rate in this transition-state limit, quantified by the rate parameter k_{ts} (21):

$$k_{\mathrm{ts}} = k_{\mathrm{cat}}[\mathrm{ES}] = \frac{k_{\mathrm{cat}}[\mathrm{E}][\mathrm{S}]}{K_{\mathrm{M}}^{\mathrm{ts}}} \quad . \quad \text{(S4)}$$

General case

When the reaction rate is limited by both the initial binding of the two molecules (encounter rate) and the transition rate, an approximation to the overall, crowding-aware reaction rate k is obtained by adding the inverse rate parameters of the two steps (21, 24, 41):

$$k^{-1} = k_{ts}^{-1} + k_{diff}^{-1} .$$

Given Eq. (S1) and (S4), this leads to (24):

$$k = k_0 \frac{(1+\theta)\Gamma \exp(-gc)}{\Gamma + \theta \exp(-gc)}$$

Here, $\theta \equiv \frac{k_{ts}}{k_{diff}}|_{\rho=0}$ quantifies the tendency of the reaction towards the diffusion-limited ($\theta \to 0$) or the transition-state limited ($\theta \to \infty$) case; k_0 is the overall rate parameter in the absence of crowding. Remember that $[E_{free}] = [E_{total}] - [ES] = [E_{total}]/(1 + [S]/K_M)$ is the concentration of free enzymes only, and thus the reaction rate is $k = k_{cat}[ES] = k_{cat}[E_{free}][S]/K_M$; from now on, we describe the reaction rate as the flux of the reaction, and denote it as v. As we assume that the reaction is irreversible and hence crowding does not influence k_{cat}, to arrive at crowding-aware Michaelis-Menten kinetics, we have to scale the Michaelis parameter, which arises from considerations on the equilibrium between the enzyme-substrate complex and its constituents. Thus, $v/v^0 = K_M^0/K_M^*$, with K_M^0 the Michaelis parameter in the low crowding limit; then the crowding-adjusted Michaelis parameter is

$$K_M^* = K_M^0 \frac{\Gamma+\theta\exp(-gc)}{(1+\theta)\Gamma\exp(-gc)} \quad \text{(S5)}$$

and so the flux can be written as

$$v = k_{cat} \frac{[E_{free}][S]}{K_M^*} \simeq k_{cat} \frac{[E_{total}][S]}{K_M^*+[S]} . \quad \text{(S6)}$$

The θ of ERK MAP kinase phosphorylation reaction was estimated to be 2.3 (24); we assume that this value is representative for cellular enzymes and use it for all modeled reactions.

The concentration of the enzyme-substrate complex, [*ES*], depends on K_M^* through $[ES] = [E_{total}][S]/([S] + K_M^*)$; at the same time, K_M^* depends on [*ES*] through Eq. (S2). To find a self-consistent solution for these two quantities, we iterate these two equations until convergence. In each iteration, we first calculate the concentration of the catalyst-substrate complexes of different reactions; we then update their K_M^* and proceed to the next iteration. We stop the interaction when all K_M^* values changed by less than 0.001% compared to the previous iteration.

Single-pathway model

We considered a simple model of a linear pathway to investigate how the size of the substrate and catalyst of a metabolic reaction affect the optimal cytosolic occupancy; here, optimality is defined as a maximal pathway flux per unit dry mass, calculated from crowding-aware kinetics. The pathway is divided into *N* steps, where E_n is the catalyst of step *n*, which converts its substrate s_n into s_{n+1}, the substrate of the next step (Fig. 2A). We assume that all internal metabolite concentrations are in steady state (i.e., producing and consuming fluxes cancel exactly), and we ignore the dilution of intermediates. Thus, all reaction fluxes have the same value, *v.* We assume that s_1 is replenished by a flux $v_{s_1} = v$, which is not modeled explicitly.

We assume that the *N* reactions are described by crowding-adjusted Michaelis Menten kinetics with identical k_{cat} and K_M^0 . We further assume that the *N* catalyst species and the *N* substrate species are spherical, with radius r_E (volume V_E) for the catalysts and radius r_s (volume V_s) for the substrate species. These assumptions simplify the solution space of the model, as in the optimal steady state, all catalysts have equal total concentrations ($[E_1]=[E_2]=...=[E_N]$, and so do all the substrates ($[s_1]=[s_2]=...=[s_N]$). We define the total substrate concentration $[s] \equiv \sum_n [s_n] = N[s_1]$ and the total catalyst concentration $[E] \equiv \sum_n [E_n] = N[E_1]$. The two variables [*s*] and [*E*] span the solution space of this model.

Substrates and catalysts, as well as the substrate-catalyst complexes, are crowders in their own right; they slow down diffusion and perturb Gibbs free energies. The cytosolic volume occupancy of dry mass in the solution, ρ ($0 \leq \rho \leq 1$), is

$$\rho = N_A \frac{4\pi}{3} ([s]r_s^3 + [E]r_E^3) \quad \text{(S7)}$$

with the Avogadro number N_A. The flux through the pathway per unit volume is

$$v = k_{cat} \frac{[s]/N\ [E]/N}{K_m^* + [s]/N} = \frac{k_{cat}}{N} \frac{[s][E]}{NK_m^* + [s]}$$

Accordingly, the flux per unit dry mass is

$$\mu = \frac{k_{cat}}{N} \frac{[s][E]}{NK_M^* + [s]} \frac{1}{\rho} . \quad \text{(S8)}$$

As we ignore crowding effects on the turnover number, k_{cat} acts only as a scaling factor, and we thus set k_{cat}=1 for simplicity. We set the Michaelis parameter at the low crowding limit to K_M^0=130μM, which is the median K_M of metabolic enzymes (26), and is also very close to the Michaelis constant of the ribosome estimated from the diffusion limit without molecular crowding (27).

Because we assume identical kinetics of all reactions and ignore the dilution of intermediates through growth, the whole pathway is equivalent to a single reaction with re-scaled kinetics. We still describe it as an *N*-steps pathway, as this more faithfully reflects the situation in the real cell, allowing us to use realistic parameter values. Moreover, while mathematically, *N* represents a scaling factor, it has an intuitive biological interpretation.

The same equations can be used to describe a system of *N* parallel enzymatic reactions with identical fluxes, with only an additional multiplication by *N* in Eq. (S8), $\mu_{parallel}=\mu \times N$ (Fig. 2B). Here, catalyst E_n of reaction *n* converts substrate s_n into the end product, and the consumption of s_n is compensated by a flux v_{s_n} that supplies s_n at an equal rate.

