Kinetics of Spatially Organized Biochemical Reactions

Optimal Strategies and Underlying Physical Principles

Florian Hinzpeter

15. Mai 2018

Kinetics of Spatially Organized Biochemical Reactions

Optimal Strategies and Underlying Physical Principles

Florian Hinzpeter

Doktorarbeit
an der Fakultät für Physik
der Ludwig–Maximilians–Universität
München

vorgelegt von
Florian Hinzpeter
aus Ulm

München, den 15. Mai 2018

Erstgutachter: Prof. Dr. Ulrich Gerland
Zweitgutachter: Prof. Dr. Erwin Frey
Tag der mündlichen Prüfung: 12. Juli 2018

Zusammenfassung

Der Stoffwechsel aller lebenden Organismen beruht auf der katalytischen Aktivität von bestimmten Proteinen – den Enzymen. Enzyme sind sehr spezifisch hinsichtlich der Stoffe, die sie umsetzen, sowie der biochemischen Reaktion, die sie beschleunigen. Um eine komplexe Umwandlungen des Stoffwechsels zu erreichen organisieren sich mehrere Enzyme in Reaktionssequenzen, den so genannten Stoffwechselwegen. In diesen wird jeder Reaktionszwischenschritt von einem bestimmten Enzym katalysiert. Zusätzlich zu der konzeptionellen Organisation von enzymatischen Reaktionen in Stoffwechselwege sind kooperierende Enzyme häufig räumlich in Cluster, Komplexe und Kompartimente organisiert. Es wird angenommen, dass durch die Bildung dieser makromolekularen Strukturen die Verarbeitungseffizienz von Zwischenprodukten verbessert wird. Dies ist für Zellen vor allem dann essentiell, wenn die Zwischenprodukte toxisch, flüchtig oder instabil sind. Die Mechanismen, die diese verbesserte Umsetzung von Zwischenprodukten ermöglichen, sind vielfältig und beinhalten Prozesse wie beispielsweise das *metabolic channeling*. Hierbei werden durch strukturelle Eigenschaften, wie intramolekulare Tunnel, Schwingarme und Ladungsverteilungen, die Zwischenprodukte direkt von einem Enzym zum nächsten transportiert. In vielen Multienzym-Clustern und -Kompartimenten sind diese strukturellen Merkmale jedoch nicht vorhanden und der Transport der Zwischenprodukte von einem Enzym zum nächsten geschieht lediglich durch Diffusion. In diesen Fällen ist unklar wie die räumliche Nähe von Enzymen die Effizienz der sequentiellen Reaktionen steigern kann. Insbesondere die Designprinzipien, die hinter solchen Multienzym-Anordnungen stehen, sind bisher wenig verstanden.

In der vorliegenden Arbeit werden physikalische Prinzipien erforscht, welche der optimalen räumlichen Organisation von Enzymen zugrunde liegen. Mit Hilfe unterschiedlicher Reaktions-Diffusions-Modelle werden effiziente Kompartimentierung- und Cluster-Strategien kooperierender Enzyme berechnet und analysiert, wie sich diese aus dem Zusammenspiel von enzymatischer Reaktion und Diffusion der Metabolite ergeben. Die Entschlüsselung dieser Strategien und der dabei zugrunde liegenden physikalischen Prinzipien trägt zu einem besseren Verständnis solcher Multienzym-Strukturen bei. Zugleich wird dadurch ermöglicht die Anordnung von Enzymen in synthetischen Bioreaktoren optimal zu koordinieren, um eine effizientere Synthese von beispielsweise Medikamenten oder Biokraftstoffen zu erzielen.

Häufig sind die Enzyme bestimmter Stoffwechselwege in semipermeablen Membranen oder Proteinhüllen eingeschlossen. Durch diese Kompartimentierung von Stoffwechselreaktionen wird das Entweichen von flüchtigen oder instabilen Zwischenprodukten verhindert, was zu einer Steigerung der Effizienz des Stoffwechselweges führt. Obwohl die Kompartimentierung von Enzymen ein allgegenwertiges Phänomen in lebenden Organismen ist und große Anstrengungen unternommen wurden, um synthetische Kompartimente herzustellen, sind die optimalen Designprinzipien bislang kaum erforscht. In Kapitel 3 wird untersucht wie die Enzymzusammensetzung und Eigenschaften der Kompartimente gewählt werden müssen, um die synergistische Aktivität der eingeschlossenen Enzyme optimal zu nutzen. Mithilfe eines quantitativen Modells konnte gezeigt werden, dass die optimale kollektive Produktivität der

Enzyme nur erreicht werden kann, wenn die Kompartimente eine bestimmte kritische Größe überschreiten und die Enzymkonzentrationen gemäß eines Potenzgesetzes in der Kompartimentgröße gewählt wird.
Zusätzlich zu der Kompartimentierung von Stoffwechselwegen akkumulieren Enzyme häufig zu großen Clustern, die nicht durch eine Membran oder Proteinhülle begrenzt sind. In den Kaptieln 4 und 5 wird analysiert, unter welchen Bedingungen die Bildung dieser Multienzym-Cluster die Produktivität von Stoffwechselwegen verbessern kann und wie die Enzyme angeordnet werden müssen um die Produktivität zu optimieren. Mit Hilfe eines Minimalmodells, in dem das erste Enzym eines zweistufigen Stoffwechselweges als lokalisierte Quelle von Zwischenprodukten beschrieben wird, wird gezeigt, dass die Kolokalisierung der zweiten Enzyme mit dieser Quelle die Umsetzung von Zwischenprodukten erheblich steigern kann. Diese verbesserte Umsetzung ist vergelichbar mit der, die durch direkte Channeling-Mechanismen erzielt wird. Die optimale Verteilung der Enzyme hängt von dem Verhältnis der katalytischen Effizienz des zweiten Enzyms zur Diffusivität der Zwischenprodukte ab. Wenn dieses Verhältnis klein ist, ist es am effizientesten alle Enzyme zu clustern. Wenn jedoch dieses Verhältnis einen kritischen Wert überschreitet, wird es vorteilhaft einige Enzyme entfernter vom Cluster zu positionieren. Dieser Übergang ist ein allgemeines Phänomen, welches bei verschiedenen Reaktionskinetiken, räumliche Dimensionen und Verlustmechanismen der Zwischenprodukte auftritt.
In dem untersuchten Modell wurden jedoch einige wichtige Details, wie die explizite erste enzymatische Reaktion sowie die räumliche Ausdehnung der Enzyme, vernachlässigt. Die Berücksichtigung dieser zusätzlichen Details führt zu zwei fundamentalen Tradeoffs, die die Strategien optimaler Enzymanordnungen bestimmen. Der erste Tradeoff zeigt sich unabhängig von der räumlichen Ausdehnung der Enzyme und ergibt sich allein aus der Reaktions-Diffusions-Dynamik der zweistufigen Reaktion. Innerhalb eines Multienzym-Clusters führt die erhöhte Konzentration der Enzyme zu einer effektiveren Verarbeitung von Zwischenprodukten. Jedoch verringert die erste Reaktion die Substratkonzentration innerhalb des Clusters, wodurch die Umsetzungsrate zum Zwischenprodukt herabgesetzt wird. Der zweite Tradeoff basiert auf der räumlichen Ausdehnung der Enzyme und den sich dadurch ergebenden sterischen Effekten, die vor allem in Enzym-Clustern relevant werden. Durch eine enge Anordnungen von Enzymen werden die Zwischenprodukte eingeschlossen, sodass ihr Entweichen verhindert wird und sie folglich effizienter umgesetzt werden können. Auf die gleiche Weise werden jedoch die Substrate daran gehindert die Enzyme zu erreichen, wodurch die Produktivität der ersten Enzymreaktion herabsetzt wird. Das Zusammenspiel dieser Tradeoffs führt zu bemerkenswerten Enzymanordnungen von Modell-Multienzym-Komplexen. Diese weisen auffallende Änlichkeiten mit den Elektronenanordnungen des berühmten Thomson-Problems der klassischen Elektrostatik auf.
Im letzten Kapitel wird untersucht, wie Zellen durch die Bildung unterschiedlicher Enzym-Cluster den Stoffwechsel an einem Verzweigungspunkt regulieren können. Die Bildung von Clustern konsekutiver Enzyme erhöht die Rate, mit der Zwischenprodukte umgesetzt werden. Hierdurch kann das Verzweigungsverhältnis in Richtung der kolokalisierten Enzyme verschoben werden. Im Gegensatz dazu ermöglicht die Sequestrierung eines der Enzyme des Verzweigungspunktes die Umsetzungsrate herabzusetzen. Dies hat wiederum eine Verschiebung des Verzweigungsverhältnisses zur Folge. Die abschließende Analyse ergab, dass beide Strategien für 75 Prozent der Enzyme des Stoffwechsels relevante Regulierungsmechanismen darstellen. Die Ergebnisse dieser Arbeit verdeutlichen, dass die räumliche Anordnung von Enzymen einen wesentlichen Einfluss auf den Stoffwechsel lebender Organismen hat.

Abstract

The metabolism of all living organism relies on the catalytic action of enzymes. To achieve complex biochemical transformations, sets of enzymes work together in so-called metabolic pathways, where each reaction step is catalyzed by a specific enzyme. Additionally to this conceptual organization of enzymatic reactions into pathways, collaborating enzymes are in many cases spatially organized into clusters, complexes, and compartments. It is believed that these macromolecular structures promote the processing of pathway intermediates, which is vital for cells when the intermediates are toxic, volatile, or unstable. The mechanisms that enable this enhanced processing of intermediates are diverse involving "metabolic channeling", which relies on structural features like intramolecular tunnels, swinging-arms, and charge distributions that direct the transfer of intermediates between active sites. Yet in many multi-enzyme clusters and compartments these structural features are not present and the transport of intermediates is mediated by diffusion. In these cases it is unclear how the spatial proximity of enzymes can enhance reaction fluxes. In particular, the design principles behind such multi-enzyme assemblies are poorly understood.

In this work, we seek to elucidate the physical principles that determine the optimal strategies of spatial enzyme organization. Using various reaction-diffusion models, we identify optimal compartmentalization and clustering strategies of collaborating enzymes and analyze how these arise from the coupled dynamics of enzymatic reactions and metabolite diffusion. Unraveling these strategies and the underlying physical principles will not only help to better understand natural multi-enzyme arrangements but will also allow synthetic biologists to optimally design enzymatic reactors for the efficient production of high-value products, pharmaceuticals, or biofuels.

A strategy that is frequently used by cells is the compartmentalization of cooperating enzymes into intracellular bodies delimited by membranes or protein shells. Confining reaction pathways within compartments allows cells to efficiently process elusive or unstable intermediate products. Although metabolic compartmentalization is an ubiquitous phenomenon and much effort has been spent in synthetically constructing these compartments, there remain many open questions regarding optimal design principles. In chapter 3 we quantitatively study how the compartment size and enzyme composition of an enzymatic compartment should be chosen in order to optimally exploit the synergistic activity of two consecutive enzymes. We find that to achieve optimal collective productivity of the enzymes, the compartments have to be larger than a certain critical size and the enzyme densities have to be adjusted according to a power-law scaling in the compartment size.

In the chapters 4 and 5 we focus on enzyme clustering and optimal enzyme arrangements in the absence of a delimiting shell or membrane. We show that enzyme coclustering can significantly improve the processing of intermediate molecules without the need for structural features that transfer intermediates between enzymes. Using simple reaction-diffusion models where the upstream enzymes are treated as a localized source of intermediates, we demonstrate that the optimal arrangement of downstream enzymes exhibits a generic transition

from a cluster to a more extended distribution as the overall enzymatic activity is increased. This transition is a general phenomenon, which occurs for different reaction kinetics, spatial dimensions, and loss mechanisms of intermediate molecules. We next extend this minimal model to also account for the spatial extension of enzymes and the explicit upstream reaction. These additional details give rise to two fundamental trade-offs, the first between efficient intermediate transfer and depletion of substrate and the second between steric confinement of intermediates and accessibility of enzymes to substrate. We characterize the optimal design principles for the arrangement of sequential enzymes that emerge from the interplay of these trade-offs. Notably, the question of optimal enzyme organization in a multi-enzyme complex is similar to the famous Thomson problem of electrostatics.
In the last chapter (Chap. 6), we ask how enzyme clustering can regulate the fluxes at a metabolic branch point. In addition to multi-enzyme coclustering, which favors the processing of intermediates towards the product of the clustered enzymes, we propose enzyme sequestration as a mechanism to downregulate the flux through the sequestered enzymatic reaction. Our analysis reveals that coclustering and sequestration represent two viable regulation strategies for more than 75% of metabolic enzymes. Collectively, the results of this thesis demonstrate that the spatial organization of enzymes can haven a strong impact on the metabolism of living organisms.

Contents

Zusammenfassung v

Abstract vii

1 Enzyme kinetics and metabolic modeling 1
- 1.1 Introduction to enzyme kinetics . . . 1
  - 1.1.1 The Michaelis-Menten equation . . . 1
  - 1.1.2 The reversible Michaelis-Menten equation . . . 3
  - 1.1.3 Enzyme inhibition . . . 4
  - 1.1.4 Elasticities . . . 5
  - 1.1.5 Impact of diffusion on enzyme kinetics . . . 5
- 1.2 Models of multi-enzyme kinetics . . . 7
  - 1.2.1 Metabolic control analysis . . . 7
  - 1.2.2 Flux Balance Analysis . . . 9

2 Metabolic channeling and spatial enzyme organization 13
- 2.1 The benefits of metabolic channeling . . . 13
- 2.2 Metabolic channeling mediated by structural features . . . 15
- 2.3 Enzyme clustering and compartmentalization . . . 16
  - 2.3.1 Membrane delimited organelles and microcompartments . . . 16
  - 2.3.2 Non-membrane bound enzyme clusters and filaments . . . 19
- 2.4 Synthetic approaches for engineering spatial enzyme organization . . . 22
  - 2.4.1 Scaffold based strategies . . . 22
  - 2.4.2 Direct enzyme conjugation . . . 25
  - 2.4.3 Compartmentalization . . . 25

3 Optimal compartmentalization strategies for metabolic compartments 29
- 3.1 Introduction . . . 29
- 3.2 Model . . . 31
  - 3.2.1 Productivity . . . 32
  - 3.2.2 Reaction-diffusion model . . . 32
- 3.3 Results . . . 33
  - 3.3.1 Optimal compartmentalization strategies for α-carboxysome parameters 33
  - 3.3.2 Well-mixed approximation . . . 36
  - 3.3.3 Breakdown of the well-mixed approximation . . . 39
  - 3.3.4 Compartmentalization strategies for highly permeable shells . . . 39
  - 3.3.5 Michaelis-Menten kinetics . . . 42
- 3.4 Discussion . . . 43

4 Enzyme clustering and optimal enzyme arrangements 47
4.1 Introduction . . . 47
4.2 Model . . . 48
4.2.1 Non-dimensionalization of the system . . . 49
4.3 Results . . . 50
4.3.1 Linear reaction kinetics . . . 50
4.3.2 Non-linear reaction kinetics . . . 58
4.4 Discussion . . . 60

5 Enhancing biocatalytic fluxes via spatial organization 63
5.1 Introduction . . . 63
5.2 Model . . . 65
5.3 Results . . . 67
5.3.1 Randomly positioned biocatalysts and correlations between arrangements and reaction fluxes . . . 67
5.3.2 Comparison of different localization strategies . . . 69
5.3.3 Optimal complex configurations . . . 73
5.4 Discussion . . . 75

6 Flux regulation via enzyme coclustering and sequestration 79
6.1 Introduction . . . 79
6.2 Model . . . 80
6.3 Results . . . 82
6.3.1 Semi well-mixed approximation for enzyme sequestration and coclustering 82
6.3.2 Regulation of the branching fraction by sequestration and coclustering 84
6.4 Discussion . . . 85

Supplementary Materials 89

A Optimal compartmentalization strategies for metabolic compartments 91
A.1 Carboxysome parameters . . . 91
A.2 Derivation of productivity for uniformly distributed enzymes . . . 92
A.3 Derivation of productivity for enzyme distribution observed in carboxysomes 93
A.4 Optimization of intra-compartment enzyme distribution . . . 94
A.5 Optimal enzyme distributions for α-carboxysome parameters . . . 95
A.6 Optimal compartmentalization strategies with intermediate decay . . . 97
A.7 Corrections to the well-mixed approximation . . . 99
A.8 High-permeability limit, $p_s, p_i \to \infty$ . . . 100
A.9 Michaelis-Menten kinetics . . . 101

B Enzyme clustering and optimal enzyme arrangements 103
B.1 Equivalence of the expressions for the reaction efficiency . . . 103
B.2 One-dimensional system with unstable intermediates . . . 104
B.2.1 Clustered enzyme profile . . . 104
B.2.2 Uniform enzyme profile . . . 105
B.3 Enzyme exposure probability distribution in one dimension . . . 105

B.4 Two- and three-dimensional systems 106
B.4.1 Clustered enzyme profile 106
B.4.2 Uniform enzyme profile 107
B.5 Enzyme exposure probability distribution in 2D and 3D 107

C Enhancing biocatalytic fluxes via spatial organization 109
C.1 Ensembles of biocatalyst arrangements 109
C.2 Optimization of the complex arrangement 109
C.3 Permeable biocatalysts 110

D Flux regulation via enzyme coclustering and sequestration 113
D.1 Supplementary figures 113
D.2 Colocalization of consecutive enzymes into pairs 114

Bibliography 117

Acknowledgements 137

List of Figures

1.1 Michaelis-Menten kinetics . . . 1
1.2 Illustration of Flux Balance Analysis . . . 10

2.1 Schematic illustration of channeling mechanisms and spatial enzyme organization 15
2.2 Natural examples for metabolic channeling and spatial enzyme organization . 17
2.3 Synthetic approaches for controlling the arrangement of enzymes . . . 23

3.1 Illustration of the compartmentalization strategies for a two-step pathway . . 31
3.2 Optimal compartmentalization strategies for α-carboxysome parameters . . . 34
3.3 Results of the well-mixed approximation . . . 38
3.4 Optimal compartmentalization strategies outside the well-mixed regime . . . 40
3.5 Optimal compartmentalization strategies with Michaelis-Menten reaction kinetics 43
3.6 Optimal compartmentalization strategies for α- and β-carboxysomes . . . 44

4.1 Illustration of the model used to study enzyme coclustering and optimal enzyme arrangements . . . 49
4.2 Schematic illustration of the concept of integrated "enzyme exposure" . . . 51
4.3 Decomposition of the efficiency into the enzyme exposure distribution and reaction probability . . . 52
4.4 Optimal E_2 profiles with unstable intermediates . . . 54
4.5 Comparison of the efficiency in higher dimensions . . . 56
4.6 Optimal enzyme profiles in higher dimensions . . . 57
4.7 Optimal enzyme profiles for nonlinear reaction kinetics . . . 58
4.8 Full model with decay and enzyme saturation . . . 59

5.1 Schematic visualization of the reaction-diffusion model with discrete biocatalysts 65
5.2 Reaction flux for delocalized biocatalysts in two dimensions . . . 68
5.3 Comparison of different localization strategies . . . 70
5.4 Illustration of shielding and confinement effects . . . 72
5.5 Optimal biocatalyst complex arrangements . . . 74

6.1 Illustration of a metabolic branch point and clustering strategies . . . 81
6.2 Branching fraction of the clustering strategies . . . 82
6.3 Capability of the spatial strategies to regulate the branching fraction . . . 85

A.1 Optimal intra-compartment enzyme distributions . . . 96
A.2 Optimal compartmentalization strategies with decay . . . 98
A.3 Optimal intra-compartment enzyme distribution for high permeabilities . . . 100

C.1 Coefficient of determination . . . 110

C.2 Impermeable vs. permeable for delocalized arrangements 111

D.1 Branching fraction only depends on the ratio e_p/e_q 113
D.2 Branching fraction for different cluster parameters 113
D.3 Approximation with $\delta_{p,q} = 1$ 114
D.4 Branching fractions 114
D.5 Branching fraction and enhancement for enzyme pairs 115

1 Enzyme kinetics and metabolic modeling

Enzymes are catalysts which accelerate biochemical reactions by reducing the activation energy. They are mostly proteins and allow reactions in living organisms, which otherwise would not occur, to proceed on sufficiently short time scales. In general, enzymatic catalysis is initiated by the binding of the so-called substrate molecule to the enzyme. In the bound substrate-enzyme-complex, the activation barrier of the reaction is reduced and the substrate is rapidly processed to the product. After this transformation, the product-enzyme-complex quickly dissociates and the enzyme is available for the next reaction. Most enzymes are very specific for the substrate they catalyze. This specificity is due to the shape and charge distribution of the binding pocket of the enzyme, the so-called active site, which is complementary to the substrate such that it fits only to a specific substrate like a key into a lock. Enzymes are crucial for cellular metabolism where they catalyze the many biochemical reactions. To achieve complex biochemical transformations enzymes are organized into pathways where each step of the pathway is catalyzed by a certain enzyme. The different pathways in cellular metabolism are interconnected forming a complex network of enzymatic reactions. In this chapter we provide background on single enzyme kinetics and illustrate two well-known mathematical frameworks for modeling the kinetics of metabolic pathways.

1.1 Introduction to enzyme kinetics

1.1.1 The Michaelis-Menten equation

Enzymatic reactions are initiated by the formation of an enzyme-substrate-complex [1]. As the substrate is bound to the enzyme it is converted to the product, leading to an enzyme-product-complex which subsequently dissociates to a free enzyme and a product. Under the assumption that the dissociation of the enzyme-product-complex is fast and the amount

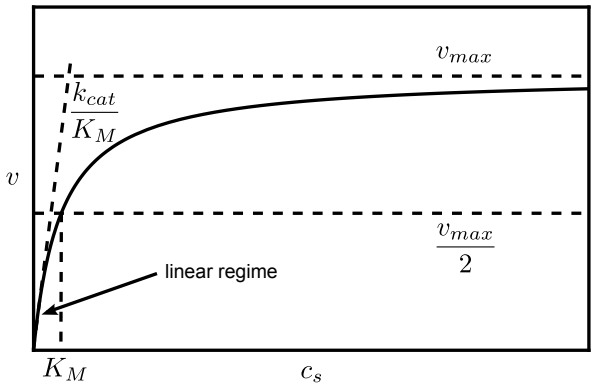

Figure 1.1: Michaelis-Menten kinetics. The reaction velocity v as a function of the substrate concentration c_s. In the low substrate concentration regime, most enzymes are not bound to a substrate and, thus, the reaction velocity increases linearly with c_s. The slope of the curve in the linear regime, $\frac{k_{cat}}{K_M}$, measures the catalytic efficiency of the enzyme. For high substrate concentrations, in contrast, all enzymes are bound to substrates and hence the reaction velocity saturates to the maximal velocity, $v_{max} = k_{cat} c_e$.

of product at the beginning of the reaction is low, the enzyme-product-complex and the recombination reaction of product with enzyme can be ignored and the enzymatic reaction can be described by the simplified reaction scheme,

$$E + S \underset{\kappa_{-1}}{\overset{\kappa_1}{\rightleftharpoons}} ES \xrightarrow{k_{\mathrm{cat}}} E + P. \tag{1.1}$$

Based on this reaction scheme, Michaelis and Menten derived an equation that connects the steady-state reaction velocity, v, with the enzyme and substrate concentrations as well as the rates of the individual reaction steps [2]. This so-called Michaelis-Menten equation captures the kinetics of many enzyme catalyzed reactions and because of its importance in this thesis, we will briefly present a derivation. We start by expressing the velocity of the catalytic reaction as,

$$v = k_{cat} c_{es}, \tag{1.2}$$

where c_{es} is the concentration of the enzyme substrate complex, and k_{cat} is the turnover number, which is defined as the number of substrate molecules converted to product per enzyme molecule and unit time. To derive an expression for the reaction velocity, we need to express c_{es} in terms of the directly known substrate concentration c_s and the total enzyme concentration c_e. In steady state the concentration c_{es} is constant and, thus, the rate of complex formation equals the rate of dissociation and catalysis,

$$\kappa_1 c_e^0 c_s = \kappa_{-1} c_{es} + k_{cat} c_{es}. \tag{1.3}$$

Here c_e^0 is the concentration of free enzymes which is equal to $c_e - c_{es}$. With Eq.1.3, we can express c_{es} as,

$$c_{es} = \left(\frac{\kappa_1}{\kappa_{-1} + k_{cat}} \right) c_e^0 c_s = \left(\frac{\kappa_1}{\kappa_{-1} + k_{cat}} \right) (c_e - c_{es})\, c_s \tag{1.4}$$

By rearranging and defining the Michaelis constant, $K_M = \frac{\kappa_{-1} + k_{cat}}{\kappa_1}$, the complex concentration reads, $c_{es} = \frac{c_e c_s}{K_M + c_s}$. With this expression and Eq.1.2 we obtain the famous Michaelis-Menten equation,

$$v = \frac{k_{cat} c_e c_s}{K_M + c_s}, \tag{1.5}$$

which has a non-linear dependence on the substrate concentration (see Fig. 1.1). The Michaelis constant K_M is an approximate measure of substrate's affinity for the enzyme. A low K_M value corresponds to a tight substrate-enzyme binding while a large K_M corresponds to a weak binding. At very high substrate concentrations, $c_s \gg K_M$, all enzymes are bound to a substrate and the velocity saturates to the maximal velocity, $v_{max} = k_{cat} c_e$, whereas at very low substrate concentrations $c_s \ll K_M$, most of the enzymes are unbound and, thus, the Michaelis-Menten equation reduces to the linear expression, $v = \frac{k_{cat}}{K_M} c_e c_s$. The ratio between turnover number and Michaelis constant $\frac{k_{cat}}{K_M}$, called catalytic efficiency, corresponds to the second order rate constant of the reaction between free enzymes and free substrates. This rate is a measure for the enzyme efficiency and the fastest enzymes like fumarase or carbonic anhydrase have catalytic efficiency of $10^8 - 10^9$ $\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ (see Sec. 1.1.5) [3, 4].

The Michaelis-Menten model provides a good phenomenological understanding of the reaction kinetics of enzymatic reactions. However, there are clearly scenarios where the model assumptions do not apply. First, in the Michaelis-Menten model the catalytic reaction is

irreversible, whereas in reality every enzymatic reaction is reversible. The assumption that the enzymatic reaction is irreversible is reasonable when the product concentration is low, e.g. at the beginning of the reaction when product molecules have not yet accumulated in the system. Second, the Michaelis-Menten model does not account for the enzyme-product-complex and only considers the reversible complex formation of enzyme and substrate. The binding of product to enzyme can be important and may lead to an inhibition of the enzyme. Thus, although the Michaelis-Menten equation provides a good phenomenological description of enzyme kinetics, in certain scenarios one may need to consider other kinetic models which include more details of the reaction.

1.1.2 The reversible Michaelis-Menten equation

As pointed out in the previous section, in many biochemical scenarios significant amounts of products exist in the reaction system such that enzymatic reactions can no longer be treated as irreversible. Analogous to the classical Michaelis-Menten equation, a rate equation can be derived which accounts for the reversible reaction. The reaction scheme 1.1 with the reverse reaction reads,

$$E + S \underset{\kappa_{-1}}{\overset{\kappa_1}{\rightleftharpoons}} ES \underset{\kappa_{-2}}{\overset{\kappa_2}{\rightleftharpoons}} E + P. \tag{1.6}$$

Following a similar derivation as for the irreversible Michealis-Menten equation, we obtain for the reversible Michealis-Menten equation,

$$v = \frac{\kappa_s c_e c_s - \kappa_p c_e c_p}{1 + c_s/K_{Ms} + c_p/K_{Mp}}, \tag{1.7}$$

where we defined the two Michaelis constants, $K_{Ms} = \frac{\kappa_{-1}\kappa_2}{\kappa_1}$ and $K_{Mp} = \frac{\kappa_{-1}+\kappa_2}{\kappa_{-2}}$. The parameters $\kappa_s = \frac{\kappa_1\kappa_2}{\kappa_{-1}+\kappa_2}$ and $\kappa_p = \frac{\kappa_{-1}\kappa_{-2}}{\kappa_{-1}\kappa_2}$ are the catalytic rates for the forward and backward reaction divided by the corresponding Michaelis constant. The reversible Michaelis-Menten equation has the advantage that its form does not depend on the specific reaction scheme. There are more complicated reaction schemes, e.g. including more intermediate complexes, which all lead to a rate equation equivalent to Eq. 1.7, only the definition of the parameters above changes according to the reaction scheme. For the reversible reaction in equilibrium, we can relate the kinetic parameters to the equilibrium constant. When an enzymatic reaction is at equilibrium, the forward and backward flux must balance and, thus, the rate Eq. 1.7 must be zero. From this it follows that

$$\kappa_s c_e c_s^{eq} = \kappa_p c_e c_p^{eq}, \tag{1.8}$$

where c_s^{eq} and c_p^{eq} are the equilibrium concentrations of the substrate and the product, respectively. With this relation, we can now express the equilibrium constant in terms of the kinetic parameters,

$$K_{eq} = \frac{c_p^{eq}}{c_s^{eq}} = \frac{\kappa_s}{\kappa_p}, \tag{1.9}$$

where K_{eq} is the equilibrium constant of the reaction. This is the Haldane relationship which is true for any reaction scheme that can be described by Eq. 1.7. More complicated reaction schemes would lead to more complicated Haldane relationships, but there is at least one relationship between the kinetic parameters and the equilibrium constant [5].

1.1.3 Enzyme inhibition

The activity of enzymes can be reduced by specific molecules, so-called inhibitors, which bind to the enzyme. This binding can be either reversible or irreversible and there are different mechanisms how the inhibitor decreases the enzyme activity. Here, we briefly discuss different mechanisms of reversible inhibition and illustrate how the Michealis-Menten equation changes in the presence of an inhibitor:

- **competitive inhibition:**
  In this case the inhibitor binds to the active site of the enzyme and thereby hinders the substrate from forming the enzyme-substrate complex. This leads to an apparent increase of K_M, since more substrate molecules are required to reach the point of half v_{max}. The resulting Michaelis-Menten equation with reversible competitive inhibition is then given by,

$$v = \frac{v_{max} c_s}{K_M(1 + c_i/K_i) + c_s}, \tag{1.10}$$

  where c_i is the concentration of inhibitor molecules and K_i is the inhibition constant. From Eq. 1.10 we can see that the Michaelis constant appears to change in the presence of an inhibitor to the apparent value $K_M^{app} = K_M(1 + c_i/K_i)$, while v_{max} remains unchanged.

- **uncompetitive inhibition:**
  In this case, the inhibitor binds only to the enzyme-substrate complex and thereby alters both the maximal reaction rate and the Michaelis constan to the apparent values, $v_{max}^{app} = v_{max}/(1 + c_i/K_i)$ and $K_M^{app} = K_M/(1 + c_i/K_i)$. The ratio of these values, however, remains unchanged.

- **non-competitive inhibition:**
  Here, the binding of inhibitor to the enzyme only reduces the activity of the enzymes but leaves the binding affinity of the substrate to the enzyme unchanged. Consequently, the Michaelis constant is not changed, only the maximal velocity is modified to the apparent value, $v_{max}^{app} = v_{max}/(1 + c_i/K_i)$.

In contrast to reversible inhibition, in irreversible inhibition the inhibitor binds covalently to the enzyme and permanently reduces its activity by specifically modifying the active site of the enzyme. This should not be confused with irreversible enzyme inactivation where the protein structure of the enzyme is destroyed by a non-specific agent.

Besides these mechanisms, enzymes can also be inhibited by their product. From the reversible Michaelis-Menten equation Eq. 1.7 we can see that the reaction rate slows down as product is accumulated in the system. This reduction of the reaction rate is only significant if the reaction is considerably reversible. However, product inhibition is also observed in reactions which are essentially irreversible. In those cases, the product can bind to the enzyme but is not processed back to the substrate. This means that the enzymes are sequestered as enzyme-product complexes, which makes them unavailable for the reaction.

1.1.4 Elasticities

The behavior of enzymatic reaction kinetics is in general described by rate equations like Eq. 1.7. Such rate equations explain how the rate changes as a function of the concentration of different substances like reactants or inhibitors. In many modeling approaches one is interested in the degree to which a small concentration change of a substance changes the reaction rate. A useful measure for this quantity is the elasticity coefficients, which is defined as,

$$\epsilon_c^v = \frac{c}{v}\frac{\partial v}{\partial c} = \frac{\partial \log v}{\partial \log c}, \tag{1.11}$$

where c is the concentration of any substance involved in the enzymatic reaction. In general this can be the substrate, the product, the enzyme itself, or any effector like an inhibitor or activator. The rescaling of the partial derivative by the factor $\frac{c}{v}$ ensures that the coefficient is independent of the units used to measure reaction rates and concentrations. The elasticities can be determined from any arbitrary rate equation. For example, from the reversible Michaelis-Menten equation Eq. 1.7, we can compute the elasticity with respect to the substrate concentration,

$$\epsilon_s^v = \frac{1}{1-\Gamma/K_{eq}} - \frac{\sigma}{1+\sigma+\pi}, \tag{1.12}$$

where K_{eq} is the equilibrium constant, $\Gamma = c_p/c_s$ is the mass action ratio, and $\sigma = c_s/K_{Ms}$ and $\pi = c_p/K_{Mp}$ are the substrate and product concentrations scaled by their Michaelis constant. The expression consists of two terms, which can be easily interpreted. The first term depends only on the degree of disequilibrium, it diverges at equilibrium. The second term measures the degree of saturation of the enzyme [5]. In general, however, one is interested in the numerical value of the elasticity rather than its algebraic form: if the elasticity is zero, v does not change with c_s, if it is positive v increases with c_s, and if it is negative v decreases with c_s.

The elasticity can be viewed as an extension of the concept of the order of a reaction. The order of a reaction is defined by constant integer values, while the elasticity can have non-integer values which change as a function of substrate concentration. For example, the Michaelis-Menten equation gradually transitions from first order at low substrate concentration to zero order at saturation. This concept of elasticities is crucial for metabolic control analysis, which we discuss in Sec. 1.2.1.

1.1.5 Impact of diffusion on enzyme kinetics

The reaction rate of an enzymatic reaction depends, on the one hand, on the rate of diffusive encounter of substrate with its enzyme and, on the other hand, on the speed of catalytic turnover once the substrate is bound to the enzyme. Standard enzyme kinetics like the Michaelis-Menten kinetics do not account for the influence of diffusion on the reaction velocity. In this section we will derive a simple modification of the Michaelis-Menten equation that accounts for the diffusion of substrates and enzymes. For that we will follow the derivations presented in [6, 7, 8], where the theory of Smoluchowsky, Collins and Kimball [9, 10] is applied on the association-dissociation-process of two molecular species. Using the rate constants of this process to describe the dynamics of enzyme-substrate-complex formation, we find an

expression for the Michaelis-Menten equation that includes the diffusive motion of reactants to the kinetics.

Let us consider substrates diffusing relative to an enzyme molecules. In any coordinate system, the diffusive motion of substrates is governed by the diffusion equation,

$$\frac{\partial \rho_s(\mathbf{r},t)}{\partial t} = D\nabla^2 \rho_s(\mathbf{r},t), \tag{1.13}$$

where the diffusion coefficient is the sum of the diffusion coefficients of the enzyme and substrate, $D = D_e + D_s$, and $\rho_s(\mathbf{r},t)$ is the concentration profile of substrates in the vicinity of the enzyme. The substrate can bind to the enzyme when the two molecules are in contact. This is the case when the distance between the molecules is $\sigma = r_e + r_s$, the sum of enzyme and substrate radius. The net reaction flux, J, of complex formation is the rate of association reactions minus the rate of dissociation reactions per unit volume. This reaction flux equals the flux at the surface of an average enzyme molecule,

$$J = 4\pi r^2 D c_e^0 \frac{\partial \rho_s(r)}{\partial r}\bigg|_{r=\sigma} = k_1 c_e \rho_s(\sigma) - k_{-1} c_{es}, \tag{1.14}$$

where c_e^0 and c_{es} are the average concentrations of free enzymes, and enzyme-substrate-complexes, respectively. The constants k_1 and k_{-1} are the intrinsic rate constants characterizing the association-dissociation-reaction upon contact of substrates with enzymes. At large distances from the enzyme, the substrate concentration relaxes to the average concentration, c_s, leading to the boundary condition, $\rho_s(r \to \infty) = c_s$.

