_computational
BIOLOGY

Enzyme clustering accelerates processing of intermediates through metabolic channeling

Michele Castellana[1,4], Maxwell Z Wilson[2,4], Yifan Xu[1,2], Preeti Joshi[2], Ileana M Cristea[2], Joshua D Rabinowitz[1,3], Zemer Gitai[2] & Ned S Wingreen[1,2]

We present a quantitative model to demonstrate that coclustering multiple enzymes into compact agglomerates accelerates the processing of intermediates, yielding the same efficiency benefits as direct channeling, a well-known mechanism in which enzymes are funneled between enzyme active sites through a physical tunnel. The model predicts the separation and size of coclusters that maximize metabolic efficiency, and this prediction is in agreement with previously reported spacings between coclusters in mammalian cells. For direct validation, we study a metabolic branch point in *Escherichia coli* and experimentally confirm the model prediction that enzyme agglomerates can accelerate the processing of a shared intermediate by one branch, and thus regulate steady-state flux division. Our studies establish a quantitative framework to understand coclustering-mediated metabolic channeling and its application to both efficiency improvement and metabolic regulation.

The small molecules of cellular metabolism are generally thought to rapidly diffuse throughout the cytoplasm. A notable exception to this occurs when the products of an enzyme active site are locally processed by a subsequent active site. This phenomenon, known as direct channeling, relies on the formation of protein tunnels that connect consecutive active sites, preventing metabolic intermediates from diffusing away (**Fig. 1a**). Indeed, numerous occurrences of direct channeling have been observed. Substrates can be channeled between active sites of two separate enzymes[1], or between two active sites within a single polyfunctional enzyme[2,3].

An alternative channeling mechanism, proximity channeling, is the subject of much speculation[4–7]. Proximity channeling involves two enzymes positioned near enough to each other such that the intermediate produced by the first enzyme is processed by the second enzyme before it can escape by diffusion, even in the absence of an actual channel (**Fig. 1b**). Proximity channeling has been invoked to explain the large improvements in product titers that occur when enzymes are colocalized on protein scaffolds[8,9].

However, there is a fundamental problem with proximity channeling in its simple form. For a diffusing substrate in the vicinity of an enzyme, processing only becomes likely if the substrate approaches the catalytic site of the enzyme within the radius of the site (~0.1–1 nm). Hence, an intermediate produced by the first member of an enzyme pair is unlikely to be processed by the second member even if the active sites of the two enzymes are only ~10 nm apart[7]. Thus, simply fusing two enzymes together will not cause productive channeling.

So how can we resolve the paradox that scaffolds empirically improve metabolic yields, yet cannot bring two enzymes close enough to mediate channeling? To address this question, we theoretically and experimentally explored the feasibility of another form of channeling, achieved by assembling multiple copies of both upstream and downstream enzymes into a functional cocluster we refer to as an 'agglomerate'. The central idea is that once an upstream enzyme produces an intermediate, even though the probability of the intermediate being processed by any individual downstream enzyme is low, the probability that the intermediate will be processed by one of the many downstream enzymes in the agglomerate can be high (**Fig. 1c**). Indeed, it has already been suggested that the spontaneous formation of multiscaffold agglomerates might account for the surprising effectiveness of engineered enzyme scaffolds[9].

Notably, evidence for enzyme clustering has been found in several organisms[10–20]. For example, the six enzymes of the human *de novo* purine biosynthetic pathway were observed to reversibly form clusters, named purinosomes, in HeLa cells in response to purine availability[17,21], though the role of expression levels and fluorescent fusions remains open[22]. Furthermore, in a recent study[18] of *Saccharomyces cerevisiae* it was observed that out of 800 GFP-tagged cytosolic proteins, 180 clearly displayed evidence of dynamic clustering, with many of these proteins involved in intermediary metabolism and stress response.

Although cluster-mediated channeling has been previously hypothesized[9], it has lacked a quantitative model to understand its limits, benefits and experimental support. Here we show quantitatively that compact agglomerates in which enzymes are coclustered offer many of the same advantages as direct channeling, in particular the acceleration of intermediate processing. Importantly, our theoretical approach allows us to find the optimal distribution of enzymes. For a representative linear pathway we are able to predict the optimal separation and size of agglomerates, as well as their detailed internal structure including enzyme ratios. Remarkably, the advantages of enzyme clustering are uniquely due to an enzyme-concentration effect. Two consecutive

enzymes in a pathway are present at high concentration in the same region of space. This mechanism induces efficient intermediate channeling without requiring any of the microscopic features of direct or proximity channeling scenarios, such as protein tunnels or extreme proximity between catalytic sites.

In support of our theoretical conclusions, we experimentally confirmed metabolic channeling by constructing an enzyme agglomerate. A robust steady-state way to detect channeling *in vivo* is to assess flux division at a metabolic branch point. Specifically, we focused on a fundamental branch point in *E. coli*, where carbamoyl phosphate synthetase (CarB) synthesizes carbamoyl phosphate, which can then be committed toward pyrimidine biosynthesis by aspartate carbamoyltransferase (PyrB) or toward arginine biosynthesis by ornithine carbamoyltransferase. We found that if CarB and PyrB are coclustered into an agglomerate, the flux is shunted toward the PyrB branch, leading to ample production of pyrimidines but low levels of arginine. We have extended our model for a linear pathway to the above branch point in *E. coli*, and the results confirm the shunting of metabolic flux owing to clustering observed in the experiments. To our knowledge, no other group has demonstrated the acceleration of intermediate processing by intentionally engineered enzyme agglomerates.

RESULTS
Model
We first describe a mathematical model describing a two-step metabolic pathway. We write the reaction-diffusion equations describing the pathway, discuss how to determine the efficiency of the pathway given a particular enzyme distribution, and how to find the enzyme distribution yielding the maximal efficiency (**Supplementary Figs. 1–5**).

Reaction-diffusion equations. We modeled a two-step metabolic pathway (**Fig. 2a**) where substrate S_0 is processed by enzyme E_1 into the intermediate S_1, which in turn can be processed by enzyme E_2 into product.

In what follows, we define a minimal model describing the above metabolic pathway. Given the large number of enzyme and substrate molecules, we represent their spatial distribution by continuous densities in the cytoplasm. The steady-state densities of enzymes E_1. E_2 are denoted by $n_1(\vec{r}), n_2(\vec{r})$, where $\vec{r}$ is the position vector. The concentration of S_0 at time t is denoted by $c_0(\vec{r}, t)$, and the concentration of S_1 at time t is denoted by $c_1(\vec{r}, t)$. We assume that the substrate S_0 is the product of upstream metabolic processes performed by enzymes that are distributed uniformly in the cytoplasm. These metabolic processes act to relax the local concentration $c_0(\vec{r}, t)$ toward a homeostatic level c_0^*. We will denote by α_0 the rate at which $c_0(\vec{r}, t)$ relaxes to c_0^*. To model intermediate decay[23] and/or incorrect processing[24], we assume $c_1(\vec{r}, t)$ decays at a rate β. For $\beta = 0$, intermediates can accumulate indefinitely curtailing the advantage of clustering, although clustering can still reduce overall metabolite concentrations (**Supplementary Fig. 6**).

Hence, the reaction-diffusion equations governing the two-step pathway are

$$\begin{cases} \dfrac{\partial c_0(\vec{r},t)}{\partial t} = -\alpha_0\left(c_0(\vec{r},t)-c_0^*\right)-k_1 n_1(\vec{r})c_0(\vec{r},t)+D\nabla^2 c_0(\vec{r},t) \\ \dfrac{\partial c_1(\vec{r},t)}{\partial t} = k_1 n_1(\vec{r})c_0(\vec{r},t)-k_2 n_2(\vec{r})c_1(\vec{r},t)-\beta c_1(\vec{r},t)+D\nabla^2 c_1(\vec{r},t) \end{cases} \quad (1)$$

where k_1, k_2 are the k_{cat}/K_M values of enzymes E_1 and E_2 from Michaelis-Menten kinetics, and D is the diffusion coefficient.
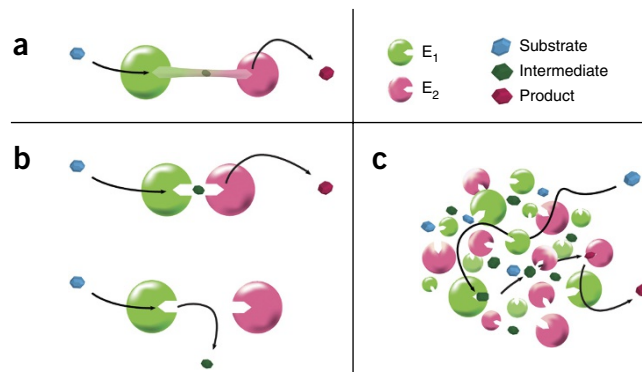

Figure 1 Different types of intermediate channeling in a two-step metabolic pathway, where a substrate is processed by enzyme E_1 and turned into intermediate, which is then processed by enzyme E_2 and turned into product. (**a**) Direct channeling. The intermediate is funneled from enzyme E_1 to enzyme E_2 by means of a protein tunnel that connects the active sites of E_1 and E_2, thus preventing the intermediate from diffusing away. (**b**) Proximity channeling. Top: E_1 and E_2 are positioned near enough to each other such that the intermediate produced by E_1 is processed by E_2 before it can escape by diffusion, even in the absence of an actual channel. Bottom: if E_1 and E_2 are not near enough to each other, an intermediate molecule produced by E_1 escapes by diffusion, and it cannot be processed by E_2. (**c**) Enzyme clustering. Once E_1 produces an intermediate molecule, even though the probability of the intermediate being processed by any individual E_2 enzyme is low, the probability that the intermediate will be processed by one of the many E_2 enzymes in the agglomerate can be high.