We consider two systems of substrate and catalyst sizes: metabolic and ribosomal. In the metabolic system, we use r_s=0.34nm for metabolites (the approximate radius of the amino acid alanine (42)) and r_E=2.4nm for globular proteins (an “average” protein in *E. coli* has a mass 40kDa (3), while a globular protein with mass 50kDa has a radius *r*=2.4nm (43); in an alternative estimation, the radius of a typical globular protein is approximately *r*=2.5nm (44)). In the ribosomal system, we use r_s=2.4nm for the ternary complexes (the gyroscopic radius of tRNA is estimated to range from 2.33nm to 2.46nm based on Eq. (7) of (45)). We use a radius of r_E=13nm for the ribosome, as the diameter of a ribosome is reported to be 20nm - 30nm (44, 46, 47). We assume that the catalyst-substrate complex is also spherical, with a volume equal to the sum of the substrate’s volume and the catalyst’s volume.

For a fixed value of the total occupancy *ρ*, we calculated the specific flux *μ* using MATLAB, while (i) varying the occupancy *ρ* in steps of 0.01 from 0.01 to 0.8, with additional, finer-grained steps of 0.001 from 0.100 to 0.360, and (ii) varying the ratio of the volume occupied by the substrates *s*, from 0.1% to 97.7%, with an increase by a factor of 1.0023 at each step.

Model cell with a metabolic and a ribosomal pathway

To more faithfully represent a complete cell and to study the tradeoff between metabolic and ribosomal reactions, we also consider a balanced growth model with a metabolic sector and a ribosomal sector (Fig. 2C). As seen from the results for the pathway models, the two types

of reactions have very different optimal conditions: the metabolic sector involves smaller catalysts and substrates than the ribosomal sector, and hence has maximal specific fluxes at a higher cytosolic occupancy.

In this model (Fig. 2C), the transporter *T* imports the external substrate s_{ext} into the cytosol, where it is now labeled s_1. The transport reaction is described by ordinary, crowding-unaware irreversible Michaelis Menten kinetics, with flux

$$v_T = k_{cat}^{T}[s_{ext}][T]/([s_{ext}] + K_M^{T}).$$

We set k_{cat}^{T}=13.7s^{-1} (the median k_{cat} of enzyme reactions (26)), and K_M^{T}=1μM (close to the growth limiting glucose concentration of *E. coli* (48)).

The metabolic sector comprises an *N*-steps linear pathway of metabolic reactions, identical to the one studied in the simple pathway model: the enzyme of metabolic reaction *n* ($1 \leq n \leq N$), denoted as M_n, converts substrate s_n into s_{n+1}; in reaction *n*=*N*, M_N converts one s_N into one precursor *p* (8). These *N* reactions follow crowding-adjusted irreversible Michaelis Menten kinetics with identical rate parameters K_M^{M0}=130μM and k_{cat}^{M}=13.7s^{-1} (the median K_M and k_{cat} of enzyme reactions (26)). As before, all *N* substrates have radius r_s=0.34nm, while all *N* enzymes have radius r_M=2.4nm; as before, we assume that the enyzme-substrate complex is spherical and occupies as much volume as one substrate plus one enzyme molecule. To facilitate the numerical solution of the model, we assume that all metabolite concentrations [s_n] and also all enzyme concentrations [M_n] are identical. This corresponds to the optimal balanced growth solution when the differential dilution of intermediate metabolites is ignored (26); thus, we treat the dilution of intermediate metabolites only approximately here. In the balanced growth condition, the production rate of each substrate is equal to its consumption rate plus its (approximate) rate of dilution through growth, so that its concentration remains stable. We define the total substrate concentration $[s] \equiv \sum_n [s_n]$ and total enzyme concentration $[M] \equiv \sum_n [M_n]$. The flux of each metabolic reaction is $\frac{k_{cat}^{M}}{N}\frac{[s][M]}{NK_M^{M*}+[s]}$.

The ribosomal sector comprises the ribosome (*R*) and the protein precursor (*p*): *R* converts *p* into the *N*+2 types of protein in the model, *N* metabolic enzymes (M_n), the ribosome (*R*), and the transporter (*T*). As before, the radius of *p* is r_p=2.4nm and the radius of *R* is r_R=13nm; their complex is assumed to be spherical and to occupy as much volume as one precursor plus one ribosome molecule. The ribosomal conversion rate is described by crowding-adjusted irreversible Michaelis Menten kinetics with parameters K_M^{R0}=120μM (27) and k_{cat}^{R}=22.0s^{-1} (8, 27), and the consumption rate of *p* by *R* to make proteins is $k_{cat}^{R}\frac{[p][R]}{K_M^{R*}+[p]}$. The ribosome converts I_T=300 precursors into one transporter, I_M=300 precursors into one metabolic enzyme, and I_R=7459 precursors into one ribosome (8).

Note that the reactions that produce and consume the precursor *p* do not conserve volume. The reason is that to model a realistic cell, we envision the precursor as a charged tRNA, only the amino acid part of which is (i) produced by the metabolic pathway and (ii) integrated

into the growing protein. The metabolic pathway only provides the amino acid, while the pool of free tRNAs is not explicitly modeled. For this reason, the size of p is substantially larger than the size of the metabolite s_N consumed in its production. Conversely, the ribosome consumes 300 precursors to produce a single transporter or enzyme, which are both substantially smaller than the combined volume of the 300 precursors. Here, we envision that the tRNA part of p is set free again and can be re-charged through M_N. This treatment assumes that the concentration of free tRNA is so low that we can ignore its dilution through growth and its contribution to the cytosolic crowding.