By solving this system in steady-state we determine the steady-state reaction flux of complex formation,

$$J = \underbrace{\frac{k_D k_1}{k_1 + k_D}}_{\kappa_1} c_e^0 c_s - \underbrace{\frac{k_D k_{-1}}{k_1 + k_D}}_{\kappa_{-1}} c_{es}. \tag{1.15}$$

From this expression we can read off the effective association and dissociation rates, κ_1 and κ_{-1}, as functions of the corresponding intrinsic rates and the Smoluchowski rate constant of diffusion limited reaction, $k_D = 4\pi\sigma D$. The considerations can be extended to also include interaction potentials between the molecules as first considered by Debye [11]. Using the expressions for the rates κ_1 and κ_{-1} in the Michaelis-Menten equation (Eq. 1.5) we obtain,

$$v = \frac{k_{cat} c_e c_s}{K_M^{in} + k_{cat}/k_D + c_s}, \tag{1.16}$$

where $K_M^{in} = \frac{k_{-1}+k_{cat}}{k_1}$ is the intrinsic Michaelis constant. In the low substrate concentration regime this equation becomes linear and the catalytic efficiency is given by, $(k_{cat}/K_M)^{-1} = \left(k_{cat}/K_M^{in}\right)^{-1} + k_D^{-1}$. In the regime, where $k_{cat}/K_M^{in} \ll k_D$, the reaction is always in binding equilibrium with the surrounding substrate pool and no substrate concentration gradients emerge. This regime is referred to as reaction-limited regime and the intrinsic rates equal the effective rates, $k_1 = \kappa_1$ and $k_{-1} = \kappa_{-1}$. In the opposite, diffusion-limited regime where $k_{cat}/K_M^{in} \gg k_D$, substrate concentration gradients emerge, and the reaction depends crucially on the diffusive transport of substrates to the enzyme. In the extreme case where every substrate-enzyme collision results in an immediate turnover of the substrate, $k_{cat}/K_M^{in} \to \infty$,

the catalytic efficiency becomes equal to k_D. Therefore, k_D sets an theoretical upper limit for the catalytic efficiency of enzymatic reactions. We can approximate this limit by assuming that the diffusion coefficient of a small metabolite is around $D = 100\,\mu m^2\,s^{-1}$ and the interaction radius is $\sigma = 1\,nm$. With this we obtain as a theoretical estimate for the upper limit of the catalytic efficiency, $k_{cat}/K_M \approx 1 \times 10^9\,M^{-1}\,s^{-1}$. Note, this approximation is in the range of the fastest measured enzymes (see Sec. 1.1.1). However, the described model omits several details of enzymatic reactions, which may need to be accounted for in order to correctly capture the kinetics of the reaction. First, enzymes are reactive only at the active site rather then over the entire molecular surface as assumed by the model. Second, electrostatic interactions between the active site and the substrate were not considered. Such interactions may enhance the encounter and binding probability. Including these details can lead to an order of magnitude difference in the value of the totally diffusion limited catalytic efficiency [12, 13, 14].

1.2 Models of multi-enzyme kinetics

So far we have focused on the kinetics of individual enzymes in isolation. However, in every biological system, enzymes function together in reaction pathways, where the product of one enzymatic reaction acts as the substrate of the next. These reaction pathways are highly interconnected, forming large reaction networks, called metabolism. Usually, metabolism is divided into two classes: First, the catabolism which involves all processes where large molecules are broken down into smaller ones to generate energy in form of ATP, and second, anabolism which includes all processes where complex molecules such as proteins or nucleic acids are build from small precursor molecules. These two classes of metabolism complement each other in the sense that the energy produced by catabolism is consumed by anabolism. To understand the kinetic behavior of metabolic networks and to predict individual fluxes, several modeling approaches have been developed. Two of these approaches, metabolic control analysis (MCA) and flux balance analysis (FBA), will be discussed in the following.

1.2.1 Metabolic control analysis

The aim of metabolic control analysis is to describe the dynamics of enzymatic pathways in terms of the kinetic properties of the individual, isolated pathway enzymes. We will briefly illustrate the main ideas of MCA following the books of Cornish-Bowden and Fell [5, 15]. In its standard form, MCA deals with the reaction dynamics of enzymatic pathways in steady-state, with at least one source of metabolites and at least one sink. The following scheme illustrates an example for such an enzymatic pathway,

$$S_0 \overset{E_1}{\rightleftharpoons} S_1 \overset{E_2}{\rightleftharpoons} S_2 \overset{E_3}{\rightleftharpoons} S_3 \overset{E_4}{\rightleftharpoons} S_4 \overset{E_5}{\rightleftharpoons} S_5, \tag{1.17}$$

where the initial substrate S_0 is the source and S_5 is the sink of the pathway. These concentrations are considered to be external and constant, while the other, internal concentrations are free to vary. Each step of the pathway is catalyzed with a rate v_i, which we consider for simplicity to be just proportional to the concentration of exactly one enzyme. Note, the

kinetic properties of the enzymes remain unchanged while v_i can vary. MCA now provides a framework to study how the kinetic behavior of a reaction pathway is determined by the characteristics of the individual enzymes. When the concentration of the metabolite S_1 is in steady-state, this implies that the production rate v_1 of the metabolite S_1 must be equal to the rate v_2 at which it is consumed. Thus, when the concentrations of all metabolites are in steady-state, all the reaction rates must be equal to one another. This global reaction rate is called the flux, J [16]. Note, for branched pathways, there can be several distinct fluxes in the same pathway and, thus, the relations between fluxes become more complex. The reaction rate of an enzyme is a local property, since it refers to a single, isolated enzyme, whereas the steady-state flux and the metabolite concentrations are systemic properties, since they refer to the pathway. Varying an external parameter of an isolated enzyme E_i leads to a change in the reaction rate v_i. The question now is, what is the corresponding effect on the system flux J, when E_i is embedded in the system. This effect is characterized by the following relation called flux control coefficient [17],

$$\mathcal{C}_i^J = \frac{\partial \log J}{\partial \log v_i}. \tag{1.18}$$

This definition allows us to identify when an enzyme catalyzes the rate-limiting step of a pathway. Any variation in the activity of the rate-limiting enzyme produces a proportional variation in the flux through the pathway. Consequently, for a rate-limiting enzyme the flux control coefficient is $C_i^J = 1$. Analogous to the flux control coefficients, the quantity that measures effects on metabolite concentrations is called concentration control coefficient. The coefficient that measures how the metabolite concentration s_j is affected when the reaction rate v_i is changed is given by,

$$\mathcal{C}_i^{s_j} = \frac{\partial \log s_j}{\partial \log v_i}. \tag{1.19}$$

Importantly, the control coefficients fulfill two summation theorems, which originate from the fact that the flux and the metabolite concentration are systemic properties and hence their control affects all reactions in the system [16, 17],

$$\sum_{i=1}^{n} \mathcal{C}_i^J = 1, \tag{1.20a}$$

$$\sum_{i=1}^{n} \mathcal{C}_i^{s_j} = 0. \tag{1.20b}$$

The latter equation expresses the fact that a simultaneous change in the enzyme activity by the same factor does not change the metabolite concentration [17]. The first relation, on the other hand, states that the control of flux through a pathway is shared by all the enzymes in the system. However, when one enzyme is rate-limiting the kinetic behavior of the whole system is determined by the step catalyzed by this enzyme. In the following we give a proof of the summation relation Eq. 1.20a. For that, we consider small changes $\mathrm{d}e_i$ in all enzyme concentrations and assume that the reaction rates v_i are all proportional to the concentrations of the enzymes that catalyze them. Thus, the simplest way to vary the reaction rates is to change the enzyme concentrations. The total change of any flux J can then be described in terms of the individual changes of the enzyme concentrations, which can be expressed by the total differential,

$$\mathrm{d}J = \sum_{i=1}^{n} \frac{\partial J}{\partial e_i} \mathrm{d}e_i. \tag{1.21}$$

Multiplying Eq. 1.21 with $\frac{e_i}{Je_i}$ we obtain,

$$\frac{\mathrm{d}J}{J} = \sum_{i=1}^{n} \frac{\partial \log J}{\partial \log e_i} \frac{\mathrm{d}e_i}{e_i}. \tag{1.22}$$

Since we assumed that the reaction rate is proportional to the enzyme concentration, Eq. 1.18 can also be written as, $\mathcal{C}_i^J = \frac{\partial \log J}{\partial \log e_i}$ and we obtain,

$$\frac{\mathrm{d}J}{J} = \sum_{i=1}^{n} \mathcal{C}_i^J \frac{\mathrm{d}e_i}{e_i}. \tag{1.23}$$

The magnitude of the small changes $\mathrm{d}e_i$ are not specified, and without limiting the generality of this expression, we can assume that $\mathrm{d}e_i$ is proportional to e_i, $\mathrm{d}e_i = \alpha e_i$. The same proportionality also holds for the change of the steady-state flux, $\mathrm{d}J$ [5]. Thus, Eq. 1.23 reduces to,

$$\alpha = \sum_{i=1}^{n} \mathcal{C}_i^J \alpha, \tag{1.24}$$

by dividing with α we obtain the summation relationship Eq. 1.20a. The same logic can be applied to derive the second summation relation.

For a linear n-step pathway without branch points and where every enzyme catalyzes one specific step, there is now one summation relation for the flux control coefficients, Eq. 1.20 and $n-1$ summation relations for the concentration control coefficients. However, there are n flux control coefficients and $n(n-1)$ concentration control coefficients and, thus, in total n^2 control coefficients. In order to determine all control coefficients, a set of further $n(n-1)$ equations is required [5]. The required equations are given by the connectivity properties,

$$\sum_{i=1}^{n} \mathcal{C}_i^J \varepsilon_{s_j}^{v_i} = 0, \tag{1.25a}$$

$$\sum_{i=1}^{n} \mathcal{C}_i^{s_k} \varepsilon_{s_j}^{v_i} = -\delta_{kj}, \tag{1.25b}$$

where $\varepsilon_{s_j}^{v_i} = \frac{\partial \log v_i}{\partial \log s_j}$ is the elasticity. In general, elasticities describe the change of the reaction rate caused by a variation of an external parameter, which in our case is the substrate concentration (see Sec. 1.1.4). For the linear n-step pathway, we obtain $n-1$ equations from Eq. 1.25a and $(n-1)^2$ equations similar to Eq. 1.25b. Together with the n summation relations these equations provide the missing $n(n-1)$ equations needed to determine all n^2 control coefficients from the elasticities [5].

1.2.2 Flux Balance Analysis

Flux Balance Analysis is a mathematical modeling approach used to determine the steady-state fluxes in complex metabolic networks. It is widely used to study genome-scale biochemical networks, which contain all known metabolic reactions in an organism. By determining the fluxes of metabolites through the network, it is possible to predict, for example, the growth rate of an organism or the production rate of metabolites in biotechnological applications [19].

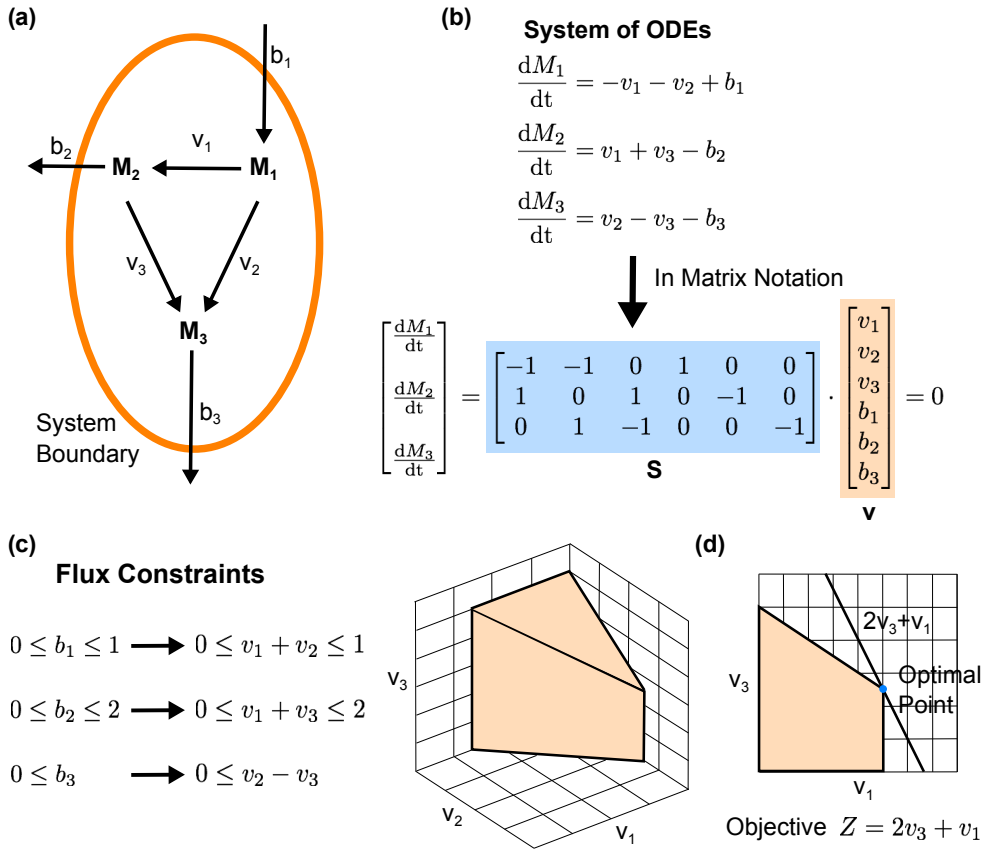

Figure 1.2: Illustration of Flux Balance Analysis. (a) Model metabolic system consisting of three metabolites (M_1, M_2, and M_3). The fluxes include the reaction fluxes, v_i, and exchange fluxes with the external environment, b_i. **(b)** Resulting system of ODEs and representation using matrix notation. In steady-state this leads to $\mathbf{S} \cdot \mathbf{v} = 0$. **(c)** Possible constraints on the fluxes due to physiological constraints. This leads to the sketched solution space. Identification of the flux distribution which optimizes the objective function Z. Illustration adapted with permission from [18]

FBA is a constraint-based approach in which the stoichiometry of the metabolic network and specific bounds on the steady-state fluxes constrain the solution.

In the following, we briefly explain the basic concepts of FBA and illustrate useful applications of the theory. The dynamics of a metabolic network, which consists of m metabolites and r reactions, can be described by a set of m ordinary differential equations (ODEs),

$$\frac{\partial M_i}{\partial t} = \sum_{j=1}^{r} S_{ij} v_j \quad \text{with} \quad i = 1, ..., m, \tag{1.26}$$

where M_i is the concentration of the i-th metabolite, v_j is the flux of the j-th reaction, andS_{ij} is the stoichiometric coefficient representing the number of moles of metabolite i produced (or consumed) in the j-th reaction [20]. By assuming that the metabolite concentrations reached steady-state and combining the stoichiometric coefficients into the so-called stoichiometric matrix $\mathbf{S}$, Eq. 1.26 reduces in matrix notation to,

$$\mathbf{S} \cdot \mathbf{v} = 0, \tag{1.27}$$

where $\mathbf{v}$ is the vector of all r fluxes. This equation incorporates the network topology and the stoichiometry of each reaction into a constraint on the possible steady-state fluxes. Since, metabolic networks typically contain more reactions than metabolites, this system of equations contains more variables than equations and consequently there is no unique solution to the system of equations. Any solution of $\mathbf{v}$ that satisfies Eq. 1.27 is said to be in the null space of $\mathbf{S}$ [19]. To obtain a unique solution, one needs to impose additional constraints on the fluxes. These constraints may be due to reaction thermodynamics, or system characteristics of the network, e.g. influxes from the environment, or other physiological limitations. By defining an upper- and lower bound for each flux, we reduce the solution space to all fluxes within the null space, which fulfill the constraints,

$$\alpha_j \leq v_j \leq \beta_j. \tag{1.28}$$

Within this allowed solution space, FBA seeks to identify the point which either maximizes or minimizes an objective function $Z = \mathbf{c}^\top \mathbf{v}$. For example, the objective can be related to a specific cellular function which is assumed to have been optimized by evolution [21]. Furthermore, often the solution of interest is a flux distribution, which leads to a maximal growth rate [19] or maximal ATP production of an organism [22].

Linear programming methods are used to determine the fluxes that optimize the objective function within the allowed solution space defined by the mass balance equations (Eq. 1.27) and the bounds on the fluxes (Eq. 1.28). The main advantage of this approach is that no knowledge about metabolite concentrations or enzyme kinetics and rate laws are required to obtain the distribution of fluxes in metabolic networks. Additionally the simulations are computationally inexpensive. For a typical metabolic network it takes about a few seconds to determine the flux distribution which optimizes biomass production. Yet, FBA has some limitations: It cannot predict metabolite concentrations because the enzyme kinetics and the parameters are not included, it is only suitable for studying fluxes at steady state, and thermodynamic aspects are not considered. Furthermore, FBA does not account for regulatory effects such as activation of enzymes by protein kinases or regulation of gene expression [19].

To address some of these limitations, additional physical constraints have be incorporated into the traditional FBA models. For example, the concept of energy balance has been added to the standard model to account for thermodynamic constraints of the reactions [23]. Furthermore, the constraint of limited cellular enzyme abundance due to spatial limitation has been added to FBA, which gave rise to three metabolic phases and hierarchical modes of substrate utilization in mixed substrate growth medium [24].

2 Metabolic channeling and spatial enzyme organization

In the previous chapter, we have seen that enzymes are organized into pathways, where the product of an enzymatic reaction acts as the substrate of the next reaction. The enzymatic pathways are interconnected forming a large network of enzymatic reactions, the so-called metabolic network. Besides this organization of enzymes into pathways and metabolic networks, metabolic enzymes also show a high degree of spatial organization inside the cell. This spatial coordination of metabolism includes the segregation of enzymes into membrane-bound compartments, the aggregation of enzymes into non-membrane bound clusters and filaments, and the organization of enzymes into macromolecular multi-enzyme complexes. Most of these spatial arrangements have been hypothesized to induce metabolic channeling, a process where an intermediate produced by an upstream enzyme is directly consumed by a downstream enzyme without first diffusing into the bulk [25, 26]. Metabolic channeling has first been introduced as a process where intermediates are transferred between consecutive enzymes through structural features like tunnels, swing arms, or electrostatic guidance [27, 28]. However, nowadays the terminology "metabolic channeling" is used for all strategies that enable rapid and efficient processing of intermediates and simultaneously prevent intermediates from diffusing into the surrounding cytoplasm. Such strategies include the compartmentalization of pathway reactions into organelles or the colocalization of consecutive enzymes into clusters, so-called metabolons [29]. In this work, we are mainly interested in how spatial organization can affect metabolic fluxes and induce metabolic channeling without the need for specific mechanisms of intermediate transfer. In the following chapter, we first shall discuss the benefits of metabolic channeling for the cell. We then describe different natural strategies of spatial enzyme organization and explain how they enable metabolic channeling. Finally, we illustrate how synthetic biologists exploit these natural strategies to engineer synthetic enzyme arrangements with the aim to efficiently produce valuable products like biofuels or pharmaceutical drugs.

2.1 The benefits of metabolic channeling

Before discussing the different natural strategies that enable metabolic channeling and the closely related spatial organization of enzymes, we first describe the beneficial effects of metabolic channeling and its importance for the regulation of metabolic fluxes.

- **Efficient processing of unstable or volatile intermediates**
  The direct transfer of intermediates between consecutive enzymes leads to efficient processing of intermediates which are either badly retained by the cellular membrane or intrinsically unstable. Without channeling, such intermediates would likely diffuse out

of the cell or decay before they encounter a downstream enzyme by diffusion and can be further processed.

- **Reduction of unwanted competing side reactions**
  Intermediates are often not specific to only one enzyme or reaction but can react in several different pathways. However, in many cases these competing reactions are unwanted and it is desirable to direct the processing of the intermediates towards one specific pathway. By channeling intermediates between consecutive enzymes, the exposer of intermediates to competing enzymes is reduced and therefore insulates intermediates from undesired side reactions.

- **Protection from toxic pathway intermediates**
  In the course of the successive processing of molecules in a metabolic pathway, intermediates can be produced which are toxic to the cell when they are dissolved in the cytoplasm. To protect the cell from intoxication, toxic intermediates need to be segregated from the cytoplasm. This is achieved by several channeling strategies like direct transfer of toxic intermediates between consecutive enzymes or the segregation of toxic intermediates by compartmentalizing entire pathways.

- **Avoiding unfavorable kinetics and equilibria in the cytoplasm**
  In some metabolic pathways, certain reactions would not proceed in the desired direction by means of classical mass action ratios if they were well mixed in the cytoplasm. Therefore, to prevent the intermediates from being back converted, they are directly channeled to the next enzyme, and the reaction pathway can proceed in the desired direction.

- **Flux enhancement**
  Channeling of intermediates may enhance the flux through metabolic pathways by reducing the time required for an intermediate to reach the active site of the next enzyme. It was argued that the main advantage comes rather from bypassing metabolite solvation and desolvation and thus saving the required time for these processes [30]. Importantly, channeling allows to achieve high fluxes without elevating intermediate concentration levels in the bulk. While flux enhancement is reasonable for direct channeling mechanisms, it is unclear whether an enhancement can be achieved only by spatial enzyme organization where diffusion mediates the transfer of intermediates. In particular, it is controversial whether proximity of sequential enzymes alone suffices to enable an efficient transfer of intermediates.

- **Regulation of metabolic fluxes**
  Metabolic fluxes at branch point can be controlled by dynamically enabling channeling of intermediates towards a specific enzyme at the branch point. Hence metabolic channeling presents a means to regulate fluxes without the need of altering enzyme expression levels or changing enzymatic activities by post-translational modifications.

The mechanisms by which metabolic channeling can be achieved are diverse and for each of the above listed cellular requirements a certain strategy is most suitable. The different strategies discussed in the following are summarized in Fig. 2.1.

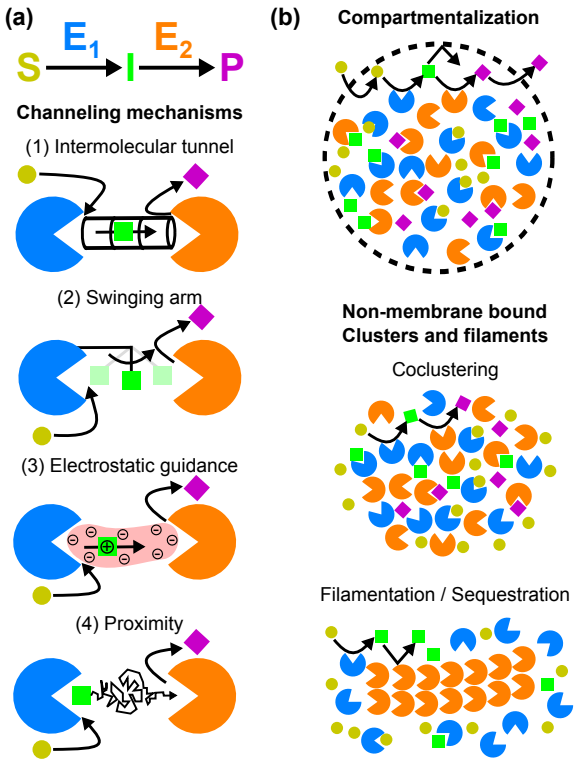

Figure 2.1: Schematic illustration of channeling mechanisms and spatial enzyme organization. **(a)** Channeling mechanisms illustrated for a simple pathway of two consecutive enzymes, which sequentially process substrate S to product P via an intermediate I. (1) Channeling through an intermolecular tunnel. (2) Intermediate transfer mediated by a swinging arm. (3) Guidance of a charged intermediate between active sites along an oppositely charged molecular surface. (4) Channeling by proximity between active sites. The intermediate transfer is mediated by free diffusion. The impact of proximity on the processing of intermediates is still controversial. **(b)** Spatial organization of enzymes. (Top) Colocalization of consecutive enzymes within membrane-bound compartments. The membrane prevents the intermediates from diffusing out of the compartment. (bottom) Spatial organization of enzymes into non-membrane bound clusters and filaments. Coclustering enhances the processing of intermediates due to the increased local concentration of enzymes, while filamentation and sequestration present mechanisms to downregulate metabolic fluxes.

2.2 Metabolic channeling mediated by structural features

A direct transfer of intermediates between enzymes can be mediated by several structural features of the enzymes or the multi-enzyme complexes. One such feature is an intramolecular tunnel between consecutive active sites, which guides the intermediate transfer and prevents intermediates from diffusing into the bulk solution [31]. The prime example of channeling by an intramolecular tunnel is the tryptophan synthase. This enzyme molecule consists of two active sites, where at the first active site indole 3-glycerolphosphate is converted to indole and glyceraldehyde-3-phosphate. The intermediate indole is subsequently transferred through the intramolecular tunnel to the second active site where it is catalyzed together with serine to the products tryptophan and water [31]. Crystal structure studies of tryptophan synthase revealed a 25 Å long tunnel through which a rapid, quasi one-dimensional diffusional transfer is possible [32]. With this channeling strategy the uncharged intermediate indole, which would easily diffuse through the cellular membrane, is prevented from escaping the reaction. Other examples where a tunnel is exploited to channel intermediates include aldolasedehydrogenase which is involved in the degradation of toxic aromatic compounds in bacteria [33] and glucosamine-6-phosphate synthase which catalyzes the rate-limiting steps in hexosamine metabolism [34].

In addition to intramolecular tunnels, channeling can be enabled by swinging arms which covalently bind and shuttle intermediates between active sites [28]. An astonishing example of a swinging arm in a multi-enzyme complex is the pyruvate dehydrogenase complex (PDC) which converts pyruvate to acetyl-CoA and thus links glycolysis to the citric acid cycle. The PDC consists of three different types of enzymes, pyruvate dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2), and dihydrolipoamide dehydrogenase (E3), which

are organized into a highly symmetric complex [35]. The E2 enzymes form a core to which the E1 and E3 enzymes non-covalently bind. The E1s and E3s are precisely positioned and oriented to ensure optimal interaction of their active sites with the lipoamide swinging arm of the E2 core during the sequential reaction [36]. Other 2-oxoacid dehydrogenases, like the 2-oxoglutarate dehydrogenase complex and branched chain 2-oxoacid dehydrogenase complex, possess similar structural features and employ the same lipoamide swinging arm to enable channeling [28, 37]. Swing arm mechanisms are also found in fatty acid synthases and polyketide synthases where intermediates are shuttled by so-called acyl carrier proteins [28].
Another channeling mechanism mediated by a structural feature is the electrostatic guidance of intermediates between active sites [38, 37]. Charged protein residues along bifunctional enzymes direct the diffusive motion of oppositely charged intermediates across the protein surface to the next active site. In the citric acid cycle, the enzymes malate dehydrogenase (MDH) and citric synthase (CS) form a complex. A positively charged region on the enzyme complex directs the diffusion of the negatively charged intermediate oxaloacetate, thereby bridging the two active sites. Surface analysis of the structure of the enzyme complex revealed a charged patch connecting the active sites [39] (see Fig. 2.2a). Modeling approaches verified the presence of channeling by electrostatic guidance in the MDH-CS complex using measured charge distributions on the enzyme surface [40]. Notably, the equilibrium constant of the MDH reaction is unfavorable in the forward direction of the citric acid cycle [41], thus channeling of oxaloacetate is necessary to achieve high fluxes through the MDH-CS pair despite the unfavorable kinetics [37]. Another example of electrostatic channeling is the transfer of dihydrofolate between the active sites of the bifunctional enzyme dihydrofolate reductase-thymidylate synthase (DHFR-TS) which catalyzes two steps of the thymidylate cycle [38, 42]. Similar to the MDH-CS complex, a negative charge distribution on the DHFR-TS complex surface leads to guided diffusion of the positively charged dihydrofolate intermediate between the active sites [42].

2.3 Enzyme clustering and compartmentalization

The concept of metabolic channeling was traditionally considered as the guided transfer of intermediates mediated by structural features. However today strategies like enzyme clustering and compartmentalization without directed transport of intermediates have similar effects as standard channeling mechanism and are thus also considered as concepts of metabolic channeling. In this thesis, we mainly focus on these latter strategies and elucidate design principles for maximal flux yield and unravel the role of clustering in flux regulation. In the following, we illustrate several examples of spatial enzyme organization and explain the functions and advantages. We classify the examples into two categories, first enzyme compartments which are delimited by membranes or shells and second non-membrane bound enzyme clusters and filaments.

2.3.1 Membrane delimited organelles and microcompartments

Intracellular encapsulation of metabolic processes in membrane-delineated compartments is a key feature not only of eukaryotic cells but also of prokaryotes. The compartmentalization of

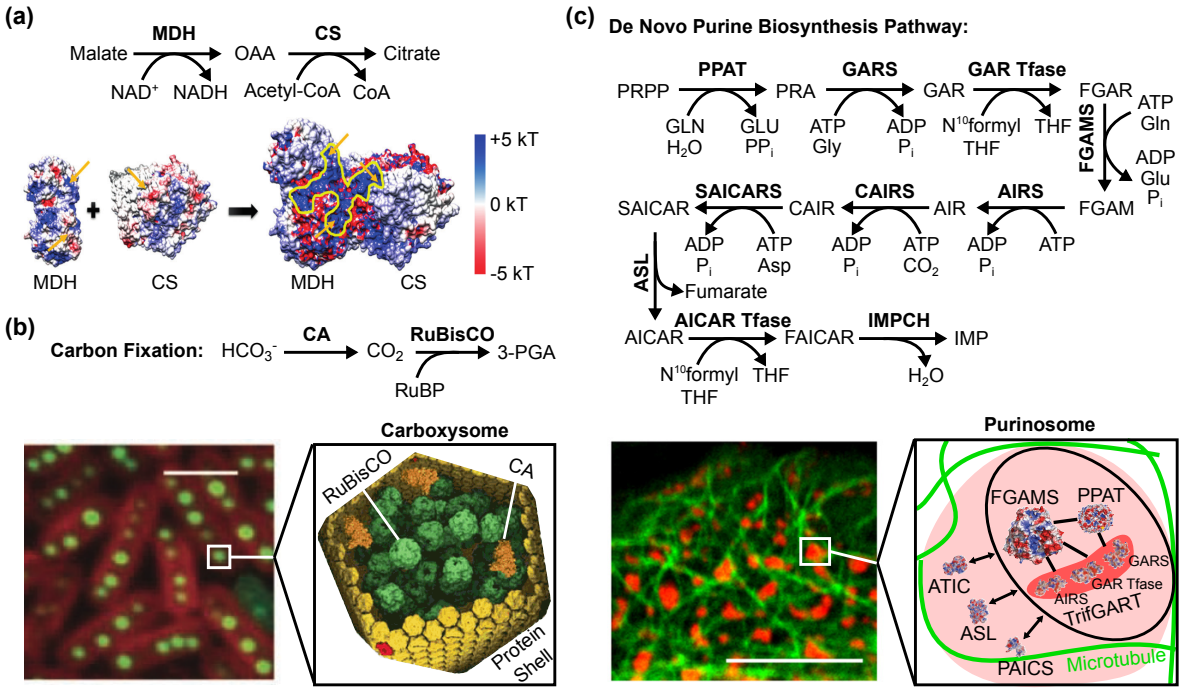

Figure 2.2: Natural examples of metabolic channeling and spatial enzyme organization. (a) Electrostatic guidance in the MDH-CS complex. (Top) MDH catalyzes the oxidation of malate to oxaloacetate (OAA). The negatively charged intermediate OAA is guided along the positively charged surface of the MDH-CS complex to the active site of CS where it is converted to citrate. (Bottom) Illustration of the surface electrostatic potential of MDH and CS in isolation and in the complex. Positive charges are shown in blue, negative charges are shown in red, and neutral charges are shown in white. The orange arrows point to the active sites and the electrostatic channel is highlighted in yellow. Figure adapted with permission from [39]. **(b)** Carbon fixation by the carboxysome. (Top) Carbon fixation pathway in cyanobacteria. (Bottom left) Carboxysomes in Cyanobacteria. The carboxysomes are fluorescently labeled in green and the bacterial membrane is labeled in red. Scale bar, 2 µm. Figure adapted with permission from [43]. (Bottom right) Schematic illustration of the carboxysome. The enzymes RuBisCO and carbonic anhydrase are confined inside a protein shell. Figure adapted with permission from [44]. **(c)** De novo purine biosynthesis performed by the purinosome. (Top) De novo purine biosynthesis pathway in humans. The pathway consists of ten steps catalyzed by six enzymes: The trifunctional enzyme TGART which consists of GARS, GAR Tfase and AIRS, the bifunctional enzyme PAIC which consists of CAIRS and SAICARS, and the bifunctional enzyme ATIC which consists of AICAR transformylase and IMPCH. (Bottom left) Localization of purinosomes (red) and microtubule filaments (green) in HeLa cells. Figure adapted with permission from [45]. (Bottom right) Enzyme-enzyme interaction in the purinosome. FGAMS, PPAT, and TGART strongly interact and form a core complex to which the other enzymes transiently bind. Schematic adapted with permission from [46].

consecutive enzymes serves as a strategy to overcome several challenges inherent to cellular metabolism. For example, toxic pathway intermediates stay segregated inside compartments or reaction fluxes of slow enzymes are boosted by confining and accumulating intermediates within compartments. Eukaryotic cells are well known for their high degree of spatial organization of metabolism. Examples of eukaryotic compartmentalization of metabolic reactions into organelles include peroxisomes, mitochondria, and chloroplasts.

Peroxisomes encapsulate, i.a., oxidative enzymes involved in the production and consumption of hydrogen peroxide and other toxic pathway intermediates. The importance of confining hydrogen peroxide inside peroxisomes becomes evident in methylotrophic yeast. Those yeast strains are capable of using methanol as its only carbon source, despite the fact that toxic hydrogen peroxide and formaldehyde are necessary intermediates for methanol metabolism [47]. When these yeast cells are grown on methanol, peroxisome biogenesis is upregulated leading to considerably increased peroxisomal structures [48, 49]. This highlights the crucial role of peroxisomes in segregating unavoidable toxic intermediates. Additionally, the alkaline pH in the peroxisome enhances the activities of peroxisomal oxidase and catalase which results in faster processing of metabolites [50]. In mitochondria, large octamers of the enzyme creatine kinase (MtCK) colocalize with the two transmembrane proteins adenine translocator (ANT) and voltage-dependent anion channel (VDAC) in the intermembrane space to efficiently catalyze the transphosphorylation of ATP and creatine (Cr) to ADP and PCr. ANT transports ATP from the mitochondrial matrix and VDAC Cr from the cytosol to the intermembrane MtCK. The spatial proximity of the transmembrane proteins and MtCK together with the confinement inside the intermembrane space enables metabolic channeling. After the reaction, ADP is recycled in the mitochondrion and stimulates oxidative phosphorylation, while PCr leaves the mitochondrion and acts as a primary high energy phosphoryl compound [51].

Besides the ubiquitous spatial organization of metabolism in eukaryotic cells, also prokaryotes, which long were believed to lack any subcellular structure, possess metabolic compartments, so-called bacterial microcompartments (BMC) [52, 53, 54, 55, 56]. In contrast to eukaryotic organelles BMCs encapsulate the metabolic reactions within protein shells, forming polyhedral compartments. Pores across the protein shell act as channels for metabolite diffusion into and out of the compartment [57, 58]. Often, specific electrostatic charges around the pores facilitate the passage of oppositely charged metabolites, thus presenting a mechanism for selective metabolite transport [54].