Because the time scale of enzyme clustering (>1 h)[18] is much longer than the time scale of intermediate processing ($\sim$l s), we will focus on the steady state of equation (1), where $\partial c_i(\vec{r},t)/\partial t = 0$, and thus set $c(\vec{r},t) = c(\vec{r})$.
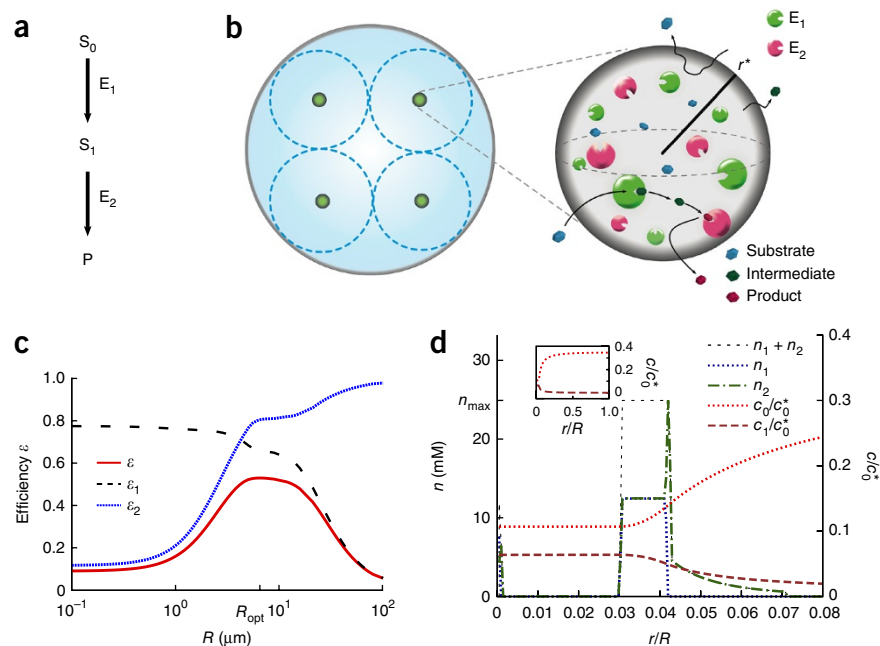
Efficiency of the pathway. An important quantity is the efficiency

$$\varepsilon[n_1,n_2] \equiv \frac{k_2 \int_V d^3 r\, n_2(\vec{r})c_1(\vec{r})}{\alpha_0 V c_0^*} \quad (2)$$

The efficiency $0 \le \varepsilon \le 1$ is the total rate of production of product in the volume V of the system divided by the maximal possible rate of production of the substrate of E_1. Given the enzyme distributions $n_1(\vec{r}), n_2(\vec{r})$, one can solve equation (1), determine the concentrations $c_0(\vec{r}), c_1(\vec{r})$, and thus compute the efficiency $\varepsilon[n_1, n_2]$ of the enzyme configuration $n_1(\vec{r}), n_2(\vec{r})$ by means of equation (2).

Before seeking the enzyme distributions $n_1(\vec{r}), n_2(\vec{r})$ that optimize the efficiency, we observe that for enzyme clustering in yeast[25], mammalian cells[17], and some bacteria[26], the overall enzyme distribution consists of one or more well-separated dense clusters. Guided by this observation, in our model we assumed that the optimal distribution $n_1(\vec{r}), n_2(\vec{r})$ is given by the identical repetition of a simple pattern: an enzyme cluster and its surrounding volume (**Fig. 2b**). We approximated this surrounding volume as a sphere of radius R, which we call the 'basin' of the cluster. Furthermore, we expect and assume spherical symmetry within each basin, that is, enzyme densities and substrate concentrations depend only on the distance r from the center of the basin $c_i(\vec{r}) = c_i(r)$, $n_i(\vec{r}) = n_i(r)$, and we fixed no-flux boundary conditions at the edge of the basin.

Figure 2 Two-step metabolic pathway with an unstable intermediate. (**a**) The two-step metabolic pathway. Substrate S_0 is processed by enzyme E_1 and turned into intermediate S_1, which is then processed by enzyme E_2 and turned into product P. (**b**) Enzyme configurations in the two-step metabolic pathway. Left: the cell cytoplasm is divided into multiple identical basins, each basin is represented by a dashed circle. Within each basin, enzymes are clustered into a central spherical agglomerate (shown in green). Right: blow-up showing the dynamics of the metabolic pathway within an agglomerate with radius r^*. Substrate S_0, which is produced throughout the cytoplasm, is processed by E_1 and turned into S_1, which is then processed by E_2. Both S_0 and S_1 may escape from the agglomerate by diffusion. (**c**) Metabolic pathway efficiency ε and efficiencies of the first and second step ε_1, ε_2 as functions of basin radius R. For each R, efficiency is optimized over enzyme densities $n_1(r),n_2(r)$, which are assumed to be spherically symmetric. Local enzyme density is constrained by $n_1(r)+n_2(r) \leq n_{max}$, and the total catalytic activity is fixed to κ_{cat}. The efficiencies of the first and second step ε_1, ε_2 are a decreasing and an increasing function of R respectively. Hence, the optimal efficiency is obtained as a tradeoff between ε_1 and ε_2, and is equal to $\varepsilon_{opt} = 0.53$ at $R_{opt} = 6.5\ \mu m$. Except where noted, $k_1 = k_2$ and parameter values are the same for all figures. The optimal efficiency $\varepsilon_{opt} = 0.53$ is about 5.9 times larger than the efficiency $\varepsilon_{delocalized} = 0.09$ of a delocalized configuration where enzymes are uniformly distributed in space. (**d**) Optimal distributions of enzymes E_1,E_2 and corresponding concentrations of substrate S_0 and intermediate S_1 as functions of r/R, where the optimal basin radius (i.e., half the optimal spacing between clusters) is $R = R_{opt} = 6.5\ \mu m$ from **b**. The local enzyme density $n_1(r) + n_2(r)$ and its maximal value n_{max} are also shown. The optimal enzyme distribution is a compact cluster with radius $r^* \approx 0.26\ \mu m$ composed of a shell of E_1 and E_2 surrounded by a halo of E_2. Inset: concentrations of substrate S_0 and intermediate S_1 as functions of r/R in the entire basin.

To compute the efficiency of an enzyme configuration $n_1(\vec{r}),n_2(\vec{r})$ and determine the enzyme configuration with maximal efficiency, we do not have to solve equation (1) in the whole volume V. We only need to identify the size R of the basin, and solve the reaction-diffusion equations within one basin. Because all basins are identical, the efficiency of the configuration $n_1(\vec{r}),n_2(\vec{r})$ will be then given by the efficiency $\varepsilon[n_1, n_2]$ computed within one single basin

$$\varepsilon\left[n_1, n_2; R\right] = \frac{k_2 \int_0^R dr\, 4\pi r^2 n_2(r) c_1(r)}{\alpha_0 (4/3)\pi R^3 c_0^*} \quad (3)$$

Determining the optimal enzyme distribution. We computed the basin radius R and the enzyme configuration $n_1(r), n_2(r)$ that maximize the efficiency in equation (3). As there are natural constraints both on enzyme catalytic constants and on the cell's enzyme production, we chose to optimize the efficiency while holding constant the total catalytic activity

$$k_1 \frac{N_1}{(4/3)\pi R^3} + k_2 \frac{N_2}{(4/3)\pi R^3} = \kappa_{cat} \quad (4)$$

where

$$N_i \equiv \int_0^R dr\, 4\pi r^2 n_i(r)$$

are the total enzyme numbers in the basin. Another physical constraint that must be taken into account is the finite size of enzymes.

This implies a maximal local enzyme density n_{max} that cannot be exceeded[27], which we modeled by the constraint

$$n_1(r) + n_2(r) \leq n_{max} \quad (5)$$

The optimization of the efficiency was performed under the constraints (4,5) (**Supplementary Fig. 1** and **Supplementary Software 1**), the values of the model parameters (diffusion coefficients, catalytic rates, etc.) being fixed from experimental data (Online Methods: Model parameters).

In the following section we present and discuss the enzyme distribution yielding the maximal efficiency.

Optimal enzyme distribution

For our chosen parameters the optimal size of the basin is $R = 6.5\ \mu m$ (**Fig. 2c**). Thus, if the cell radius is large enough, the optimal enzyme distribution consists of multiple enzyme clusters spaced by ~6.5 μm. Notably, this distance is similar to the spacing between purine biosynthesis enzyme clusters in human cells[17].