The solution space of this model cell spans five dimensions: [s], [p], [T], [M], and [R]. The corresponding molecules, along with their complexes, are also the crowders that slow down diffusion and disturb Gibbs free energies. The cytosolic occupancy ρ, $0 \leq \rho \leq 1$, is

$$\rho = N_A \times \frac{4\pi}{3}([s]r_s^3 + [M]r_M^3 + [p]r_p^3 + [R]r_R^3) \tag{S9}$$

with the Avogadro number N_A. The growth rate μ of this model cell can be expressed as the flux through the ribosome reaction divided by the total protein concentration,

$$\mu = k_{cat}^R \frac{[p][R]}{K_M^{R*}+[p]} \frac{1}{N_A(l_M[M]+l_R[R]+l_T[T])} \tag{S10}$$

In the balanced growth state, the production of s and p is offset by their consumption and dilution by growth,

$$\begin{aligned} k_{cat}^T \frac{[s_{ext}][T]}{K_M^T+[s_{ext}]} - \frac{k_{cat}^M}{N} \frac{[s][M]}{NK_M^{M*}+[s]} - \mu[s] &= 0 \\ \frac{k_{cat}^M}{N} \frac{[s][M]}{NK_M^{M*}+[s]} - k_{cat}^R \frac{[p][R]}{K_M^{R*}+[p]} - \mu[p] &= 0 \end{aligned} \tag{S11}$$

For a given number of enzymes N and occupancy ρ, we solved this model numerically. As a preliminary step, we used the BARON algorithm (49) implemented in Pyomo (50, 51) and assumed normal crowding-unaware irreversible MIchaelis Menten kinetics, maximizing the growth rate μ over the space of concentrations ([s],[p],[T],[M],[R]), subject to the constraints of Eq. (S9) and (S11). Using the BARON solution as a starting point, we then applied the SLSQP algorithm within the function "*minimize*" in SciPy (52), now using crowding-adjusted Michaelis-Menten kinetics. It is not clear if the optimization problem has a unique solution; to increase the probability that the solution is a global maximum, we repeatedly ran the SLSQP algorithm at least 20 times for each (N,ρ) combination, and picked the solution with the highest μ. For each simulation, at least half of the independent runs supported the same, maximal optimum.

Estimating the number of enzymes in the metabolic pathway

To obtain a realistic estimate of the number of simultaneously active metabolic reactions in a bacterial cell, we performed flux balance analysis simulations accounting for molecular crowding in terms of a hard limit on the total cellular protein concentration. Simulations were run using "sybil", an R library for efficient constraint-based analyses (29, 53, 54). We used sybilccFBA (55, 56), a re-implementation of the MOMENT algorithm with an improved

treatment of multifunctional enzymes, which maximizes the biomass production rate while constraining the sum of cytosolic metabolic enzyme concentrations at the experimentally observed level.

We used a sybilccFBA implementation of the iAF1260 stoichiometric model (57), parameterized with turnover numbers for *E. coli* (55, 56). We considered four different nutritional environments. In each case, we counted the number of active metabolic reactions with both substrates and products located in the cytosol, and with a flux $>10^{-6}$ mmol/(gram Dry Weight)/h to filter out numerical noise; we enumerated the enzymes supporting these reactions, filtering out those with proteome fraction lower than a cutoff (10^{-9}; for reference, the most abundant enzyme has a density $\sim 10^{-3}$).

The estimated numbers of active metabolic enzymes and reactions in the different conditions are as follows: (i) 259 active enzymes and 349 active reactions in a minimal medium with glucose as the sole carbon source, corresponding to slow to intermediate growth with a cytosol dominated by the metabolic sector (56, 58) (Supplementary Table S1); (ii) 206 active enzymes and 288 active reactions in the same minimal glucose medium supplemented with 20 amino acids, simulating intermediate growth (56, 58) (Supplementary Table S2); (iii) 174 active enzymes and 250 active reactions in a rich medium, corresponding to fast growth and a cytosol dominated by the ribosome and its substrates (56, 59) (Supplementary Table S3); and (iv) 140 active enzymes and 234 active reactions in an extremely rich medium, where all exchange reactions in the iAF1260 *E. coli* model are allowed to be active.

Acknowledgements

We thank Deya Alzoubi and David Heckmann for providing the environments used in the ccFBA simulations. We thank Deniz Sezer for helpful discussions. This work was supported by DFG Grants CRC 1310 to T.Y.P. and M.J.L. and by a grant of the Volkswagen Foundation in the “Life?” initiative to M.J.L..

Author contributions

T.Y.P. developed the methodology and performed all analyses. M.J.L. and T.Y.P. conceived of the study, interpreted the results, and wrote the manuscript.

References

1. M. Scott, C. W. Gunderson, E. M. Mateescu, Z. Zhang, T. Hwa, Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science* **330**, 1099–1102 (2010).
2. J. Rabl, M. Leibundgut, S. F. Ataide, A. Haag, N. Ban, Crystal Structure of the Eukaryotic 40S Ribosomal Subunit in Complex with Initiation Factor 1. *Science* **331**, 730–736 (2011).
3. F. C. Neidhardt, Ed., *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd Ed. (ASM Press, 1996).
4. D. P. Nierlich, Regulation of ribonucleic acid synthesis in growing bacterial cells: II. Control over the composition of the newly made RNA. *J. Mol. Biol.* **72**, 765–777 (1972).