The best studied BMC, the carboxysome, is the key element of the carbon concentrating mechanism (CCM) in cyanobacteria [59, 60, 61, 62]. The CCM is crucial for efficient carbon fixation in the Calvin cycle, which relies on the inefficient enzyme ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO). Besides having a low turnover rate, RuBisCO also catalyzes the futile fixation of oxygen in addition to the desired fixation of carbon dioxide. To overcome these challenges, RuBisCO is compartmentalized together with a carbonic anhydrase (CA) in a proteinaceous shell, the carboxysome (see Fig. 2.2b). CA converts incoming bicarbonate HCO_3^- to CO_2 which is subsequently used by RuBisCO in the carbon fixation reaction. Positively charged pores promote the diffusion of negatively charged HCO_3^- into the compartment while the escape of uncharged CO_2 intermediates is not favored. This leads to an accumulation of CO_2 around RuBisCO, which enhances the carbon fixation and reduces the futile fixation of oxygen. This mechanism is necessary since the cellular membrane presents no diffusion barrier for small uncharged molecules like CO_2 and, thus, an accumulation in the cytosol is not possible. Additionally, to further promote carbon fixation in the carboxysome, the negatively charged ion HCO_3^-, which does not diffuse through the cell membrane, is actively accumulated inside the cytosol by transport mechanisms and active conversion of CO_2 to HCO_3^- [63, 64, 62]. Quantitative models confirmed the ability of the described mechanism to concentrate CO_2 inside the carboxysome and around RuBisCO [65, 66, 67].

Besides carbon fixation, bacterial microcompartments were also found to play crucial roles in

the degradation of 1,2-propanediol (Pdu microcompartment) in enteric bacteria and ethanolamine (Eut microcompartment) in mammalian gut bacteria [68, 69]. The Pdu microcompartment is essential for the catabolism of 1,2-propanediol, a product from plant sugars like fucose and rhamnose [52]. 1,2-propanediol diffuses into the microcompartment where it is processed in a two-step coenzyme-B_{12}-dependent reaction to propionyl-CoA via the intermediate propionaldehyde [53]. The segregation of these reactions protects the cell from toxic propionaldehyde and ensures an efficient processing of propionaldehyde which is badly-retained by the cell membrane [70].

The Eut microcompartment fulfills a similar function in the efficient usage of ethanolamine as a carbon, nitrogen, and energy source [52]. In a process similar to the 1,2-propanediol degradation, ethanolamine transverses the shell membrane into the microcompartment where it is converted to ethanol and acetate by a series of reactions [71]. Analogous to the function of the Pdu microcompartment, the Eut microcompartment segregates the toxic and volatile intermediate acetaldehyde, thereby ensuring efficient processing while minimizing acetaldehyde toxicity [72].

Inspired by these strategies, synthetic biologists have begun to construct synthetic compartmentalization that mimics natural systems. Many of these systems have been applied to sequential reactions and shown to significantly increase the production of industrially and commercially important chemicals [73]. We will later come back to the discussion of synthetic approaches of enzyme compartmentalization.

In Chap. 3, we quantitatively study optimal compartmentalization strategies for metabolic microcompartments. In particular, we ask how to optimally distribute a given quantity of enzymes over several microcompartments of a certain size to generate maximal productivity of a two-step enzymatic reaction.

2.3.2 Non-membrane bound enzyme clusters and filaments

Many metabolic enzymes have been found to dynamically assemble into non-membrane bound clusters and filaments under nutrient starvation or other external stimuli [74, 75]. The assembly into these subcellular structures is intimately connected with the regulation of key metabolic processes. One strategy proposed to upregulate pathway activity without directed intermediate transfer is the colocalization of sequential enzymes into clusters (see Chap. 4 and 5). Intermediates produced by upstream enzymes in the clusters are rapidly consumed by downstream enzymes due to the high local enzyme and intermediate concentrations inside the cluster.

For example, in HeLa cells, the enzymes performing the de novo synthesis of purine were found to colocalize into clusters, called purinosomes (see Fig. 2.2c). Purinosome formation was induced in purine depleted medium while replenishing the medium with purine resulted in a dissolution of purinosomes [76]. Additionally, the presence of purinosomes was found to be associated with elevated levels of purine metabolites compared to cells lacking purinosomes [77]. This indicates that purinosome formation is tightly linked to the upregulation of de novo purine biosynthesis [45, 78, 79]. Biochemical studies suggest that three enzymes involved in the first half of the ten-step de novo purine biosynthesis form the core structure of the

purinosome, while the enzymes processing the other half of the pathway are dynamically associated with the core complex via protein-protein interactions [78, 80, 81]. Notably, the knockout of any purinosome enzyme led to the disruption of purinosome assembly and resulted in an accumulation of the substrate of the knocked out enzyme [82]. Also crucial for the formation of functional purinosomes is the microtubule network which additionally controls the spatial distribution of purinosomes inside the cell [45].

Interestingly, the sequestration of a single purinosome enzyme was recently proposed as a downregulation mechanism of de novo purine biosynthesis. It was demonstrated that in the absence of purinosomes a certain level of metabolic activity is maintained [77, 45]. A strategy by which this basal activity could be downregulated is the sequestration of one of the core enzymes into clusters [83]. Collectively, this indicates that two spatial strategies regulate de novo purine biosynthesis. First, the coclustering of all pathway enzymes which leads to an upregulation of the pathway activity and second, the sequestration of a single enzyme which leads to a downregulation of the pathway activity (see Fig. 2.1b bottom). In Chap. 6, we quantitatively characterize the ability of these two spatial strategies to regulate metabolic fluxes.

The enzymes involved in de novo pyrimidine biosynthesis also show a certain degree of spatial organization. Although no evidence has been presented to date for a multi-enzyme cluster catalyzing all nine steps of de novo pyrimidine biosynthesis, certain enzymes of the pathway have been found to form aggregates [78, 84]. For example, the multifunctional enzyme catalyzing the first three steps of pyrimidine biosynthesis, CAD (carbamoyl-phosphate synthetase 2, aspartate transcarbamylase, and dihydroorotase), has been found to form oligomers within mammalian cells. Phosphorylation of CAD induced by mammalian target of rapamycin complex 1 (mTORC1) through S6 kinase promoted oligomerization and resulted in an upregulation of de novo synthesis of pyrimidines [85].

Furthermore, the enzyme cytidine triphosphate synthase (CTPS) which catalyzes the final and rate-limiting step of pyrimidine de novo biosynthesis [86, 87, 88] has been found to form filamentous structures in bacteria [89], yeast [90], Drosophila [91], and human cells [92]. The presence of CTPS filaments across different species suggests that polymerization of CTPS into subcellular structures has an important biological function [93]. CTPS filaments range in size from $\sim 0.4\,\mu m$ in bacteria to $\sim 2\,\mu m$-$5\,\mu m$ in drosophila [89, 94]. The assembly process of these large structures is divided into several phases of fusion and aggregation. In the first phase several spherical CTPS clusters form simultaneously in the cytoplasm. These clusters then assemble into small filaments which subsequently fuse to middle-sized filaments by actively sliding towards each other. This directed movement suggests that cytoskeletal structures are involved in guiding the assembly process. In the last phase, middle-sized filaments aggregate to long and thick filamentous structures which sometimes bend to C or ring-shaped geometries [95, 96]. The assembly of CTPS filaments is induced by CTPS overexpression while deprivation of the CTPS-levels led to filament disassembly [94, 97, 95, 98]. Besides this, various other stimuli that control filament formation have been reported, which appear to be highly organism dependent. These stimuli include nutrient depletion, inhibition of nucleotide or one-carbon metabolic pathways, and the differential regulation of the transcription factor Myc [84, 99, 97, 100, 101]. In contrast to previously discussed active enzyme assemblies which promote intermediate channeling and enhanced pathway activity, polymerization of CTPS was found to inhibit enzymatic activity [102, 103]. Hence, the

assembly and disassembly of CTPS filaments could function as a strategy to rapidly regulate CTPS activity. Storing inactive enzymes in the filaments and only releasing required amounts may present a mechanism to fine-tune metabolic regulation [104]. In addition to this putative function, filamentation could also serve as a stabilization mechanism for CTPS preventing it from degradation by proteases [96]. Furthermore, filaments may have a cytoskeleton-like function in bacteria helping to maintain cell shape [105].

Similar to CTPS the rate-limiting enzyme of de novo guanine biosynthesis, inosine monophosphate dehydrogenase (IMPDH), has been found to form filaments when cells were treated with the inhibitor mycophenolic acid (MPA), MgATP, or ribavirin [106, 107, 92]. Interestingly, CTPS and IMPDH form two independent filaments which sometimes were found to colocalize [108, 109]. However, the function of this partial colocalization and the coordination of filament formation is not well understood. Additionally, in contrast to CTPS, the enzyme activity of IMPDH is upregulated by filament assembly [109]. Collectively, filamentation of key enzymes in de novo nucleotide biosynthesis may present a sophisticated strategy to fine-tune the regulation of metabolic processes. Yet there are many open questions concerning regulation mechanisms and assembly procedures of these subcellular structures [96].

Recently, multi-enzyme clusters were found to also play an important role in the regulation of glycolysis. Glycolysis is an evolutionarily conserved pathway that generates energy in form of ATP by degrading glucose into pyruvate. Several glycolytic enzymes were reported to spatially organize into non-membrane bound clusters under hypoxic or high energy demanding conditions in yeast, *C. Elegans* neurons, and mammalian red blood cells [110, 111, 112, 113]. In yeast, The presence of these clusters was associated with increased glucose consumption and reduced levels of glycolytic intermediates. Cells incapable of cluster formation showed defective cell proliferation under oxygen deprived conditions [110]. This suggests that coclustering of glycolytic enzymes enables efficient processing of intermediates and thereby upregulates glycolysis activity. Furthermore, in *C. Elegans* neurons under energy stress, glycolytic enzymes dynamically colocalized near presynaptic sites into multi-enzyme clusters. The formation of these clusters has been found to be crucial for proper synaptic function and locomotion under high energy demanding conditions. This suggests that energy demands are locally met by coclustering glycolytic enzymes [112]. Interestingly, in human cancer cells, enzymes involved in glycolysis as well as gluconeogenesis, were found to organize spatially. It was proposed that the dynamic organization of enzymes into clusters is a mechanism to regulate the flux and allocation of glucose derived pathway intermediates [114]. Despite the ubiquitous presence of glycolytic enzyme clusters and their putative role in regulating glucose metabolism, the assembly mechanisms of enzyme clusters remain poorly understood. Recently, it was shown that glycolytic enzymes follow the gradient of their specific substrate [115]. This phenomenon was proposed to induce aggregation of pathway enzymes. A cluster of upstream enzymes creates a locally high concentration of intermediates which directs the motion of downstream enzymes towards the cluster.

2.4 Synthetic approaches for engineering spatial enzyme organization

Inspired by the natural strategies of spatial enzyme organization, synthetic biologist follow various approaches to spatially arrange enzymes with the aim to enhance pathway activity. In these approaches, synthetic or natural scaffolds and compartments are used to position sequential enzymes into proximity in order to enhance the efficiency of intermediate transfer between the enzymes. Increasing production rates of complicated biochemical pathways are of high interest for the synthesis of pharmaceutically useful compounds or biofuels. In the following we discuss some of the endeavors to engineer the spatial organization of enzymatic pathways.

2.4.1 Scaffold based strategies

A common approach to control the spatial organization of enzymes is to use natural or synthetic scaffolds onto which the enzymes are positioned. The scaffolds can be constructed from proteins or nucleic acids where certain functionalized domains allow for binding of the enzymes to the scaffold. The ability to construct large scale structures from nucleic acids allowed synthetic biologist to arrange consecutive enzymes on DNA or RNA-based scaffolds. In several studies, the consecutive enzymes glucose oxidase (GOx) and horseradish peroxidase (HRP) functionalized with DNA oligonucleotides were tethered to DNA-based scaffolds. In one example, hexagon-like DNA strips were designed where overhanging 'hinges' allowed for binding of the functionalized enzymes to the DNA structure. The relative localization of the two enzymes was controlled by using either two- or four-hexagon DNA strips leading to 13 nm or 33 nm inter-enzyme separation. A ∼1.2-fold higher overall activity was measured for the two-hexagon structure compared to the four-hexagon structure (see Fig. 2.3a). This was explained by the different distances intermediates have to diffuse to reach the next enzyme [116]. Using the same enzyme pair, a two dimensional DNA origami scaffold was used to probe the overall pathway activity as a function of the inter-enzyme distance. An almost 20-fold activity enhancement was observed when the enzymes were positioned at 10 nm separation. At larger distances, between 20 nm and 65 nm interenzyme separation, the activity dropped to ∼3 fold enhancement compared to the free enzyme case [117] (see Fig. 2.3b). In another study, the flat DNA origami sheet equipped with the enzyme pair was rolled into a tube confining the enzymes. The confinement of intermediates inside the tube led to a further enhancement of the activity compared to enzymes on flat DNA origami tiles [120, 121]. In a more elaborate approach, a swinging arm was introduced between two consecutive dehydrogenases colocalized on a DNA nanostructure. The swinging arm mediated efficient hydride transfer between the enzymes resulting in an almost 100-fold activity enhancement as compared to the assembly lacking the swinging arm. Moreover, adding more upstream enzymes with swinging arms around the downstream enzyme further increased the activity [122]. Unfortunately, these approaches are not well suited for *in vivo* application. The formation of DNA nanostructures and the conjugation of proteins to these structures is still hard to achieve *in vivo*. Furthermore, the binding of single-stranded oligonucleotides to enzymes may lead to a loss of enzymatic activity. To circumvent some of these challenges, specific nucleic acid binding proteins were used to arrange enzymes on DNA. By fusing the enzymes to zinc finger motifs,

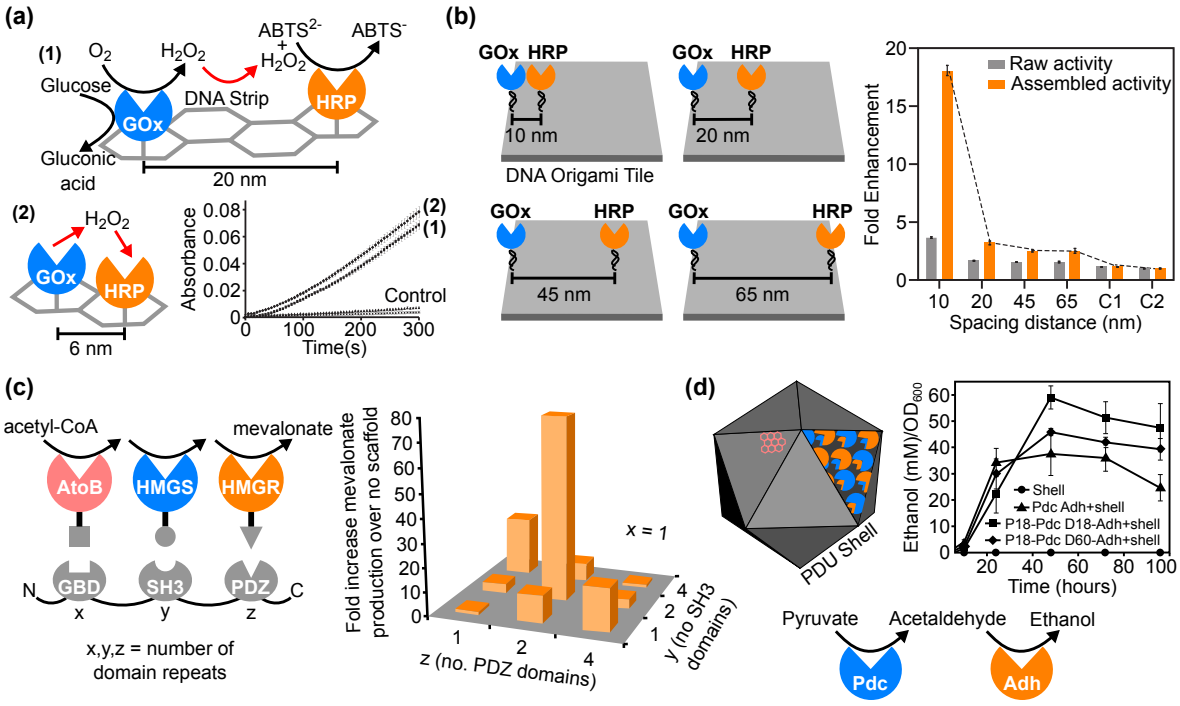

Figure 2.3: Synthetic approaches for controlling the arrangement of enzymes. **(a)** Assembly of GOx and HRP on hexagonal DNA strips. The enzymes can be positioned either 6 nm or 20 nm apart from each other. The plot shows the change of absorbance over time resulting from the oxidation of $ABTS^{2-}$ by the GOx–HRP pair for (1) a four hexagon scaffold, (2) a two hexagon scaffold, and in the absence of any DNA and in the presence of non-functional DNA (Control). Figure adapted with permission from [116]. **(b)** Colocalization of GOx/HRP pairs on DNA origami tiles. The interenzyme distance was changed from 10 nm to 65 nm. The plot shows the activity enhancement of the enzyme pairs on DNA origami scaffolds compared to free enzymes in solution. The raw activity, not corrected for the yield of the completely assembled nanostructures, is plotted in gray, and the yield corrected activity is plotted in orange. Figure adapted with permission from [117]. **(c)** Synthetic protein scaffold for spatial enzyme organization. The protein scaffold was constructed using three protein-protein interaction domains (GBD, SH3, and PDZ) connected by flexible nine-residue glycine serine linkers. The repeats of each domain are represented by x (no. GBD), y (no SH3), and z(no PDZ). Different combinations of numbers of repeats exhibited dramatic differences in mevalonate product titers. The Optimal combination of repeats (x=1, y=2, z=2) resulted in a 77-fold increase of product titer compared to the non-scaffolded pathway (x=0, y=0, z=0). Figure adapted with permission from [118]. **(d)** Encapsulation of an ethanol producing pathway into the PDU microcompartment shell. The enzymes pyruvate decarboxylase (Pdc) and Alcohol dehydrogenase (Adh), which sequentially convert pyruvate to ethanol via the intermediate Acetaldehyde are targeted to the microcompartment shell by fusing the enzymes to targeting peptides of the native system. An increase of ethanol production was observed for compartmentalized enzymes. Figure adapted with permission from [119].

which recognize and bind to specific 9-18 base pair sequences of double-stranded DNA, sequential enzymes were arranged to rationally designed DNA scaffold in *Escherichia coli*. Cells with spatially arranged enzymes yielded five fold larger product titers compared to cells with free enzymes [123]. In another *in vivo* approach RNA structures with distinct protein-docking sites were used to organize the sequential enzymes [FeFe]-hydrogenase and ferredoxin which catalyze the reduction of protons to hydrogen through electron transfer. The enhancement

of hydrogen production was measured for different scaffold architectures. Enzymes arranged on two-dimensional structures yielded an order of magnitude higher enhancement compared to enzymes on one-dimensional structures [124].

A recent study challenges the picture that spatial proximity of sequential enzymes colocalized on nucleic acid based scaffolds enhances pathway activity by the reduced diffusive distance between the enzymes [125, 126]. In this study, the activity enhancing effect was ascribed to the increased enzymatic activity close to the DNA nanostructures induced by the locally reduced pH around DNA. Hence, it remains to be understood how and when the spatial proximity of enzymes can enhance pathway activities and when the effects need to be attributed to certain factors that change intrinsic enzymatic efficiencies.

In addition to nucleic acids, also proteins can be used to construct scaffolds for spatial enzyme organization. Instead of rationally designing such protein structures, researchers often repurpose naturally occurring protein scaffolds of large enzyme complexes. For example, the scaffold of the cellulosome enzyme complex is utilized for a modular spatial organization of various enzymatic pathways. Cellulosomes are large enzyme complexes in anaerobic cellulolytic microorganisms responsible for the degradation of insoluble cellulose. These complexes are localized at the outer surface of the bacteria and the proximity of multiple enzymes in the complexes increases their synergistic catalytic activity [127, 128]. The scaffold of the cellulosome is composed of several cohesin-dockerin pairs, where the cohesins are integrated into a large glycoprotein. The dockerin domain of the cellulases and xylanases then bind to the cohesins to form the enzyme complex [129]. Additionally, cellulose-binding modules (CBMs) attach the cellulosome to cellulose and, thus, localize the multi-enzyme complex in proximity to its substrate [130]. Hence, the cellulosome serves two functions, it colocalizes cellulases and other enzymes for an efficient breakdown of cellulose and simultaneously attaches the bacterial cell to the plant cell wall giving the cell direct access to degradation products. Using orthogonal cohesin-dockerin pairs from different species allowed for a rational design of synthetic mini-cellulosomes in different species [131, 132, 133, 134]. In yeast cells heterologously expressed mini cellulosomes localized to the cell surface and a combined cellulose hydrolysis and ethanol production pathway yielded almost three fold more ethanol than compared to free enzymes [133]. The modular nature of the orthogonal cohesin-dockerin pairs also enables researchers to assemble heterologous enzymatic pathways to the cellulosome scaffold. This was demonstrated by organizing three dehydrogenases catalyzing the oxidation of methanol to carbon dioxide. The assembly of all three dehydrogenases to the scaffold resulted in a 5.1-fold increase in methanol consumption compared to unassembled enzymes [135].

In another approach, protein scaffolds were designed using three different protein-protein interaction domains: The GTPase binding domain (GBD) from the actin polymerization switch N-WASP, the Src homology three domain (SH3) from the adaptor protein CRK, and the PSD95/DlgA/Zo-1 (PDZ) domain from the adaptor protein syntrophin. The interaction domains were combined using flexible nine-residues glycine-serine linkers (see Fig. 2.3c). Using this scaffold the three enzymes catalyzing the production of mevalonate from acetyl-CoA were colocalized via small ligands attached to the enzymes that bind to the interaction domains. Varying the number of repeats of the different interaction domains provided control over the ratio of the individual enzymes colocalized to the assembly. For the optimal stoichiometry, the enzyme assembly on the scaffold achieved a 77-fold enhancement of product titer compared to free enzymes [118]. To demonstrate the modular nature of the scaffold design, the enzymes of

another pathway were colocalized using the same scaffold. For a synthetic pathway producing glucaric acid, the spatial organization of enzymes on the scaffold improved product titers by 3-5 fold over the non-scaffolded control [118, 136].

2.4.2 Direct enzyme conjugation

Instead of using scaffolds for enzyme colocalization, metabolic enzymes can also be brought into proximity by direct conjugation. This can be achieved by forming fusion proteins where small polypeptide linkers connect the enzymes. In a first attempt to form chimeric proteins consisting of two consecutive enzymes, the enzymes beta-galactosidase (LacZ) and galactokinase (GalK) which sequentially convert lactose to galactose-1-phosphate were genetically fused together. The resulting fusion protein showed the enzymatic activity of both enzymes however with considerably reduced individual activities [137]. In a follow up to this study, LacZ was fused to galactose dehydrogenase (GalDH) for the two-step production of galactonolactone from lactose. In this instance, the bifunctional enzyme displayed kinetic advantages over the identical native system in the conversion of lactose to galactonolactone, especially at low substrate concentrations. The Michaelis-Menten constant, K_M, for lactose was decreased by two-fold and the transient time was 2-4 fold shorter [138, 139]. In another example, a fusion protein was expressed in yeast comprising the enzymes farnesyl diphosphate synthase (FPPS) from yeast and patchoulol synthase (PTS) from patchouli, which sequentially produce patchoulol. The fusion of these enzymes increased patchoulol production up to 2-fold [140]. Moreover, it was recently reported that the orientation of active sites relative to each other has a strong impact on the activity improvement of fused enzymes. When the active sites were directed to each other a four-fold higher efficiency enhancement was observed compared to the case where the active sites were directed away from each other [141].

However, in recent years, the view has been challenged that the spatial proximity of consecutive active sites in these fusion proteins enables efficient processing of intermediates without directed transfer [142, 125]. Instead it was suggested that efficient processing of intermediates could only be achieved if several active sites of the sequential enzymes are colocalized in clusters or compartments. Thus the major disadvantage of the enzyme fusion approach is that only a pair of enzymes can be brought into proximity. Furthermore, fusing enzymes together often leads to a reduction of the intrinsic enzymatic efficiency of the individual enzymes caused by incorrect folding of the proteins.

In Chap. 5, we will quantify under which conditions the spatial proximity of enzymes in pair arrangements can enhance pathway fluxes and when other spatial arrangements are desirable.

2.4.3 Compartmentalization

The natural strategy of confining sequential enzymatic reactions inside compartments inspired synthetic biologists to engineer metabolic compartmentalization to improve pathway activity and to protect host organisms from potentially toxic intermediates. Compartmentalization is achieved using a variety of natural and synthetic compartments such as eukaryotic organelles, bacterial microcompartments, viral capsids, and polymersomes. Several eukaryotic organelles

were successfully repurposed for the compartmentalization of heterologous enzymes. Avalos *et al.* engineered the mitochondrion to contain all five biosynthetic enzymes catalyzing isobutanol in *Saccharomyces cerevisiae*. In wild type cells the first three enzymes of this pathway are located in the mitochondrion while the two downstream enzymes are only present in the cytosol. By targeting all pathway enzymes to the mitochondrion a 3.25-fold enhancement of isobutanol production was achieved compared to wild type cells. It was concluded that the colocalization of the entire pathway inside the mitochondrion enabled efficient processing of the intermediate, α-ketoisovalerate, which otherwise would need to diffuse out of the mitochondrion to be available for further processing to isobutanol by the two downstream enzymes [143].

Furthermore, the peroxisome was used to compartmentalize exogenous pathways. Using an optimized Peroxisomal Targeting Sequence (PTS1), DeLoache *et al.* achieved rapid targeting of the three consecutive enzymes (VioA, VioB, and VioE), which convert tryptophan to the green pigment prodeoxyviolacein (PDV), to the peroxisome. The substrate tryptophan, the first intermediate IPA imine, and the first enzyme VioA easily traverse the peroxisomal membrane while the second intermediate, IPA imine dimer, and the last two enzymes do not cross the membrane and stay confined in the peroxisome. By exploiting this confinement, it was demonstrated that compartmentalizing the last two enzymes within the peroxisome generated a 35% enhancement of PDV production compared to cells where the enzymes were in the cytoplasm. This enhancement was generated when the last enzyme was limiting, indicating that compartmentalization enabled efficient processing of IPA imine dimers which also spontaneously degrades to the dead-end side product chromopyrrolic acid (CPA) [144]. Moreover targeting exogenous enzymes to the peroxisome was shown to increase the production of fatty-acid-derived fatty alcohols, alkanes, and olefins in both *Saccharomyces cerevisiae* and *Yarrowia lipolytica* [145, 146, 147]. These endeavors present promising directions for engineering high yield biofuel producing cells.

In addition to mitochondria and peroxisomes also chloroplasts in plant cells have been employed to compartmentalize exogenous enzymes. In a recent study, the translation machinery of chloroplasts was exploited to localize the enzymes of the mevalonate pathway (MEV) to the chloroplast lumen [148]. The chloroplast genome was transfected with the six genes encoding for the MEV enzymes and their over-expression was induced. This strategy resulted in a 10-fold increase in squalene production, which was mainly attributed to the higher concentration of pathway intermediates inside the chloroplasts [148, 149].

Despite these successful studies of targeting exogenous enzymes to eukaryotic organelles, there are still many challenges associated with this approach. For example, pathway enzymes may be inactive in the organelle interior because of unfavorable pH values or redox environments. Additionally, organelles that do not have a genome and translation machinery cannot be genetically engineered to express the exogenous enzymes inside the lumen. Hence, to target enzymes to such organelles certain tags are required which direct the incorporation of the enzymes into the organelle. In many cases incorporation can be slow and additional engineering of the targeting tags may be necessary. Finally, the substrate of the compartmentalized pathway should be able to permeate through the organelle membrane while the intermediates should stay confined. Hence compartmentalization for pathway flux enhancement is not suited for pathways consisting of metabolites that do not fulfill these requirements.

Besides targeting heterologous enzymes to eukaryotic organelles, in recent years much effort was made to reuse BMCs for compartmentalization. BMCs are well suited for bioengineering since they can be entirely expressed in heterologous systems by incorporating the BMC genes into the host genome [150, 44]. Nevertheless, to be able to target heterologous cargo to BMCs a thorough understanding of the natural assembly pathways is necessary. Most BMCs were predicted to form from the inside out, first a core of enzyme proteins forms around which the shell assembles [151, 152]. A short N-terminal sequence found at the lumen enzymes, the encapsulation peptide (EP), facilitates the aggregation of the enzymes into the core and their subsequent encapsulation by the shell proteins [153, 154, 155, 152]. Attaching EPs to heterologous proteins indeed led to an encapsulation inside the BMCs, which was verified using fluorescent proteins as cargo [156, 157, 43]. By exploiting this mechanism, exogenous enzymes were targeted to the interior of BMCs. For example, the EPs from the native enzymes of the Pdu microcompartment were used to compartmentalize the exogenous enzymes pyruvate decarboxylase and alcohol dehydrogenase into the Pdu shell generating an *in vivo* bioreactor that converts pyruvate to ethanol in *E. coli*. The strains containing this engineered BMC were not only able to produce ethanol from pyruvate but also produced elevated levels of ethanol compared to the uncompartmentalized control [119] (see Fig. 2.3d). Furthermore, the encapsulation of enzymes inside BMC shells was demonstrated to insulate enzymes from pH stress [158] and to protect cells from cytotoxic enzymes [159]. A possibility to further optimize the activity of synthetic BMCs is to tune shell permeability by modifying the residues that surround the pores of the shell. Ideally, the shell should allow for a facile permeation of substrates and final products while impeding the escape of intermediates. In preliminary studies, it has been shown that a mutation of residues surrounding the pores alters shell permeability [160, 161].

Protein-based nano-reactors can not only be engineered by repurposing BMCs but also by using viral capsids as shells for enzyme compartmentalization. It was shown that several enzymes can be simultaneously targeted to the capsid of the bacteriophage P22. A model pathway consisting of two or three consecutive enzymes was encapsulated by fusing the enzymes together and tagging the multi-enzyme fusion with a scaffold protein domain (SP). This domain is recognized by the coat proteins derived from the virus, which then leads to the compartmentalization of the multi-enzyme gene product [162]. However, encapsulation of the pathway did not display any kinetic advantages [162]. Using the same approach it was shown that encapsulation of enzymes into viral capsids protects the enzymes from proteases and stress conditions [163]

3 Optimal compartmentalization strategies for metabolic compartments *

The compartmentalization of metabolic pathways is a ubiquitous strategy used by cells to alleviate challenges such as toxic pathway intermediates, competing metabolic reactions, and slow reaction rates (see Sec. 2.1). Inspired by nature, synthetic biologists seek to engineer enzyme compartmentalization within vesicles or proteinaceous shells to enhance the yield of industrially and pharmaceutically valuable products. Although enzymatic compartments have been extensively studied experimentally, a quantitative understanding of the underlying design principles is still lacking.

In this chapter, we use a theoretical approach to elucidate optimal design principles for metabolic microcompartments that apply both to natural and synthetic system. In particular, we study how the size and enzymatic composition of compartments should be chosen so as to maximize the productivity of a model metabolic pathway. By analyzing a simple reaction-diffusion model, we find that maximizing productivity requires compartments larger than a critical size, where the enzyme density within each compartment should be tuned according to a power-law scaling in the compartment size. This behavior will be explained using an analytically solvable, well-mixed approximation. In regimes where this approximation breaks down, qualitatively similar behaviors are observed. Our results suggest that the experimentally observed different sizes and enzyme packings of α- and β-carboxysomes each constitute an optimal compartmentalization strategy given the property of their respective protein shells.

3.1 Introduction

We have seen in the previous chapter (Chap. 2) that metabolic processes in living organisms are often coordinated in space. This spatial organization is believed to have evolved as a strategy to overcome certain challenges inherent to intracellular metabolism. Intermediate products of a pathway can be toxic for the cell, poorly retained by the cell membrane or undergo undesired side reactions [70, 47]. Furthermore, enzymes can be inefficient or have a low specificity for the desired reaction [165] (see Sec. 2.1). These limitations can be alleviated by spatially orchestrating metabolic reactions. The mechanisms by which this is achieved are diverse, ranging from direct enzyme conjugation and enzyme localization on scaffolds to spatial confinement of enzymes within subcellular microcompartments.

*This chapter is adapted from the publication: Optimal Compartmentalization Strategies for Metabolic Microcompartments, by F. Hinzpeter, U. Gerland, and F. Tostevin, published in *Biophysical Journal* in 2017. See also [164].

In this chapter the focus is on metabolic reactions taking place within subcellular microcompartments. The compartmentalization of metabolic processes into organelles is a hallmark of eukaryotic cells. Examples include peroxisomes, glycosomes, and mitochondria. The discovery of bacterial microcompartments (BMC) demonstrated that metabolic compartments also exist in prokaryotes [166, 167]. In contrast to the lipid membranes that typically define eukaryotic organelles, BMCs have protein shells reminiscent of viral capsids. Small pores in the shell allow metabolites, but not enzymes, to pass into and out of the compartment [57].

The best-studied BMC, the carboxysome, is a key element of so-called "carbon concentrating mechanisms" that are crucial for efficient carbon fixation by cyanobacteria [168]. Carbon fixation in the Calvin cycle relies on the enzyme RuBisCO, which has a low turnover rate and, in addition to the fixation of carbon dioxide, also catalyzes the unproductive fixation of oxygen. RuBisCO is encapsulated within a protein shell together with carbonic anhydrase (CA). CA converts incoming bicarbonate (HCO_3^-) to CO_2 that accumulates within the compartment, thereby enhancing the carbon fixation reaction of RuBisCO and reducing the futile reaction with oxygen. Carboxysomes in different organisms are subdivided into α- and β-carboxysomes on the basis of the type of RuBisCO (IA vs. IB) present in the particular species [169]. The two classes of carboxysome also differ in their shell proteins, size and internal organization, and are thought to be the result of convergent evolution [168].

In recent years much effort has been made to engineer carbon concentrating mechanisms in plants, with the goal of increasing crop yields by enhancing the photosynthetic efficiency [170, 171, 172]. To this end, the introduction of carboxysomal RuBisCO into tobacco plants supported autotrophic growth [173]. More broadly, there is increasing interest in using synthetic enzyme compartments for efficient synthesis of industrial chemicals and biofuels [174, 73, 143]. However, despite the ubiquity of metabolic compartmentalization and its potential applications in synthetic biology, a quantitative understanding of the design principles and functional trade-offs of such compartments is still lacking.

Here, we use mathematical modeling to study how the size and enzymatic composition of compartments affect an encapsulated metabolic pathway. In particular, we ask how compartments should be constructed so as to maximize the enzymatic productivity. In the parameter regime of α-carboxysomes, we find that the optimal compartmentalization strategy is to form multiple compartments, each with at least a certain critical size, and each with less than maximal enzyme occupancy. Interestingly, the same productivity can be achieved for any compartment larger than a critical size, provided the enzyme densities are chosen appropriately. The enzyme arrangement within the compartment plays little role over the biologically-relevant range of compartment sizes. We explain this behavior with the aid of an analytically-solvable model in which metabolites are well mixed within the compartment. We also characterize the qualitatively different compartmentalization strategies that emerge outside the well-mixed regime. The results suggest that the different sizes and packing densities of α- and β-carboxysomes each represent an optimal strategy given the structures of their corresponding shell proteins.