For the optimal basin radius, **Figure 2d** shows the distribution of enzymes that optimizes metabolic efficiency. E_2 is colocalized with E_1 in a shell to be able to process the intermediate where S_1 is produced, and the extra shell and halo of E_2 help to process molecules of S_1 that might otherwise diffuse away. The outer radius $r^* \approx 0.26\ \mu m$ of the spherical shell of E_1 and E_2 is only about 4% of the basin radius R. Within the shell, the total enzyme density is the maximum value allowed by crowding n_{max}, and both of the enzyme distributions are nearly uniform. The fraction of enzyme E_1 in this optimal configuration with $k_1 = k_2$ is $N_1/(N_1 + N_2) = 0.33$.

Figure 3 Two-step metabolic pathway with an unstable intermediate for uniform enzyme spheres with equal density. (**a**) Metabolic pathway efficiency for uniform enzyme spheres of enzymes E_1 and E_2 with $N_1 = N_2$ compared to the optimal case from **Figure 2c**, and efficiencies of the first and second step of the pathway for uniform enzyme spheres as functions of basin radius R. For each R, efficiency is optimized over enzyme densities $n_1(r), n_2(r)$, which are assumed to be spherically symmetric. Local enzyme density is constrained by $n_1(r)+n_2(r) \leq n_{max}$, and the total catalytic activity is fixed to κ_{cat}. The optimal efficiency and radius for uniform enzyme spheres are $\varepsilon_{opt}^{spheres} = 0.50$ and $R_{opt}^{spheres} = 6.15\ \mu m$,

close to the values $\varepsilon_{opt} = 0.53$ and $R_{opt} = 6.5\ \mu m$ obtained for the fully optimized case. (**b**) Optimal distributions of enzymes E_1, E_2 and corresponding concentrations of substrate S_0 and intermediate S_1 for uniform enzyme spheres of enzymes E_1 and E_2 as functions of r/R, where the basin radius is $R = R_{opt}^{spheres} = 6.15\ \mu m$ from **a**. The local enzyme density $n_1(r)+n_2(r)$ and its maximal value n_{max} are also shown. The optimal enzyme distribution is a compact cluster with radius $r^* \approx 0.25\ \mu m$ composed of a sphere uniformly filled with enzymes E_1 and E_2. Inset: concentrations of substrate S_0 and intermediate S_1 as functions of r/R in the entire basin region.
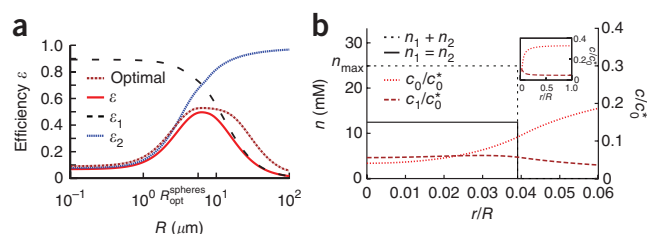
How advantageous is clustering? The optimal efficiency, $\varepsilon_{opt}=0.53$ (**Fig. 2**) for our choice of intermediate decay rate $\beta = 10/s$, should be compared with the efficiency in the delocalized case, for example, the case where enzymes uniformly fill the whole cytoplasm. The latter was found to be $\varepsilon_{delocalized} = 0.09$, showing that enzyme clustering improved the efficiency of the two-step pathway by almost sixfold.

Notably, the detailed internal structure of the enzyme distribution (**Fig. 2d**) was not essential to achieve a high metabolic efficiency. A much simpler enzyme configuration where both E_1 and E_2 were uniformly distributed in a maximally dense enzyme sphere of radius $\sim r^*$ (**Fig. 3**) yielded an efficiency within 6% of the optimal efficiency $\varepsilon_{opt} = 0.53$ of the enzyme configuration shown in **Figure 2d** (**Supplementary Software 2**).

It can be shown that the optimal basin radius R_{opt} is the best compromise between the efficiency of the first and second steps of the pathway (**Fig. 2c**). For large R, diffusion of S_0 to the enzyme cluster is limiting and the maximum possible rate of production of S_0 greatly exceeds the rate of processing by E_1 leading to small ε_1. By contrast, for small R the enzyme agglomerate is so small that much of the intermediate S_1 escapes by diffusion, resulting in small ε_2. Note that for a cell whose radius is much smaller than the optimal basin radius $R_{opt} = 6.5\ \mu m$, the highest achievable efficiency was substantially reduced compared to the optimal value $\varepsilon_{opt} = 0.53$ for the pathway parameters used in **Figure 2**.

Finally, we performed the optimization in **Figure 2c,d** for a range of values of each relevant parameter (**Supplementary Fig. 4**), and showed that these results were robust with respect to the assumptions used to build the model, that is spherical symmetry, instability of intermediates and absence of enzyme saturation (**Supplementary Figs. 5–7**).
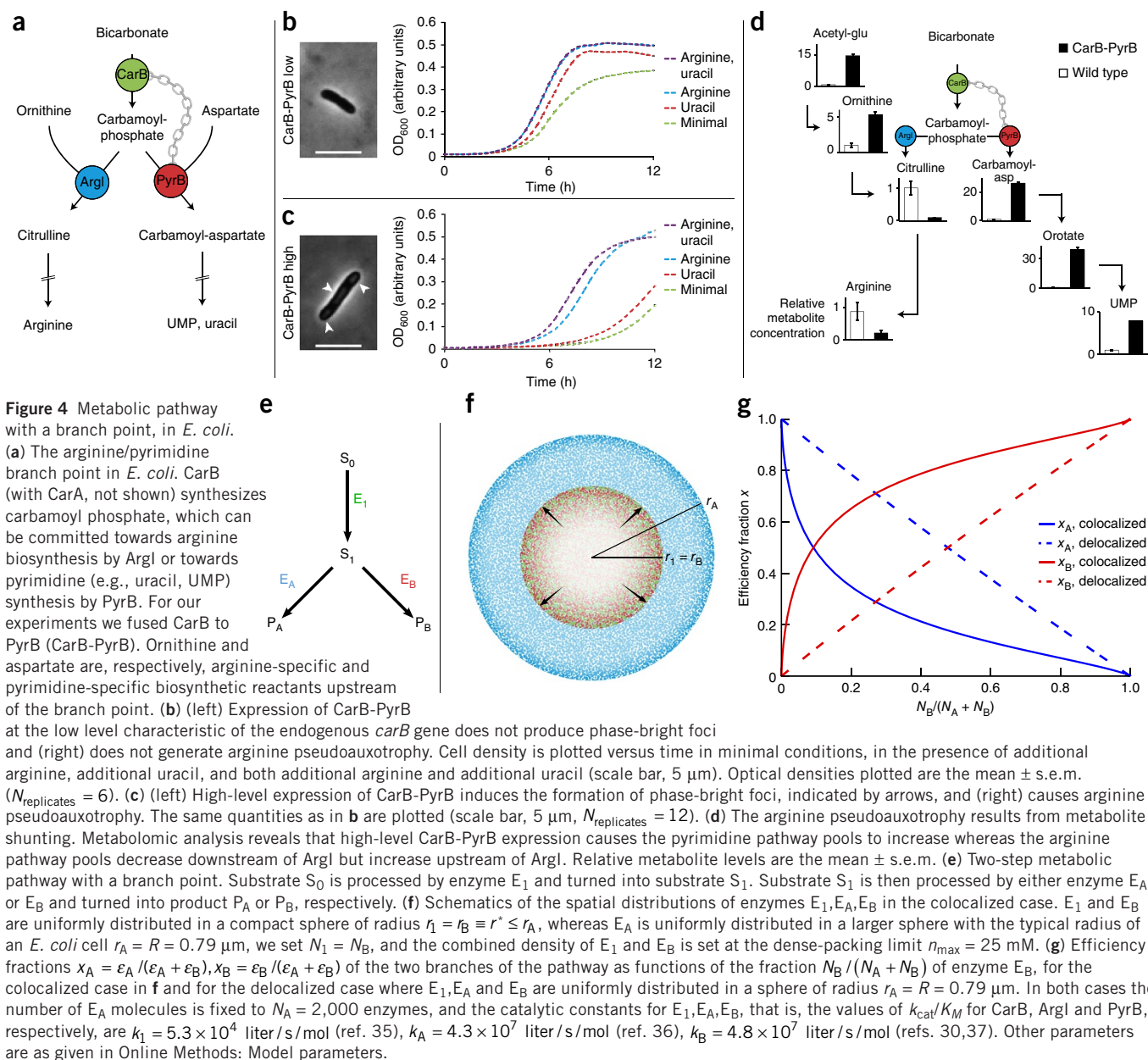
Branch-point regulation by enzyme clustering in *E. coli*