5. X.-P. Hu, H. Dourado, P. Schubert, M. J. Lercher, The protein translation machinery is expressed for maximal efficiency in Escherichia coli. *Nat. Commun.* **11**, 5260 (2020).
6. E. R. Oldewurtel, Y. Kitahara, B. Cordier, G. Özbaykal, S. van Teeffelen, Bacteria control cell volume by coupling cell-surface expansion to dry-mass growth. *bioRxiv*, 769786 (2019).
7. Q. K. Beg, *et al.*, Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 12663–12668 (2007).
8. A. Y. Weiße, D. A. Oyarzún, V. Danos, P. S. Swain, Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci.* **112**, E1038–E1047 (2015).
9. D. Molenaar, R. van Berlo, D. de Ridder, B. Teusink, Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol. Syst. Biol.* **5** (2009).
10. H. Dourado, M. J. Lercher, An analytical theory of balanced cellular growth. *Nat. Commun.* **11**, 1226 (2020).
11. A. P. Minton, Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolym. Orig. Res. Biomol.* **20**, 2093–2120 (1981).
12. S. B. Zimmerman, A. P. Minton, Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **22**, 27–65 (1993).
13. G. B. Ralston, Effects of" crowding" in protein solutions. *J. Chem. Educ.* **67**, 857 (1990).
14. I. M. Kuznetsova, K. K. Turoverov, V. N. Uversky, What macromolecular crowding can do to a protein. *Int. J. Mol. Sci.* **15**, 23090–23140 (2014).
15. M. G. Norris, N. Malys, What is the true enzyme kinetics in the biological system? An investigation of macromolecular crowding effect upon enzyme kinetics of glucose-6-phosphate dehydrogenase. *Biochem. Biophys. Res. Commun.* **405**, 388–392 (2011).
16. S. N. Olsen, Applications of isothermal titration calorimetry to measure enzyme kinetics and activity in complex solutions. *Thermochim. Acta* **448**, 12–18 (2006).
17. K. Totani, Y. Ihara, I. Matsuo, Y. Ito, Effects of macromolecular crowding on glycoprotein processing enzymes. *J. Am. Chem. Soc.* **130**, 2101–2107 (2008).
18. I. Pozdnyakova, P. Wittung-Stafshede, Non-linear effects of macromolecular crowding on enzymatic activity of multi-copper oxidase. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **1804**, 740–744 (2010).
19. T. Kalwarczyk, M. Tabaka, R. Holyst, Biologistics–diffusion coefficients for complete proteome of Escherichia coli. *Bioinforma. Oxf. Engl.* **28**, 2971–2978 (2012).
20. T. Kihara, Virial Coefficients and Models of Molecules in Gases. *Rev. Mod. Phys.* **25**, 831–843 (1953).
21. A. P. Minton, Molecular crowding: analysis of effects of high concentrations of inert cosolutes on biochemical equilibria and rates in terms of volume exclusion. *Methods Enzymol.* **295**, 127–149 (1998).
22. A. Vazquez, Optimal cytoplasmatic density and flux balance model under macromolecular crowding effects. *J. Theor. Biol.* **264**, 356–359 (2010).
23. R. A. Fisher, *The Genetical Theory Of Natural Selection* (Clarendon Press, 1930).
24. K. Aoki, K. Takahashi, K. Kaizu, M. Matsuda, A quantitative model of ERK MAP kinase phosphorylation in crowded media. *Sci. Rep.* **3** (2013).
25. B. Lee, Calculation of volume fluctuation for globular protein models. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 622–626 (1983).
26. A. Bar-Even, *et al.*, The Moderately Efficient Enzyme: Evolutionary and Physicochemical Trends Shaping Enzyme Parameters. *Biochemistry* **50**, 4402–4410 (2011).

27. S. Klumpp, M. Scott, S. Pedersen, T. Hwa, Molecular crowding limits translation and cell growth. *Proc. Natl. Acad. Sci.* **110**, 16754–16759 (2013).
28. R. Adadi, B. Volkmer, R. Milo, M. Heinemann, T. Shlomi, Prediction of Microbial Growth Rate versus Biomass Yield by a Metabolic Network with Kinetic Parameters. *PLOS Comput. Biol.* **8**, e1002575 (2012).
29. A. Desouki, "Algorithms for Improving the Predictive Power of Flux Balance Analysis," Heinrich Heine University Düsseldorf, Düsseldorf. (2016) (October 10, 2020).
30. S. Hui, *et al.*, Quantitative proteomic analysis reveals a simple strategy of global resource allocation in bacteria. *Mol. Syst. Biol.* **11**, 784 (2015).
31. X. Dai, *et al.*, Reduction of translating ribosomes enables Escherichia coli to maintain elongation rates during slow growth. *Nat. Microbiol.* **2**, 1–9 (2016).
32. H. Fischer, I. Polikarpov, A. F. Craievich, Average protein density is a molecular-weight-dependent function. *Protein Sci. Publ. Protein Soc.* **13**, 2825–2828 (2004).
33. L.-M. Bobay, H. Ochman, Factors driving effective population size and pan-genome evolution in bacteria. *BMC Evol. Biol.* **18**, 153 (2018).
34. D. Sezer, P. Schubert, M. J. Lercher, Structure of autosynthetic models of balanced cell growth and numerical optimization of their growth rate. *bioRxiv*, 2020.09.19.304998 (2020).
35. C. L. Woldringh, N. Nanninga, "Structure of the Nucleoid and Cytoplasm in the Intact Cell" in *Molecular Cytology of Escherichia Coli*, (Academic Press Inc, 1985), pp. 161–197.
36. A. Schmidt, *et al.*, The quantitative and condition-dependent Escherichia coli proteome. *Nat. Biotechnol.* **34**, 104–110 (2015).
37. F. Trovato, V. Tozzini, Diffusion within the cytoplasm: a mesoscale model of interacting macromolecules. *Biophys. J.* **107**, 2579–2591 (2014).
38. W. G. McMillan, J. E. Mayer, The Statistical Thermodynamics of Multicomponent Systems. *J. Chem. Phys.* **13**, 276–305 (1945).
39. J. O. Hirschfelder, C. F. Curtiss, R. B. Bird, *The Molecular Theory of Gases and Liquids*, Revised Auflage (Wiley-Interscience, 1964).
40. T. Boublík, Statistical thermodynamics of convex molecule fluids. *Mol. Phys.* **27**, 1415–1427 (1974).
41. A. Minton, *Structural and Organizational Aspects of Metabolic Regulation; Chapter Holobiochemistry: An integrated Approach to the Understanding of Biochemical Mechanisms That Emerges from the Study of Proteins and Protein Associations in Volume-Occupied Solutions* (Wiley: Hoboken, NJ, USA, 1990).
42. M. Chwastyk, A. P. Bernaola, M. Cieplak, Statistical radii associated with amino acids to determine the contact map: fixing the structure of a type I cohesin domain in the Clostridium thermocellum cellulosome. *Phys. Biol.* **12**, 046002 (2015).
43. H. P. Erickson, Size and Shape of Protein Molecules at the Nanometer Level Determined by Sedimentation, Gel Filtration, and Electron Microscopy. *Biol. Proced. Online* **11**, 32–51 (2009).
44. J. Roskams, L. Rodgers, Eds., *Lab Ref: A Handbook of Recipes, Reagents, and Other Reference Tools for Use at the Bench* (Cold Spring Harbor Laboratory Press, 2002).
45. C. Hyeon, R. I. Dima, D. Thirumalai, Size, shape, and flexibility of RNA structures. *J. Chem. Phys.* **125**, 194905 (2006).
46. C. G. Kurland, Molecular characterization of ribonucleic acid from Escherichia coli ribosomes: I. Isolation and molecular weights. *J. Mol. Biol.* **2**, 83–91 (1960).
47. R. A. Freitas, *Kinematic Self-Replicating Machines*, 1st Ed. (CRC Press, 2004).
48. T. E. Shehata, A. G. Marr, Effect of Nutrient Concentration on the Growth of Escherichia coli. *J. Bacteriol.* **107**, 210–216 (1971).