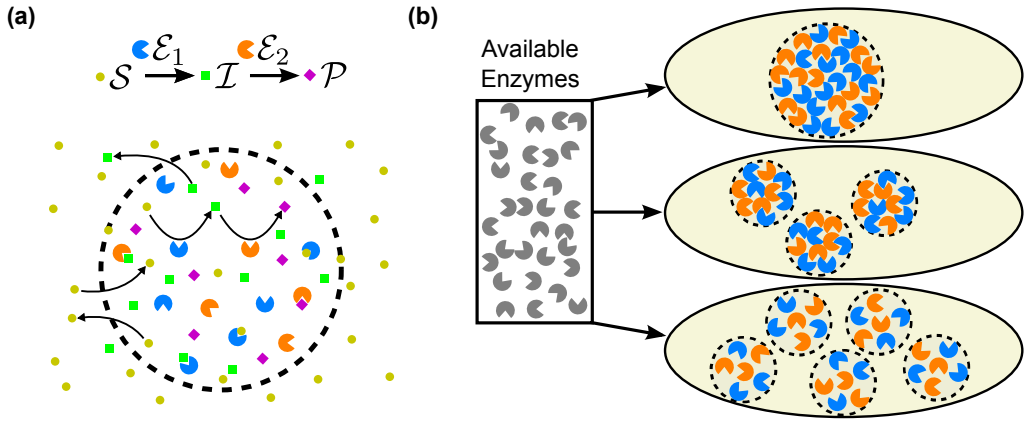

Figure 3.1: Illustration of the compartmentalization strategies for a two-step pathway. (a) Model two-step enzymatic pathway with enzymes contained within a microcompartment. The permeable compartment shell allows for the exchange of metabolites $\mathcal{S}$ and $\mathcal{I}$ with the cytoplasm. **(b)** A fixed number of enzymes could be distributed according to many different compartmentalization strategies, each characterized by a particular compartment size, enzyme density and ratio of $\mathcal{E}_1$ to $\mathcal{E}_2$ enzymes. Each such strategy would lead to a different pathway flux, and therefore a different productivity.

3.2 Model

Let us consider a subcellular compartment as shown schematically in Fig. 3.1a, consisting of a selectively permeable shell that encapsulates enzymes of two types, $\mathcal{E}_1$ and $\mathcal{E}_2$. This shell may be composed of proteins, like in bacterial microcompartments, or a lipid bilayer, as is typical of eukaryotic organelles. Our model two-step pathway has the reaction scheme $\mathcal{S} \xrightarrow{\mathcal{E}_1} \mathcal{I} \xrightarrow{\mathcal{E}_2} \mathcal{P}$. Substrate molecules $\mathcal{S}$ enter the compartment from the cytoplasm through the boundary shell, and are converted by enzyme $\mathcal{E}_1$ into intermediate product $\mathcal{I}$, which is converted into final product $\mathcal{P}$ by enzyme $\mathcal{E}_2$. However, both $\mathcal{S}$ and $\mathcal{I}$ can also be lost from the compartment through the shell before reacting further. In principle, intermediate could also be unstable or consumed by side reactions within the compartment. The discussion of this scenario can be found in the SM Sec. A.6.

The total rate at which substrate $\mathcal{S}$ is converted into product $\mathcal{P}$ will depend on the number and size of compartments, the densities of the two enzymes, and the spatial arrangement of enzymes within the compartment. We refer to a particular choice for this collection of properties as a "compartmentalization strategy" (Fig. 3.1b). Which compartmentalization strategy is the optimal one? Clearly, the answer to this question must depend on the objective or "design goal" of the system, as well as potential additional constraints. The biological imperative to achieve a high production flux while limiting the expenditure of resources in enzyme production [175] suggests maximal enzymatic productivity as the design goal, where productivity is defined as the total rate of $\mathcal{P}$ production per enzyme. In this theoretical study we assume that the choice of compartmentalization strategy is otherwise not constrained in any significant way.

3.2.1 Productivity

Let us suppose that a cell contains a total of $N_{\mathcal{E}} = N_{\mathcal{E}_1} + N_{\mathcal{E}_2}$ copies of the enzymes $\mathcal{E}_1$ and $\mathcal{E}_2$, distributed equally over a number N_c of identical compartments, such that each compartment contains $E_c = N_{\mathcal{E}}/N_c$ enzymes. The total rate of production of $\mathcal{P}$ is given by $J_{\mathcal{P}} = N_c J_c$ where J_c is the rate of production per compartment. Then, the total production can be written as

$$J_{\mathcal{P}} = N_{\mathcal{E}} \frac{J_c}{E_c} \equiv N_{\mathcal{E}} \mathrm{P}, \tag{3.1}$$

where we define the productivity P as the flux of $\mathcal{P}$ production in one compartment divided by the total number of enzymes contained within the compartment. In steady-state, the productivity can also be written as $\mathrm{P} = J_{\mathrm{in}}\ \epsilon/E_c$ where J_{in} is the total influx of substrate $\mathcal{S}$ across the compartment boundary and the factor ϵ, which we term the "conversion efficiency", is the probability that an $\mathcal{S}$ molecule entering the compartment will be converted into product $\mathcal{P}$, rather than leaving the compartment in the unreacted $\mathcal{S}$ or partially-reacted $\mathcal{I}$ form.

The productivity allows us to quantify the performance of a given compartmentalization strategy, taking into account both the rate at which each compartment produces $\mathcal{P}$ and the necessary enzyme investment (how many such compartments can be constructed with the available enzymes). To isolate the effect of the compartmentalization strategy from processes happening at the scale of the cell as a whole, we will assume that the compartments do not affect one another via the cellular pools of metabolites $\mathcal{S}$ and $\mathcal{I}$, which we will take to be fixed. Under this assumption, the strategy that maximizes the productivity of a single compartment will also maximize the total production of $\mathcal{P}$ in the cell for a fixed number of enzymes $N_{\mathcal{E}}$.

3.2.2 Reaction-diffusion model

To determine the productivity, we must specify the dynamics of $\mathcal{S}$ and $\mathcal{I}$ inside the compartment. Here, the reactions of $\mathcal{S}$ with $\mathcal{E}_1$ and $\mathcal{I}$ with $\mathcal{E}_2$ is modeled using Michaelis-Menten kinetics, with $k_{\mathrm{cat}}^{(1)}$ and $K_{\mathrm{M}}^{(1)}$ the catalytic rate and the Michaelis constant of $\mathcal{E}_1$, and similarly for $\mathcal{E}_2$. The density of enzymes of type $\mathcal{E}_1$ and $\mathcal{E}_2$ at position $\mathbf{r}$ within the compartment is denoted by $e_1(\mathbf{r})$ and $e_2(\mathbf{r})$, respectively. Enzymes are taken to be fixed in position and thus the density distributions do not vary over time. However, to account for the finite size of the enzymes, a maximal packing density of enzymes, $e_1(r) + e_2(r) \leq e_{\max}$, is imposed as a constraint. Within the compartment $\mathcal{S}$ and $\mathcal{I}$ are assumed to move by diffusion, with equal diffusion constants. The concentrations of $\mathcal{S}$ and $\mathcal{I}$, $s(\mathbf{r},t)$ and $i(\mathbf{r},t)$, respectively, therefore follow the coupled reaction-diffusion equations,

$$\frac{\partial s(\mathbf{r},t)}{\partial t} = D\nabla^2 s(r,t) - \frac{k_{\mathrm{cat}}^{(1)} e_1(\mathbf{r}) s(\mathbf{r},t)}{K_{\mathrm{M}}^{(1)} + s(\mathbf{r},t)} \tag{3.2}$$

$$\frac{\partial i(\mathbf{r},t)}{\partial t} = D\nabla^2 i(\mathbf{r},t) + \frac{k_{\mathrm{cat}}^{(1)} e_1(\mathbf{r}) s(\mathbf{r},t)}{K_{\mathrm{M}}^{(1)} + s(\mathbf{r},t)} - \frac{k_{\mathrm{cat}}^{(2)} e_2(\mathbf{r}) i(\mathbf{r},t)}{K_{\mathrm{M}}^{(2)} + i(\mathbf{r},t)}. \tag{3.3}$$

The last term in Eq. 3.2 represents the conversion of $\mathcal{S}$ into $\mathcal{I}$ by $\mathcal{E}_1$, which appears as a local source of $\mathcal{I}$ in Eq. 3.3; the final term in Eq. 3.3 describes the conversion of $\mathcal{I}$ into $\mathcal{P}$ by $\mathcal{E}_2$.

For simplicity, we assume compartments that are spherical with radius R, and concentration profiles that are spherically symmetric.

Let us assume that through regulation of uptake and export rates, the cell maintains a homeostatic concentration s_0 of substrate in the cytoplasm. Intermediates, in contrast, are assumed to be poorly retained by the cell membrane or rapidly degraded in the cytosol, such that their concentration in the cytoplasm is negligible. For carboxysomes in cyanobacteria this assumption is reasonable, since detailed models of carboxysomes [176, 67] predict that active conversion of CO_2 to HCO_3^- in the cytoplasm results in a concentration of CO_2 outside the compartment about 1000-fold lower than inside. The exchange of metabolites across the compartment shell is described by the boundary conditions,

$$D\frac{\partial s(r,t)}{\partial r}\Big|_{r=R} = p_s\left[s_0 - s(R,t)\right] \tag{3.4}$$

$$D\frac{\partial i(r,t)}{\partial r}\Big|_{r=R} = -p_i i(R,t), \tag{3.5}$$

where p_s and p_i are the permeabilities of the shell to $\mathcal{S}$ and $\mathcal{I}$. In this model, a compartmentalization strategy consists of a particular choice for the radial enzyme concentration profiles $e_1(r)$, $e_2(r)$ and the compartment radius R. For each such strategy, the steady-state productivity will be

$$\mathrm{P}\big\{e_1(r), e_2(r); R\big\} = \frac{\int_0^R 4\pi r^2 \frac{k_{\mathrm{cat}}^{(2)} e_2(r) i(r)}{K_{\mathrm{M}}^{(2)} + i(r)}\,\mathrm{d}r}{\int_0^R 4\pi r^2\left[e_1(r) + e_2(r)\right]\mathrm{d}r}, \tag{3.6}$$

where $i(r)$ is the corresponding steady-state solution of Eqs. 3.2-3.5. In Eq. 3.6, the numerator is the total production flux of $\mathcal{P}$ per compartment, J_c, and the denominator is the total number of enzymes in one compartment, E_c (c.f. Eq. 3.1). In the following, we will seek the compartmentalization strategy that optimizes P by maximizing Eq. 3.6 with respect to $e_1(r)$, $e_2(r)$ and R.

3.3 Results

The aim of this study is to investigate generic features of optimal compartmentalization strategies that apply to microcompartments under both biological and synthetic conditions. We first study in detail the behavior of the model for parameters representative of one of the best-studied biological microcompartments, the α-carboxysome of cyanobacteria. Subsequently, we return to the question of how optimal compartmentalization strategies change in different parameter regimes. Additionally, the focus is first on the linear regime of the Michaelis-Menten kinetics. Later, we will see that including the full non-linear kinetics does not qualitatively alter the results of our analysis. The catalytic efficiencies of the two enzymes in the linear regime is denoted by, $\kappa_1 = k_{\mathrm{cat}}^{(1)}/K_{\mathrm{M}}^{(1)}$ and $\kappa_2 = k_{\mathrm{cat}}^{(2)}/K_{\mathrm{M}}^{(2)}$.

3.3.1 Optimal compartmentalization strategies for α-carboxysome parameters

The catalytic efficiencies of the two carboxysome enzymes, CA and RuBisCO, have been measured to be $\kappa_1 = 5\,(\mu\mathrm{M\,s})^{-1}$ and $\kappa_2 = 0.06\,(\mu\mathrm{M\,s})^{-1}$, respectively [177, 178]. Since there

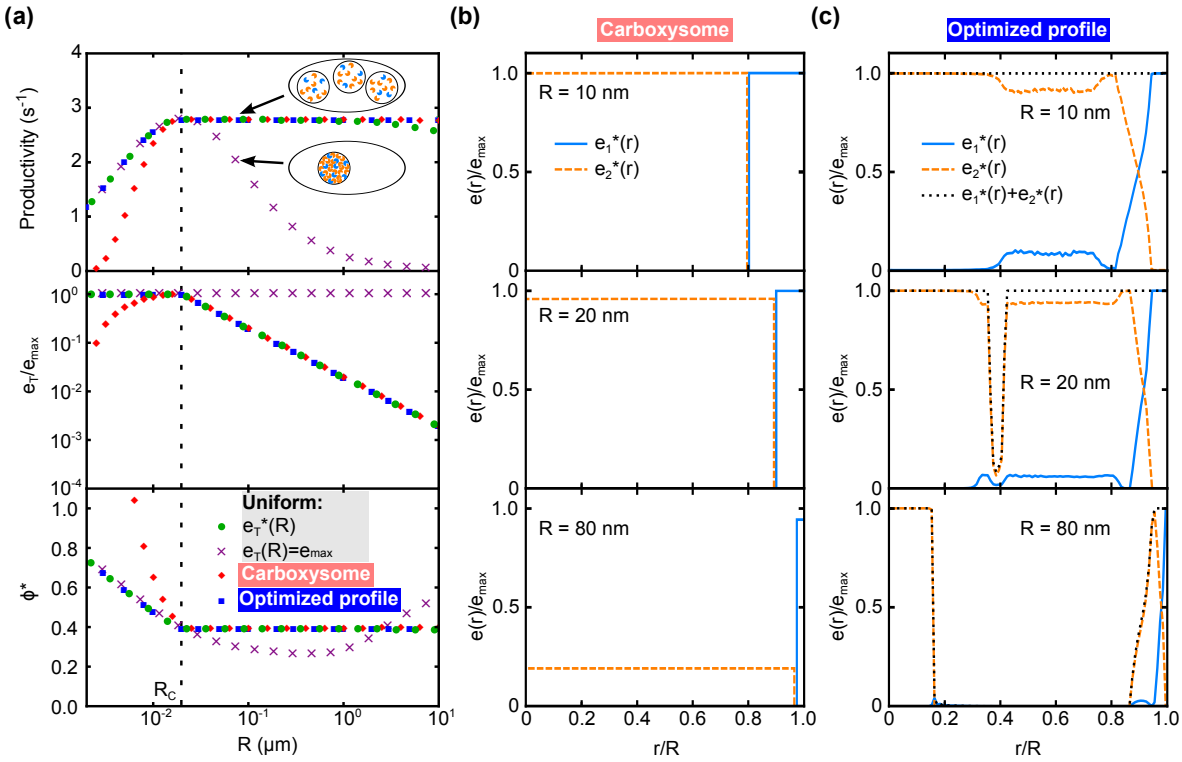

Figure 3.2: Optimal compartmentalization strategies for α-carboxysome parameters.
(a) Optimal productivity (top), total enzyme density $e_T^* = e_1^* + e_2^*$ (middle), and enzyme abundance ratio $\phi^* = e_1^*/e_2^*$ (bottom) for four different arrangements of enzymes: uniform distribution at optimal density (green circles); uniform distribution at maximal density, $e_T = e_{\max}$ (purple crosses); enzyme arrangement as observed in carboxysomes, Eq. 3.7 (red diamonds), and optimized intracompartment enzyme arrangements (blue squares) as plotted in Fig. A.1 in the SM. **(b)** Examples of the enzyme arrangements for the carboxysome configuration Eq. 3.7 (with optimized e_1 and e_2 values) for different compartment sizes. **(c)** Examples of the globally-optimized intra-compartment enzyme distribution for different compartment sizes.

are no direct measurements of permeability for the α-carboxysome shell, we estimated these parameters from the known structure of the shell proteins [179, 180] ($p_s = 90\,\mu\mathrm{m\,s^{-1}}$ and $p_i = 18\,\mu\mathrm{m\,s^{-1}}$, see SM Sec. A.1 for details). The assumed metabolite diffusion coefficient is, $D = 1000\,\mu\mathrm{m^2\,s^{-1}}$ and the maximal enzyme concentration was approximated to $e_{\max} = 25\,\mathrm{mM}$. For calculating the productivity a reference substrate concentration of $s_0 = 25\,\mu\mathrm{M}$ was assumed.

Let us first consider the scenario where both types of enzymes are uniformly distributed throughout the compartment, $e_1(r) = e_1$ and $e_2(r) = e_2$. Such a distribution may represent the case where the enzymes are freely diffusing. Under this assumption, the steady state of Eqs. 3.2-3.3 can be solved exactly to find an analytical expression for the productivity $\mathrm{P}(e_1, e_2; R)$ (see SM Sec. A.2). However, it is not possible to extract from this expression a closed form for the optimal values of e_1, e_2 and R. Therefore, we used numerical methods to optimize the productivity. The results are visualized by plotting the optimal productivity, P^*, the optimal total enzyme concentration, $e_T^* = e_1^* + e_2^*$, and the optimal ratio of enzymes, $\phi^* = e_1^*/e_2^*$, for different values of R (see Fig. 3.2a).

Examining first the optimal productivity (Fig. 3.2a top, green circles), we see that $P^*(R)$ increases with R up to a critical radius R_c. Above R_c we see a broad plateau in the value of $P^*(R)$, indicating that similar productivity can be achieved over a wide range of compartment sizes that encompasses the entire biologically-relevant range (from $\sim 10\,\text{nm}$ to $\sim 3\,\mu\text{m}$, comparable to the size of cyanobacterial cells). For extremely large compartments, the productivity decreases gradually with increasing R.

The plateau in productivity is associated with a qualitative change in the optimal enzyme concentrations (Fig. 3.2a, middle & bottom). For small compartments, $R < R_c$, maximizing productivity requires that compartments are maximally-packed with enzymes, $e_T^* = e_1^* + e_2^* = e_{\max}$. In contrast, for $R > R_c$, productivity is largest when $e_T^*(R) < e_{\max}$; the optimal density shows a power-law dependence $e_T^*(R) \sim R^{-1}$. This tuning of the total enzyme density to the compartment size is crucial for generating the plateau in productivity observed in Fig. 3.2a (top): if instead compartments are always maximally filled with enzymes then the productivity shows a pronounced peak at $R = R_c$ and rapidly decreases for $R > R_c$ (Fig. 3.2a, purple crosses). The optimal ratio of abundances of the two types of enzymes, $\phi^*(R) = e_1^*(R)/e_2^*(R)$, decreases with increasing R for $R < R_c$ before taking a constant value, ϕ_c^*, for $R > R_c$.

The optimal compartmentalization strategy for α-carboxysome parameters is therefore to produce enzymes of type $\mathcal{E}_1$ and $\mathcal{E}_2$ in the ratio ϕ_c^*, and to assign these to compartments of size $R > R_c$ such that the total enzyme density is $e_T^*(R) = e_{\max} R_c/R$. For a compartment radius of $R = 60\,\text{nm}$, the typical radius of an α-carboxysome [181], our analysis predicts that the optimal enzyme density is $\approx 30\%$ of the maximal packing density $e_{\max}$, resulting in a productivity that is approximately 30% higher than similarly-sized compartments that are maximally packed (Fig. 3.2a, top). Interestingly, α-cyanobacteria indeed contain several α-carboxysomes [182], each with just a quarter of the compartment volume occupied with RuBisCO enzymes [181]. Thus, these cyanobacteria may have evolved to form a surplus of α-carboxysomes that are not fully packed in order to optimally exploit the cooperative activity of the encapsulated CA and RuBisCO.

So far it was assumed that the enzymes are uniformly distributed throughout the compartment. However, in naturally occurring microcompartments the enzymes are in general not uniformly distributed. For example, in carboxysomes, CA is located at the inner surface of the protein shell whereas RuBisCO is distributed throughout the compartment interior [183]. Similarly, the locations of enzymes in synthetic microcompartments are often constrained by the method of enzyme encapsulation. For example, enzymes may be incorporated into the microcompartment by tethering them to shell proteins [184]. This naturally raises the question of how the enzyme arrangement inside the compartment affects the productivity.

Focusing first on the enzyme arrangement observed in carboxysomes, it is examined how the optimal compartmentalization strategy changes when $\mathcal{E}_1$ is restricted to a layer of thickness $\delta = 2\,\text{nm}$ (roughly the size of a small protein) at the compartment boundary, while $\mathcal{E}_2$ is distributed throughout the remaining volume. For that, we solved the Eqs. 3.2-3.5 with

$$e_1(r) = \begin{cases} 0 & r \leq R - \delta \\ e_1 & r > R - \delta \end{cases}, \; e_2(r) = \begin{cases} e_2 & r \leq R - \delta \\ 0 & r > R - \delta \end{cases} \tag{3.7}$$

(see SM Sec. A.3) and again optimized the productivity with respect to e_1 and e_2, subject to

the constraint $e_1, e_2 \leq e_{\max}$ (Fig. 3.2b).

For $R > R_c$, the optimal productivity and compartment-averaged enzyme densities ($e_T^*(R) = V^{-1} \int_V [e_1(r) + e_2(r)] \, \mathrm{d}^3 r$, $\phi^*(R) = \int_V e_1(r) \, \mathrm{d}^3 r / \int_V e_2(r) \, \mathrm{d}^3 r$) are almost identical to those found for uniformly distributed enzymes (Fig. 3.2a, red diamonds). Thus, the optimal compartmentalization strategy is not changed when the enzyme distribution is of the form in Eq. 3.7. For $R < R_c$, $e_T^*(R)$ and $\phi^*(R)$ differ from the optimal uniform enzyme distribution, and a lower productivity is achieved. However, this deviation arises simply because for fixed δ it is impossible to maximally pack the compartment while simultaneously maintaining a specific ratio of $\mathcal{E}_1$ to $\mathcal{E}_2$ enzymes.

Next we determined numerically the full enzyme distributions $e_1^*(r)$ and $e_2^*(r)$ that maximize the productivity, subject only to the constraint $e_1(r) + e_2(r) \leq e_{\max}$ (see SM Sec. A.4 for details). The resulting optimal profiles change with the compartment radius R and differ substantially from a uniform enzyme distribution (see Fig. 3.2c). A discussion of the optimal concentration profiles can be found in SM Sec. A.5. Notably, though, the corresponding productivity is barely increased compared to that achieved by uniformly distributed enzymes (Fig. 3.2a). Furthermore, the average densities of $\mathcal{E}_1$ and $\mathcal{E}_2$ over the entire compartment are the same as for uniform enzymes.

Together these results show that, for the chosen parameters, the precise arrangement of enzymes within the compartment does not significantly affect the productivity. What is most important is rather to ensure that the average density of enzymes within compartments is appropriately chosen according to their size.

3.3.2 Well-mixed approximation

The observation that the intra-compartment enzyme arrangement has little impact on the productivity suggests that the spatial distribution of metabolites within the compartment is rather uniform. Indeed, we found that for $R \lesssim 1\,\mu\mathrm{m}$, where the productivities for different enzyme arrangements coincide, the densities $s(r)$ and $i(r)$ vary within the compartment by at most $\sim 10\%$ and $\sim 1\%$, respectively. To understand for which parameter regimes such a uniform distribution of metabolites will apply, it is instructive to consider the various timescales that characterize the behavior of the system. Diffusive mixing of the metabolites throughout the compartment occurs on a timescale $\tau_D \sim R^2/D$. The reactions of metabolites with enzymes occur on the timescales $\tau_1 = 1/(\kappa_1 e_1)$ and $\tau_2 = 1/(\kappa_2 e_2)$. Finally, we identify two timescales associated with the exchange of metabolites across the compartment boundary, $\tau_s \sim R/p_s$ and $\tau_i \sim R/p_i$. The R-dependence of these latter timescales can be understood by considering the rate of exchange across the boundary to consist of the intrinsic (R-independent) crossing rate for a molecule that is located at the boundary, multiplied by the fraction of the compartment volume that can be considered close to the boundary, which decreases as R^{-1}.

Metabolites within the compartment can be considered well mixed when the diffusive timescale is short compared to those of reaction and exchange, $\tau_D \ll \tau_{1,2}, \tau_{s,i}$. In this limit, metabolites effectively sample the entire pool of enzymes before they leak through the boundary or react. It is then plausible that the pathway flux does not depend on the precise arrangement of

enzymes within the compartment. From the R-dependence of the different timescales, we can see that τ_D will be smaller than the other timescales when R is sufficiently small. For the parameters and optimal enzyme concentrations of Fig. 3.2, it can be estimated that this will be the case for $R \leq 1.4\,\mu\text{m}$, which is much larger than the typical size of BMCs and the critical radius R_c. Hence, assuming that the contents of BMCs are well mixed appears to be a good approximation for biologically-relevant compartment sizes.

To gain further insight into the optimal compartmentalization strategy, a simplified version of the previously discussed model is analyzed, where the reactants $\mathcal{S}$ and $\mathcal{I}$ are assumed to be well mixed within the compartment. Supposing that the reactants flowing across the boundary are rapidly distributed throughout the compartment, the steady-state equations for s and i become

$$0 = 4\pi R^2 p_s(s_0 - s) - \frac{4}{3}\pi R^3 \kappa_1 e_1 s \tag{3.8}$$

$$0 = 4\pi R^2 p_i(-i) + \frac{4}{3}\pi R^3 \left(\kappa_1 e_1 s - \kappa_2 e_2 i\right), \tag{3.9}$$

where the first term in each equation describes the net transport flux across the compartment boundary, and the second term corresponds to the reaction flux. This simplified model can now be compared to the results in Fig. 3.2. Solving Eqs. 3.8-3.9 for the productivity as a function of the total enzyme concentration, $e_T = e_1 + e_2$, and the ratio of enzyme densities, $\phi = e_1/e_2$ we obtain,

$$\mathrm{P}\big(e_T, \phi, R\big) = \frac{3p_s s_0}{e_T R} \underbrace{\left[\frac{\phi e_T R}{\frac{3p_s}{\kappa_1}(1+\phi) + \phi e_T R}\right]\left[\frac{e_T R}{\frac{3p_i}{\kappa_2}(1+\phi) + e_T R}\right]}_{\epsilon(e_T,\phi,R)}. \tag{3.10}$$

The total enzyme density e_T and radius R appear in the productivity only as the product $e_T R$, which immediately implies a scaling $e_T^*(R) \sim R^{-1}$. From Eqs. 3.8-3.10 we see that this scaling originates in the need to balance metabolite transport across the compartment boundary, which depends on the surface area $j_{s,i} \propto p_{s,i} A \sim R^2$, against the flux of enzymatic reactions, which occur throughout the compartment volume $j_{1,2} \propto \kappa_{1,2} e_{1,2} V \sim R^3$. By choosing $e_{1,2} \sim A/V \sim R^{-1}$, the ratio of transport to reaction fluxes can be held fixed at the particular level that achieves maximal productivity. However, since the enzyme density e_T cannot exceed $e_{\max}$, this scaling cannot be satisfied for all R.

Maximizing the productivity with the constraint $e_T(R) \leq e_{\max}$, we obtain

$$e^*_{T,\mathrm{wm}}(R) = e_{\max}\begin{cases} 1 & R \leq R_{c,\mathrm{wm}} \\ R_{c,\mathrm{wm}}/R & R > R_{c,\mathrm{wm}} \end{cases}, \quad \phi^*_{\mathrm{wm}}(R) = \begin{cases} \left[\frac{1+R/\lambda_2}{1+R/\lambda_1}\right]^{1/2} & R \leq R_{c,\mathrm{wm}} \\ \left[\frac{\lambda_1}{\lambda_2}\right]^{1/3} & R > R_{c,\mathrm{wm}} \end{cases}. \tag{3.11}$$

where the critical radius $R_{c,\mathrm{wm}} = \lambda_1^{2/3}\lambda_2^{1/3} + \lambda_2^{2/3}\lambda_1^{1/3}$ and the length scale parameters $\lambda_1 = 3p_s/(\kappa_1 e_{\max})$ and $\lambda_2 = 3p_i/(\kappa_2 e_{\max})$ correspond to the compartment radii at which the maximal turnover of substrate/intermediate via enzymatic reaction equals the rate of loss across the compartment boundary. The optimal productivity is then

$$\mathrm{P}^*_{\mathrm{wm}}(R) = \begin{cases} \frac{\gamma\lambda_1\lambda_2}{R(R+\lambda_1+\lambda_2)^2}\left[\left\{1 + \frac{R(R+\lambda_1+\lambda_2)}{\lambda_1\lambda_2}\right\}^{1/2} - 1\right]^2 & R \leq R_{c,\mathrm{wm}} \\ \frac{\gamma}{\left[\lambda_1^{1/3}+\lambda_2^{1/3}\right]^3} & R > R_{c,\mathrm{wm}} \end{cases}, \tag{3.12}$$

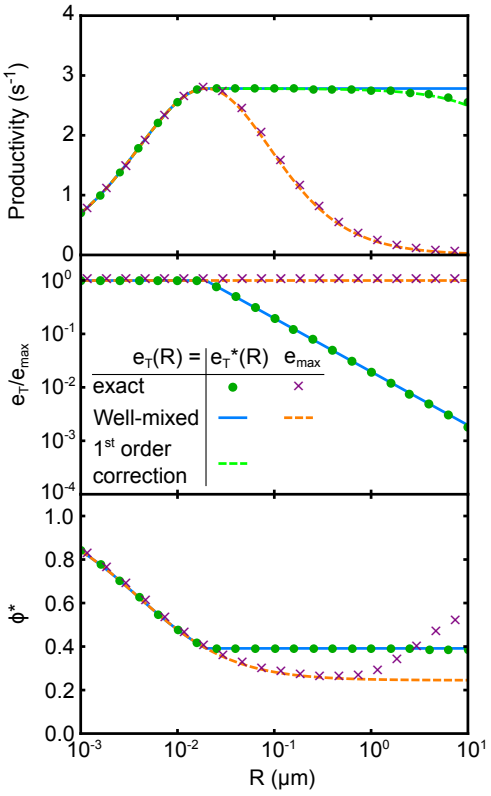

Figure 3.3: Results of the well-mixed approximation. (Top) Optimal productivity, (middle) enzyme density, e_T^*, and (bottom) ratio of $\mathcal{E}_1$ to $\mathcal{E}_2$ enzymes, ϕ^*. Data points show numerical optimization of the analytical expression for the productivity for uniform enzymes, as in Fig. 3.2a. Blue and orange dashed lines are the corresponding optima in the well-mixed approximation (Eqs. 3.11-3.12). The dashed green line shows the leading-order correction to the optimal productivity, Eq. 3.13.

with $\gamma = 3p_s s_0/e_{\max}$. Equations 3.11-3.12 are in excellent agreement with the numerical optimization of the full reaction-diffusion model (Fig. 3.3), except for large R values. Thus the well-mixed model correctly describes the optimal compartmentalization strategy for α-carboxysome parameters.

We can now also understand the trends in $e_T^*(R)$ and $\phi^*(R)$ observed in Fig. 3.3 across the full range of compartment sizes. The densities of substrate and intermediate inside the compartment are determined by two effects: exchange across the boundary, and production and consumption in enzymatic reactions. In particular, reactions deplete the available pool of metabolites, reducing the efficiency with which enzymes can operate. In small compartments, $R < R_c$, exchange dominates over the effects of enzymatic reactions. Here, the most productive strategy is to pack as many enzymes as possible into the compartment; the ratio of $\mathcal{E}_1$ to $\mathcal{E}_2$ enzymes should be chosen so as to balance the relative depletion of $\mathcal{S}$ and $\mathcal{I}$, thereby maximizing the efficiency with which $\mathcal{S}$ is converted into $\mathcal{P}$. As R increases towards R_c the larger number of enzymes in the compartment increases the impact of enzymatic reactions on the metabolite concentrations, which in turn leads to diminishing returns in conversion efficiency (reflected in the saturation of ϵ at large R). At the same time, the investment of enzymes needed to maintain fully packed compartments, represented by the prefactor $(e_T R)^{-1}$ in the productivity, continues to increase with R. The critical radius $R_{c,\mathrm{wm}}$ represents the compartment size at which the rate of increase of the product flux exactly matches the increasing cost of maintaining fully-packed compartments. For $R > R_{c,\mathrm{wm}}$, the productivity of maximally-packed compartments becomes smaller than could be achieved by constructing a larger number of partially-filled compartments.

Note that the $\sim R^{-1}$ scaling of the optimal enzyme concentrations and the plateau in optimal productivity rely not only on the fast diffusive mixing but also on the loss of intermediate exclusively through the compartment boundary. If intermediates were also to decay within the compartment volume the optimal compartmentalization strategy, including the optimal enzyme densities, would change. The analysis of this scenario can be found in the SM Sec. A.6.

3.3.3 Breakdown of the well-mixed approximation

For large compartment radii we see that the numerically-optimized productivity deviates from the prediction of the well-mixed model (Fig. 3.3 top, $R \gtrsim 10\,\mu\text{m}$). This is to be expected because the mixing timescale $\tau_D \sim R^2$ increases more rapidly for large R than the other timescales in the system, which scale at most linearly with R. The deviation from the fast-diffusion limited regime can be systematically examined by performing a series expansion of P^* in powers of D^{-1},

$$\mathrm{P}^*(R) = \mathrm{P}^*_{\mathrm{wm}} \left[1 + \frac{(p_i - p_s)R}{5\,(1+\phi^*_{\mathrm{wm}})\,D} + \mathcal{O}(D^{-2}) \right]. \tag{3.13}$$

(See SM Sec. A.7 for further details and the case $p_i \approx p_s$.) The first-order correction to the productivity (see Fig. 3.3 top, green dashed line) is due to the appearance of spatial gradients of $\mathcal{S}$ and $\mathcal{I}$. Changes in the optimal enzyme concentrations away from their well-mixed values enter only in higher-order terms.

From Eq. 3.13 we can identify the compartment size, $R_x \sim 5D(1+\phi^*_{\mathrm{wm}})/|p_i - p_s|$ (for the case $p_s = p_i = p$, $R_x \sim \sqrt{175/3}(D/p)$, see SM Sec. A.7), at which the optimal productivity is expected to deviate significantly from the well-mixed result. Interestingly, P^* can either increase or decrease relative to its well-mixed value depending on which of p_s or p_i is larger. This is because the explicit diffusion of metabolites has opposing effects on the efficiencies of the two reaction steps. The rate of $\mathcal{S}$-$\mathcal{E}_1$ reactions is reduced in the center of the compartment because fewer of the $\mathcal{S}$ molecules that are introduced at the boundary diffuse into the center. Conversely, those $\mathcal{I}$ molecules that are produced away from the compartment boundary have a significantly increased probability of reacting with $\mathcal{E}_2$ rather than escaping compared to the well-mixed scenario, leading to a higher probability of conversion of $\mathcal{I}$ to $\mathcal{P}$. The relative values of D/p_s and D/p_i, which represent the length scales over which $\mathcal{S}$ and $\mathcal{I}$ molecules can be considered "close" to the boundary, determine which of these effects has a larger impact on the pathway flux. In most biological contexts one would expect compartments to primarily confine intermediates, rather than limiting the influx of substrates, suggesting that $p_s > p_i$ is the more natural condition. In this case, as in Fig 3.3, the productivity decreases at large R.

If alternative biochemical processes that give rise to concentration gradients, such as the spontaneous decay of intermediates, were also to take place in the system, this may cause both greater deviation from the well-mixed approximation and a break down of the approximation at smaller values of R (see SM Sec. A.6).

3.3.4 Compartmentalization strategies for highly permeable shells

If the shell permeability is sufficiently high, the well-mixed approximation can break down at compartment sizes comparable to or even smaller than the compartment size at which

the transition to sub-maximal packing would occur, $R_x \lesssim R_{c,\mathrm{wm}}$. In such cases, we should not necessarily expect a plateau in productivity as predicted by the well-mixed model. We therefore investigated how the optimal compartmentalization strategies change when one or both of the shell permeabilities was set to a value of $p = 10^4\,\mu\mathrm{m\,s^{-1}}$, such that R_x lies in the range of physically-realistic compartment sizes ($R = 0.01\,\mu\mathrm{m}$ - $1\,\mu\mathrm{m}$).