An important question is whether enzyme clustering can do more inside cells than increase metabolic efficiency. An intriguing possibility is that clustering might prove effective in regulating flux division, specifically at metabolic branch points. To experimentally test this possibility, we focused on a fundamental branch point in *E. coli*, where carbamoyl phosphate synthetase (CarB) synthesizes carbamoyl phosphate, an intermediate that can then be committed toward pyrimidine biosynthesis by aspartate carbamoyltransferase (PyrB) or toward arginine biosynthesis by ornithine carbamoyltransferase (ArgI). Analysis of the catalytic parameters specific to this branch point suggests that a large shunting of the metabolic flux can be obtained by coclustering the upstream enzyme with the downstream enzyme for one branch. In this regard, suppose that we colocalize a typical number[28,29] of CarB and PyrB molecules $N_{CarB} = N_{PyrB} \sim 2,000$ in a cluster with radius $r^* \sim 40\ nm$, so that the combined density of CarB

and PyrB is equal to the dense-packing limit $n_{max} = 25\ mM$. Given the catalytic constant of PyrB $k_{PyrB} = 4.8 \times 10^7$ liter/s/mol (ref. 30) and an intermediate diffusion coefficient $D = 100\ \mu m^2/s$ (ref. 31), the rate at which an intermediate molecule produced within the cluster is processed by PyrB is $k_{PyrB}N_{PyrB}/(4/3\pi r^{*3}) \sim 5.9 \times 10^5/s$, whereas the rate at which the intermediate molecule escapes from the cluster is $D/r^{*2} \sim 6.2 \times 10^4/s$. As the processing rate is much larger than the escape rate, we expect that by colocalizing CarB and PyrB in a compact cluster we can efficiently channel the intermediate (carbamoyl phosphate) toward the PyrB branch, and thus obtain substantial shunting of flux. Such flux shunting toward the PyrB branch should manifest in high pyrimidine production and low arginine production. The growth of *E. coli* cells would be consequently stimulated by the presence of additional arginine and insensitive to the presence of additional uracil (a pyrimidine precursor). We refer to such a growth dependence on arginine as arginine pseudoauxotrophy.

Simply fusing CarB to PyrB does not induce channeling. We translationally fused CarB to PyrB (hereafter called CarB-PyrB, **Fig. 4**) and confirmed that each of the two enzymes in the resulting fusion is functional by demonstrating that the fusion complements the growth of a $\Delta carB \Delta pyrB$ double-deletion mutant. We replaced the native *carB* gene with the *carB-pyrB* fusion in an NCM3722 $\Delta pyrB$ background and assayed growth in minimal media with or without additional arginine or uracil. In all conditions, CarB-PyrB supported wild-type growth with no signs of auxotrophy for either pyrimidines or arginine (**Fig. 4b**). Thus, there was no evidence that simply fusing CarB and PyrB increased flux of carbamoyl phosphate toward PyrB at the expense of ArgI.

Channeling occurs when CarB-PyrB forms large clusters. We expected that flux shunting at the CarB-PyrB-ArgI metabolic branch point could be achieved when enzymes were coclustered into functional agglomerates. To test this prediction, we engineered many-enzyme clusters by overexpressing the CarB-PyrB fusion protein, and examined the resulting effect on flux shunting. We conditionally expressed the *carB-pyrB* gene fusion from the tetracycline-inducible $P_{LtetO-1}$ promoter and assayed CarB-PyrB agglomeration and function in a $\Delta carB \Delta pyrB$ background. At low induction levels (0–0.1 nM anhydrotetracycline (aTc)) cells exhibited generally homogeneous cytoplasm and wild-type growth patterns (**Fig. 5a,b**). At higher induction levels, however (above 0.5 nM aTc), cells began to exhibit phase-bright cytoplasmic structures typical of protein-dense clusters (**Fig. 5b,c**), and became strikingly pseudoauxotrophic for arginine: cell growth was insensitive to the presence of additional uracil, but was strongly stimulated by the addition of arginine (**Fig. 5**). The growth dependence on arginine became increasingly exaggerated upon the addition

Figure 4 Metabolic pathway with a branch point, in *E. coli*. (**a**) The arginine/pyrimidine branch point in *E. coli*. CarB (with CarA, not shown) synthesizes carbamoyl phosphate, which can be committed towards arginine biosynthesis by ArgI or towards pyrimidine (e.g., uracil, UMP) synthesis by PyrB. For our experiments we fused CarB to PyrB (CarB-PyrB). Ornithine and aspartate are, respectively, arginine-specific and pyrimidine-specific biosynthetic reactants upstream of the branch point. (**b**) (left) Expression of CarB-PyrB at the low level characteristic of the endogenous *carB* gene does not produce phase-bright foci and (right) does not generate arginine pseudoauxotrophy. Cell density is plotted versus time in minimal conditions, in the presence of additional arginine, additional uracil, and both additional arginine and additional uracil (scale bar, 5 μm). Optical densities plotted are the mean ± s.e.m. ($N_{replicates}$ = 6). (**c**) (left) High-level expression of CarB-PyrB induces the formation of phase-bright foci, indicated by arrows, and (right) causes arginine pseudoauxotrophy. The same quantities as in **b** are plotted (scale bar, 5 μm, $N_{replicates}$ = 12). (**d**) The arginine pseudoauxotrophy results from metabolite shunting. Metabolomic analysis reveals that high-level CarB-PyrB expression causes the pyrimidine pathway pools to increase whereas the arginine pathway pools decrease downstream of ArgI but increase upstream of ArgI. Relative metabolite levels are the mean ± s.e.m. (**e**) Two-step metabolic pathway with a branch point. Substrate S_0 is processed by enzyme E_1 and turned into substrate S_1. Substrate S_1 is then processed by either enzyme E_A or E_B and turned into product P_A or P_B, respectively. (**f**) Schematics of the spatial distributions of enzymes E_1, E_A, E_B in the colocalized case. E_1 and E_B are uniformly distributed in a compact sphere of radius $r_1 = r_B \equiv r^* \leq r_A$, whereas E_A is uniformly distributed in a larger sphere with the typical radius of an *E. coli* cell $r_A = R = 0.79$ μm, we set $N_1 = N_B$, and the combined density of E_1 and E_B is set at the dense-packing limit $n_{max} = 25$ mM. (**g**) Efficiency fractions $x_A = \varepsilon_A /(\varepsilon_A + \varepsilon_B), x_B = \varepsilon_B /(\varepsilon_A + \varepsilon_B)$ of the two branches of the pathway as functions of the fraction $N_B /(N_A + N_B)$ of enzyme E_B, for the colocalized case in **f** and for the delocalized case where E_1, E_A and E_B are uniformly distributed in a sphere of radius $r_A = R = 0.79$ μm. In both cases the number of E_A molecules is fixed to $N_A = 2,000$ enzymes, and the catalytic constants for E_1, E_A, E_B, that is, the values of k_{cat}/K_M for CarB, ArgI and PyrB, respectively, are $k_1 = 5.3 \times 10^4$ liter/s/mol (ref. 35), $k_A = 4.3 \times 10^7$ liter/s/mol (ref. 36), $k_B = 4.8 \times 10^7$ liter/s/mol (refs. 30,37). Other parameters are as given in Online Methods: Model parameters.

of higher concentrations of aTc inducer (**Fig. 5a**), and was most pronounced in an unrepressed (maximally expressed) strain lacking the *tetR* repressor gene altogether (**Fig. 4c**). For this unrepressed strain, the exponential phase doubling time in minimal media supplemented with arginine was faster by 26.6 ± 0.9 min, or about 30%, compared to the doubling time in minimal media alone (**Supplementary Table 1**). If we assume that the fusion strain's reduction in growth rate in comparison with wild type reflects the fraction of carbamoyl phosphate flux being channeled away from arginine biosynthesis, then we infer that the ratio of arginine biosynthetic fluxes between the clustered and unclustered strains is about 1:2 (from doubling times of 89.4 ± 0.6 min and 47.9 ± 0.6 min; **Supplementary Table 1**).

We have also shown that the arginine pseudoauxotrophy was not simply a consequence of PyrB overexpression, but rather required CarB and PyrB to be coclustered and PyrB to be functional (**Supplementary Figs. 8–9**, Online Methods: Arginine pseudoauxotrophy results from coclustering of CarB and PyrB, and **Supplementary Video 1**).

The hallmark of branch-point shunting was observed downstream of CarB-PyrB with large increases in pyrimidine pathway metabolite pools (as large as 50×) and a concomitant drop in arginine pathway metabolite pools (**Fig. 4d** and **Supplementary Fig. 10**). Furthermore, the arginine-specific biosynthetic intermediates upstream of the branch point, such as ornithine, accumulated significantly ($P = 4.0 \times 10^{-4}$) (**Fig. 4d**). This result indicates that ArgI was unable to access its carbamoyl phosphate substrate, which it requires to produce citrulline by carbamoylating ornithine (for evidence that clusters contained active enzyme see **Supplementary Fig. 11**).