49. N. V. Sahinidis, BARON: A general purpose global optimization software package. *J. Glob. Optim.* **8**, 201–205 (1996).
50. W. E. Hart, J.-P. Watson, D. L. Woodruff, Pyomo: modeling and solving mathematical programs in Python. *Math. Program. Comput.* **3**, 219 (2011).
51. W. E. Hart, *et al.*, *Pyomo — Optimization Modeling in Python*, 2nd Ed. (Springer International Publishing, 2017) https:/doi.org/10.1007/978-3-319-58821-6 (October 9, 2020).
52. P. Virtanen, *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
53. G. Gelius-Dietrich, A. A. Desouki, C. J. Fritzemeier, M. J. Lercher, sybil – Efficient constraint-based modelling in R. *BMC Syst. Biol.* **7**, 125 (2013).
54. A. A. Desouki, F. Jarre, G. Gelius-Dietrich, M. J. Lercher, CycleFreeFlux: efficient removal of thermodynamically infeasible loops from flux distributions. *Bioinformatics* **31**, 2159–2165 (2015).
55. D. Heckmann, *et al.*, Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. *Nat. Commun.* **9**, 5252 (2018).
56. D. Alzoubi, A. A. Desouki, M. J. Lercher, Flux balance analysis with or without molecular crowding fails to predict two thirds of experimentally observed epistasis in yeast. *Sci. Rep.* **9**, 11837 (2019).
57. A. M. Feist, *et al.*, A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.* **3**, 121 (2007).
58. T. Fuhrer, M. Zampieri, D. C. Sévin, U. Sauer, N. Zamboni, Genomewide landscape of gene–metabolome associations in Escherichia coli. *Mol. Syst. Biol.* **13**, 907 (2017).
59. R. Takeuchi, *et al.*, Colony-live — a high-throughput method for measuring microbial colony growth kinetics— reveals diverse growth effects of gene knockouts in Escherichia coli. *BMC Microbiol.* **14**, 171 (2014).

Figures

Figure 1. Limiting cases of crowding effects on kinetics. (A) Transition state limit, valid if the conversion of enzyme (E) plus substrate (S) to the complex ES is much faster than the conversion of ES to product (P). **(B)** Diffusion limit, where the conversion of ES to P is much faster than the formation of ES.

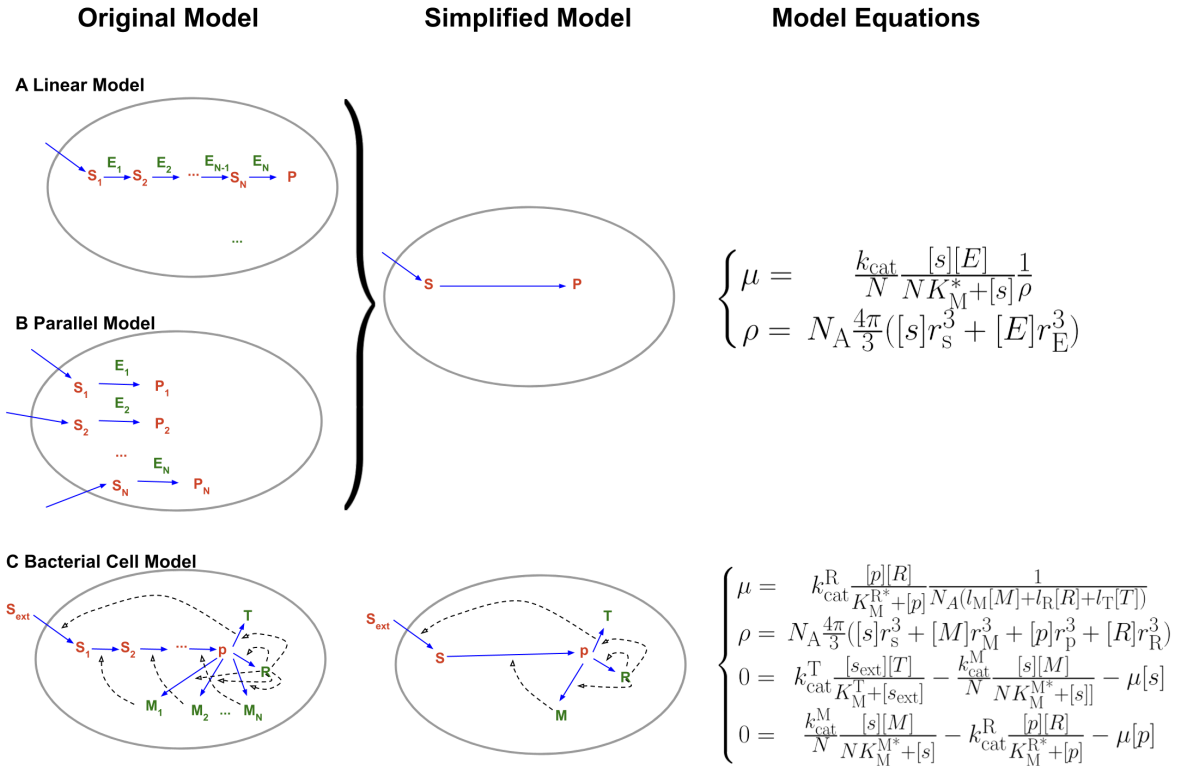

Figure 2. Illustration of the models. (A) Reaction system in an *N*-steps linear pathway, representing a metabolic system. The initial substrate, s_1, is replenished by a transport process not included in the model, and is converted by the pathway into the product *p*. **(B) Reaction system of *N* parallel reactions**, representing ribosomes presenting *N* distinct anticodons; s_1, ... , s_N are the corresponding ternary complexes, p_1, …, p_N are the extending amino acid chains. Assuming that (i) all reactions follow identical kinetics, (ii) all substrate concentrations s_i are identical, and (iii) all enzyme concentrations are identical, the fluxes of the models in (A) and (B) are both mathematically identical to the simplified model to the right of the two panels, with re-scaled concentrations $[s]=[s_1]+[s_2]+...+[s_N]$ and $[E]=[E_1]+[E_2]+...+[E_N]$. **(C) GBA model simulating the balanced growth of a bacterial cell.** Transporter *T* imports nutrient s_1, which is converted to the precursor for protein production *p* by a metabolic pathway with *N* consecutive enzymes, M_1, …, M_N, via intermediate substrates s_2, …, s_N. The ribosome *R* synthesizes the *N*+2 proteins (*T*, *R*, M_1, …, M_N) from *p*. Assuming identical concentrations of metabolic enzymes, $[M_1]=...[M_N]=:[M]$, and of metabolites, $[s_1]=...[s_N]=:[s]$, the solution space of this model cell is spanned by the five concentrations $[s]$,$[p]$,$[T]$,$[M]$,$[R]$.