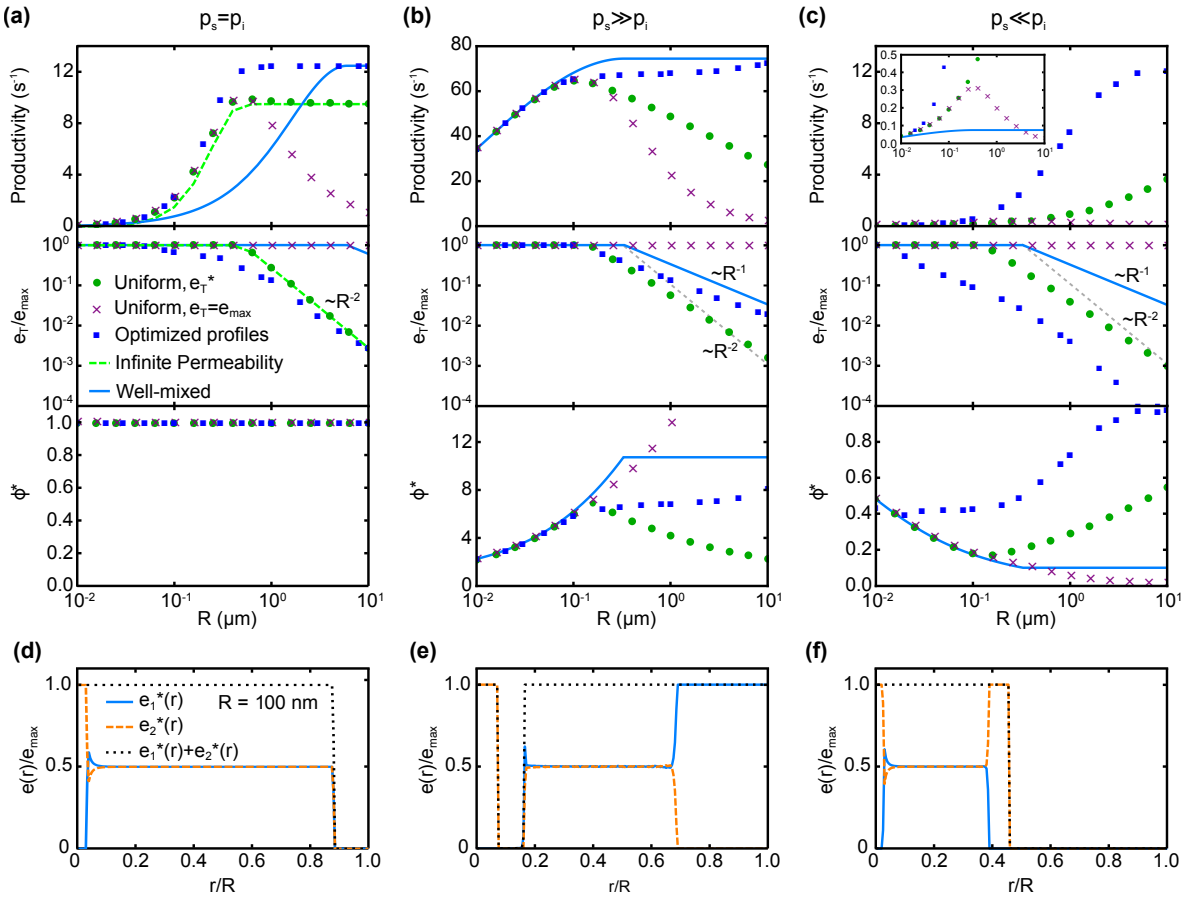

Figure 3.4: Optimal compartmentalization strategies outside the well-mixed regime.
(a-c) Optimal productivity and corresponding enzyme densities as a function of the compartment radius for different boundary permeabilities. **(a)** $p_s = p_i = 10^4\,\mu\mathrm{m\,s^{-1}}$. The limit $p_s, p_i \to \infty$ (see SM Sec. A.8) is shown as the green dashed line. **(b)** $p_s = 10^4\,\mu\mathrm{m\,s^{-1}}$, $p_i = 10\,\mu\mathrm{m\,s^{-1}}$. **(c)** $p_s = 10\,\mu\mathrm{m\,s^{-1}}$, $p_i = 10^4\,\mu\mathrm{m\,s^{-1}}$. Other parameters are fixed at $\kappa_1 = \kappa_2 = 0.4\,(\mu\mathrm{M\,s})^{-1}$, $e_{\max} = 25\,\mathrm{mM}$, s_0 =250 μM and $D = 100\,\mu\mathrm{m^2\,s^{-1}}$. **(d-f)** Optimal concentration profiles for the three cases above at a compartment radius of $R = 100\,\mathrm{nm}$.

Let us again first consider the case of uniformly distributed enzymes. For this enzyme arrangement, some common features are shared between the optimal compartmentalization strategies in the different parameter regimes (Fig. 3.4a-c, green circles). In all cases we can again identify a critical radius R_c at which a distinct transition in the optimal enzyme density occurs. For $R < R_c$ the optimal compartment is maximally packed; for $R > R_c$, the optimal enzyme density, $e_T^* < e_{\max}$, and the productivity again exceeds that of maximally-packed compartments of similar size (Fig. 3.4, purple crosses). However, the critical radii differ from those predicted by the well-mixed model (Fig. 3.4a-c, blue lines), $R_{c,\mathrm{wm}}$.

Beyond these similarities, qualitative differences can be observed between the optimal compartmentalization strategies. When both p_s and p_i are large (Fig. 3.4a), the productivity shows a plateau above R_c. In this regime, the optimal enzyme density in compartments decreases as $e_T^* \sim R^{-2}$, in contrast to the low-permeability regime where $e_T^* \sim R^{-1}$. Furthermore, examining the solution for the productivity in the limit $p_{s,i} \to \infty$, it turns out that the quantities e_T and R appear only as the product $e_T R^2$ (see SM Sec. A.8). This result allows us to identify the origin of the different scaling in the different processes governing exchange of reactants with the environment. In the low-p limit, the overall exchange rate is determined by the fraction of the compartment volume that is proximal to the boundary, which decreases as $\sim R^{-1}$. On the other hand, in the high-p regime exchange is limited by the time taken to diffuse to the boundary, which scales as $\tau_D \sim R^2$. Yet, despite the different underlying dynamics, the optimal compartmentalization strategy resembles that in the low-permeability regime, consisting of compartments larger than R_c with the appropriately chosen $e_T^*(R)$ and ϕ^*.

In contrast, when only one of p_s and p_i is large the optimal productivity does not exhibit a plateau for $R > R_c$. Rather, when p_s is small and p_i is large (Fig. 3.4c), the optimal productivity steadily increases with the compartment size. Thus, large sparsely-occupied compartments are the most productive. Since in this parameter regime productivity is limited by the rapid escape of $\mathcal{I}$ across the boundary, compartments should contain more $\mathcal{E}_2$ than $\mathcal{E}_1$ ($\phi^* < 1$). Conversely, when p_s is large and p_i is small, which is expected to be the more realistic scenario, the productivity exhibits a pronounced maximum at a certain radius $R^* \approx R_c$ (Fig. 3.4b). The optimal compartmentalization strategy here is to produce compartments of size $R = R^*$ that are maximally packed, $e_T^*(R^*) = e_{\max}$. More enzymes of type $\mathcal{E}_1$ than of type $\mathcal{E}_2$ are required ($\phi^* > 1$) to exploit the $\mathcal{S}$ that diffuses rapidly across the compartment boundary, whereas the slow leakage of $\mathcal{I}$ means this metabolite will accumulate inside the compartment.

In the well-mixed regime, changing the intra-compartment arrangement of enzymes has little effect on the productivity (Fig. 3.2a), since metabolites explore the entire compartment before reacting. However, if the compartment is not well mixed, we expect that the specific choice of enzyme arrangement will significantly affect the reaction flux. We therefore investigated numerically the enzyme profiles $e_1(r)$ and $e_2(r)$ that maximized productivity. Examples of the resulting optimal enzyme profiles are shown in Fig. 3.4d-f. The corresponding metabolite density profiles are shown in Fig. A.3 in the SM.

Figure 3.4a-c (blue squares) confirms that by a suitable choice of $e_1(r)$ and $e_2(r)$ it is possible to achieve a higher productivity than with uniformly distributed enzymes. The advantage is more significant when only one of p_s and p_i is large, rather than when both permeabilities are high. We can understand how the increased productivity comes about by examining the main features of the optimal enzyme profiles (the fine-scale details of the enzyme distributions, in particular at small r values, are susceptible to discretization artifacts and do not significantly affect the value of P). When p_s is large and p_i is small (Fig. 3.4e), enzymes of type $\mathcal{E}_1$ form a shell at the outer boundary of the compartment so as to maximize reactions with substrate $\mathcal{S}$ as it enters the compartment. Since the leakage of intermediate is slow, $\mathcal{E}_2$ enzymes can be localized to the interior side of this $\mathcal{E}_1$ band without seriously compromising the pathway flux. This enzyme arrangement requires more enzymes per compartment (larger e_T^* [Fig 3.4b, middle]) than the best uniform-enzyme strategy, meaning that fewer compartments with more

enzymes yield a larger product formation rate. When p_s is small and p_i is large, enzymes of both types are concentrated in the center of the compartment (Fig. 3.4f). Localizing $\mathcal{E}_1$ enzymes in this way ensures that $\mathcal{I}$ is produced far from the compartment boundary. $\mathcal{E}_2$ enzymes colocalize with $\mathcal{E}_1$ and also form a shell at the outer edge of the $\mathcal{E}_1$ domain so as to maximize the probability of capturing $\mathcal{I}$ molecules before they diffuse out of the compartment. By arranging enzymes in this way, each compartment can achieve a similar reaction with significantly fewer enzymes per compartment (smaller e_T^* [Fig 3.4c, middle]), such that a larger number of similar compartments can be constructed with the pool of available enzymes. Thus, the optimal compartmentalization strategies deviate qualitatively from those in the uniform-enzyme scenario.

3.3.5 Michaelis-Menten kinetics

So far, we have considered only the regime of low metabolite concentrations where the Michaelis-Menten reaction kinetics become linear in substrate concentration. However, this may not be the case in general. In this section, we will see how the full, non-linear reaction kinetics leads to changes in the optimal compartmentalization strategies. Previously, we saw that the breakdown of the well-mixed approximation is due to the appearance of metabolite gradients. The inclusion of non-linear reaction kinetics is likely to mitigate this breakdown through two effects. First, relative to linear kinetics with the same value of $\kappa = k_{\mathrm{cat}}/K_{\mathrm{M}}$, the effective reaction rate will be reduced by a factor of $(1 + s/K_{\mathrm{M}}^{(1)})^{-1}$ or $(1 + i/K_{\mathrm{M}}^{(2)})^{-1}$. This suppresses the consumption of metabolites in metabolic reactions, and therefore the appearance of metabolite concentration gradients. Secondly, even if gradients of metabolites form, they would only lead to differences in reaction rate at different positions if the absolute concentration were also to drop below the K_{M} value of the enzyme. For reactions that are in the zero-order regime, there will be no difference in reaction flux regardless of where in the compartment the enzymes are placed.

To test this hypothesis, we analytically determined the productivity $\mathrm{P}(e_1, e_2, R)$ in a well-mixed model taking into account the full Michaelis-Menten kinetics (see SM Sec. A.9), which we again optimized numerically with respect to $e_T(R)$ and $\phi(R)$. Additionally, we determined the optimal $e_T(R)$, $\phi(R)$, and productivity for the full non-linear reaction-diffusion system by solving Eqs. 3.2-3.3 with uniform enzymes. Since an analytic solution of the non-linear boundary-value problem is not possible, the steady-state metabolite concentrations were also calculated numerically in this case (see SM Sec. A.9). The extent of enzyme saturation was varied by changing the external substrate concentration s_0.

The qualitative characteristics of the optimal compartments are similar to the linear reaction regime (see Fig. 3.5). In particular, we again observe a plateau in $\mathrm{P}^*(R)$ above a critical radius R_c, accompanied by a decrease in the total enzyme density $e_T^*(R) \sim R^{-1}$. However, as the extent of saturation is increased, the maximal productivity increases sub-linearly with s_0 and R_c also increases. Note, in Fig. 3.5 the productivity is plotted relative to the external substrate concentration s_0. As in the case of linear reactions, the transition to sub-maximally packed compartments occurs when depletion of metabolites by reactions, which increases with the number of enzymes in a compartment, reaches such an extent that the gain in product formation from maintaining fully-packed compartments is less than the corresponding cost in terms of available enzymes. However, when the encapsulated enzymes are saturated, depletion

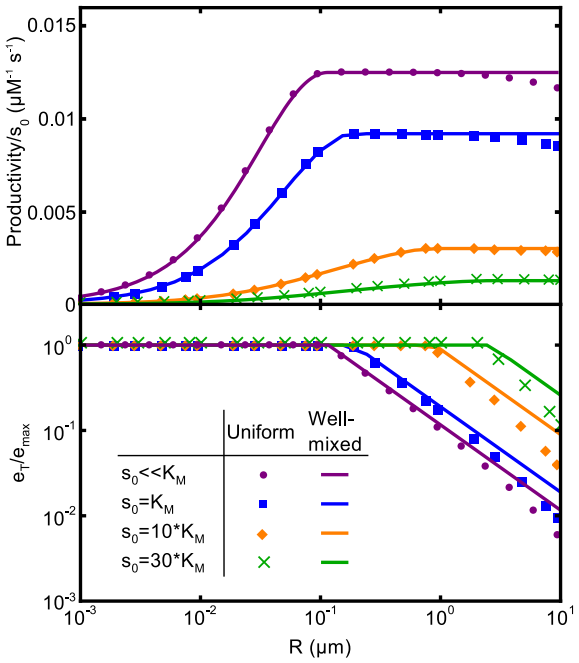

Figure 3.5: Optimal compartmentalization strategies with Michaelis-Menten reaction kinetics. Optimal productivity (top) and total enzyme concentration (bottom) for different external substrate concentrations s_0. Data points show results of numerical optimization for the reaction-diffusion model (Eqs. 3.2-3.3) with a uniform distribution of enzymes. Continuous lines show the corresponding optimization in the well-mixed approximation. Reaction parameters were chosen to resemble the activity of an "average" enzyme, $k_{\mathrm{cat}}^{(1)} = k_{\mathrm{cat}}^{(2)} = 10\,\mathrm{s}^{-1}$ and $K_{\mathrm{M}}^{(1)} = K_{\mathrm{M}}^{(2)} = 100\,\mu\mathrm{M}$ [185]. Other parameters were $D = 100\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, $p_s = p_i = 50\,\mu\mathrm{m}\,\mathrm{s}^{-1}$.

of metabolites will have a negligible effect on the reaction flux. In order for there to be a significant impetus to reduce the enzyme density inside the compartment, the number of enzymes must become so large as to deplete the metabolite pool into the range where enzymes are no longer saturated. Thus, the transition to sub-maximally packed compartments occurs at larger compartment sizes as the supply of $\mathcal{S}$ is increased.

Under biologically-realistic low-permeability conditions, excellent agreement is found between the well-mixed approximation and the solution of the full reaction-diffusion model. Hence, the well-mixed approximation also provides an excellent model for compartmentalization in the non-linear regime. At large $R > 1\,\mu\mathrm{m}$, we observe some deviation in the productivity for unsaturated conditions, $s_0 \lesssim K_{\mathrm{M}}$, in the same way as for the linear regime. However for $s_0 \gg K_{\mathrm{M}}$, the same radius falls in the range $R < R_c$, and the optimal compartmentalization strategy still appears to track the well-mixed result. In particular, a difference in the optimal enzyme densities is only observed for $R > R_c$, corresponding to the point at which the enzymes within the compartment first become less than fully saturated.

3.4 Discussion

In this chapter, we investigated the design principles underlying enzyme compartmentalization for a simple model metabolic pathway. We found that the strategy of packing enzymes into compartments as densely as possible only provides the highest productivity if compartments are constrained to be very small. On the other hand, if compartments can exceed a critical size, the productivity is optimized by constructing large compartments, each of which is less than maximally packed with enzymes. In the parameter regime of low membrane permeabilities that includes α-carboxysomes, it is reasonable to assume that metabolites are well-mixed throughout the compartment. Therefore, the precise arrangement of enzymes within the

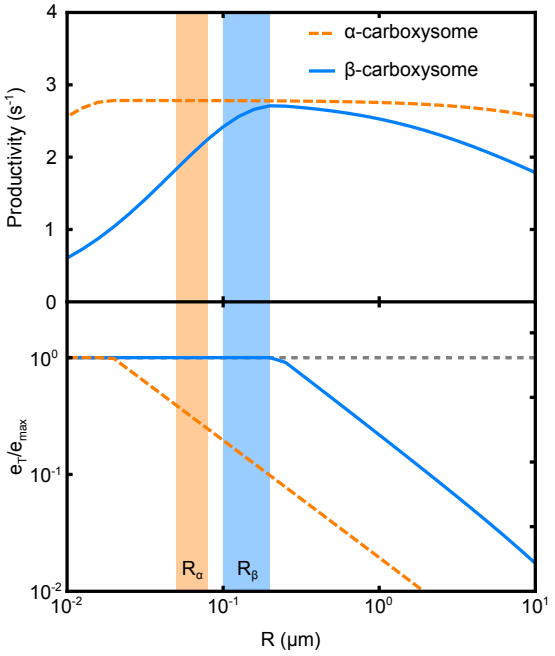

Figure 3.6: Optimal compartmentalization strategies for α- and β-carboxysomes. The higher shell permeability of β-carboxysomes (blue line) (p_s=1080 µm s^{-1}, p_i=215 µm s^{-1}) shifts the optimal productivity curve towards larger compartment radii and leads to a more pronounced productivity maximum compared to α-carboxysomes (orange dashed line) (p_s=90 µm s^{-1}, p_i=18 µm s^{-1}). Typical size ranges for the two carboxysome types are denoted by the shaded regions.

compartments is of little importance in determining the productivity. Furthermore, similar productivity can be achieved across a wide range of compartment sizes, provided that the enzyme concentration is appropriately chosen according to the compartment radius. These results provide a guide for the construction of efficient synthetic bioreactors.

The focus in the first part of this chapter was on parameters that resemble α-carboxysome microcompartments, one of the best-studied metabolic microcompartments. Interestingly, β-carboxysome shells were found to have larger pores than α-carboxysome shells [57, 179]. We estimate that the shell permeability of β-carboxysomes is approximately 12-fold higher than that of α-carboxysomes (see SM Sec. A.1). This leads to a larger critical radius R_c at which the optimal strategy switches from maximal- to partial-packing of enzymes, but also means that the well-mixed regime for β-carboxysomes is limited to smaller compartments (corresponding to a smaller R_x), resulting in a more pronounced maximum in the optimized productivity (Fig. 3.6). Notably, the typical size of β-carboxysomes, which is larger than that of α-carboxysomes (150 nm vs. 60 nm [186]), lies close to the radius at which the studied model predicts maximal productivity. The optimal strategy at this compartment size is maximal packing of enzymes. In contrast, α-carboxysomes fall in the regime $R > R_c$ for which the model predicts partial packing as the optimal strategy. Interestingly, the arrangement of RuBisCO within β-carboxysomes was described as densely-packed or paracrystalline [187], in contrast to the less organized packing observed within α-carboxysomes [181, 188]. Our analysis suggests that the observed differences in carboxysome properties may reflect the different optimal compartmentalization solutions that arise given the specific properties of each protein shell.

The aim in this chapter has been to explore some of the generic properties of compartmentalization strategies. Certainly, the studied reaction-diffusion model omits a number of biochemical details of metabolic reactions. For example, product inhibition will tend to reduce the efficiency of enzymes as the concentration of $\mathcal{P}$ becomes larger, thereby providing a

further incentive to reduce the enzyme concentration within a compartment in order to reduce the rate of $\mathcal{P}$ production and hence also its concentration. This potentially leads to a shift of the critical radius to smaller radii and a change in the scaling of e_T, but the qualitative transition from maximally packed to sparsely packed compartments will persist. Furthermore, in this study, the discrete nature and spatial structure of enzyme molecules has been neglected. In Chap. 5, we will study how the obstruction of metabolite diffusion, due to the spatial extension of enzymes, affects reaction fluxes and gives rise to certain trade-offs. Besides such effects, the discreteness of enzymes means that there will inevitably be fluctuations in the number of enzymes between compartments. It has previously been shown that such variability can alter the optimal strategy for partitioning signaling proteins among different sub-domains [189]. It remains to be seen whether a similar effect occurs in the context of metabolic reactions.

It is also possible for the reactions within compartments to become coupled through exchange of intermediate or depletion of substrate in the cytoplasm. It will be interesting to see under which conditions such inter-compartment interactions, which will be governed by the same physical processes of diffusion and exchange across compartment and cell boundaries as our intra-compartment model, alter the optimal compartmentalization strategies for the cell as a whole.

Although, in this study, we have considered only the steady-state productivity of static compartment configurations, metabolism is a highly dynamic and tightly regulated collection of processes. In scenarios where maximizing productivity is not the principal design goal, our model nevertheless demonstrates that metabolic fluxes may be controlled by changing the number and size of organelles in which reactions take place. Multiple examples are known where such changes in subcellular organelles are induced in response to varying metabolic demands or environmental cues [190, 191, 192]. That changing compartmentalization can alter metabolic flux has been demonstrated in the fungus *Penicillium chrysogenum*. Overproduction of a peroxisomal membrane protein resulted in a significant increase in the number of peroxisomes [193], which in turn led to a 2.5-fold increase in the level of penicillin in the culture medium. Interestingly, the amounts of peroxisomal enzymes involved in penicillin biosynthesis were unchanged compared to the control strain; rather, enzymes were distributed over a larger number of smaller organelles. It would be of interest to observe penicillin production if the peroxisomes are further divided into even smaller organelles. Our analysis predicts that at a certain organelle size the production of penicillin would decrease again due to the increasing loss of intermediates.

In addition to changing compartment size and number, cells are also able to regulate the morphology of organelles. For example, studies have shown that mitochondria change their shape according to the energy demands of the cell [194, 195]. This morphological change was suggested to be an active mechanism to increase bio-energetic efficiency [195]. Proteinaceous microcompartments could also potentially be constructed with different morphologies depending on the abundance of different shell components [196]. Our model predicts that for rapidly diffusing metabolites, productivity depends crucially on the surface-area-to-volume ratio of compartments, while the specific geometry will be less important. Thus adapting an organelle's surface-to-volume ratio by altering its shape presents an alternative possibility for tuning metabolic fluxes without altering enzyme expression levels.

4 Enzyme clustering and optimal enzyme arrangements *

We have seen in Chap. 2 that the spatial coordination of sequential enzymes often plays a crucial role in metabolic processes. Placing enzymes into complexes or macromolecular assemblies enables efficient processing of intermediates and thereby enhances metabolic fluxes. This enhancement often relies on structural features, like intramolecular tunnels, swing arms, or charge distributions on the molecular surface, that direct the transfer of intermediate substrates between the active sites of consecutive enzymes (see Sec. 2.2). Similar effects are achieved by segregating pathway reactions into compartments, where the compartment membrane hinders intermediates from escaping the reaction. However, there are many examples of multi-enzyme assemblies where such structural features are not present but for which enhanced metabolic fluxes have been hypothesized (see Sec. 2.3.2). In such cases, the transfer of intermediates between consecutive active sites is mediated by free diffusion and intermediates are not prevented from being lost to the surrounding bulk solution.

To understand how, in such systems, metabolic fluxes are enhanced only by enzyme organization and in the absence of direct intermediate transfer mechanisms, we use a minimal reaction-diffusion model that allows us to study the impact of the relative localization of consecutive enzymes on the efficiency of intermediate processing. We show that the clustering of sequential enzymes achieves the highest reaction flux when the catalytic efficiency of the downstream enzyme is low compared to the diffusion of intermediate substrates. However, when the catalytic efficiency is high, it becomes advantageous to redistribute some of the downstream enzymes away from the cluster into the bulk. This transition is a general phenomenon which occurs for different reaction kinetics, spatial dimensions, and loss mechanisms of intermediate substrate molecules. We explain this behavior in terms of the underlying stochastic nature of the reaction and diffusion process.

4.1 Introduction

In living organisms, the catalytic action of enzymes is essential for all metabolic and signaling processes. Enzymes are often found to be organized into large, non-membrane bound multi-molecular assemblies (see Sec. 2.3.2). These enzyme assemblies are believed to enhance metabolic fluxes by locally increasing the concentration of reactants. Examples include the purinosome [76], the cellulosome [198], and glycolytic enzyme clusters [110, 112]. In some

*This chapter is in large parts adapted from the publication: Optimization of collective enzyme activity via spatial localization, by A. Buchner, F. Tostevin, F. Hinzpeter, and U. Gerland, published in the *Journal of Chemical Physics* in 2013. See also [197].

cases, such as the cellulosome, enzyme assembly is directed by an inert scaffold to which the enzymes bind. In other cases, such as the purinosome, direct enzyme-enzyme interactions enable enzyme colocalization. Besides the impact on metabolic fluxes, enzyme colocalization also influences the functioning of signaling cascades. For example, enzymes in a cluster can generate a higher signal amplification than randomly distributed enzymes [199, 200]. In contrast to metabolic systems, where the aim is to enhance reaction fluxes, in signaling systems the design goal is to discriminate between different inputs.

Despite the ubiquity of multi-enzyme assemblies, a deep understanding of the consequences of certain enzyme arrangements on the reaction flux of a pathway is still lacking. In contrast to certain enzyme complexes, where intermediates are "channeled" between active sites via structural components, for multi-enzyme clusters, where such mechanisms are missing and intermediates freely diffuse, it is less clear how the spatial coordination of enzymes alone can improve reaction fluxes.

As reviewed in Sec. 2.4, in the last ten years, there has been growing interest in the experimental study of enzyme colocalization both *in vitro* and *in vivo* [201, 124]. The aim of these studies is to enhance pathway reactions with the goal of unraveling design principles which allow for an efficient synthesis of valuable products such as pharmaceuticals or biofuels. To be able to rationally design optimal enzyme arrangements using such approaches, a thorough understanding of the underlying physical principles will be crucial.

Here, we study the impact of enzyme positioning on the processing of intermediate products using a modeling approach. In a previous study [202] it has been shown, using a minimal one-dimensional reaction-diffusion model, that enzyme coclustering is not always the optimal strategy, rather when enzymatic reactions become fast compared to intermediate diffusion, it becomes advantageous to redistribute downstream enzymes away from the cluster. In this chapter, we will generalize the model to higher dimensions, non-linear reaction kinetics, and different mechanisms of intermediate leakage. We show that these diverse models exhibit similar behavior. In particular, the optimal enzyme distribution of downstream enzymes around a localized cluster of upstream enzymes transitions from cocluster to a more extended profile as a function of the effective reaction rate. While this transition is generic, the precise shape of the optimal enzyme distributions depends on the particular model. We demonstrate that the universal transition originates from the stochastic nature of reaction and diffusion, which we explain using the concept of enzyme exposure first introduced in [202].

4.2 Model

Let us consider a model reaction pathway consisting of two enzymes which sequentially convert substrate S to product P via an intermediate I. We assume that the upstream enzymes, E_1, are localized at a specific position and examine the impact of the arrangement of the downstream enzymes E_2 relative to the E_1 cluster on the efficiency of intermediate processing. For simplicity, we do not include the first reaction explicitly, but rather model the E_1 enzymes as a source of intermediates, with a total production rate J_1. Instead of being converted to P, intermediates can also be intrinsically unstable or degraded by other reactive molecules. Under the assumption that these processes are position independent, we model them as

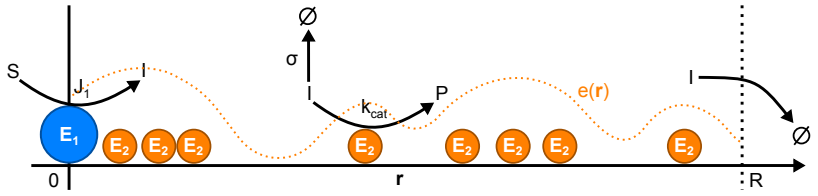

Figure 4.1: Illustration of the model used to study enzyme coclustering and optimal enzyme arrangements. A cluster of E_1 enzymes at the center acts as a source of intermediates. Intermediates diffuse and either react to the desired product or be lost due to decay or leakage through the system boundary. In this illustration, we depicted the enzymes as discrete particles. In our model, however, we coarse-grain the spatial positions of E_2 into the continuous concentration profile $e(\mathbf{r})$. Additionally to the depicted one-dimensional model, we consider systems in all spatial dimensions as well as different leakage mechanisms.

a first-order reaction with a constant rate σ. The transfer of intermediates between the enzymes is mediated by simple diffusion, and the arrangement of E_2 enzymes is described by a mean-field concentration profile. The dynamics of intermediates is then governed by the reaction-diffusion equation,

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} = D\nabla^2\rho(\mathbf{r},t) - \frac{k_{\mathrm{cat}} e(\mathbf{r})\rho(\mathbf{r},t)}{K_M + \rho(\mathbf{r},t)} - \sigma\rho(\mathbf{r},t), \tag{4.1}$$

where D is the diffusion coefficient of the intermediate and $e(\mathbf{r})$ is the concentration profile of E_2 enzymes. For the enzymatic reactions we have assumed standard Michaelis-Menten kinetics with k_{cat} the catalytic efficiency of E_2 and K_M the Michaelis constant. The production of intermediates by E_1 is implemented by a uniform source at the inner boundary of the system, $-DA_{in}\mathbf{n}\nabla\rho(\mathbf{r}_{in}) = J_1$, where $\mathbf{n}$ is the unit vector normal to the boundary and A_{in} is the area of the inner boundary. At the outer boundary, we assume either reflective ($\mathbf{n}\nabla\rho(\mathbf{r}_{out}) = 0$) or absorbing ($\rho(\mathbf{r}_{out}) = 0$) boundary conditions. The latter condition may describe the loss of intermediates to the extracellular environment. The model is illustrated in Fig. 4.1.

4.2.1 Non-dimensionalization of the system

Since the time scale on which enzyme assemblies form is much longer than the time scale of intermediate diffusion and processing, we will focus on the steady-state of Eq. 4.1. In steady state, we can recast the reaction-diffusion equation into the dimensionless form,

$$0 = \nabla^2\rho'(\mathbf{r}') - \frac{\alpha e'(\mathbf{r}')\rho'(\mathbf{r}')}{1+\gamma\rho'(\mathbf{r}')} - \beta\rho'(\mathbf{r}'), \tag{4.2}$$

where $\mathbf{r}'$ denotes the spatial coordinate rescaled by the system size R, and ρ' indicates that the intermediate concentration has been rescaled such that the total influx of intermediates is 1. Furthermore, we define the rescaled E_2 concentration $e'(\mathbf{r}) = e(\mathbf{r})/\bar{e}$, where $\bar{e} = V^{-1}\int_V e(\mathbf{r})\mathrm{d}\mathbf{r}$ is the enzyme concentration averaged over the system volume V. The dimensionless parameter $\alpha = (k_{\mathrm{cat}}\bar{e}/K_M)(R^2/D)$ captures the relative timescales of reactions with E_2 compared to the typical time to explore the system by diffusion. Similarly, the parameter $\beta = \sigma(R^2/D)$ measures the timescale of decay compared to the time to diffuse a distance R. The last

parameter $\gamma = J_1 R/(K_M D)$ represents the rate of influx of intermediate relative to the level at which E_2 enzymes become saturated. In the following, we omit the prime notation and exclusively consider the dimensionless system.

We can derive a flux-conservation equation by integrating Eq. 4.2 over the entire system and using the specified conditions at the boundary,

$$1 = \underbrace{\int_V \frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})}\mathrm{d}\mathbf{r}}_{J_2/J_1} + \underbrace{\int_V \beta\rho(\mathbf{r})\mathrm{d}\mathbf{r} - \int_{\partial V} \nabla\rho(\mathbf{r})\cdot\mathbf{n}(\mathbf{r})\mathrm{d}\mathbf{r}}_{J_{\mathrm{loss}}/J_1}. \tag{4.3}$$

The rescaled influx of intermediates on the left-hand side must be balanced by the flux of reactions by E_2, J_2/J_1, plus the loss of intermediates due to decay or the absorbing boundary, J_{loss}/J_1. How efficiently the E_2 enzymes process intermediates is then given by the ratio J_2/J_1, measuring the fraction of intermediates that are converted into the desired product P. In this chapter, we study how changing $e(\mathbf{r})$ affects the efficiency of intermediate processing. To only compare enzyme arrangements with the same overall enzyme abundance, we keep the total amount of E_2 enzymes in the system constant, which is assured by the condition $V^{-1}\int_V e(\mathbf{r})\mathrm{d}\mathbf{r} = 1$.

4.3 Results

In this chapter, we extend the minimal model studied in [202] to higher dimensional systems with different reaction kinetics and leakage mechanisms. We start by focusing on the low intermediate concentration limit, where the Michaelis-Menten kinetics becomes linear in $\rho(\mathbf{r})$. Later, we will see how including the full non-linear reaction kinetics will affect the behavior of the system. In all scenarios we observe a transition of the optimal profile from a cocluster to more extended profiles, a behavior which has first been reported in [202] for a minimal one-dimensional model. This demonstrates that the transition is a general phenomenon, which arises from the generic physical processes of enzymatic reaction and metabolite diffusion. Besides these similarities, we also find qualitative differences in the behavior of the different systems.

4.3.1 Linear reaction kinetics

Enzyme exposure

We start by recapitulating the concept of integrated "enzyme exposure" introduced in [202], which helps to understand the impact of different enzyme arrangements on the efficiency of intermediate processing. For systems with linear reaction kinetics, this concept allows us to decompose the processing efficiency into two factors, one spatial factor that only depends on the enzyme distribution, $e(\mathbf{r})$, and captures the diffusive motion of intermediates, and one local factor that is independent of $e(\mathbf{r})$ and describes the reaction dynamics. To explain this concept, let us consider a single intermediate molecule that is produced at time $t = 0$.

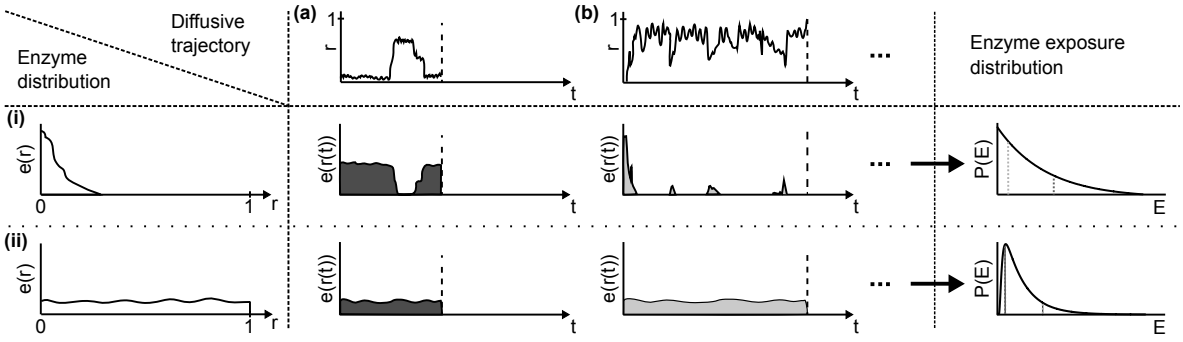

Figure 4.2: Schematic illustration of the concept of integrated "enzyme exposure". The diffusive motion of a set of intermediate molecules forms an ensemble of stochastic trajectories (top row), which have different enzyme exposure values in the presence of different enzyme distributions (left column). For example, the diffusive trajectory in **(a)** spends a relatively long time close to the source. This generates a higher enzyme exposure value $E = \int_0^{\tau_{\text{esc}}} e(\mathbf{r}(t))\mathrm{d}t$ (shaded area) for an enzyme distribution clustered near the source (e.g., distribution **(i)** middle row). In contrast, in **(b)** where the intermediate rapidly diffuses away from the source and spends most of the time in the bulk, the clustered distribution generates a smaller enzyme exposure value. For a rather uniform enzyme distribution (distribution **(ii)** bottom row) the enzyme exposure value is determined essentially by how long the particle stays in the system. The ensemble of diffusive trajectories generates a distribution of enzyme exposures, $P(E)$, which only depends on $e(r)$.