Model for branch-point regulation by clustering in *E. coli*

Our model predicts that enzyme clustering yields high metabolic efficiency only for cells with radius larger than ~6.5 μm (**Fig. 2c**), with the effectiveness of enzyme clustering progressively reduced for smaller cells. Nevertheless, the above experiments for a branch point in *E. coli* show that enzyme clustering may also be advantageous in
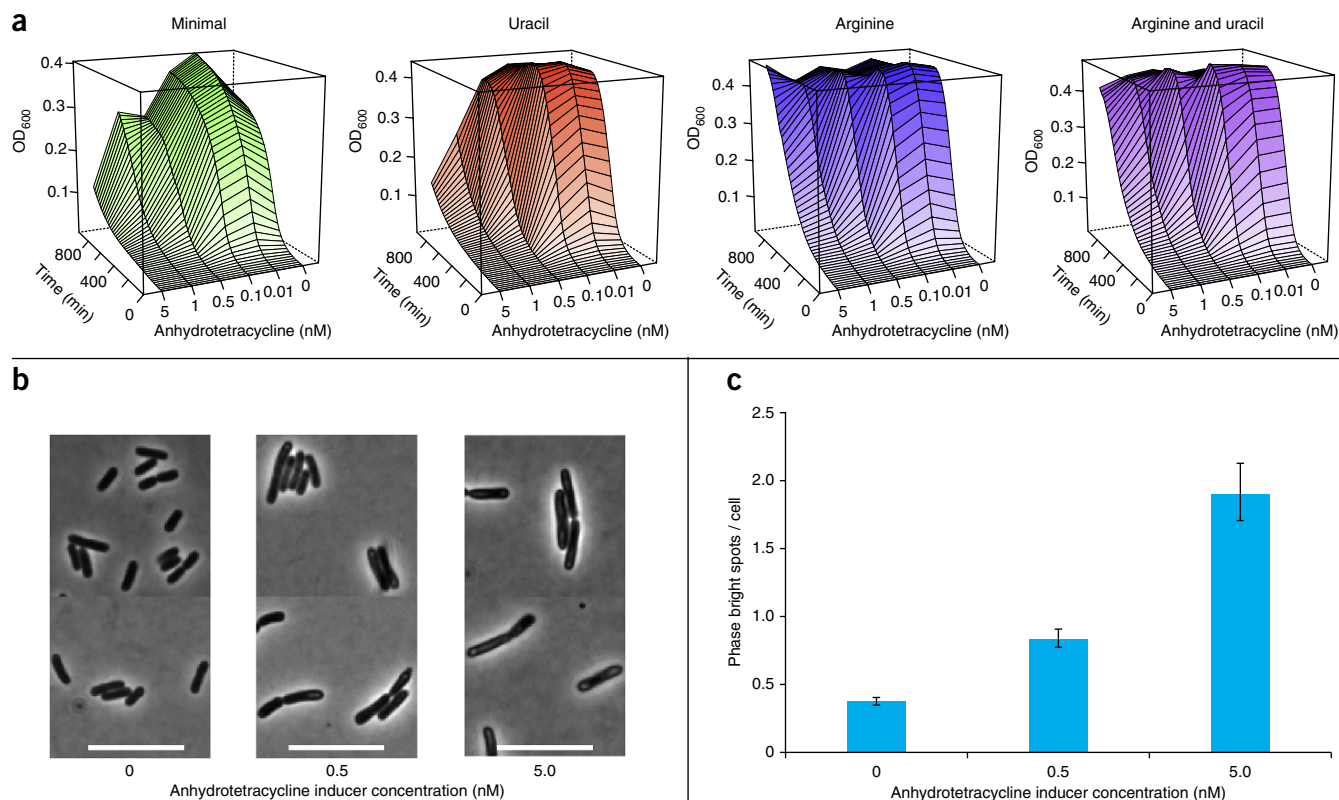
Figure 5 Phase-bright clusters and arginine pseudoauxotrophy increase with increasing CarB-PyrB overexpression. (**a**) Cell density with inducible CarB-PyrB expression as a function of time for different levels of anhydrotetracyline (aTc) inducer concentration and in different media: in minimal conditions, in the presence of additional arginine, additional uracil and both additional arginine, and additional uracil. Mean of $N_{replicates} = 4$ plotted. (**b,c**) Phase images (**b**) (scale bars, 5 µm) and quantification (**c**) of phase-bright foci as functions of aTc inducer concentration. Number of phase-bright spots per cell followed an exponential distribution. Mean and 95% confidence intervals are shown as the s.e.m. ($N_{0 nM} = 760$, $N_{0.5 nM} = 632$, $N_{5 nM} = 317$).

smaller cells. Importantly, compared to the 'typical' catalytic constants $k_1 = k_2 = 10^6$ liter/s/mol used in **Figure 2c**, the catalytic constant of PyrB $k_{PyrB} = 4.8 \times 10^7$ liter/s/mol is much larger. As a consequence, the rate at which an intermediate molecule is processed within the cluster substantially exceeds the rate at which the molecule escapes from the cluster, and so enzyme clustering is able to channel intermediates to the PyrB branch of the pathway.

Next, we compared the predictions of our model directly with the experimental results for the CarB, PyrB, ArgI branch point in *E. coli*. Specifically, we applied our modeling formalism to the metabolic branch point shown in **Figure 4e** (**Supplementary Software 3** and **4**). A substrate S_0 is processed by enzyme E_1 into intermediate S_1, which

Figure 6 Ratio between the efficiency ε_A in the colocalized case and the efficiency ε_A in the delocalized case as a function of the fraction $N_B/(N_A + N_B)$ of enzyme E_B for the two-step metabolic pathway with a branch point for the same geometry and parameters as in **Figure 4**.

can subsequently be processed by enzyme E_A into product P_A or by enzyme E_B into product P_B, or S_1 can decay. We modeled this pathway using reaction-diffusion equations similar to equation (1). The efficiencies, $\varepsilon_A, \varepsilon_B$, of processing by enzymes E_A, E_B, respectively, are the total rates of production of products P_A, P_B divided by the maximal possible rate of production of the substrate S_0.

A natural control parameter in our experiments in *E. coli* the expression level of the CarB-PyrB fusion protein. With this in mind, in we fixed N_A and we varied $N_1 = N_B$ together as if these two enzymes were fused (**Fig. 4f,g**). We computed the efficiency fractions $x_A = \varepsilon_A/(\varepsilon_A + \varepsilon_B), x_B = \varepsilon_B/(\varepsilon_A + \varepsilon_B)$ of the two branches of the pathway as functions of the enzyme fraction $N_B/(N_A + N_B)$. We took E_A to be uniformly distributed in a sphere with the typical radius of an *E. coli* cell $r_A = R = 0.79$ µm, with a number of molecules $N_A = 2,000$, and we assumed no-flux boundary conditions at the boundary of the basin $r = R = r_A$. We then considered two cases: the colocalized case and the delocalized case (**Fig. 4f,g** and **Supplementary Software 3** and **4**).

Moderate expression levels of E_1 and E_B ($N_B/(N_A + N_B) \sim 0.5$) produced considerable flux shunting toward branch B if E_1 and E_B were colocalized, whereas little to no flux shunting occurred if E_1 and E_B were delocalized (**Fig. 4g**). This result agrees with the experimental observation that significant flux shunting occurred within an agglomerate at enzyme concentrations corresponding to moderate induction levels (**Fig. 4b–d**). In particular, for moderate expression levels the efficiency ε_A in the colocalized case was about half of ε_A in the delocalized case (**Fig. 6**). This prediction is in fairly good agreement with the experimental observation that
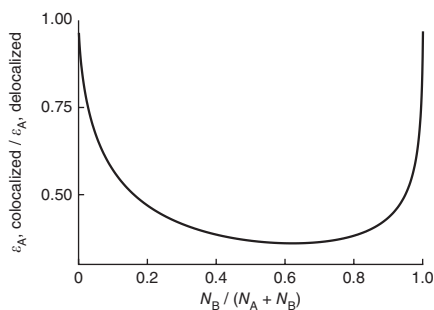
the ratio of arginine biosynthetic flux between the clustered and unclustered strains is about 1:2 (see also **Supplementary Fig. 12** for details).

DISCUSSION

Motivated in part by the recent observation that enzyme scaffolds, presumably bound together into compact agglomerates, can improve metabolic efficiency[8,9,32,33], and in part by *in vivo* observations of enzyme clusters[17,18], we developed a quantitative model to assess the benefits of enzyme clustering. We found that enzyme clustering can provide benefits by colocalizing many copies of the enzyme that processes an intermediate with many copies of the enzyme that produces it. By thus increasing the rate of intermediate processing, enzyme clustering can increase the metabolic efficiency of pathways with unstable intermediates[17].

Our model achieves computational speed by treating both enzymes and substrates as continuous densities with spherical symmetry, allowing us to find globally optimal enzyme distributions. For a simple two-step metabolic pathway with a short-lived intermediate, we found that the optimal distribution of enzymes in a large cell is given by multiple enzyme clusters, and is almost 6 times more efficient than a delocalized (uniform) distribution (**Fig. 2**), and 110 times more efficient than a delocalized distribution for a three-step pathway (**Supplementary Fig. 13**). Moreover, our method allows us to find the optimal spacing between clusters. For our choice of parameters (such as diffusion coefficient, enzyme catalytic rates) in the simple two-step pathway, we found a spacing of 6.5 μm, comparable to the range observed for purinosome spacings in human cells[17]. Notably, a cluster spacing of 6.5 μm implies that multiple clusters would be optimal only in large cells, such as human cells[17]. Still, our analysis shows that for enzymes with high catalytic constants, clustering can also be effective in smaller cells, such as bacteria.