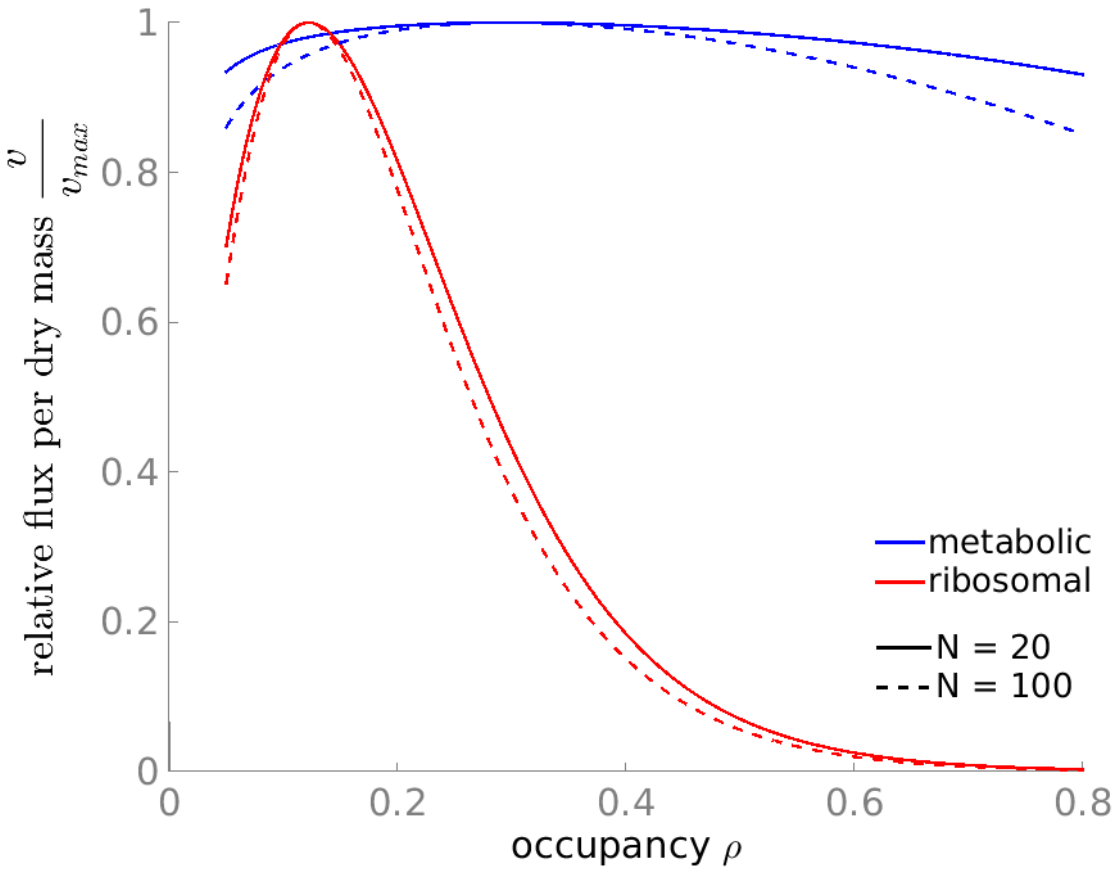

Figure 3. The cytosolic occupancy that facilitates maximal biochemical reaction fluxes is lower for ribosomal than for metabolic systems. Blue lines represent metabolic systems with small catalysts and substrates, red lines represent ribosomal systems with much larger molecules. Solid lines are for systems of *N*=20 consecutive (metabolic, blue) or parallel (ribosomal, red) reactions, dashed lines are for larger systems of *N*=100 reactions.