In the absence of the E_2 enzymes, the intermediate would follow a diffusive trajectory until it escapes through the system boundary or decays. We denote the time at which these processes happen and the trajectory ends as τ_{esc}. Because of the stochastic nature of diffusion, the possible paths intermediates can take form an ensemble of trajectories, $\mathbf{r}(t)$, which is independent of the distribution of E_2 enzymes. Now, let us assume that we were to reintroduce the E_2 enzymes according to the distribution $e(\mathbf{r})$. For each of the diffusive trajectories, the local propensity to react with an E_2 enzyme in the linear reaction regime is given by $\alpha e(\mathbf{r}(t))$. Therefore, for each trajectory, the survival probability $S(t)$ that an intermediate has not reacted until the time t is governed by the differential equation, $\dot{S}(t) = -\alpha e(\mathbf{r}(t))S(t)$. Consequently, the probability that a reaction would have occurred along the trajectory is given by, $1 - \exp\left[-\alpha \int_0^{\tau_{\text{esc}}} e(\mathbf{r}(t))\mathrm{d}t\right]$. Thus, the probability of reacting with the E_2 enzymes can be decoupled into the likelihood of diffusing a certain trajectory and the probability of reacting along this trajectory. We summarize all trajectories which are exposed to the same enzyme concentration, and thus have the same reaction probability, by introducing the integrated enzyme exposure,

$$E = \int_0^{\tau_{\text{esc}}} e(\mathbf{r}(t))\mathrm{d}t. \tag{4.4}$$

The ensemble of trajectories then forms a distribution of enzyme exposure values, $P(E)$ (see Fig. 4.2). Note, this distribution is independent of the reactivity of E_2 enzymes and only depends on the enzyme distribution $e(\mathbf{r})$ through Eq. 4.4. The probability to react depends on the enzyme exposure and is given by $p_r(E) = 1 - \exp[-\alpha E]$. This probability depends on the reactivity of E_2 via the parameter α but is independent of the E_2 distribution itself. Using this decomposition, we can express the reaction efficiency as,

$$\frac{J_2}{J_1} = \int_0^{\infty} P(E)p_r(E)\mathrm{d}E. \tag{4.5}$$

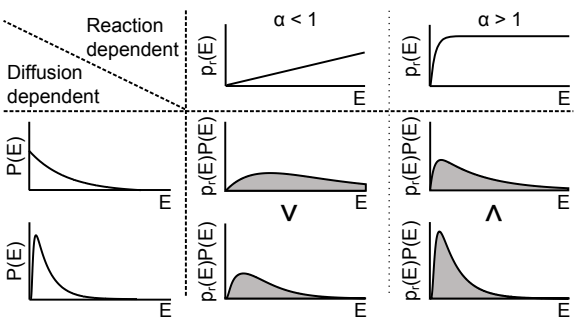

Figure 4.3: Decomposition of the efficiency into the enzyme exposure distribution and reaction probability. Enzyme exposure distribution P(E) (left column) and α-dependent reaction probability, $p_r(E)$ (top row). The conversion efficiency is given by the overlap integral of $P(E)$ with $p_r(E)$ (gray shading). Exposure distributions with large tails are preferable when α is small (middle column), whereas exposure distributions concentrated around small E values are favorable when α is large (right column).

For a derivation showing the equivalence of Eq. 4.5 with the efficiency defined in Eq. 4.3 see SM Sec. B.1. Figure 4.3 illustrates the decomposition of the reaction-diffusion dynamics into one diffusion- and $e(\mathbf{r})$-dependent component $P(E)$ and one reaction dependent component $p_r(E)$.

Instead of calculating the exposure distribution from individual diffusive trajectories, we can calculate $P(E)$ for a given profile $e(\mathbf{r})$ using the corresponding solution for the efficiency. For that, we recast Eq. 4.5 by expressing the efficiency in terms of the loss of intermediates and using the definition of the reaction probability,

$$\frac{J_2}{J_1} = 1 - \frac{J_{\text{loss}}}{J_1} = 1 - \int_0^\infty P(E) \exp\left[-\alpha E\right] \mathrm{d}E. \tag{4.6}$$

The last term on the right-hand side is the Laplace transform of $P(E)$ with transform variable α. Thus, the exposure distribution can be computed by performing an inverse Laplace transform of the loss flux, J_{loss}/J_1.

Unstable intermediate molecules

In the minimal model studied in [202], intermediates were assumed to leak out of the system via the boundary. Here, we seek to understand how the system behaves when intermediates are unstable and decay but do not escape through the system boundary. This scenario may represent the case of a negatively charged intermediate which cannot permeate through the cell membrane but is either intrinsically unstable or degraded by other reactive species. We start by studying the one-dimensional system with linear reaction kinetics. The reaction-diffusion equation Eq. 4.2 then reads,

$$0 = \partial_x^2 \rho(x) - \alpha e(x) \rho(x) - \beta \rho(x), \tag{4.7}$$

where the spatial coordinate is denoted by x in one dimension. On the left end of the system intermediates are produced, which is described by the boundary condition $\partial_x \rho(x)|_{x=0} = -1$. On the right end intermediates are reflected, leading to the condition $\partial_x \rho(x)|_{x=1} = 0$. The parameters α and β measure the timescale of reaction with E_2 and decay compared to the time to explore the entire system by diffusion. Alternatively, α^{-1} and β^{-1} can also be interpreted as measures for the typical distance from the source at $x = 0$ in units of the system size an intermediate molecule diffuses before it reacts with E_2 or decays. For $\beta \ll 1$, intermediates

typically explore the entire system before they react, and hence the spatial arrangement of enzymes will have little effect on the processing efficiency, since intermediates will be exposed to each enzyme irrespectively of its position. On the other hand, when $\beta \gg 1$ most intermediates rapidly decay and do not diffuse far from the source. Here we expect that positioning enzymes close to the source will have a strong effect on the processing efficiency.

We start by comparing two scenarios: a uniform distribution of E_2 enzymes over the domain $x \in [0, 1]$, ($e_{\mathrm{u}}(x) = 1$), and a clustered distribution where the E_2 enzymes are colocalized with the E_1s, $e_c(x) = \delta(x)$. Solving Eq. 4.7 for both enzyme distributions and computing the integral over the reaction term (Eq. 4.3), we obtain the efficiencies for uniformly distributed and coclustered enzymes,

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{u}} = \frac{\alpha}{\alpha+\beta}, \qquad \left(\frac{J_2}{J_1}\right)_{\mathrm{c}} = \frac{\alpha}{\alpha+\beta^{1/2}\tanh\beta^{1/2}}, \tag{4.8}$$

where the index u denotes the uniform configuration and the index c denotes the clustered configuration (for the derivation see SM Sec. B.2). The efficiency for a uniform E_2 distribution can be understood as the ratio between the rate of P production divided by the total rate of intermediate consumption.

Comparing the expressions in Eq. 4.8, we notice that the clustered E_2 arrangement always achieves a higher efficiency than the uniform arrangement, since $\tanh\beta^{1/2} \leq \beta^{1/2}$. This is no longer the case when we additionally assume intermediate loss via the system boundary. This behavior is due to the spatially independent rate of intermediate decay, measured by β, which determines how far intermediates diffuse away from the source at $x = 0$ before they decay. For uniformly distributed enzymes, the amount of E_2 enzymes intermediates are exposed to is reduced compared to the clustered arrangement. This can be quantified by studying the enzyme exposure distributions of the two enzyme arrangements, which are given by,

$$P_{\mathrm{u}}(E) = \beta\exp\left[-E\beta\right], \qquad P_{\mathrm{c}}(E) = \beta^{1/2}\tanh\beta^{1/2}\exp\left[-E\beta^{1/2}\tanh\beta^{1/2}\right], \tag{4.9}$$

(see SM Sec. B.3 for the derivation). For uniformly distributed enzymes, the enzyme exposure distribution, $P_{\mathrm{u}}(E)$, is more concentrated around $E = 0$ than $P_{\mathrm{c}}(E)$, meaning that a higher fraction of trajectories rapidly decays before being exposed to a significant amount of E_2 enzymes. These trajectories have a low probability of reacting with E_2 and, thus, lead to a low processing efficiency.

Next, we ask what is the enzyme distribution that globally maximizes the efficiency. Despite the absence of a transition from a clustered to a uniform enzyme distribution, the optimal distribution still may exhibit a transition away from the pure cluster. To study this, we numerically determine the optimal enzyme configuration by discretizing the domain into a lattice and solving the reaction-diffusion system on the lattice. Each optimization step is initialized with a trial E_2 profile on the lattice. From this, we generate 50 variants by randomly selecting one lattice site and moving a random amount of E_2 from this site to another randomly chosen site. For each of these modified profiles, we solve Eq. 4.7 on a lattice and compute the efficiency, J_2/J_1, by numerically integrating the reaction term over the domain. For the next iteration, we use the mean profile of the ten most efficient configurations found in the previous round. This procedure turned out to be more robust and to converge faster than a simple random exploration of the possible configuration space. Figure 4.4a,b shows the

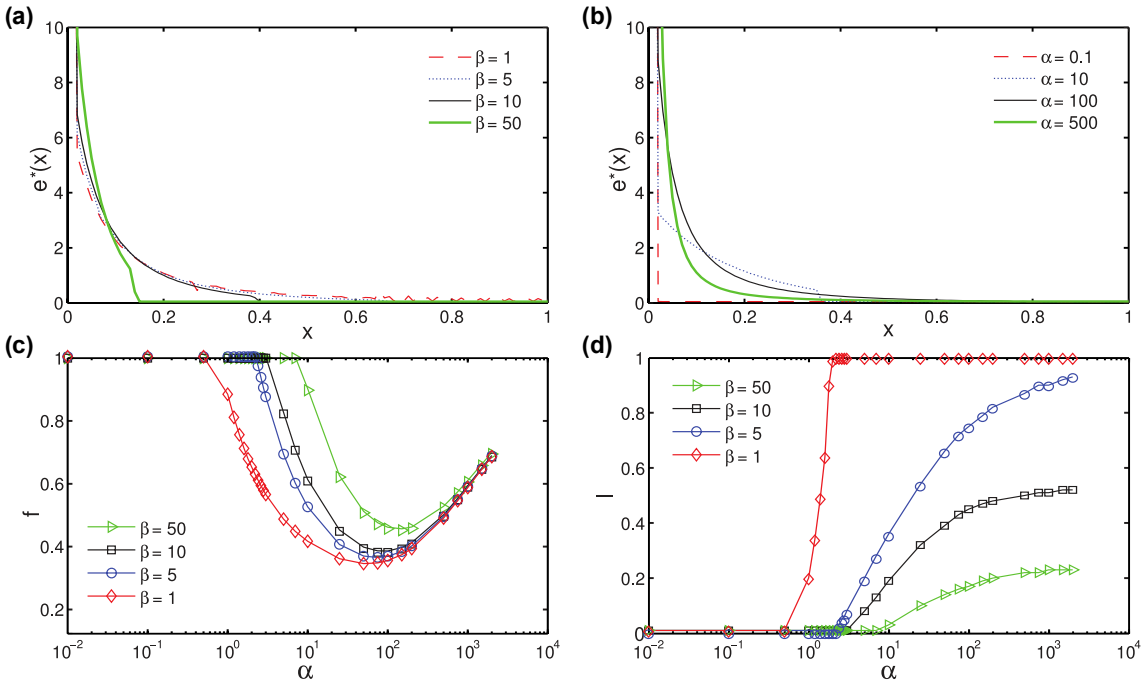

Figure 4.4: Optimal E_2 profiles with unstable intermediates. **(a)** Optimal profiles for different values of β at $\alpha = 50$. **(b)** Optimal profiles for different values of α at $\beta = 5$. All optimal profiles were computed numerically using a lattice with 100 lattice sites. **(c)** Fraction of total E_2, f, that are clustered at the source of intermediates. **(d)** Extension of the profiles, l, for different β values as a function of α.

optimal profiles for different combinations of the parameters α and β. Notably, we find that enzyme clustering is not always the optimal strategy, rather in a certain parameter regime, it is advantageous to redistribute enzymes from the cluster into the bulk. This behavior is reminiscent of the transition observed in [202] where intermediate leakage is modeled via an absorbing boundary condition. However, the shape of the optimal profiles is different, and there is no parameter regime where distributing all enzymes uniformly over the system is optimal.

We quantify the optimal enzyme profiles using the fraction of E_2 enzymes clustered at the source, denoted by f, and the extension of the tail of the distribution defined as the distance, l, from the source over which the enzyme concentration is above the threshold 10^{-3}. Focusing first on the behavior of the optimal profiles for different β (Fig. 4.4a), we find that as β is increased it becomes favorable to cluster a growing fraction of enzymes at the source. This tendency is due to the decreased distance intermediate molecules diffuse before they decay. Examining now the behavior of the optimal profiles as a function of α (see Fig. 4.4b), we observe a sharp transition from a pure cluster ($f = 1$ and $l = 0$) to more extended profiles ($f < 1$ and $l > 0$) as α is increased above a critical value. Interestingly, as α becomes very large, the enzymes are relocated from the bulk back to the source rather than further into the bulk. This behavior is shown in Fig. 4.4c, where f passes through a minimum and starts increasing again. However, the distance, l, over which enzymes are distributed does not decrease (see Fig. 4.4d), only the amount of enzymes close to the source increases at the expense of decreasing the amount far away from the source.

We can understand the behavior of the optimal profiles as follows: when α is small enzyme clustering achieves the highest efficiency, since this configuration generates the largest number of high enzyme exposure trajectories. At moderate α above the critical value, it becomes favorable to move some of the enzymes from the cluster into the bulk. This enhances the probability of catching intermediates that rapidly diffuse away from the cluster at diminishing returns of reactions at the cluster. For very large α, the need for distributing enzymes is reduced, since intermediates rapidly react and rarely diffuse away from the cluster. Therefore, by again clustering most of the enzymes at the source the probability of processing intermediates that stay a very short time in the system is maximized. This is the main difference to the behavior observed in [202], where intermediates must diffuse past all E_2 enzymes before they can escape the system via the boundary at $x = 1$. While in the system studied here, intermediates can be lost before being exposed to enzymes positioned too far from the source.

Higher-dimensional systems

In the following section, we study the system in two and three dimensions in a spherical geometry. By positioning the source of intermediates at the system center, the concentration profiles become spherically symmetric depending only on the radial position, r. We start by studying the two-and three-dimensional analog of the system studied in [202], where intermediates are lost only at the system boundary and do not decay ($\beta = 0$). In this scenario, the steady-state reaction-diffusion equation Eq. 4.2 reads

$$0 = \nabla^2 \rho(r) - \alpha e(r)\rho(r), \tag{4.10}$$

with the source boundary condition, $[A_{\text{in}}(r)\partial_r \rho(r)]_{r=0} = -1$ (where $A_{\text{in}}(r) = 2\pi r$ in 2-D and $A_{\text{in}}(r) = 4\pi r^2$ in 3-D), accounting for the production of intermediates by E_1, and the absorbing boundary condition $\rho(1) = 0$ describing the loss of intermediates. We again first compare the efficiencies generated by a clustered and uniform enzyme distribution. The clustered distribution is modeled by a shell of E_2 enzymes localized close to the source at $r = r_0$. The corresponding distribution in two and three dimensions are then given by, $e_{\text{c}}(r) = \delta(r - r_0)/(2r_0)$ and $e_{\text{c}}(r) = \delta(r - r_0)/(3r_0^2)$, respectively. The uniform distribution is in all spatial dimension given by, $e_{\text{u}}(r) = 1$. Note, the distributions are scaled by the average enzyme density such that $V^{-1}\int_V e(\mathbf{r})\mathrm{d}\mathbf{r} = 1$. Solving Eq. 4.10 and integrating over the reaction term, we obtain for the efficiencies in two and three dimensions,

$$\left(\frac{J_2}{J_1}\right)_{\text{c}} = \frac{\alpha \log r_0}{\alpha \log r_0 - 2}\,, \qquad \left(\frac{J_2}{J_1}\right)_{\text{u}} = 1 - \mathrm{I}_0\left(\alpha^{1/2}\right)^{-1} \qquad \text{in 2-D}, \tag{4.11a}$$

$$\left(\frac{J_2}{J_1}\right)_{\text{c}} = \frac{\frac{\alpha}{3}(1 - r_0)}{r_0 + \frac{\alpha}{3}(1 - r_0)}\,, \qquad \left(\frac{J_2}{J_1}\right)_{\text{u}} = 1 - \alpha^{1/2}\,\mathrm{csch}\left(\alpha^{1/2}\right) \qquad \text{in 3-D}, \tag{4.11b}$$

where $\mathrm{I}_k(x)$ is the modified Bessel function of the first kind (see SM Sec. B.4). As the radius r_0 of the enzyme cluster approaches zero, the local enzyme concentration becomes infinite, and the efficiency approaches one. This situation is not physically realistic, as enzymes are molecules of finite size. Thus, there will be an upper bound for the concentration of enzymes in the cluster. Other effects emerging from the discrete nature of enzymes, like obstruction

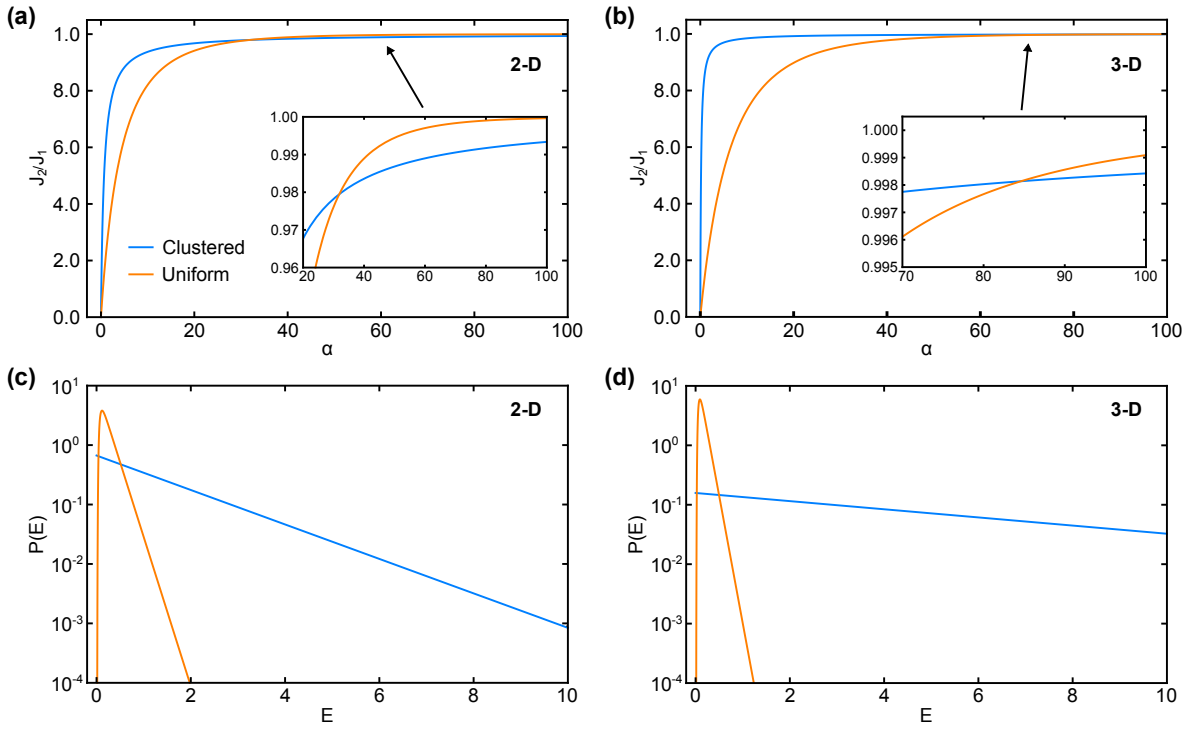

Figure 4.5: Comparison of the efficiency in higher dimensions. (a,b) Comparison of the processing efficiencies for clustered (blue curve, $r_0 = 0.05$) and uniformly distributed (orange curve) E_2 profiles in **(a)** two dimensions and **(b)** three dimensions. A transition is observed at $\alpha \approx 30$ in two dimensions and $\alpha \approx 85$ in three dimensions **(a,b inset)**. **(c,d)** Enzyme exposure distribution for the clustered (blue curve) and uniform (orange curve) E_2 distribution in (a) two dimensions and (b) three dimensions.

of metabolite diffusion, will be neglected here and we will return to this question in Chap. 5. If r_0 is small but finite and for small α values, the clustered configuration is more efficient in processing intermediates than the uniform distribution. However, above a critical α value, which depends on r_0, the uniform distribution becomes more efficient (see Fig 4.5a,b). This crossover is similar to the one observed in [202]. However, the critical α is shifted to higher values from $\alpha \approx 9$ in the one-dimensional case to $\alpha \approx 30$ and $\alpha \approx 85$ in the two- and three-dimensional case, respectively. Since for such large critical α values almost all intermediates are processed, the difference between reaction efficiencies is very small (see Fig. 4.5a,b inset). However, in the low α regime, the advantage of clustering is much more significant than in the one-dimensional case with the largest enhancement observed in three-dimensions. We will later come back to the question of how strongly clustering can enhance and regulate fluxes (see Chap. 6).

To further examine the behavior of the system in higher dimensions, we compute the enzyme exposure distributions for both enzyme arrangements in two and three dimensions (see SM Sec. B.5 for details),

$$P_{\mathrm{u}}(E) = 2\sum_{n=1}^{\infty} \mathrm{J}_1(j_{0,n})^{-1} \exp\left[-j_{0,n}^2 E\right] j_{0,n} \ , \qquad P_{\mathrm{c}}(E) = -\frac{2\exp\left[\frac{2E}{\log r_0}\right]}{\log r_0} \qquad \text{in 2-D,} \quad (4.12a)$$

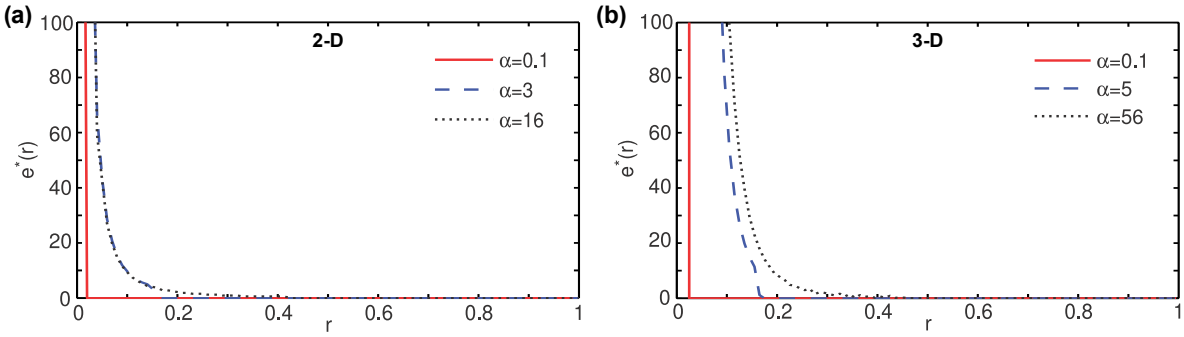

Figure 4.6: Optimal enzyme profiles in higher dimensions. Optimal enzyme profiles in **(a)** two dimensions and **(b)** three dimensions. The critical α value above which the optimal enzyme distribution contains a fraction of enzymes distributed away from the cluster is $\alpha_c \approx 0.5$ in two dimensions and $\alpha_c \approx 0.05$ in three dimensions.

$$P_{\mathrm{u}}(E) = 2\sum_{n=1}^{\infty}(-1)^{n+1}(\pi n)^2 \exp\left[-(\pi n)^2 E\right] , \quad P_{\mathrm{c}}(E) = \frac{3r_0 \exp\left[-\frac{3r_0}{1-r_0}E\right]}{1-r_0} \quad \text{in 3-D,} \quad (4.12b)$$

where $\mathrm{J}_k(x)$ is the Bessel function of the first kind and $j_{0,n}$ its n^{th} zero point. In Figure 4.5c,d, we see that $P_{\mathrm{u}}(E)$ is sharply peaked around small but finite E values, while $P_{\mathrm{c}}(E)$ is exponentially distributed. Thus, $P_{\mathrm{u}}(E)$ is concentrated in the regime where for small α, $p_r(E)$ is small, whereas the exponential tail of $P_{\mathrm{c}}(E)$ results in a larger mean value of E and, hence has more trajectories in the region where $p_r(E)$ is significant. Consequently, in the small α regime the clustered configuration is more efficient. In contrast, when α becomes large, the peak of $P_{\mathrm{u}}(E)$ lies in the region where $p_r(E) \approx 1$, while in the region $E \ll 1$, where $P_{\mathrm{c}}(E)$ is larger, $p_r(E)$ is small. These trajectories, which correspond to intermediates that rapidly diffuse away without returning to the cluster, will escape and, therefore, reduce the efficiency of the clustered configuration.

Why does clustering in higher dimensions lead to a larger enhancement over the uniform distribution for small α values? In higher spatial dimensions, diffusive intermediate molecules have more spatial directs to explore and, thus, the spatial fraction of the system an intermediate molecule covers before it escapes decreases with the spatial dimension. If the same number of E_2 enzymes are equally distributed at a radial position further away from the center, the effective "reaction cross section" becomes smaller for higher dimensions, because the fraction of enzymes at this radial position an intermediate will typically explore decreases, and with it the fraction of enzymes in the system an intermediate will be exposed to. This is in contrast to the one-dimensional case, where an intermediate is exposed to all enzymes before it escapes at the boundary. Therefore, to gain a benefit from distributing enzymes away from the cluster, a larger α value is required in higher dimensions to compensate for the reduction in E_2 enzymes to which intermediate molecules are exposed to.

Now, we turn to the question of what is the enzyme profile that optimizes the processing efficiency as a function of the control parameter α. As a constraint for the optimal profiles, we impose a limit on the possible packing concentration by assuming that E_2 enzymes cannot be positioned closer to the source than a minimal distance r_0. We adapt the optimization procedure described above by solving Eq. 4.10 on a radial lattice where the lattice sites are at

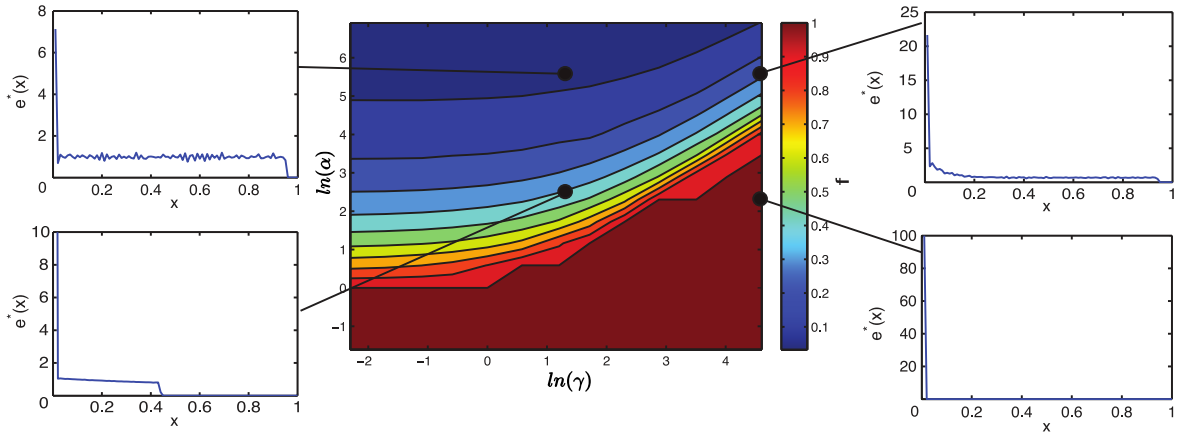

Figure 4.7: Optimal enzyme profiles and fraction of enzymes in the cluster for nonlinear reaction kinetics. The central panel shows the fraction, f, of E_2 enzymes in the cluster as a function of α and γ. The outer panels show the optimal enzyme profiles in different regimes of α and γ. Optimization were performed with $N = 100$ lattice sites, 2×10^5 reaction steps and a solution error-tolerance of 10^{-7}

equidistant radial positions. The minimal radius is then fixed by the total number of lattice sites, N, through $r_0 = (2N)^{-1}$ as the mid-point of the first site. With the constraint of a minimal clustering radius, we again find that the optimal profile transitions from a pure cluster to more extended profiles when α exceeds the critical value. The critical α value depends strongly on r_0; for $r_0 = 0.05$ the transition point occurs at $\alpha \approx 0.5$ in two dimensions and $\alpha \approx 0.05$ in three dimensions ($\alpha = 1$ in one dimensions), showing that the critical α decreases with dimension. This behavior is in contrast to the finding that distributing enzymes away from the cluster is less efficient in higher dimensions. The increased penalty of moving enzymes to larger values of r, due to the reduced reaction cross-section in higher dimensional systems, is reflected in the shape of the extended enzyme profile. In contrast to the one-dimensional case where the extended profile is constant, the profiles decrease as $e^*(r) \sim r^{-2}$ in two dimensions and $e^*(r) \sim r^{-4}$ in three dimensions. Collectively, the emergence of a robust transition from clustered to more extended configurations shows that the underlying physics is generic, and is not dependent on the different statistics of diffusion in higher dimensions.

4.3.2 Non-linear reaction kinetics

In the following section, we will study how enzyme saturation affects the generic results presented above. As mentioned in Sec. 4.2, we include enzyme saturation by assuming standard Michaelis-Menten kinetics. We start by considering the one-dimensional case with $\beta = 0$ and intermediate loss occurring only at the system boundary. The corresponding reaction-diffusion equation then reads,

$$0 = \partial_x^2 \rho(x) - \frac{\alpha e(x) \rho(x)}{1 + \gamma \rho(x)}, \tag{4.13}$$

with the source-sink boundary conditions $\partial_x \rho(x)|_{x=0} = -1$ and $\rho(1) = 0$, where γ measures the effective saturation of the E_2 enzymes (see definition in Sec. 4.2.1).

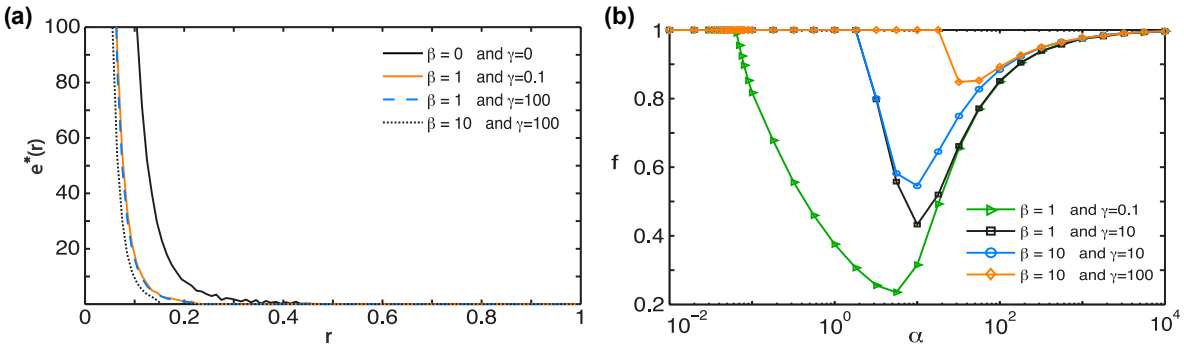

Figure 4.8: Full model with decay and enzyme saturation. Optimal enzyme distribution and fraction, f, for the full model at different β and γ parameters and with $\alpha = 56$ fixed. **(a)** Optimal distribution, **(b)** Optimal fraction of E_2 enzymes in the cluster.

We again determine the optimal enzyme profile as a function of α and γ. Due to the nonlinearity, the discretized equation cannot be transformed into a linear system that can be solved directly. We therefore solved Eq. 4.13 for a given enzyme profile using a shooting approach. In this approach, a trial solution at the rightmost lattice site is selected. This trial solution is then used to successively solve Eq. 4.13 on the remaining lattice sites. The equation for any lattice site depends on the solution of the adjacent sites right and left of it. Once a trial solution was found, it is tested against the reaction-diffusion equation at the first site which includes the source boundary condition. When at all lattice sites Eq. 4.13 is satisfied to a certain tolerance, then the solution is accepted. The optimization procedure described above remained unchanged.

The numerical solutions for the optimal profiles are shown in Fig. 4.7 for different parameters α and γ. In the limit of small γ, we see that the numerical procedure reproduces the optimal profiles found in [202]. For larger values of γ, we find the same qualitative behavior, namely a transition from clustered to more extended profiles. The critical α at which the transition occurs increases with γ, since saturation, measured by γ, reduces the effective reaction rate. Furthermore, we find that the shape and extension of the optimal profile change with the degree of saturation. The tail of the profiles is not constant and the extension, l, decreases with γ. This behavior arises from the position-dependent saturation of E_2 enzymes. With increasing distance from the source, the reaction changes from fully saturated to unsaturated. Finally, it appears that after the transition to a more extended profile occurred, the clustered E_2 fraction, f, decreases more quickly with α when γ is large. For increasing α, the intermediate concentration in the system is reduced and, thus, the reactions transition from saturated to unsaturated. Where this transition occurs is determined by γ. Indeed, for very large α the reaction kinetics become effectively linear and the optimal profiles become independent of γ.

For nonlinear reaction kinetics the decomposition of the reaction-diffusion dynamics, using the concept of enzyme exposure, is no longer possible. This is because the reaction probability of an individual intermediate molecule now depends not only on the enzyme concentration but also on the concentration of intermediate molecules. One could define an effective enzyme exposure taking into account the effect of saturation, but this does not lead to a decomposition of the reaction efficiency into one diffusion and one reaction dependent component.

Finally, we study the full model introduced in Sec. 4.2 including enzyme saturation ($\gamma \neq 0$)

and intermediate decay ($\beta \neq 0$) in a radial symmetric three-dimensional system. We numerically optimize the E_2 profile for different parameter combinations of β and γ at constant α. The solutions for the optimal profiles (see Fig. 4.8a) are qualitatively similar to the three-dimensional system with linear kinetics and without decay (compare to Fig. 4.6b). However, the fraction of clustered enzymes is always larger than in the case where ($\beta = \gamma = 0$) (see Fig. 4.8b). This is consistent with the previous results showing that large β or γ increase the tendency of clustering enzymes at the source. The non-monotonic behavior of f as a function of α is similar to that observed in the one-dimensional system with decay. Similarly to Fig. 4.7 when the saturation effect becomes stronger (larger γ), the critical α is shifted to larger values. Collectively this shows that the individual effects arising from the spatial dimension, decay, or nonlinear reaction kinetics are also present in the full, three-dimensional model.

4.4 Discussion

In this chapter, we have shown that the transition from a pure clustered to a more extended optimal enzyme profile is a generic feature of a class of reaction-diffusion systems, where intermediates are produced at a localized source. However, the exact shape of the optimal profiles and the critical α parameter depend on the loss mechanism, spatial dimension, and reaction kinetics. We saw that the transition arises from the interplay of the stochastic dynamics of individual intermediate molecules in terms of enzyme exposure and the probability to react for a given enzyme exposure value. By decomposing the reaction efficiency using this concept, we were able to give an intuitive explanation of the observed behavior. Clustering all enzymes close to the source leads to high enzyme exposure times, since this is the region where intermediates are most likely to spend a significant amount of time. Therefore, for slow reactions, clustering is the optimal strategy, since intermediates will spend a comparably long time in the system. In contrast, when the reactions are fast, most intermediates instantly react and distributing a fraction of enzymes away from the cluster increases the probability of catching those intermediates that rapidly escape from the cluster.

Furthermore, for the different systems, we have found some interesting qualitative differences to the results presented in [202]. For unstable intermediates, it becomes advantageous to move enzymes back to the cluster when the reaction rate is increased beyond a certain value. This behavior results from the fact that intermediates can also be lost close to the source due to decay rather than only at the boundary. We also have seen that increasing the spatial dimensions increases the enhancement of the efficiency generated by the cluster relative to the uniform enzyme configuration. The increased space intermediates can explore leads to a smaller fraction of distributed enzymes to which intermediates are exposed along their diffusive trajectory. Finally, we have seen that including enzyme saturation further favors enzyme clustering. This suggests that, depending on the influx rate of intermediates, it may be desirable to dynamically adjust the localization of enzymes in order to maintain optimal intermediate processing. An example where a dynamic reorganization of enzymes has been observed is the reversible localization of the mammalian hexokinase isoform (HKII) to the outer membrane of mitochondria, which depends on factors like glucose-6-P and GSK3 [203, 204]. This dynamic strategy is believed to regulate the relative flux of glucose through different metabolic pathways.