In addition, we provide a simple analytical expression for the metabolic efficiency of a clustered enzyme configuration (**Supplementary Fig. 4**). This expression can be used to obtain the predictions of our model (such as the efficiency of a clustered enzyme configuration, the optimal cluster radius and intercluster spacing) for other metabolic pathways by simply substituting the values of the relevant parameters (such as diffusion coefficient, catalytic constants).

The central prediction of our model is that enzyme clustering within cells can achieve rapid processing of intermediates. From an experimental point of view, measuring the rate of intermediate processing in a linear pathway such as the one shown in **Figure 2a** is not an easy task, generally requiring kinetic measurements that are hard to carry out *in vivo*. In contrast, if we add a second branch to the pathway—for example, if we consider a metabolic branch point as shown in **Figure 4e**—the relative rate of intermediate processing by the two branches can be determined by steady-state measurement of the flux branching ratio. Therefore, an ideal place to experimentally verify our prediction of rapid intermediate processing by an enzyme agglomerate is at a metabolic branch point. We tested this prediction for a fundamental branch point in *E. coli*. At this branch point, carbamoyl phosphate synthetase (CarB) synthesizes carbamoyl phosphate, which can then be driven into pyrimidine biosynthesis by aspartate carbamoyltransferase (PyrB) or toward arginine biosynthesis by ornithine carbamoyltransferase. A simple estimate based on the catalytic constants of the downstream enzymes suggested that enzyme clustering could yield considerable shunting of flux at this metabolic branch point. Indeed, we showed that if CarB and PyrB are coclustered into an agglomerate, flux is strongly shunted toward the PyrB branch.

To compare the predictions of our model with the above experimental results, we applied the model to the specific metabolic branch point studied in the experiments. The model confirms that if many copies of the upstream enzyme CarB and the downstream enzyme PyrB are suitably colocalized into a cluster, the carbamoyl phosphate produced by CarB is processed by PyrB before decaying or diffusing out of the cluster, resulting in substantial shunting of flux to the pyrimidine pathway.

Although previous studies[8] on enzyme scaffolds demonstrated an increase in product titer, leading to speculation that enzyme agglomerates were formed[9], ours is the first demonstration of acceleration of intermediate processing as a direct consequence of enzyme-agglomerate formation, without any specific microscopic arrangement of the enzyme molecules.

Overall, our results provide some general guidance for cluster engineering efforts—for example, highlighting the importance of achieving maximum density but the relative unimportance of internal cluster organization. Moreover, our approach provides a computationally tractable means of targeting *de novo* engineered clusters to the right sizes, stoichiometries and intercluster spacings. These tools will provide for designable control of cluster-regulated metabolic networks to produce economically viable product titers and meet demands for several biotechnology applications, such as therapeutics, drugs and biofuels. Interesting directions for the future will be to determine the mechanisms by which natural enzyme clusters are formed[34].

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

AUTHOR CONTRIBUTIONS
M.C., M.Z.W., Z.G. and N.S.W. conceived the study. M.C. and N.S.W. developed the theory. M.C. performed the analytical and numerical calculations. M.Z.W. and Z.G. designed the experiments. M.Z.W. performed the experiments, assisted by Y.X., P.J., J.D.R. and I.M.C.

1. James, C.L. & Viola, R.E. Production and characterization of bifunctional enzymes. Substrate channeling in the aspartate pathway. *Biochemistry* **41**, 3726–3731 (2002).
2. Dunn, M.F. *et al*. The tryptophan synthase bienzyme complex transfers indole between the alpha and beta sites via a 25–30 ANG long tunnel. *Biochemistry* **29**, 8598–8607 (1990).
3. Huang, X., Holden, H.M. & Raushel, F.M. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu. Rev. Biochem.* **70**, 149–180 (2001).
4. Welch, G.R. On the role of organized multienzyme systems in cellular metabolism: A general synthesis. *Prog. Biophys. Mol. Biol.* **32**, 103–191 (1978).
5. Bernhard, S.A. & Srivastava, D.K. Direct transfer of metabolites via enzyme-enzyme complexes: evidence and physiological significance. *NATO ASI Series* **127**, 143–163 (1987).
6. Srere, P.A. Complexes of sequential metabolic enzymes. *Annu. Rev. Biochem.* **56**, 89–124 (1987).

7. Bauler, P., Huber, G., Leyh, T. & McCammon, J.A. Channeling by proximity: the catalytic advantages of active site colocalization using Brownian dynamics. *J. Phys. Chem. Lett.* **1**, 1332–1335 (2010).
8. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
9. Lee, H., DeLoache, W.C. & Dueber, J.E. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.* **14**, 242–251 (2012).
10. Ginsburg, A. & Stadtman, E.R. Multienzyme systems. *Annu. Rev. Biochem.* **39**, 429–472 (1970).
11. Srere, P.A. & Mosbach, K. Metabolic compartmentation: Symbiotic, organellar, multienzymic, and microenvironmental. *Annu. Rev. Microbiol.* **28**, 61–84 (1974).
12. Smith, G.K., Mueller, W.T., Wasserman, G.F., Taylor, W.D. & Benkovic, S.J. Characterization of the enzyme complex involving the folate-requiring enzymes of *de novo* purine biosynthesis. *Biochemistry* **19**, 4313–4321 (1980).
13. Friedrich, P. *Supramolecular Enzyme Organization* (Pergamon, New York, 1984).
14. Welch, G.R. *Organized Multienzyme Systems: Catalytic Properties* (Academic, New York, 1985).
15. Hrazdina, G. & Jensen, R.A. Spatial organization of enzymes in plant metabolic pathways. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**, 241–267 (1992).
16. Campanella, M.E., Chu, H. & Low, P.S. Assembly and regulation of a glycolytic enzyme complex on the human erythrocyte membrane. *Proc. Natl. Acad. Sci. USA* **102**, 2402–2407 (2005).
17. An, S., Kumar, R., Sheets, E.D. & Benkovic, S.J. Reversible compartmentalization of *de novo* purine biosynthetic complexes in living cells. *Science* **320**, 103–106 (2008).
18. Narayanaswamy, R. *et al.* Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl. Acad. Sci. USA* **106**, 10147–10152 (2009).
19. Deng, Y. *et al.* Mapping protein-protein proximity in the purinosome. *J. Biol. Chem.* **287**, 36201–36207 (2012).
20. French, J.B. *et al.* Hsp70/Hsp90 chaperone machinery is involved in the assembly of the purinosome. *Proc. Natl. Acad. Sci. USA* **110**, 2528–2533 (2013).
21. An, S., Deng, Y., Tomsho, J.W., Kyoung, M. & Benkovic, S.J. Microtubule-assisted mechanism for functional metabolic macromolecular complex formation. *Proc. Natl. Acad. Sci. USA* **107**, 12872–12876 (2010).
22. Zhao, A., *et al.* Transiently transfected purine biosynthetic enzymes form stress bodies. *PLoS ONE* **8**, 2 (2013).
23. Perham, R.N. Self-assembly of biological macromolecules. *Phyl. Trans. R. Soc. Lon.* **272**, 123–136 (1975).
24. Reed, L.J. Multienzyme complexes. *Acc. Chem. Res.* **7**, 40–46 (1973).
25. Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gen knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2** 2006.0008 (2006).
26. Yeates, T.O., Crowley, C.S. & Tanaka, S. Bacterial microcompartment organelles: protein shell structure and evolution. *Annu. Rev. Biophys.* **39**, 185–205 (2010).
27. Johnson, S.G. *The NLopt nonlinear-optimization package,* http://ab-initio.mit.edu/nlopt, (2010).
28. Taniguchi, Y. *et al.* Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
29. Müller-Hill, B. *The Lac Operon: A Short History of a Genetic Paradigm* (Walter de Gruyter, Berlin, 1996).
30. Schomburg, I. *et al.* BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA. *Nucleic Acids Res.* **41**, D764–D772 (2013).
31. García-Pérez, A.I. *et al.* Molecular crowding and viscosity as determinants of translational diffusion of metabolites in subcellular organelles. *Arch. Biochem. Biophys.* **362**, 329–338 (1999).
32. Moon, T.S., Dueber, J.E., Shiue, E. & Prather, K.L. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered *E. coli. Metab. Eng.* **12**, 298–305 (2010).
33. Delebecque, C.J., Silver, P.A. & Lindner, A.B. Designing and using RNA scaffolds to assemble proteins *in vivo. Nat. Protoc.* **7**, 1797–1807 (2012).
34. Brangwynne, C. *et al.* Granules are liquid droplets that localize by controlled dissolution/condensation. *Science* **324**, 1729–1732 (2009).
35. Guillou, F., Liao, M., Garcia-Espana, A. & Lusty, C.J. Mutational analysis of carbamyl phosphate synthetase. Substitution of Glu84 1 leads to loss of functional coupling between the two catalytic domains of the synthetase subunit. *Biochemistry* **31**, 1656–1664 (1992).
36. Kuo, L.C., Miller, A.W., Lee, S. & Kozuma, C. Site-directed mutagenesis of *Escherichia coli* ornithine transcarbamoylase: role of arginine-57 in substrate binding and catalysis. *Biochemistry* **27**, 8823–8832 (1988).
37. Shepherdson, M. & Pardee, A.B. Production and crystallization of aspartate transcarbamylase. *J. Biol. Chem.* **235**, 3233–3237 (1960).