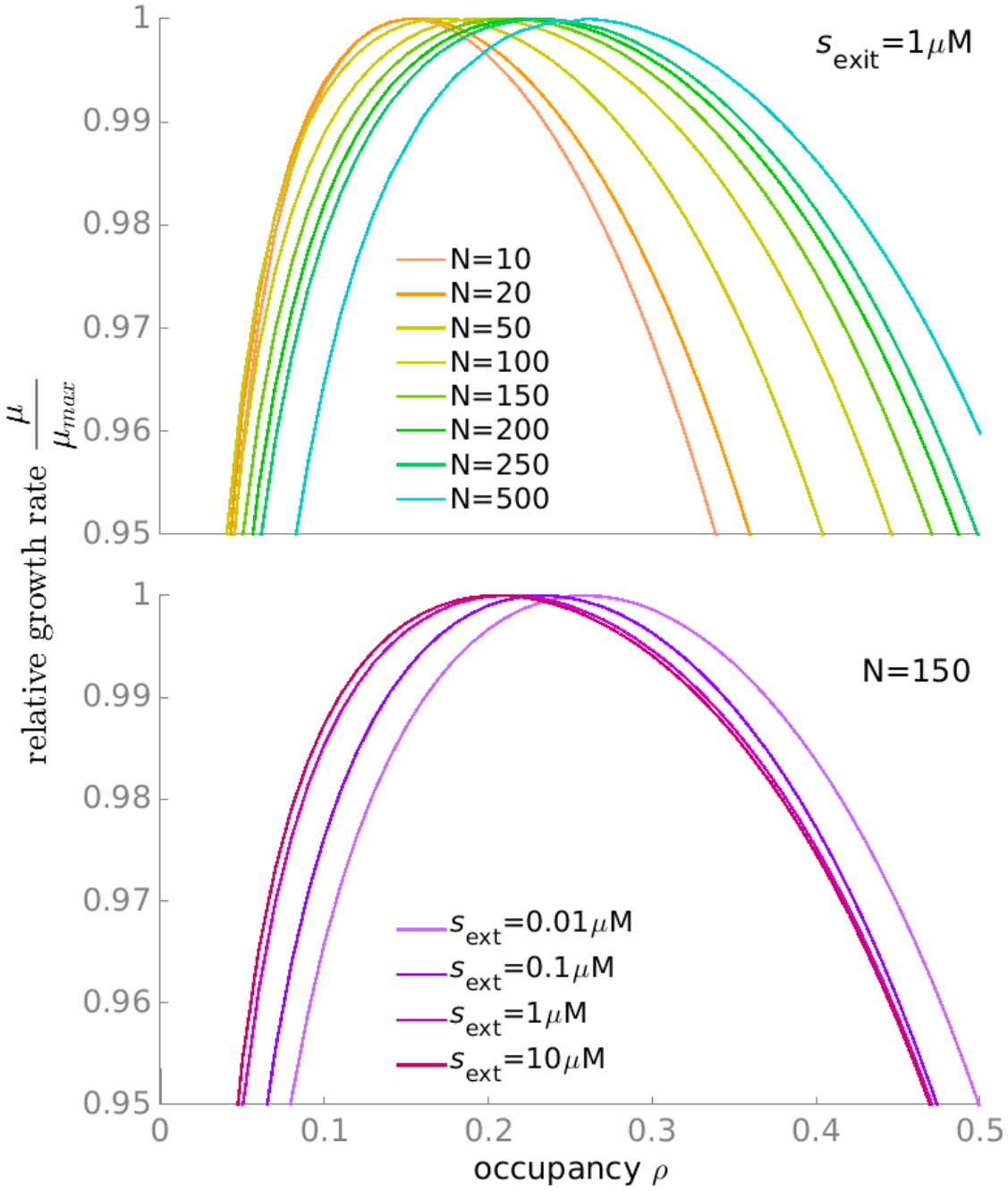

Figure 4. The growth rate *μ* of the crowding-aware whole-cell model constrained at different cytosolic occupancies *ρ*. The optimal cytosolic occupancy ρ_{opt} **(A)** increases with *N* (the number of enzymes in the metabolic pathway) and **(B)** decreases with external nutrient concentration [s_{ext}]. μ_{max} is the maximal growth rate for each curve across occupancies.

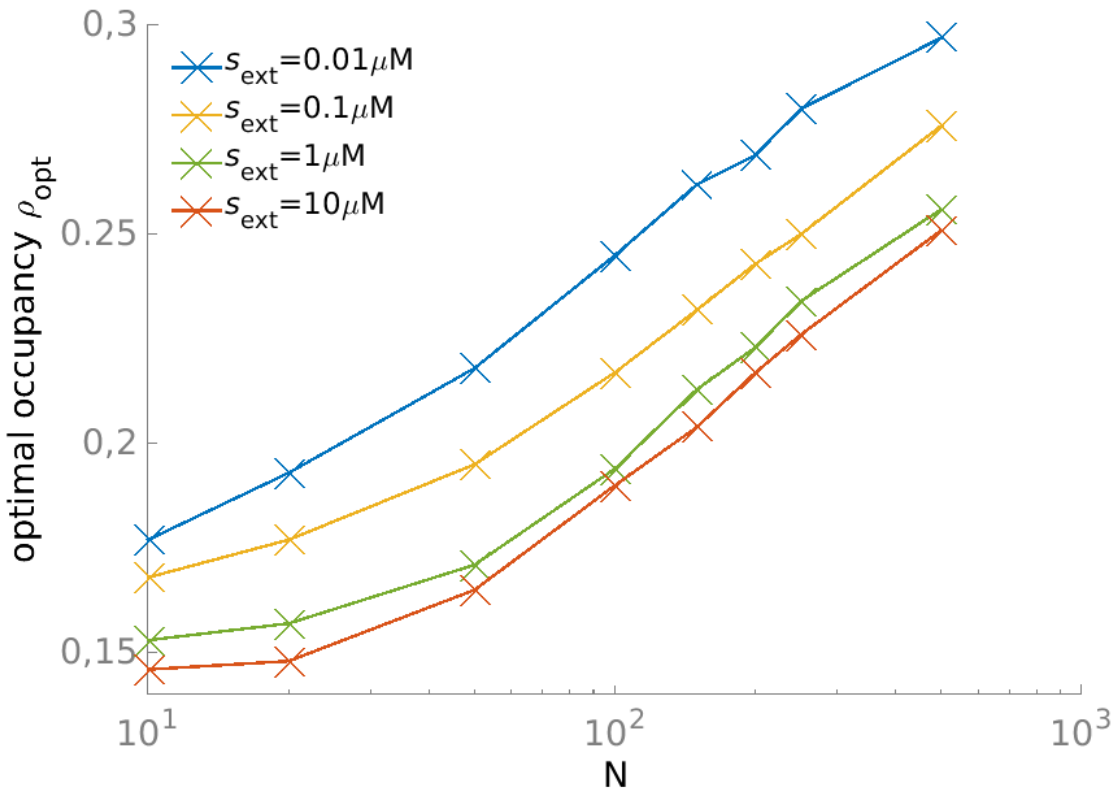

Figure 5. The optimal cytosolic occupancy increases with metabolic pathway length *N* and decreasing external nutrient concentration s_{ext}.