In a subsequent theoretical study by Castellana *et al.*, qualitatively similar effects of enzyme clustering were found [142]. In this study, the conversion of substrates to intermediates by the first enzyme was explicitly included and, furthermore, the first enzymes were not restricted to a specific location but rather allowed to be arbitrarily distributed in the system. It was found that enzyme clustering can significantly enhance the conversion of substrates to products catalyzed by two consecutive enzymes. The distribution of enzymes that optimizes the processing also consisted of a cocluster and a certain fraction of the second enzyme distributed away from the cluster. This additionally shows that the results presented in this chapter are very generic and apply to an even broader range of reaction-diffusion systems than considered here.

Despite the interesting physical principles underlying the transition in optimal enzyme distribution, the difference between the reaction efficiency of a pure cluster and the optimal profile is small. Therefore, for most biological contexts, the relevant conclusion of this chapter is that enzyme coclustering can considerably enhance intermediate processing without the need for direct transfer of intermediates or confinement within a membrane. This major finding could be tested experimentally using the various techniques available to localize individual molecules such as "single-molecule cut-and-paste" [205] or scaffolding strategies based on DNA, RNA, and protein nanostructures [117, 206, 207].

The mesoscopic reaction-diffusion models studied in this chapter neglect some details that can be important on very small length scales. First and foremost, the spatial extensions of both enzyme and intermediate molecules have been neglected. In real systems, the size of enzyme molecules sets an upper limit on the clustering density. Furthermore, the tight packing of enzymes in a cluster will lead to obstruction of diffusive intermediate trajectories. This may lead to physical confinement of intermediates inside the cluster. However, the same steric hindrance will prevent the access of initial substrates to the first enzyme. In Chap. 5 we will study these effects by explicitly considering the enzyme molecules as discrete particles. This also allows us to investigate enzyme arrangements on a microscopic scale that consist of few enzyme molecules such as simple pairs of consecutive enzymes fused together. We will furthermore investigate in Chap. 6 how enzyme coclustering can regulate the fluxes at a metabolic branch point and introduce enzyme sequestration as an alternative strategy for flux regulation.

5 Enhancing biocatalytic fluxes via spatial organization *

The spatial organization of biochemical reactions is ubiquitous in biological systems across different length scales, from clusters of enzymes to collaborating species in microbial consortia. How this organization affects the throughput of metabolic pathways, and the general principles directing biocatalyst organization, remain poorly understood. In particular, biocatalysts often form dense clusters, within which crowding leads to complex spatial networks of diffusive fluxes that exchange metabolites between biocatalysts and with the environment. In this chapter we extend the models studied in Chap. 4 and analyze how the spatial arrangement of discrete biocatalysts affects the flux of a model reaction pathway. We study how the spatial extensions of biocatalysts leads to crowding effects and how these can be exploited or mitigated to modify the reaction dynamics. Using a reaction-diffusion model with biocatalysts treated as discrete reaction centers and metabolites described via continuous concentration profiles, we studied how different localization strategies are affected by the relative rates of the reactions and metabolite diffusion. In the reaction-limited regime, where the diffusion is fast compared to reactions, the formation of large clusters is most efficient, whereas for diffusion-limited biocatalysts it is advantageous to form pairs or small complexes of a few of each type of biocatalyst. This design principle arises from two fundamental trade-offs in the reaction-diffusion dynamics of metabolites, the first between efficient inter-biocatalyst transfer of intermediates and depletion of substrates, and the second between steric confinement of intermediates versus accessibility of biocatalysts to substrates. We conjecture that the class of optimal geometrical arrangements of biocatalysts within a complex includes those of the famous Thomson problem in electrostatics.

5.1 Introduction

In almost all metabolic process the action of biochemical catalysts is crucial for the processing of reactions that otherwise would not occur on a sufficiently short time scale as to sustain life. These biocatalysts, which are highly interconnected in complex metabolic networks, exhibit a striking degree of spatial organization across all the length-scales of living systems [47], from individual molecules to collections of cells.

At the molecular scale, enzymes that catalyze consecutive reactions in a biochemical pathway are often organized into multi-enzyme complexes, micro-compartments, or other assemblages

*This chapter is adapted from the manuscript: Enhancing biocatalytic fluxes via spatial organization: trade-offs and design principles, by F. Hinzpeter, F. Tostevin, A. Buchner, and U. Gerland, which has been submitted for publication.

[208, 209, 78, 52]. Well known examples include cellulosomes [128], the purinosome [76], and carboxysomes [54]. Such arrangements are thought to enable efficient processing of metabolic intermediates, a process known as metabolic channeling [37] (see Chap. 2). In recent years there have been extensive efforts to design efficient spatially-organized multi-enzyme reactions *in vitro* using a variety of scaffolding and confinement approaches [210, 211, 212]. Similar ideas of efficiently arranging consecutive catalysts are pursued in the realm of concurrent tandem catalysis to improve the yield and specificity of sequential chemical reactions [213, 214, 215]. Understanding the physical principles underlying channeling is crucial for engineering such efficient multi-catalyst systems. However, it remains controversial whether proximity of consecutive enzymes alone is sufficient to allow channeling of diffusing intermediates [216, 116, 117, 125], or whether additional confinement of intermediates or active mechanisms are required [37].

At a higher level of organization, enzyme compartments and complexes themselves can be seen as catalytic reaction centers. Notably, these structures also function synergistically and some have been found to spatially colocalize [217, 218, 219]. For example, purinosomes in HeLa cells were found to localize to mitochondria [219]. It was concluded that the spatial proximity ensures that mitochondrial-derived metabolic intermediates are efficiently captured by purinosomes to enhance nucleotide production [219, 220].

On an even larger scale, whole cells can be considered as catalytic reaction centers. By taking up, processing, and secreting chemicals, cells effectively work as catalysts that alter the chemical composition of their environment. Interestingly, even on this length scale, different cells work together to sequentially process metabolites [221, 222, 223, 224]. For example, biological nitrification, the conversion of ammonia to nitrate via the intermediate nitrite, is performed by two specialized microbes [225]. The first step, the oxidation of ammonia to nitrite, is catalyzed by ammonia-oxidizing bacteria/archaea, while the second step, the conversion of nitrite to nitrate, is performed by nitrite-oxidizing bacteria. These synergistic microbes are found to grow together in spatially structured biofilms [226].

Despite the differences in length scale, the behavior of these systems is often governed by common underlying physical processes. The metabolites of interest are typically small molecules that move by diffusion, while the biocatalysts are typically much larger and are spatially localized or move on a much slower timescale. The reaction fluxes are determined by the kinetic interplay between diffusive transport of metabolites and the reaction kinetics at the specific locations of the biocatalysts.

In previous studies (see Chap. 4) the reaction-diffusion dynamics of spatially arranged biocatalysts have been investigated using continuum models, which do not account for the discrete nature of biocatalysts but describe their arrangement by density profiles on mesoscopic length scales [202, 197, 142]. This prior work analyzed the impact of the large-scale enzyme arrangement on the overall reaction efficiency. However, it is currently unclear how the discrete nature of the biocatalysts affects this efficiency. If multiple biocatalysts are placed in close proximity, as in enzyme clusters or microbial biofilms, the resulting "crowded" geometries lead to a complex spatial network of diffusive fluxes that exchange the participating metabolites between the biocatalysts and with the environment. Many previous studies have characterized the effects of "random crowding" on the diffusion and reaction dynamics of molecules [227, 228, 229, 230, 231]. In contrast, the possibility for "designed crowding", in

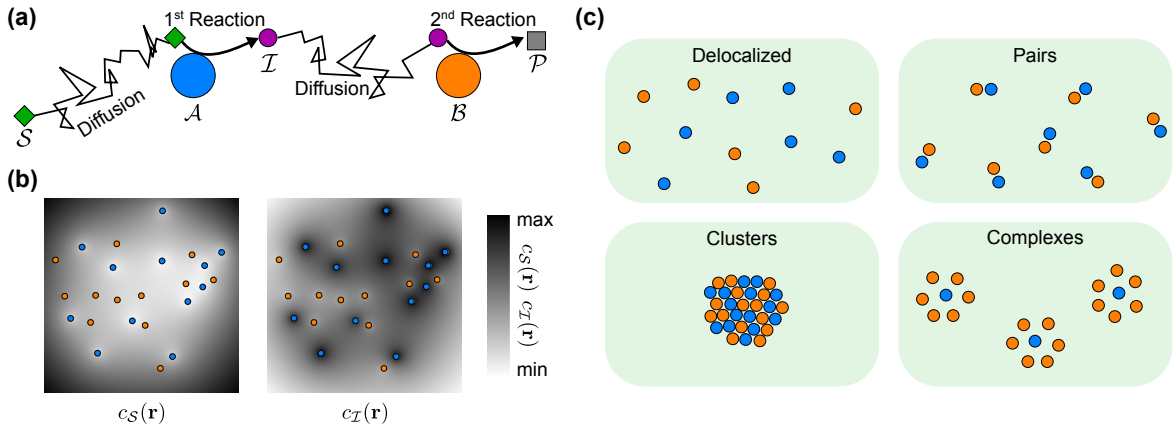

Figure 5.1: Schematic visualization of the reaction-diffusion model with discrete biocatalysts. **(a)** Two-step reaction pathway. Metabolites diffuse and react when they come into contact with the biocatalysts. **(b)** The biocatalysts are modeled as spatially extended objects (colored circles), while the metabolites are represented by their continuous concentration profiles, $c_{\mathcal{S}}(\mathbf{r})$ and $c_{\mathcal{I}}(\mathbf{r})$ (shading). **(c)** Classes of spatial organization considered in this chapter. (1) random, delocalized distribution, (2) Pairs of one $\mathcal{A}$ and $\mathcal{B}$, (3) a cluster of randomly arranged biocatalysts, and (4) complexes with a defined stoichiometry and geometry.

which arrangements of objects are chosen such as to selectively block or direct the diffusion of molecules, and simultaneously to catalyze their biochemical conversion, remains largely unexplored. When and how should the arrangement of individual biocatalysts be tuned such as to promote reactions along a reaction pathway? Which trade-offs and design principles emerge from the interplay of the physical processes described above?

In this chapter, we compare different strategies for spatial biocatalyst organization using a model of discrete biocatalysts together with continuous reaction-diffusion dynamics for metabolites. We identify localization strategies that are advantageous in different parameter regimes of biocatalyst reactivity and metabolite diffusion. We find that in the reaction-limited regime, where the catalytic reaction is slow compared to diffusion, the best strategy is to colocalize the biocatalysts into large clusters. In contrast, in the diffusion-limited regime it is beneficial to form pairs or small complexes of biocatalysts. The enhancement of the reaction flux due to these localization strategies compared to unordered, delocalized arrangements is highest when the biocatalyst concentrations are low. We find that this change of the optimal localization strategy arises from two trade-offs. The first trade-off is the compromise between efficient transfer of intermediates and competition for substrates. The second arises between steric shielding and confinement of metabolites. The interplay of these effects gives rise to non-trivial symmetries of the optimal configurations of a model multi-biocatalyst complex.

5.2 Model

We consider a model two-step catalytic reaction performed by two biocatalysts, described by the scheme, $\mathcal{S} \xrightarrow{\mathcal{A}} \mathcal{I} \xrightarrow{\mathcal{B}} \mathcal{P}$ (see Fig. 5.1a). The first biocatalyst $\mathcal{A}$ converts a substrate $\mathcal{S}$ to an intermediate $\mathcal{I}$, which is subsequently converted to the product $\mathcal{P}$ by the second biocatalyst $\mathcal{B}$.

We assume that the metabolites $\mathcal{S}$ and $\mathcal{I}$ are small molecules that move by diffusion. Their concentrations, $C_{\mathcal{S}}(\mathbf{r},t)$ and $C_{\mathcal{I}}(\mathbf{r},t)$, follow the diffusion equation,

$$\frac{\partial C_{\mathcal{S},\mathcal{I}}(\mathbf{r},t)}{\partial t} = D\nabla^2 C_{\mathcal{S},\mathcal{I}}(\mathbf{r},t), \tag{5.1}$$

where D is the diffusion coefficient, which we assume to be the same for both metabolites. In contrast to the metabolites, biocatalysts are mostly mesoscopic objects (macromolecules, organelles, or cells), which, owing to their large size, are modeled as discrete spatially-extended reaction centers (see Fig 5.1b). We implement the two consecutive reactions through boundary conditions imposed on the metabolite concentrations at the surface, $\partial\mathcal{A}$ and $\partial\mathcal{B}$ of the respective biocatalyst,

$$\left.\begin{aligned} D\nabla C_{\mathcal{S}}(\mathbf{r},t)\cdot\mathbf{n} - \tfrac{k_{\mathcal{A}}}{A_{\mathcal{A}}}C_{\mathcal{S}}(\mathbf{r},t) = 0 \\ D\nabla C_{\mathcal{I}}(\mathbf{r},t)\cdot\mathbf{n} + \tfrac{k_{\mathcal{A}}}{A_{\mathcal{A}}}C_{\mathcal{S}}(\mathbf{r},t) = 0 \end{aligned}\right\} \forall\mathbf{r}\in\partial\mathcal{A}, \tag{5.2}$$

$$\left.\begin{aligned} D\nabla C_{\mathcal{S}}(\mathbf{r},t)\cdot\mathbf{n} = 0 \\ D\nabla C_{\mathcal{I}}(\mathbf{r},t)\cdot\mathbf{n} - \tfrac{k_{\mathcal{B}}}{A_{\mathcal{B}}}C_{\mathcal{I}}(\mathbf{r},t) = 0 \end{aligned}\right\} \forall\mathbf{r}\in\partial\mathcal{B}, \tag{5.3}$$

where $k_{\mathcal{A}}$ and $k_{\mathcal{B}}$ are the intrinsic catalytic activities of $\mathcal{A}$ and $\mathcal{B}$, $A_{\mathcal{A}}$ and $A_{\mathcal{B}}$ are the surface areas of biocatalyst $\mathcal{A}$ and $\mathcal{B}$, respectively, and $\mathbf{n}$ is a unit vector normal to the surface. Equation 5.2 represents the conversion of $\mathcal{S}$ to $\mathcal{I}$ catalyzed by $\mathcal{A}$. Equation 5.3 describes the consumption of $\mathcal{I}$ by $\mathcal{B}$, as well as a no flux condition for $\mathcal{S}$ at the surface of $\mathcal{B}$, since there is no reactive interaction between $\mathcal{S}$ and $\mathcal{B}$. In Eqs. 5.2-5.3 we have neglected saturation of the biocatalysts, assuming we are in the low metabolite concentration regime. Here the catalytic activities k_l, $l\in\mathcal{A},\mathcal{B}$, are effective parameters that represent the rate of reactions for metabolites at the surface of a biocatalyst, taking into account short diffusive excursions between unreactive collisions. They differ from the macroscopically observable catalytic activities, $\kappa_l = k_{\mathrm{cat}}/K_{\mathrm{M}}$, which also include the effective timescale of metabolite-biocatalyst interactions via diffusion. Such macroscopic rates κ_l are usually modeled as a combination of the effective intrinsic reaction rate, k_l and the rate at which substrate arrives at biocatalysts via diffusion, k_D, $\kappa_l^{-1} = k_l^{-1} + k_D^{-1}$ [232, 233] (see Sec. 1.1.5). In particular, while the macroscopic reaction rate approaches a finite constant value $\kappa_l \to k_D$ in the diffusion-limited regime, k_l can become arbitrarily large.

The reaction cascade is supplied with $\mathcal{S}$ from the surrounding environment, which we modeled as a homeostatic level C_0 of $\mathcal{S}$ at the system boundary. Intermediates were assumed not to be present in the environment and, thus, their concentration is negligible at the system boundary, $\partial\Omega$. These conditions can be written as

$$\left.\begin{aligned} C_{\mathcal{S}}(\mathbf{r},t) = C_0 \\ C_{\mathcal{I}}(\mathbf{r},t) = 0 \end{aligned}\right\} \forall\mathbf{r}\in\partial\Omega. \tag{5.4}$$

As a measure of the collective performance of the biocatalysts we focused on the steady-state production rate of $\mathcal{P}$,

$$J_{\mathcal{P}} = \frac{k_{\mathcal{B}}}{A_{\mathcal{B}}}\int_{\partial\mathcal{B}} C_{\mathcal{I}}(\mathbf{r})\mathrm{d}\mathbf{r}, \tag{5.5}$$

where the integral is taken over the surface of all $\mathcal{B}$'s. Analogously, we can define the rate of $\mathcal{I}$ production as,

$$J_{\mathcal{I}} = \frac{k_{\mathcal{A}}}{A_{\mathcal{A}}} \int_{\partial \mathcal{A}} C_{\mathcal{S}}(\mathbf{r}) \mathrm{d}\mathbf{r}. \tag{5.6}$$

In this chapter we are primarily interested in how the spatial organization of the synergistic biocatalysts influences the pathway flux for a given number of biocatalysts $\mathcal{A}$ and $\mathcal{B}$ ($N_{\mathcal{A}}$ and $N_{\mathcal{B}}$). In addition to all the model parameters, $J_{\mathcal{P}}$ will also depend on the shape and size of the biocatalysts and the system geometry $\delta\Omega$. For simplicity we assume that $\mathcal{A}$ and $\mathcal{B}$ have the same size and a spherical shape, such that $A_{\mathcal{A}} = A_{\mathcal{B}} = S_d r_c^{d-1}$ in a model of dimension d with S_d the surface area of a d-dimensional unit sphere. We also assume a spherical geometry for the system with a fixed radius $100 r_c$. Note that taking $r_c = 2\,\mathrm{nm}$ on the typical scale of a single enzyme molecule results in an effective enzyme concentration in the higher nanomolar range for the small values of $N_{\mathcal{A}}$ and $N_{\mathcal{B}}$ considered here, which is within the range expected under intracellular conditions. Rescaling all lengths with the biocatalyst radius, r_c, we can identify from Eqs. 5.2-5.3 two dimensionless reaction-diffusion parameters, $\alpha_{\mathcal{A}} = k_{\mathcal{A}} / \left(S_d D r_c^{d-2}\right)$ and $\alpha_{\mathcal{B}} = k_{\mathcal{B}} / \left(S_d D r_c^{d-2}\right)$, that together with the dimensionless system radius determine the metabolite concentration profiles. Finally, since Eqs. 5.1-5.3 are linear in the metabolite concentrations, these can be normalized by C_0: $c_{\mathcal{S},\mathcal{I}}(\mathbf{r}) = C_{\mathcal{S},\mathcal{I}}(\mathbf{r})/C_0$. Rewriting $J_{\mathcal{I},\mathcal{P}}$ in terms of these dimensionless variables and rescaling the time by r_c^2/D, we can identify dimensionless reaction fluxes,

$$j_{\mathcal{I}} = \frac{J_{\mathcal{I}}}{J_D} = \alpha_{\mathcal{A}} \int_{\partial \mathcal{A}} c_{\mathcal{S}}(\mathbf{r}) d\mathbf{r}, \qquad j_{\mathcal{P}} = \frac{J_{\mathcal{P}}}{J_D} = \alpha_{\mathcal{B}} \int_{\partial \mathcal{B}} c_{\mathcal{I}}(\mathbf{r}) d\mathbf{r}, \tag{5.7}$$

where $J_D = D r_c^{d-2} C_0$.

The reaction fluxes for different biocatalyst arrangements are calculated by numerically solving the steady-state nondimensionalized versions of Eq. 5.1-5.4 using COMSOL Multiphysics. To obtain a comprehensive understanding that applies for different biological scenarios like biofilms of syntrophic bacteria attached to a surface or clusters of enzymes in the cellular cytoplasm, we study both the two- and three-dimensional cases.

5.3 Results

5.3.1 Randomly positioned biocatalysts and correlations between arrangements and reaction fluxes

As a reference against which specific localization strategies can be compared, we begin by considering an ensemble of random biocatalyst arrangements, with the aim of identifying characteristics of these arrangements that correlate with changes in the reaction flux. Additionally, such an ensemble approximates the scenario of freely-diffusing biocatalysts: due to their different sizes, metabolites will typically diffuse much faster than the biocatalysts, and it is reasonable to assume that the biocatalysts are fixed in position on a time scale the metabolites need to reach steady state. Therefore, the expected pathway flux in a system of diffusing biocatalysts is equivalent to an ensemble average over all configurations of biocatalysts.

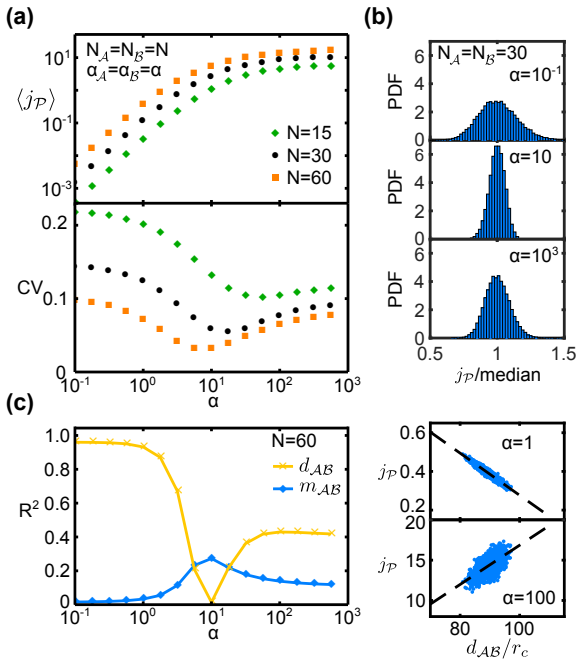

Figure 5.2: Reaction flux for delocalized biocatalysts in two dimensions. Parameters are chosen symmetrically with $N_{\mathcal{A}} = N_{\mathcal{B}} = N$ and $\alpha_{\mathcal{A}} = \alpha_{\mathcal{B}} = \alpha$. **(a)** The mean flux $\langle j_{\mathcal{P}} \rangle$ (top) and coefficient of variation CV (bottom) as α is varied for different values of N, computed from an ensemble of 3000 randomly generated biocatalyst arrangements. **(b)** Histogram of reaction fluxes for $N = 30$ at three different values of α. **(c)** Coefficient of determination R^2 of the linear regression of $j_{\mathcal{P}}$ against $d_{\mathcal{AB}} = \langle |\mathbf{r}_i - \mathbf{r}_j| \rangle_{i \in \mathcal{A}, j \in \mathcal{B}}$ (crosses) and $m_{\mathcal{AB}} = \langle \min_{j \in \mathcal{B}} |\mathbf{r}_i - \mathbf{r}_j| \rangle_{i \in \mathcal{A}}$ (diamonds), and regression plots against $d_{\mathcal{AB}}$ at low and high values of α.

Figure 5.2a (top) shows the mean pathway flux from 3000 randomly-generated biocatalyst arrangements (see SM Sec. C.1) for different values of the dimensionless reaction-diffusion parameters and biocatalyst abundances, which for simplicity we chose symmetrically ($\alpha_{\mathcal{A}} = \alpha_{\mathcal{B}} = \alpha$ and $N_{\mathcal{A}} = N_{\mathcal{B}} = N$). Here we limited ourselves to a two-dimensional geometry, where the smaller configuration space allows for a more thorough sampling of possible arrangements than in three dimensions. Nevertheless, we expect similar behavior in three dimensions.

In Fig. 5.2a we see that for $\alpha \lesssim 1$ the mean flux increases as $\langle j_{\mathcal{P}} \rangle \sim \alpha^2$, while for $\alpha \gg 10$ the mean flux saturates at a constant value. This behavior reflects the transition of the system from a reaction-limited regime at small α, in which each reaction rate is limited by the probability $p \sim \alpha \ll 1$ that a metabolite-biocatalyst encounter will result in a reaction (with the quadratic increase in $\langle j_{\mathcal{P}} \rangle$ coming from the fact that the pathway consists of the two reaction that each scale with α), to a diffusion-limited regime at large α in which most encounters are reactive, $p \approx 1$, and the rate of reactions is instead set by the frequency of such encounters due to diffusion.

To investigate how sensitive the pathway flux is to the arrangement of biocatalysts we examined the distribution of $j_{\mathcal{P}}$ values at different values of α (Fig. 5.2b). Interestingly, we found that the distribution was significantly narrower around $\alpha \approx 10$ than at smaller or larger α values. The coefficient of variation, $CV = \langle (j_{\mathcal{P}} - \langle j_{\mathcal{P}} \rangle)^2 \rangle^{1/2} / \langle j_{\mathcal{P}} \rangle$, shows a non-monotonic dependence on the reaction rate parameter α, with a minimum at the transition between reaction- to diffusion-limited regimes (see Fig. 5.2a, bottom). At this point, the configuration of the biocatalysts has the least impact on the reaction flux. On the other hand, at small or large α the CV is larger, indicating that the reaction flux is more sensitive to the particular arrangement of biocatalysts.

We next sought properties of the biocatalyst arrangements that correlated with changes in $j_{\mathcal{P}}$ in different α regimes. For small α (the reaction-limited regime), the pathway flux of a

given configuration was strongly anti-correlated with the mean distance between $\mathcal{A}$s and $\mathcal{B}$s, $d_{\mathcal{AB}} = \langle|\mathbf{r}_i - \mathbf{r}_j|\rangle_{i\in\mathcal{A},j\in\mathcal{B}}$ ($R^2 > 0.95$, Fig. 5.2c). For large α (the diffusion-limited regime), there was a somewhat weaker ($R^2 \approx 0.4$) but positive correlation between these quantities. However, in the transition region, the correlation became very small ($R^2 \approx 0.01$). Similar trends, albeit with weaker correlations, were also observed for $d_{\mathcal{AA}}$, $d_{\mathcal{BB}}$, and the mean radial coordinates $r_{\mathcal{A}} = \langle|\mathbf{r}_i|\rangle_{i\in\mathcal{A}}$ and $r_{\mathcal{B}} = \langle|\mathbf{r}_i|\rangle_{i\in\mathcal{B}}$ (see SM Fig. C.1). In contrast, in the transition region we observed a significant (positive) correlation between the flux and the minimal distance between $\mathcal{A}$s and $\mathcal{B}$s, $m_{\mathcal{AB}} = \langle\min_{j\in\mathcal{B}}|\mathbf{r}_i - \mathbf{r}_j|\rangle_{i\in\mathcal{A}}$ ($R^2 \approx 0.27$).

Together these data indicate that for small α the highest flux is generated by configurations in which the biocatalysts are generally placed closer together, and closer to the center of the system. In contrast, for large α the highest flux comes from configurations where the reaction centers are further apart, and closer to the periphery of the system. At the point where the variability in the flux is smallest, both the best- and worst-performing configurations have similar mean separations. However there remains an advantage to placing biocatalysts such that each is in close proximity to at least one reaction center of the other type.

5.3.2 Comparison of different localization strategies

Having seen that the reaction flux depends on such quantities as the average and minimal distance between biocatalysts, we now investigate in more detail two extreme localization strategies that emphasize these properties (see Fig. 5.1c). These are (i) a single dense but disordered cluster of biocatalysts; and (ii) pairs of biocatalysts consisting of one $\mathcal{A}$ and one $\mathcal{B}$ at a separation of r_c. We again consider mean fluxes averaged over an ensemble of configurations where the biocatalysts are either paired or clustered (see SM Sec. C.1). Figure 5.3a and b show the expected reaction flux of the different spatial organizations as a function of α in two and three dimensions.

Comparing first pair arrangements with delocalized arrangements, in which the two types of biocatalysts are positioned independently (see Fig. 5.1c), the mean flux of the pair arrangements was always larger. This reflects the fact that placing $\mathcal{A}$ and $\mathcal{B}$ in close proximity increases the probability that each $\mathcal{I}$ that is produced will encounter a $\mathcal{B}$ and react before diffusing to the boundary of the system. The magnitude of improvement was largest at low biocatalyst concentrations (small N) and in three dimensions (see Fig. 5.3c and d), reaching an enhancement of almost 7-fold for $N = 20$, which for an enzymatic system corresponds to a biocatalyst concentration of $\sim$1 µM. For biocatalysts free in solution, a low concentration corresponds to a large mean separation between consecutive biocatalysts. It is therefore relatively unlikely that a given $\mathcal{I}$ molecule will encounter a $\mathcal{B}$ before diffusing out of the system. However, at high concentrations the free biocatalysts are on average close together, and hence the increase in encounter probability due to positioning biocatalysts in close proximity becomes less significant.

Turning now to the clustering strategy, we found that for small α the clustered configurations achieve a significantly higher mean flux than either the delocalized and the pair arrangements. This enhancement was approximately ten-fold in two-dimensions and hundred-fold in three-dimensions for similar numbers of biocatalysts (see Fig. 5.3a and b). However, as α was increased, we observe a transition into a regime where the pair arrangements produce a

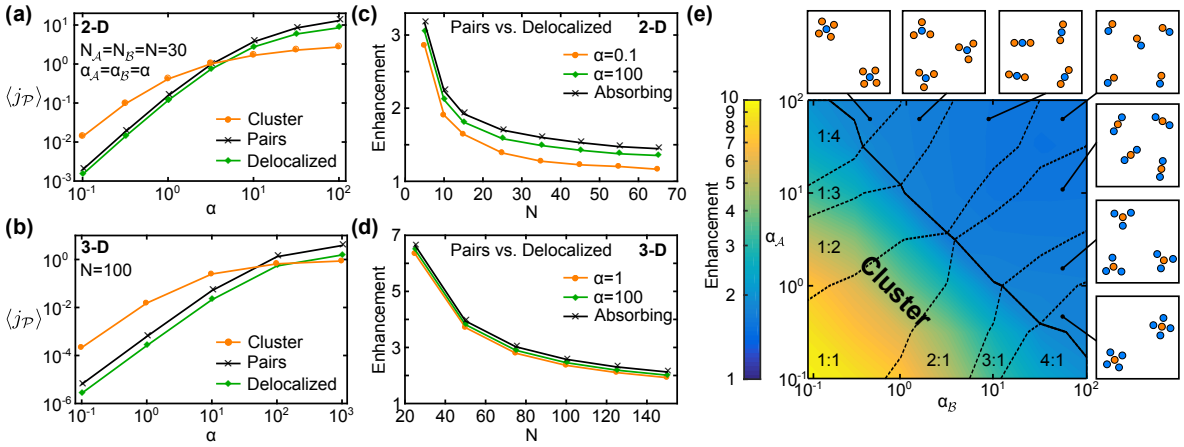

Figure 5.3: Comparison of different localization strategies. (a-b) Comparison of the mean reaction fluxes of clustered arrangements, $\mathcal{A}$-$\mathcal{B}$-pair arrangements, and delocalized arrangements for $\alpha_{\mathcal{A}} = \alpha_{\mathcal{B}} = \alpha$ and $N_{\mathcal{A}} = N_{\mathcal{B}} = N$. **(c-d)** Enhancement of the mean flux in pair arrangements relative to delocalized arrangements as the biocatalyst number is varied. **(a)** & **(c)** show results for a two-dimensional system, **(b)** & **(d)** for a three-dimensional system. **(e)** Phase diagram showing the optimal spatial arrangement and biocatalyst stoichiometry as $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{B}}$ are varied, while the total number of biocatalysts $N_{\mathcal{A}} + N_{\mathcal{B}} = 60$ remains constant. The solid line separates the regime where clustering is the optimal strategy from that in which small complexes produce a higher flux; dashed lines denote changes in the ratio $N_{\mathcal{A}}$:$N_{\mathcal{B}}$. Color shows the enhancement of the mean flux relative to delocalized arrangements.

higher mean flux than the cluster. Notably, as α is further increased, even the delocalized arrangements outperformed the cluster.

To generalize these observations we considered scenarios where the α values of the two reaction steps and the number of biocatalysts of the two types were not equal. Figure 5.3e summarizes these results in the form of a phase diagram showing the configuration that produced the highest expected flux for different $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{B}}$ values, subject to the constraint $N_{\mathcal{A}} + N_{\mathcal{B}} = 60$. We see that in the region defined approximately by $\alpha_{\mathcal{A}}\alpha_{\mathcal{B}} \leq 10$, the highest flux was produced by a single cluster (Fig. 5.3e, solid line). For larger $\alpha_{\mathcal{A}}\alpha_{\mathcal{B}}$, on the other hand, small complexes produced a higher flux. In both regimes, the relative values of $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{B}}$ determined the optimal biocatalyst stoichiometry, favoring a larger number of $\mathcal{B}$ than $\mathcal{A}$ in regions with $\alpha_{\mathcal{A}} \gg \alpha_{\mathcal{B}}$, but more $\mathcal{A}$ than $\mathcal{B}$ when $\alpha_{\mathcal{A}} \ll \alpha_{\mathcal{B}}$. Interestingly, despite the difference in boundary conditions that apply for $\mathcal{S}$ and $\mathcal{I}$, the phase diagram appears to be approximately symmetrical about the line $\alpha_{\mathcal{A}} = \alpha_{\mathcal{B}}$.

The transition between a clustered arrangement being optimal at small α but a more disperse arrangement being preferable at large α is reminiscent of the crossover described in Chap. 4 for the optimal enzyme profile around a localized source (see also [202, 197]). However the model considered in this chapter includes many physical effects that were not considered in [202, 197]. These include the impact of the biocatalyst arrangement on the first reaction flux, as well as steric effects due to the discreteness of biocatalysts. What are the contributions of these different physical effects to the behavior observed in Fig. 5.3?

Trade-off between substrate depletion and efficient intermediate exchange

We first considered how the pathway flux would vary if we were to eliminate steric effects. In such a scenario, changing the relative positioning of biocatalyst can still affect the pathway flux via the combined effects of the localized reaction centers on the metabolite density profiles.

We have seen above that placing one $\mathcal{A}$ and one $\mathcal{B}$ in close proximity to one another (pair arrangement) is always beneficial compared to the delocalized arrangement. Placing several $\mathcal{A}$s and $\mathcal{B}$s into a large cluster is even more favorable for small α. In this regime, an $\mathcal{I}$ molecule is very likely to escape to the absorbing boundary without reacting with $\mathcal{B}$ even if the biocatalysts are arranged in pairs, since the probability of reaction at each $\mathcal{I}$-$\mathcal{B}$ encounter is low. This loss can be attenuated by clustering several copies of $\mathcal{A}$ and $\mathcal{B}$. As previously described by [142], an $\mathcal{I}$ molecule produced in such a cluster has a higher probability to be processed by any of the proximal $\mathcal{B}$s, even if the probability of reaction with each individual $\mathcal{B}$ is low. As α becomes larger the probability of reaction at each $\mathcal{I}-\mathcal{B}$ encounter increases, which reduces the benefit of clustering many $\mathcal{B}$ in the vicinity of a single $\mathcal{A}$. At the same time, since each $\mathcal{A}$ consumes more of the $\mathcal{S}$ molecules that approach it, steeper concentration gradients develop around the cluster. This depletion of substrate reduces the rate with which the many $\mathcal{A}$s in the cluster produce $\mathcal{I}$ as they are effectively forced to compete for substrate. This makes it increasingly unfavorable to position the $\mathcal{A}$s into close proximity to one another [234, 235, 236]. Therefore a more disperse arrangement in which the distance between biocatalysts of the same type is large, becomes preferable.