ONLINE METHODS

Bacterial strain and plasmid construction. The bacterial strains, plasmids, and oligonucleotide primers used in this study are listed in **Supplementary Table 2**. All gene deletions were generated by P1vir-transducing the kanamycin cassette insertions from the Keio Collection[25] into WT NCM 3722 or other strain of interest and selecting for kanamycin resistance. Kanamycin insertions were then removed through the transformation of the heat-inducible flippase-expressing plasmid, pCP20 (ref. 38). Transformed strains were incubated overnight and patched on to LB, LB + Kan, and LB + Carb to assure loss of plasmid and gene insertion. Synthetic linear DNA fragments were used to carry out chromosomal integrations via the lambda-red recombineering method, which has been described[39]. Deletions and integrations were verified for their resulting genomic scar and modification, respectively, by PCR, followed by sequencing when deemed necessary. All plasmids and synthetic linear DNA fragments were constructed by incubating purified PCR products, flanked by at least 20 bp overlapping ends, together with 1X Gibson Assembly Master Mix (NEB Labs) at 50 °C for 1 h[40]. Isolated constructs were then verified by sequencing.

Bacterial growth conditions. *E. coli* strain NCM3722 and its derivatives were grown at 37 °C in Luria Broth, Gutnick Minimal Media (0.4% glucose (w/v))[41], or Gutnick Minimal Media supplemented with either 0.2 mM uracil (Sigma) or 0.5 mM arginine (Sigma). Antibiotics were added at the following concentrations (to liquid media/to agarose plates): carbenicillin (Omega Scientific Inc.) (50 g/ml/100 g/ml), kanamaycin (Gibco by Life Technologies) (30 g/ml/50 g/ml), chloramphenicol (Fisher) (50 g/ml/100 g/ml).

Microscopy and image analysis. Images were taken with a Nikon 90i upright microscope equipped with a Nikon Plan Apo 100X/1.4 phase-contrast objective. Images were collected with a Rolera XR cooled CCD camera and initially processed by NIS-Elements Advanced Research software. Images were further analyzed for inclusion-body content with either custom Matlab code or ImageJ. Samples were spotted onto 1% agarose (Invitrogen) pads, resting on glass slides, made with the appropriate medium. Coverslips were sealed with valap (1:1:1, lanolin:paraffin:petroleum jelly).

Growth assays. Cell growth was assayed in flat-bottom Costar 96-well polystyrene plates using a Biotek Synergy HT. Wells were filled with 150 µl of the appropriate liquid media inoculated from overnight cultures back-diluted 1:145, and covered with 50 µl mineral oil (Sigma). Plates were incubated at 37 °C with continuous shaking and the optical density (OD) at 600 nm was read every 20 min.

Single-cell inclusion body lysis assay. Overnight cultures of inclusion-body-containing strains were diluted 1:50 into Gutnick minimal media and grown at 37 °C for 2 h. Cells were then spotted onto 1% agarose pads containing minimal media, 0.1% Triton X-100 (Sigma), 10 mM EDTA (Sigma), and 1 mg/ml chicken egg lysozyme (Sigma).

Inclusion-body isolation. Cell-free inclusion bodies were collected by following a modified isolation protocol detailed elsewhere[42]. Briefly, mid-log phase cells were pelleted and resuspended in ice-cold lysis buffer made from 50 mM TrisHCl (pH 8.8) (Fisher), 100 mM NaCl (Sigma), 1.5 mM EDTA (Sigma) and distilled water. This suspension was flash frozen to −80 °C, thawed at room temperature, and combined with chicken egg lysozyme (to 1 mg/ml) and PMSF (to 200 µM) (Sigma). After incubating at 37 °C for 1 h Triton-100× (to 1 µl/ml), NP-40 (to 0.1 µl/ml) (as IGEPAL from Sigma), DNAse (to 0.3 g/ml) (Sigma), and MgSO$_4$ (to 0.15 mM) (Fisher) were added; this suspension was incubated for 1 h at 37 °C. Finally, the lysate was pelleted at 15,000*g* for 15 min and washed twice in lysis buffer before being resuspended in distilled water + 5% glycerol and stored in aliquots at −80 °C.

Inclusion-body proteomics. Cell-free inclusion bodies were reduced in 1× NuPage Sample Buffer (Invitrogen), incubated at 70 °C for 10 min, then alkylated with iodoacetamide (100 mM) at room temperature for 30 min before being heated to 95 °C for 2 min. Soluble protein was resolved by 1 dimensional gel-electrophoresis (4–12% Bis-Tris NuPAGE gel) and digested in-gel with trypsin, as previously described[43]. Digested peptides were concentrated by vacuum centrifugation, desalted using StageTips[44] constructed using Empore C$_{18}$ extraction discs (3M Analytical Biotechnologies). Desalted peptides were analyzed by nanoliquid chromatography–tandem mass spectrometry using a Dionex Ultimate 3000 nRSLC coupled to an LTQ-Orbitrap XL mass spectrometer (ThermoFisher Scientific, San Jose, CA), as previously described[9]. MS/MS spectra were extracted by Proteome Discoverer and analyzed using SEQUEST by searching *E. coli* and contaminant protein databases. Probabilistic calculation of false-positive rates (<1% FDR) was performed by Scaffold/X! Tandem (Proteome Software) using the PeptideProphet and ProteinProphet algorithms[45].

Metabolite measurement. The metabolome of batch culture *E. coli* was quantitated by liquid chromatography–mass spectrometry. Briefly, saturated overnight cultures were diluted 1:50 and grown in liquid media in a shaking flask to OD$_{600}$ of ~0.3. A portion of the cells (3 ml) were filtered onto a 50-mm nylon membrane filter, which was immediately transferred into −20 °C extraction solvent (40:40:20 acetonitrile/methanol/water). Cell extracts were analyzed by reversed phase ion-pairing liquid chromatography (LC) coupled by electrospray ionization (ESI) (negative mode) to a high-resolution, high-accuracy mass spectrometer (Exactive; Thermo Fisher Scientific) operated in full scan mode at 1 s scan time, 10^5 resolution, with compound identities verified by mass and retention time match to authenticated standard[46]. Quantation of low abundance metabolites such as arginine and citrulline was also confirmed by carbobenzyloxy (CBZ) derivitization followed by LC-MS analysis. Briefly, 200 µl of cell extract was mixed with 5 µl of triethylamine (Sigma) and 1 µl benzylchloroformate (Sigma). Resulting samples were analyzed by reversed phase ion-pairing liquid chromatography (LC) coupled to a Thermo TSQ Quantum triple quadrupole mass spectrometer operating in multiple reaction monitoring mode with compound identities verified by mass spectrometry and retention time match to authenticated standards. Day-to-day variation of absolute spectral counts prevents useful biological replicate comparison; therefore all data presented were from the same sequence with blanks run between each set of four samples. Three technical replicates were taken to measure reproducibility of the extraction procedures and quantifications via the mass spectrometer. Sample replicates typically varied between 13% and maximally by 45% as calculated by the coefficient of variation. Replicates were averaged before fold changes were calculated.

Colorometric assay of aspartate transcarbamylase activity. Standard aspartate transcarbamylase assay conditions have been detailed[47]. All reactions were carried out at pH 7 at 37 °C for 1 h in 1 ml reaction volumes. One activity unit, defined as 10 µl of isolated inclusion body material, was assayed per reaction. Briefly, TrisHCl (100 mM pH 7) (Fisher), L-aspartate (100 mM pH 7) (Sigma), ATP (2 mM) (Sigma), lithium carbamoyl phosphate (10 mM, prepared fresh) (Sigma) were added to distilled water and equilibrated to 37 °C. To begin the reaction 100 µl of water containing 1 enzyme activity unit was added to the reaction volume. The reaction was halted by the addition of 2 ml of 5% (w/v) trichloroacetic acid solution (Sigma). Color development was carried out as detailed by Prescott and Jones[48]. Developed samples were assayed for the production of carbamoyl-aspartate by measuring the absorbance at 466 nm.

Arginine pseudoauxotrophy results from coclustering of CarB and PyrB. To establish that the arginine pseudoauxotrophy was not simply a consequence of PyrB overexpression, but rather required CarB and PyrB being coclustered, we constructed a synthetic operon where CarB and PyrB were similarly co-overexpressed as two separate proteins. Without fusing CarB and PyrB, their co-overexpression resulted in wild-type growth with no detectable auxotrophy at maximal induction (**Supplementary Fig. 8a**), in agreement with the model prediction that no significant flux shunting occurs if CarB and PyrB are moderately overexpressed and delocalized (**Fig. 4g**).