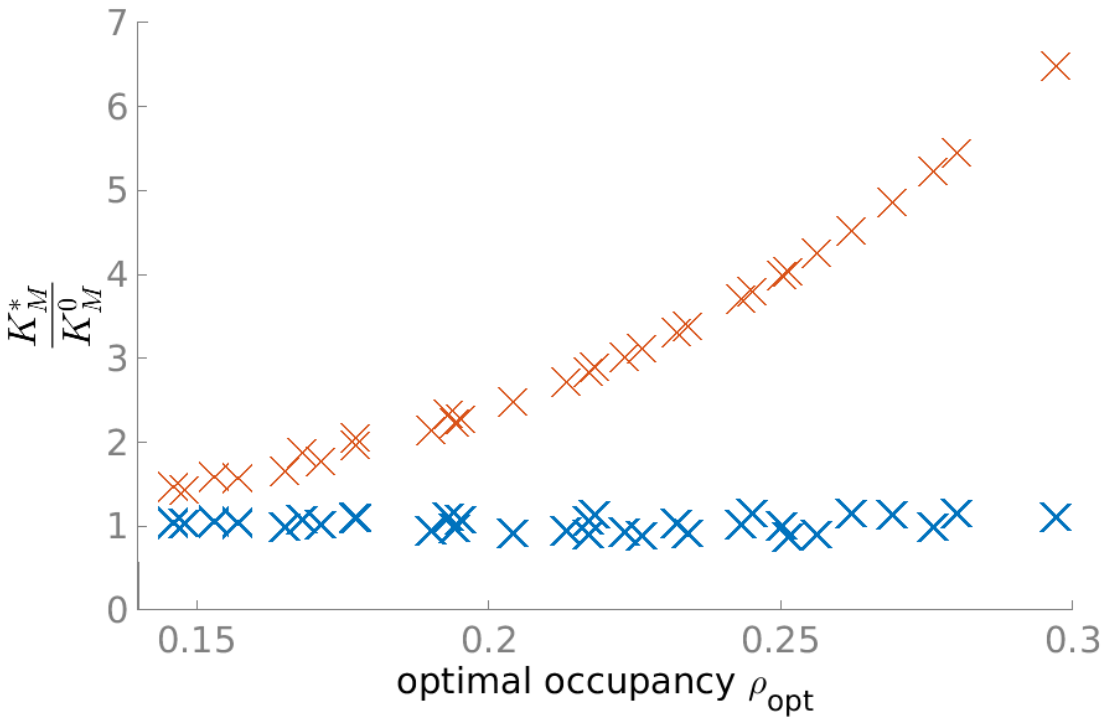

Figure 6. The optimal occupancy strongly influences the effective Michaelis parameter K_M^* of the ribosomal (red) but not of the metabolic (blue) reactions in the whole-cell model. Each data point corresponds to a different combination of external nutrient concentration s_{ext} and number of active metabolic reactions N. While the K_M^* of the metabolic reactions does not correlate with ρ_{opt} (blue markers; two-sided Spearman's rank correlation coefficient r=0.034, P=0.85), the K_M^* of the ribosomal reactions correlates with ρ_{opt} (red markers; r=0.998, $P<10^{-15}$). The discrete distribution of points along the x-axis reflects the step size used for ρ in the simulations.

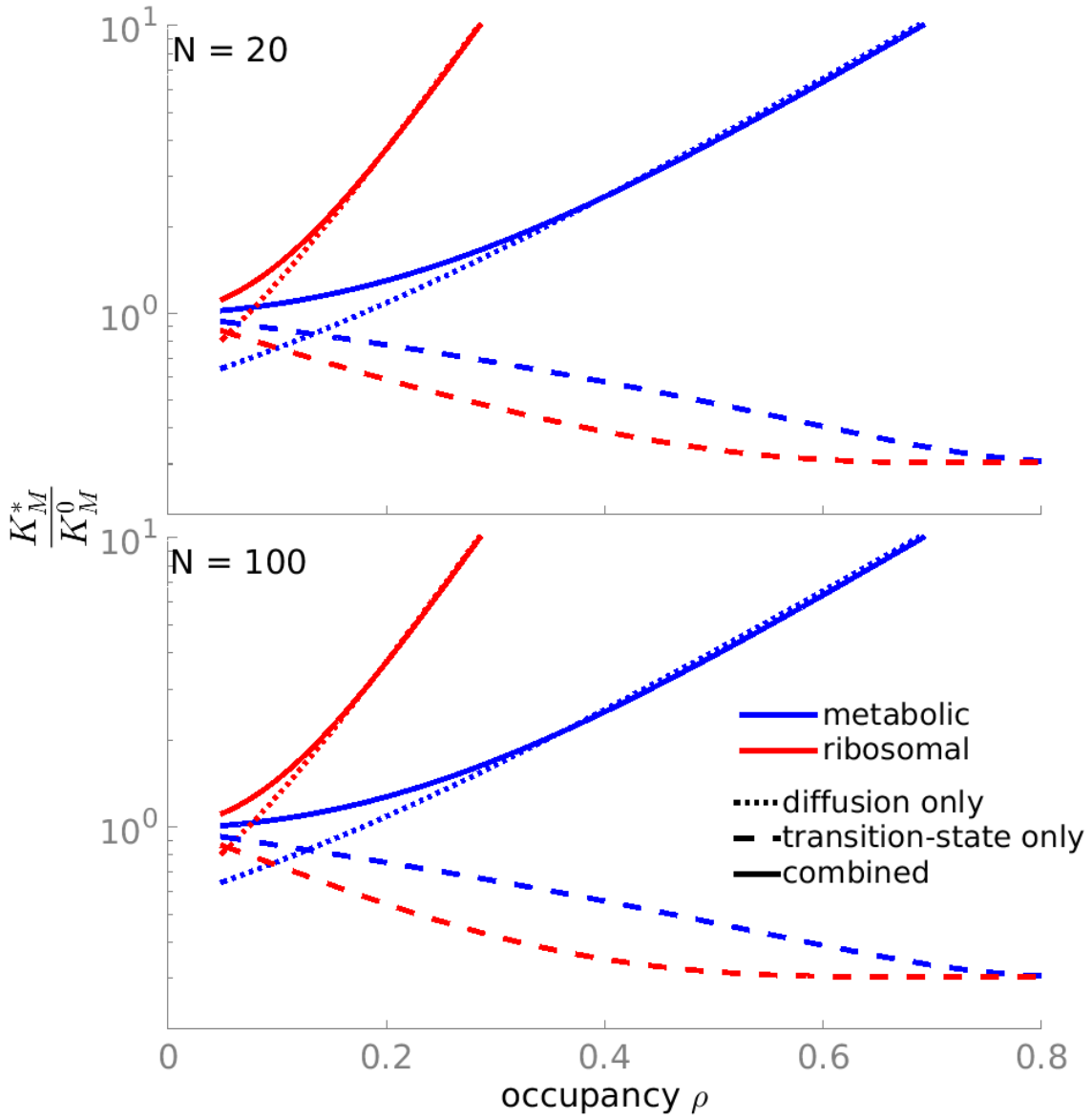

Figure 7. Opposing effects of molecular crowding on the Michaelis parameter K_M^* through perturbations of diffusion and of Gibbs free energies. Simulations considered only the respective effect for systems of (linear, blue) metabolic and (parallel, red) ribosomal reactions with pathway sizes of **(A)** *N*=20 and **(B)** *N*=100.