To establish that the arginine pseudoauxotrophy did not result from a specific effect caused by the linker region, we sampled serine-glycine repeat linkers from 3 to 25 amino acids in length as well as the TEV protease linker region and found that the phenotype did not depend on the composition of the chain used to tether CarB to PyrB.

To establish that the arginine pseudoauxotrophy required functional PyrB, we constructed a CarB fusion to an enzymatically dead form of PyrB (CarB-PyrB(R54A)) and found that it did not display arginine pseudoauxotrophy (**Supplementary Fig. 8b**). The arginine pseudoauxotrophy thus requires not only colocalization of CarB with PyrB, but also functional PyrB, and cannot be attributed to a dominant-negative effect of the CarB-PyrB fusion on the activity of ArgI.

One final concern might be that the CarB-PyrB fusion somehow "hyperactivates" PyrB, reducing flux through ArgI by processing an increased fraction of the cellular pool of carbamoyl phosphate. To eliminate this possibility, we introduced the CarB-PyrB fusion into a strain expressing functional CarB-msfGFP from the native *carB* locus. CarB-msfGFP did not become incorporated into the phase-bright foci (**Supplementary Fig. 9a**), indicating that it synthesized a delocalized, cellular pool of carbamoyl phosphate. This unclustered CarB-msfGFP eliminated the arginine auxotrophy of CarB-PyrB overexpression (**Supplementary Fig. 9b**). Thus, auxotrophy must involve a local flux shunting within a CarB-PyrB cluster, rather than an overall effect of the fusion on cellular metabolite pools. Our modeling results taken together with these experiments led us to hypothesize that the observed pseudoauxotrophy for arginine results from flux shunting to PyrB away from ArgI due to coclustering of CarB and PyrB.

To support the coclustering-mediated shunting hypothesis, we performed a metabolomic analysis of the ΔcarB ΔpyrB strain expressing the unrepressed CarB-PyrB fusion protein. Inspection of the relative metabolite concentrations in the unrepressed fusion-containing cells revealed large alterations of the metabolite pools only within the metabolic network of the carbamoyl phosphate branch point, suggesting the pseudoauxotrophy resulted from local modifications to the network (**Supplementary Fig. 10**). In this regard, it is important to point out that although CTP is an allosteric inhibitor of PyrB[49] we find that the CTP pool size in the fusion strain does not change substantially relative to wild type (**Supplementary Fig. 10**). This observation is in agreement with the findings that CTP levels are tightly regulated by multiple means, including a recently described directed overflow mechanism[50].

Our hypothesis of flux shunting via coclustering requires that CarB-PyrB form large enzyme clusters. Indeed, we observed a correlation between the onset of metabolic shunting and the accumulation of phase-bright structures in cells expressing CarB-PyrB (**Fig. 5**). To establish that the agglomerates observed above a certain critical concentration are composed of CarB-PyrB we used two approaches. First, we made a three-part CarB-PyrB-msfGFP fusion and overexpressed it as was done for CarB-PyrB. We again observed the concentration-dependent development of phase-bright structures (**Supplementary Fig. 11a**). These structures also displayed extremely high levels of msfGFP fluorescence (**Supplementary Fig. 11a**), indicating that the compound protein was present in these structures. As a second assay, we used real-time imaging to determine that the phase-bright structures remained intact upon cell lysis (**Supplementary Video 1**), which enabled us to purify these structures from cell lysates by ultracentrifugation. Analysis of the protein content of these structures by mass spectrometry revealed that they were overwhelmingly composed of the CarB-PyrB fusion (**Supplementary Fig. 11b**).

Our hypothesis that the CarB-PyrB agglomerates are responsible for metabolic shunting also depends on the agglomerates being enzymatically active. Although insoluble protein clusters have generally been assumed to be dominated by misfolded protein[51], there are several reports of enzymatically active clusters[52]. Consistently, we note that our cell-free msfGFP-containing inclusion bodies are highly fluorescent (containing 91 ± 7% of the total cellular fluorescence). Because msfGFP fluorescence requires its proper folding, the CarB-PyrB structures must contain a substantial population of properly folded protein (**Supplementary Fig. 11a**). Furthermore, we subjected purified inclusion bodies composed of the CarB-PyrB fusion to a well-established *in vitro* aspartate carbamoyltransferase assay[48], and we found that they readily produced carbamoyl aspartate when supplied with carbamoyl phosphate and aspartate (**Supplementary Fig. 11c**). This result confirmed that our purified inclusion bodies indeed contained active enzyme and support our clustering-mediated metabolic-channeling model.

Although we were unable to achieve metabolic shunting of carbamoyl phosphate toward the arginine side of the branch point using our simple translational fusion method, we note that the lack of such a phenotype could be caused by partial inactivation of the fused form of ornithine transcarbamylase or a relative decrease in the density of active sites within a cluster. This negative result does not alter the fact that our experiments on the pyrimidine side of the carbamoyl phosphate branch in *E. coli* qualitatively confirm a central prediction of our model.

Model parameters. Except where otherwise noted, the parameters in the reaction-diffusion equations are chosen as follows. The diffusion coefficient is taken to be $D = 10^2 \, \mu m^2 / s$ (ref. 53) for all metabolites. The rate $\alpha_0 = 10^{-1} / s$ at which $c_0(r,t)$ relaxes to $c_0^\star$ is estimated from the typical timescale of substrate turnover in bacteria. The enzyme k_{cat}/K_M values $k_1 = k_2 = 10^6 \, liter / s / mol$ (ref. 54) are chosen to be within the measured range for metabolic enzymes. We assume the single-enzyme size to be typical of a globular protein[55], that is, radius ≈ 2 nm. Taking into account that active enzymes are solvated by water with a roughly 50% volume fraction, we obtain a maximum enzyme density $n_{max} = 0.5 \times 1/((4/3)\pi(2 \, nm)^3) \approx 25 \, mM$. The intermediate decay rate was chosen to be $\beta = 10/s$ (**Supplementary Fig. 14** for details regarding the optimization for different values of β). The total catalytic activity is $\kappa_{cat} = 1.66 / s$, and corresponds to 1,000 enzymes per μm^3 for our choice of k_1, k_2. Finally, the homeostatic value $c_0^\star$ of the concentration of the substrate S_0 is arbitrary because rescaling $c_0^\star$ by a constant factor does not change the optimal enzyme densities or the basin radius or the efficiency.

38. Cherepanov, P.P. & Wackernagel, W. Gene disruption in *Escherichia coli*: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. *Gene* **158**, 9–14 (1995).
39. Datsenko, K.A. & Wanner, B.L. One-step inactivation of chromasomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).
40. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
41. Gutnick, D., Calvo, J.M., Klopotowski, T. & Ames, B.N. Compounds which serve as the sole source of carbon or nitrogen for Salmonella typhimurium LT-2. *J. Bacteriol.* **100**, 215–219 (1969).
42. Rodríguez-Carmona, E., Cano-Garrido, O., Seras-Franzoso, J., Villaverde, A. & García-Fruitós, E. Isolation of cell-free bacterial inclusion bodies. *Microb. Cell Fact.* **9**, 71 (2010).
43. Kramer, T., Greco, T.M., Enquist, L.W. & Cristea, I.M. Proteomic characterization of pseudorabies virus extracellular virions. *J. Virol.* **85**, 6427–6441 (2011).
44. Rappsilber, J., Mann, M. & Ishihama, Y. Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* **2**, 1896–1906 (2007).
45. Joshi, P. *et al.* The functional interaction landscape of the human histone deacetylase family. *Mol. Sys. Biol.* **9**, 672 (2013).
46. Lu, W. *et al.* Metabolomic analysis via reversed-phase ion-pairing liquid chromotography coupled to a stand alone orbitrap mass spectrometer. *Anal. Chem.* **82**, 3212–3221 (2010).
47. Shepherdson, M. & Pardee, A. Product and crystallization of aspartate transcarbamylase. *J. Biol. Chem.* **235**, 3233–3237 (1960).
48. Prescott, L.M. & Jones, M.E. Modified methods for the determination of carbamoyl aspartate. *Anal. Biochem.* **32**, 408–419 (1969).
49. Gerhart, J.C. & Pardee, A.B. The enzymology of control by feedback inhibition. *J. Biol. Chem.* **237**, 891–896 (1962).
50. Reaves, M.L., Young, B.D., Hosios, A.M., Xu, Y.-F. & Rabinowitz, J. Pyrimidine homeostasis is accomplished by directed overflow metabolism. *Nature* **500**, 237–241 (2013).
51. Fahnert, B., Lilie, H. & Neubauer, P. Inclusion bodies: formation and utilization. *Adv. Biochem. Eng. Biotechnol.* **89**, 93–142 (2004).
52. García-Fruitós, E. *et al.* Aggregation as bacterial inclusion bodies does not imply inactivation of enzymes and fluorescent proteins. *Microb. Cell Fact.* **4**, 27 (2005).
53. García-Pérez, A.I. *et al.* Molecular crowding and viscosity as determinants of translational diffusion of metabolites in subcellular organelles. *Arch. Biochem. Biophys.* **362**, 329–338 (1999).
54. Fersht, A. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W. H. Freeman and Company, 1999).
55. Erickson, H.P. 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).