To study quantitatively to what extent these two conflicting effects of substrate depletion and efficient transfer of intermediate are responsible for the transition observed in Fig. 5.3, we considered a model wherein metabolites were able to diffuse through the space occupied by biocatalysts, and reactions occurred throughout their volume rather than on their surfaces (see SM Sec. C.3). In this way we eliminate any influence from steric exclusion by biocatalysts. In Fig. 5.4a-b (crosses) we plot the flux of the full model with impermeable biocatalysts divided by the corresponding flux for permeable biocatalysts. We see that there is little difference in the fluxes for delocalized configurations (see SM Fig. C.2). For the clustered configurations the impermeability of reaction centers leads to an increase of $30-40\%$ in $\langle j_{\mathcal{P}} \rangle$ at small α but a decrease of up to $15-25\%$ at large α. While these changes are significant, they are nonetheless small compared to the order-of-magnitude differences observed between the different arrangement strategies in Fig. 5.3. Hence the pathway fluxes in this modified model showed qualitatively similar behavior to that shown in Fig. 5.3a-b, including a crossover from the cluster to pairs being the preferred biocatalyst arrangement. Therefore, the transition from a cluster to small complexes being preferable does not result from steric effects, but can be understood as a trade-off between the efficiency of the two reactions in the pathway, as described above. However, the position of the transition and the values of the fluxes are not correctly captured by this reduced model. Thus, steric effects must nevertheless be included for a full quantitative understanding of the system.

Trade-off between metabolite shielding and confinement

To understand the relevant steric effects in more detail we considered the two reactions individually. Additionally to the fluxes defined in Eq. 5.7, we define the efficiency of the second

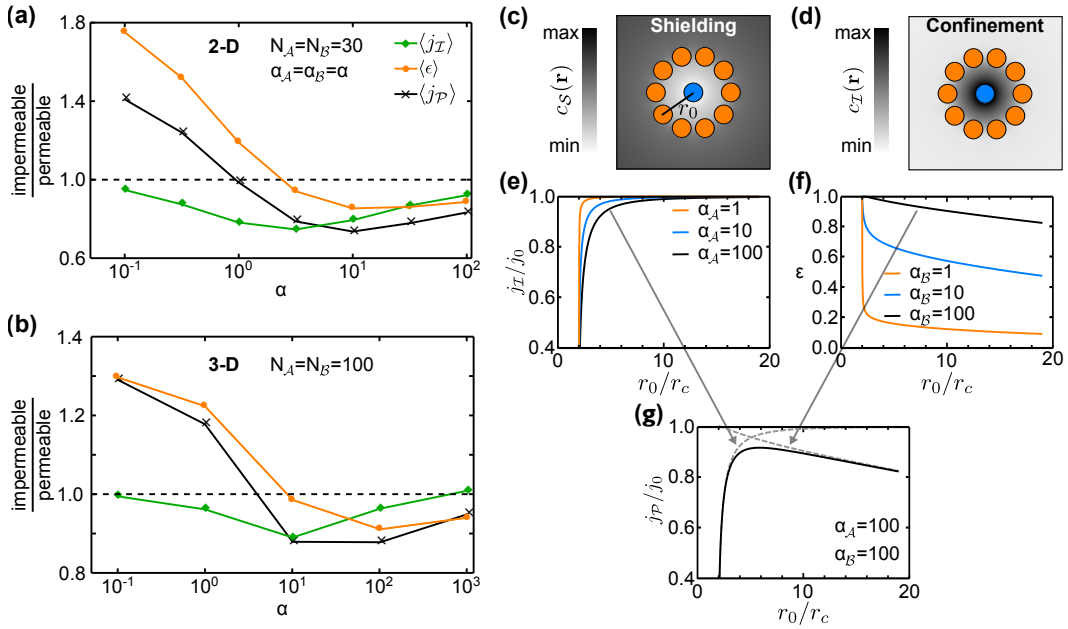

Figure 5.4: Illustration of shielding and confinement effects. (a-b) Effect of biocatalyst impermeability on $\langle j_\mathcal{P} \rangle$, $\langle j_\mathcal{I} \rangle$, and $\langle \epsilon \rangle$, for clustered arrangements. Plotted are the ratios (impermeable:permeable) of the respective quantities for **(a)** a two-dimensional system and **(b)** a three-dimensional system. Lines are included as a guide to the eye. **(c)** Shielding of diffusing $\mathcal{S}$ from a central $\mathcal{A}$ by a surrounding ring of 10 $\mathcal{B}$s. **(d)** Confinement of $\mathcal{I}$ within a ring of $\mathcal{B}$s. **(e)** $j_\mathcal{I}$, **(f)** ϵ, and **(g)** $j_\mathcal{P}$ as a function of the ring radius for a ring consisting of 6 $\mathcal{B}$s. Fluxes are plotted relative to $j_\mathcal{I}$ in the absence of a $\mathcal{B}$ ring, denoted by j_0. The trade-off between $\mathcal{S}$ shielding and $\mathcal{I}$ confinement leads to an optimal ring radius that maximizes the pathway flux $j_\mathcal{P}$.

reaction as $\epsilon = j_\mathcal{P}/j_\mathcal{I}$, the fraction of the produced $\mathcal{I}$'s that are converted into $\mathcal{P}$. Figures 5.4a,b show that for the first reaction in the pathway, when reaction centers obstruct diffusion in the clustered configuration, $\langle j_\mathcal{I} \rangle$ is reduced across the full range of α, but this effect is strongest at intermediate α values. In contrast, $\langle \epsilon \rangle = \langle j_\mathcal{P}/j_\mathcal{I} \rangle$ displays the same qualitative behavior as $\langle j_\mathcal{P} \rangle$ but with larger amplitude. These results arise from two effects, namely the "shielding" of biocatalysts within the cluster from metabolites, and "confinement" or diffusive trapping of intermediates within the cluster.

To demonstrate these effects, we considered the special arrangement of several $\mathcal{B}$s arranged on a ring of radius r_0 around a central $\mathcal{A}$. Such an arrangement could approximate the environment around a single $\mathcal{A}$ in a cluster, and allows us to monitor how the flux of each reaction varies as a function of the clustering density, measured by r_0, and $\alpha_\mathcal{A}$, $\alpha_\mathcal{B}$.

Focusing initially on the first reaction, we see that $j_\mathcal{I}$ decreases sharply as r_0 is decreased below a critical radius (Fig. 5.4e). Below this radius, the $\mathcal{B}$ ring blocks the diffusion of $\mathcal{S}$ into the vicinity of the $\mathcal{A}$, thereby leading to a sharp reduction in $c_\mathcal{S}$ (Fig. 5.4c). The ring of $\mathcal{B}$s effectively shields the central $\mathcal{A}$ from substrate. The magnitude of the reduction in $j_\mathcal{I}$, and the radius at which $j_\mathcal{I}$ starts to decrease, both increase with $\alpha_\mathcal{A}$. This is because the flux depends on the product of the probability of an $\mathcal{S}$ to diffuse through the ring, which decreases with decreasing r_0, and the probability of an $\mathcal{S}$ within the ring to react with the central $\mathcal{A}$. The latter is an increasing function of $\alpha_\mathcal{A}$. For the second reaction, the presence

of the $\mathcal{B}$ ring restricts the diffusion of $\mathcal{I}$ out of the ring, increasing the local concentration of $\mathcal{I}$ within the ring (Fig. 5.4d) and therefore the efficiency with which $\mathcal{I}$ is converted to $\mathcal{P}$ (Fig. 5.4f). This confinement effect is largest at small $\alpha_{\mathcal{B}}$, when the probability of reaction in each $\mathcal{I}-\mathcal{B}$ encounter is lowest. The trade-off between shielding of substrate and confinement of $\mathcal{I}$ leads to the emergence of an optimal ring radius at the point where the decline of $j_{\mathcal{I}}$ due to shielding is exactly balanced by the increased efficiency of $\mathcal{I}$-processing achieved by the confinement effect (see Fig. 5.4g).

Returning now to the scenario of a dense cluster containing both $\mathcal{A}$ and $\mathcal{B}$, we conclude from Figs. 5.4a-b that shielding tends to reduce $\langle j_{\mathcal{I}} \rangle$ most significantly at intermediate values of α. When α is small, reactions are slow and $\mathcal{S}$ molecules nevertheless have sufficient time to diffuse throughout the cluster. At the opposite extreme of large α and fast reactions, $\mathcal{A}$ biocatalysts on the periphery of the cluster are effectively able to consume most of the available $\mathcal{S}$, such that little substrate reaches the center of the cluster even when diffusion is unimpeded.

For the second reaction, the confinement of $\mathcal{I}$ that is produced within the cluster increases the conversion efficiency ϵ predominantly at small α. In the large-α regime, however, shielding also dominates and reduces the efficiency of the second reaction. Here, since $\mathcal{I}$ is produced primarily at the periphery of the cluster, it is effectively shielded from $\mathcal{B}$s within the cluster, increasing the chance of it diffusing out of the system rather than reacting.

5.3.3 Optimal complex configurations

Besides the largely disordered aggregation of biocatalysts into large clusters, as we have considered in our random cluster model, it is known that collections of enzymes in cells organize into highly specific arrangements, with each type of enzyme located at defined positions [237]. A well known example of such a multi-enzyme complex is the pyruvate dehydrogenase complex (PDC), which consists of three consecutive enzymes that transform pyruvate to acetyl-CoA. Depending on the cell type, the PDC comprises different copy numbers of the sequential enzymes, yet these are arranged in a highly symmetric fashion [35, 238]. The precise spatial organization of the enzymes is required for efficient function of such biocatalytic machines [238].

Inspired by such highly-efficient multi-enzyme complexes we asked how the biocatalysts in a model multi-biocatalyst complex of a single $\mathcal{A}$ surrounded by several $\mathcal{B}$s should be arranged in order to globally maximize $j_{\mathcal{P}}$. As we found above for ring arrangements, we expect that the optimal geometry will arise as a compromise between the advantageous effects of proximity and confinement and the detrimental effects of shielding and intermediate depletion. Since the relative magnitudes of these different effects depend on the parameters $\alpha_{\mathcal{A}}$ and $\alpha_{\mathcal{B}}$, the optimal complex geometry will also crucially depend on these parameters. Therefore, we determined numerically the optimal configurations for different $\mathcal{B}$ numbers, $N_{\mathcal{B}}$, and for different $\alpha_{\mathcal{B}}$ with $\alpha_{\mathcal{A}}$ fixed (see SM Sec. C.2).

The resulting optimal configurations are shown in Fig. 5.5. Surprisingly, we found that the symmetries of the optimal configurations change qualitatively with $N_{\mathcal{B}}$ and $\alpha_{\mathcal{B}}$. In two-dimensions and for $N_{\mathcal{B}} \leq 6$ we found that it is always optimal to arrange the $\mathcal{B}$s at equidistant positions on a concentric ring (point symmetry group $D_{N_{\mathcal{B}}}$), whose radius increases with $\alpha_{\mathcal{B}}$

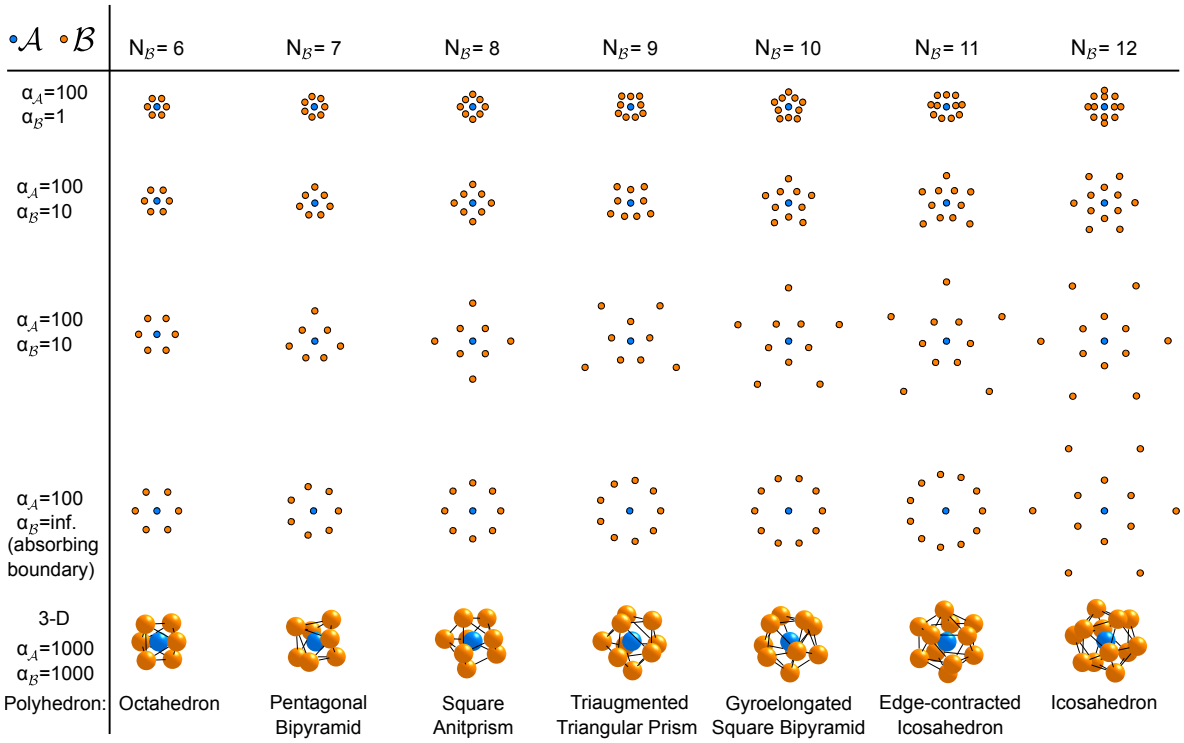

Figure 5.5: Optimal biocatalyst complex arrangements. Optimal arrangement of biocatalyst complexes with several $\mathcal{B}$s (orange) surrounding a single $\mathcal{A}$ (blue) in two dimensions, as the reaction rate $\alpha_\mathcal{B}$ and the complex stoichiometry are varied. Bottom line shows the conjectured optimal arrangements in three dimensions which, we conjecture, match the solutions of the Thomson problem [239].

as described above. In contrast, for $N_\mathcal{B} > 6$ at finite $\alpha_\mathcal{B}$ the $\mathcal{B}$s are divided between an inner ring and an outer population that are arranged at angular positions corresponding to the gaps in the inner ring. This appears to be a strategy to attenuate the shielding effect while maintaining efficient processing of $\mathcal{I}$ to $\mathcal{P}$. For even $N_\mathcal{B}$, these arrangements are concentric rings that are rotated by $2\pi/N_\mathcal{B}$ with respect to one another (point group $D_{N_\mathcal{B}/2}$). For odd $N_\mathcal{B}$, it is not possible to form two full rings and thus the arrangement shows only a single reflection symmetry axis (point group D_1). Interestingly, in the limit of extremely large $\alpha_\mathcal{B}$, the optimal arrangement changes from these star-like arrangements back to a single ring (point group $D_{N_\mathcal{B}}$), provided $N_\mathcal{B} < 12$.

We also sought to identify the optimal configurations in three dimensions. However, the increase in the space of possible configurations made the optimization procedure very slow and prone to getting stuck in local optima, which makes a global optimization computationally difficult. Nevertheless, the best-performing configurations that we identified consisted of $\mathcal{B}$s arranged on the surface of a sphere concentric with the $\mathcal{A}$, with radius that increases with $N_\mathcal{B}$ (see Fig. 5.5). In contrast to the two-dimensional case, in three dimensions we never observed the division of a second outer $\mathcal{B}$ population, although we cannot rule out that this still occurs at higher $N_\mathcal{B}$. Except in special cases ($N_\mathcal{B} = 2, 3, 4, 6, 8, 12$) it is not possible to arrange points on the sphere such that all edges are of equal length; thus the $\mathcal{B}$s are not all equidistant from their neighbors. Interestingly, the optimal configurations generated by our numerical approach are similar to solutions of the well-known Thomson problem from classical electrostatics

[239], where the objective is to minimize the electrostatic interaction energy of identical point charges on the surface of a sphere. Here, however, instead of interactions being defined by an identical local potential around each charge, the biocatalysts in our model effectively interact via the metabolite concentration fields, which depend on the positions of all biocatalysts. Interestingly, when we calculated the reaction flux for the exactly known configurations of the Thomson problem, we found that these always achieved a slightly higher reaction flux than any configuration found during our numerical optimization. Additionally, when we initialized the random search optimization procedure with the configuration of the Thomson Problem, the algorithm was not able to identify any better configuration. We therefore conjecture that in three dimensions, the Thomson problem configurations also optimize the reaction flux for the model described here, provided that the radius of the sphere on which the $\mathcal{B}$s are arranged is chosen optimally.

5.4 Discussion

We have investigated how the spatial coordination of two model biocatalysts affects their ability to synergistically convert substrate to product via an intermediate molecule. Our model predicts that choosing the most favorable localization strategy can provide a several-fold increase in pathway flux. In the reaction-limited regime of slow reactions or fast diffusion, it is beneficial to form large clusters of biocatalysts. On the other hand in the diffusion limited regime, when reactions are fast compared to diffusion, it is preferable to form pairs or small complexes of a few of each type of biocatalyst in close proximity. Biocatalyst systems on different scales are likely to span these different regimes. For example, for a small intracellular substrate with diffusion coefficient $D \approx 100\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, a typical enzyme with catalytic efficiency $k_l \approx k_{\mathrm{cat}}/K_{\mathrm{M}} \sim 10^6\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ and radius $r_c \approx 2\,\mathrm{nm}$ will lie well within the regime where clustering is preferred, with $\alpha \sim 10^{-3}$. However, a fast, diffusion-limited enzyme, for which $k_l \gg k_{\mathrm{cat}}/K_{\mathrm{M}} \sim 10^9\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$, will be in the regime $\alpha \gg 1$ where pairs or small complexes of enzymes lead to a larger flux. We expect systems of synergistic bacteria to be mostly in the reaction-limited regime, while for fast metabolite uptake the transition region from reaction to diffusion limited may be reached. Taking as an example parameters for the uptake of ammonia by *N. maritimus*, which functions as an ammonia oxidizer in two-step nitrification ($k_{\mathrm{cat}}/K_{\mathrm{M}} \simeq 2.6 \times 10^{13}\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$ [240, 241], $r_c = 0.4\,\mu\mathrm{m}$, $D = 1000\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$), we obtain $\alpha \simeq 10$, which lies within the transition region.

The general phenomenology that we have observed arises from two trade-offs in the reaction-diffusion dynamics of metabolites. In the first case, placing many biocatalysts of both types into a large cluster increases the efficiency with which intermediate is converted into product. However, clustering many biocatalysts of the first type leads to local depletion of substrate, limiting the rate of the first reaction. In the second trade-off, clustering of biocatalysts leads to shielding of interior biocatalysts, further limiting their access to substrate. On the other hand, confinement of intermediates produced within the cluster increases the number of potential interactions with biocatalysts of the second type. Notably, the second trade-off is the result of steric effects arising from the discrete nature of biocatalysts. We showed that this trade-off can have a strong impact, altering reaction fluxes up to 40%. For clusters containing a larger number of biocatalysts, we expect that this effect becomes even stronger. Indeed to alleviate this effect it has been found that bacteria in a biofilm collectively form channels into the

biofilm to supply the bacteria in the interior with nutrients [242].

Inspired by the natural strategies of spatial biocatalyst organization, synthetic biologists are seeking to engineer precisely controlled arrangements of consecutive enzymes with the aim to increase pathway fluxes. In several experiments it has been found that tethering pairs of enzymes to small DNA scaffolds increases the rate of product formation [117, 116, 243, 118]. However, in a recent experiment where the consecutive enzymes were directly fused together, no enhancement was observed [125]; earlier reports of enhancement were instead attributed to local changes in pH around the DNA-based scaffolds that increase the enzymatic activities of the individual enzymes. Importantly, in many experiments the intermediates are not subject to any loss to the surrounding environment or to competing reactions. In such scenarios any enhancement effect would be only transient and steady-state fluxes will not be affected by spatial organization. Consequently, the observed enhancements are most likely due to changes in the intrinsic activity of the enzymes. In our model, we assumed that intermediates are lost through the system boundary. With this loss mechanism a considerable enhancement in metabolic flux arose solely from proximity, without the need for any other physical effects and assuming that the properties of the enzymes themselves are unchanged. While this strong loss mechanism may lead to an overestimation of the enhancement generated by the localization strategies, the qualitative behavior that the enhancement is largest when the overall concentration of enzymes is low remains unchanged. This finding is in good agreement with [118], where the highest improvement of product titer was observed at low enzyme expression levels.

We saw that the interplay of two fundamental trade-offs leads to complicated and varied optimal geometries for model multi-biocatalyst complexes with a single biocatalyst of the first type surrounded by several biocatalysts of the second type. Notably, the resulting optimal configurations show striking similarities to the well known Thomson problem of classical electrostatics. However, in contrast to the Thomson problem where electrons interact via the Coulomb potential and the total energy function is determined as a sum of the individual pair interactions, for our model of multi-biocatalyst complexes the metabolite concentration fields mediate effective many-body interactions between biocatalysts. Consequently, the total reaction flux cannot be expressed simply as a superposition of the individual contributions of the biocatalysts. Therefore, the problem of finding the optimal biocatalyst complex arrangements adds an additional level of complexity to the class of generalized Thomson problems. Although these problems are very easy to pose they are notoriously difficult to solve rigorously. For the standard Thomson problem the symmetries have only be rigorously identified for a few electron numbers [244, 245, 246] and solving the general case has been included in the list by Steven Smale, inspired by the famous Hilbert problems, of eighteen unsolved mathematical problems for the 21^{st} century [247]. Besides the interesting mathematical nature of the stated problem and the symmetries that emerge, the optimal structure of the model complex may also provide valuable insights into the design principles of such complexes. Indeed, similar optimization problems, yet with very different objectives, have previously been found to serve as useful minimal models for understanding the structure and geometry of biological materials ranging from proteins and viral capsids to plant phyllotaxis and honeycombs in beehives [248, 249, 250, 251, 252, 253, 254].

In this work we have focused on general features that are applicable to reaction-diffusion systems on different length scales, and therefore we have omitted a number of details that

may play a role in specific systems. For example, in microbial communities where different microbes divide the labor of processing certain metabolites, the biocatalysts, in this case microbes, not only consume and secrete metabolites but also grow, divide, and actively move. In our model these processes were neglected since their dynamics are expected to be slow compared to the considered reaction-diffusion dynamics. However, they will certainly affect the possible arrangements of biocatalysts. Furthermore, in enzymatic systems the biocatalysts are reactive only at a specific active site rather than over their entire surface. Introducing such an active site adds another degree of freedom, namely the orientation of the active site, to the different spatial strategies. While this may slightly alter the reaction flux in some cases, in initial investigations we found that localization effects were much more significant in determining differences between configurations. Furthermore, it is unlikely that orientation can be controlled against thermal fluctuations in natural and synthetic systems unless enzymes are tightly constrained in macromolecular complexes; in such cases transport mechanisms other than free diffusion, such as swinging arms or channels [37], may be more realistic.

6 Flux regulation via enzyme coclustering and sequestration *

The metabolism of living organisms responds dynamically to changing metabolic needs or environmental conditions. In addition to regulating metabolic fluxes by changing the abundance or activities of metabolic enzymes, cells also use spatial organization strategies such as enzyme colocalization and sequestration to regulate pathway fluxes. In this chapter we study two spatial strategies, the sequestration of a single pathway enzyme into clusters and the coclustering of consecutive enzymes, and characterize their capability of regulating the flux distribution at a metabolic branch point. By employing a semi well-mixed approximation we find that the fluxes depend on only two non-dimensional parameters, the ratio of activities of the downstream enzymes and the ratio of diffusion and reaction timescales within a cluster. Our analysis reveals that sequestration and coclustering both represent viable regulation strategies when the catalytic efficiencies of the downstream enzymes are larger than $k_{\mathrm{cat}}/K_{\mathrm{M}} \sim 10^4\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$, a range that includes more than 75% of metabolic enzymes.

6.1 Introduction

Cells are able to quickly and precisely control key metabolic pathways in response to changing metabolic conditions. Such regulation can be achieved by changing either the enzyme abundance, via the rates of transcription, translation and proteolysis, or enzymatic activities, via allosteric mechanisms or post-translational modifications [15, 255, 256, 257]. In addition to these mechanisms, it has been hypothesized that cells can control metabolic fluxes by coordinating the spatial organization of enzymes [96, 258]. Many enzymes have been found experimentally to aggregate into intracellular, non-membrane bound bodies [75, 96, 259], through diverse mechanisms that include direct interactions between the enzymes or with scaffolds, adsorption to specific spots on the cellular membrane or the cytoskeleton, or liquid-droplet phase separation [260, 75, 209, 261, 45, 262, 263]. Importantly, assembly and disassembly of enzyme clusters is often reversible upon depletion or supply of metabolites or other environmental stimuli [96, 74, 76]. This suggests that cells may use the spatial coordination of enzymes as a strategy to regulate metabolism dynamically as metabolic demands on the cell change, according to whether or not the catalytic activity of the enzymes in question is needed [75, 78].

Colocalization of enzymes that operate within the same metabolic pathway is generally understood to promote the processing of intermediate products, and thereby upregulate the

*This chapter is adapted from the manuscript: Post-translational flux regulation via enzyme coclustering and sequestration, by F. Hinzpeter, F. Tostevin, and U. Gerland, which is in preparation for publication.

pathway flux [37]. A well-known example for this strategy is the purinosome, a multi-enzyme cluster comprising enzymes of the *de novo* purine biosynthesis pathway [76, 79]. Coclustering has also been observed for enzymes involved in glucose metabolism and *de novo* pyrimidine biosynthesis [114, 110, 84]. In some such assemblies, intermediates are directly "channeled" or transferred between enzymes [37]. However, theoretical analyses (see Chap. 4 and 5) have shown that proximity of consecutive enzymes may increase the efficiency of intermediate turnover even without direct or guided intermediate transfer [202, 197, 142].

Sequestration of a single enzyme into clusters has received less attention as a potential regulation strategy. However, the reversible sequestration of individual enzymes into filaments or clusters has been observed in the *de novo* CTP and purine biosynthesis pathways [96, 78], and was associated with downregulation of the corresponding pathway [102, 83]. Thus, the purine biosynthesis pathway appears to have the capacity for two spatial strategies in response to changing environmental conditions: the assembly of all pathway enzymes into the purinosome as a mechanism for pathway upregulation, and the sequestration of a single enzyme into a cluster as a mechanism for pathway downregulation [78]. While the origin of this downregulation remains unclear, the spatial sequestration alone may suffice as a mechanism of downregulation without the need for reducing the catalytic activities of the enzymes.

Here we propose the sequestration of a single enzyme into non-membrane bound clusters as a mechanism to regulate metabolic fluxes without changing the inherent catalytic efficiencies of the enzymes. By considering a metabolic branch point, we quantify the capability of this sequestration strategy to regulate the flux distribution at the branch point and compare it to the strategy of multi-enzyme coclustering. We find a crossover between the two strategies from a regime where enzyme coclustering has the strongest impact on the fluxes to a regime where enzyme segregation has the largest effect. Employing a semi well-mixed approximation, we find that the impact of the two strategies on the fluxes depends on only two non-dimensional parameters, one measuring the ratio of the activities of the downstream enzymes and the other measuring the ratio of diffusion and reaction timescales within a cluster. With this we find that a significant flux regulation by sequestration or coclustering is not possible when the catalytic efficiencies (k_{cat}/K_M) of the two competing downstream enzymes are smaller than $\sim 10^4\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$.

6.2 Model

We consider a branch point in a metabolic network, shown schematically in Fig. 6.1a, where intermediates I, produced by an upstream enzyme E_I, are processed by two downstream enzymes E_P and E_Q into two distinct products P and Q. To quantitatively describe the flux-regulation capacity of enzyme coclustering and sequestration, we define the branching fraction J_p as the fraction of I that is converted into P, and similarly $J_q = 1 - J_p$. We denote the catalytic efficiency and average concentration of each enzyme as $k_{p,q}$ and $e_{p,q}$, respectively. For simplicity we will assume that we are in the regime of low intermediate concentration, such that the reaction kinetics are linear. If both enzymes are uniformly-distributed, then the branching fraction depends only on the ratio of the overall activities of the two enzymes, $\alpha = (k_p e_p) \,/\, (k_q e_q)$, according to

$$J_p^0 = \frac{\alpha}{1+\alpha}. \tag{6.1}$$

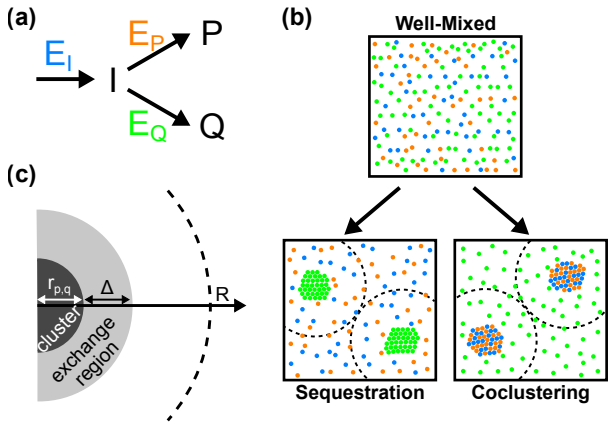

Figure 6.1: Illustration of a metabolic branch point and clustering strategies for flux regulation. (a) Reaction scheme of a metabolic branch point. An upstream enzyme E_I produces intermediates I, which are processed by two competing downstream enzymes, E_P and E_Q, to two distinct products P and Q. **(b)** Strategies of spatial enzyme organization for flux regulation. (left) Sequestration of E_Q into clusters. (right) Coclustering of E_Ps together with E_Is. **(c)** Schematic of the piecewise well-mixed approximation.

Compared to this basal branch point fraction, the flux can be biased towards P and away from Q either by the coclustering of E_I and E_P, or by the sequestration of E_Q into single-enzyme clusters, as shown in Fig. 6.1b. These localization strategies either decrease the average distance between E_I and E_P enzymes or increase the average distance between E_I and E_Qs. We model these arrangements as a spherically symmetric domain containing a single central cluster, representative of the region around a typical example of one of many such clusters within a larger cell. To illustrate and compare the branching fraction of the different spatial strategies, we focus on the effect of changing the mean enzyme concentrations $e_{p,q}$. We will assume that the typical cluster radii and clustered enzyme densities (r_q and c_q for E_Q sequestration clusters, r_p and c_p for the E_P–E_I coclusters) are fixed intrinsic properties of the particular enzymes and their assembly processes. Varying the mean density of the clustered enzyme alters the typical separation of clusters, which we describe by a $e_{p,q}$-dependent system size, $R(e_{p,q}) = r_{p,q}(c_{p,q}/e_{p,q})^{1/3}$, but will not change the properties of clusters themselves.

Assuming that intermediates move by diffusion, the steady-state density of intermediates, $\rho(r)$, follows

$$D\nabla^2\rho(r) + F\left[\rho(r), r\right] = 0, \tag{6.2}$$

where D is the diffusion coefficient and $F\left[\rho(r), r\right]$ is a reaction term that specifies the spatial strategy. For sequestration of E_Q, the corresponding reaction term is

$$F^{\text{seq}}\left[\rho(r), r\right] = \begin{cases} j_i - k_p e_p' \rho(r) & r > r_q \\ -k_q c_q \rho(r) & r \le r_q \end{cases}, \tag{6.3}$$

where $e_p' = e_p\left[1 - r_q^3/R(e_q)^3\right]^{-1}$ is the density of E_P in the free space between the E_Q clusters, and j_i represents the production flux of intermediates by E_I. For E_P-E_I coclusters we have

$$F^{\text{cocl}}\left[\rho(r), r\right] = \begin{cases} j_i - k_p c_p \rho(r) & r \le r_p \\ -k_q e_q' \rho(r) & r > r_p \end{cases}, \tag{6.4}$$

where $e_q' = e_q\left[1 - r_p^3/R(e_p)^3\right]^{-1}$ is similarly the density of E_Q between the clusters. Conservation of mass requires that $\rho(r)$ is continuous and smooth at the interface $r = r_{p,q}$. The model is completed by no-flux boundary conditions, $\rho'(0) = \rho'(R) = 0$. At the outer boundary, this represents symmetric exchange of intermediate with similar regions around neighboring

clusters. These equations can be solved analytically to determine the steady state fluxes of P and Q for each of the different strategies. In particular, the branching fractions towards the product P are given by

$$J_p^{\text{seq}} = \frac{\int_{\mathcal{V}\setminus\mathcal{V}_c} k_p e'_p \rho(\mathbf{r}) d\mathbf{r}}{\int_{\mathcal{V}\setminus\mathcal{V}_c} j_i d\mathbf{r}}, \qquad J_p^{\text{cocl}} = \frac{\int_{\mathcal{V}_c} k_p c_p \rho(\mathbf{r}) d\mathbf{r}}{\int_{\mathcal{V}_c} j_i d\mathbf{r}}, \tag{6.5}$$

with the integrals taken over the region outside the cluster ($\mathcal{V}\setminus\mathcal{V}_c$) or over the cluster volume ($\mathcal{V}_c$) according to where E_P and E_I are present in each scenario.

6.3 Results

6.3.1 Semi well-mixed approximation for enzyme sequestration and coclustering

Figure 6.2 shows the branching fraction for uniformly distributed enzymes Eq. 6.1, which only depends on α. The distribution of intermediates between the two branches is symmetric about $\alpha = 1$, with $J_p^0 < J_q^0$ when the activity of E_Q is greater than that of E_P ($\alpha < 1$) and $J_p^0 > J_q^0$ for $\alpha > 1$. However, the symmetry between the two downstream pathways will be broken when the different spatial arrangements of the enzymes are considered.

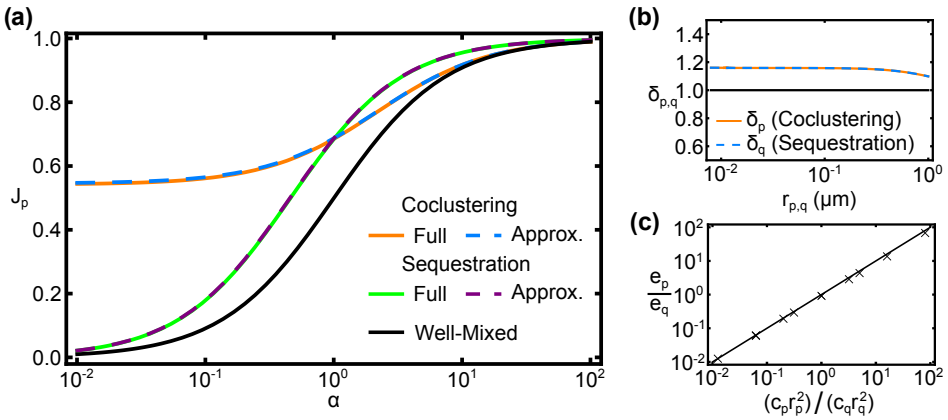

Figure 6.2: Branching fraction of the clustering strategies. **(a)** Branching fraction J_p in the reaction diffusion model for the enzyme sequestration strategy (green solid curve), the enzyme coclustering strategy (orange solid curve), and for a well-mixed system of uniformly distributed enzymes (black curve). Dashed lines are approximations of Eqs. 6.8 and 6.11. Other parameters are $k_p = k_q = 7.5 \times 10^4\,\text{M}^{-1}\,\text{s}^{-1}$, $c_p = c_q = 25\,\text{mM}$, $r_p = r_q = 0.4\,\mu\text{m}$, and $D = 100\,\mu\text{m}^2\,\text{s}^{-1}$. **(b)** The fitted exchange lengths $\delta_{p,q} = \Delta/r_{p,q}$ as cluster sizes $r_{p,q}$ are varied. The mean enzyme concentrations are $e_{p,q} = 10\,\text{nM}$. Other parameters are as above. **(c)** Value of e_p/e_q at the crossover position $J_p^{\text{seq}} = J_p^{\text{cocl}}$ for different combinations of, $r_{p,q} = 0.1$, 0.4 μm and $c_{p,q} = 10$, 25 mM (crosses). Solid line shows Eq. 6.12.

We found that for realistic biological parameters with $c_{p,q}$ (typically in the millimolar range) much larger than $e_{p,q}$ (nano- to micromolar), the dependence of J_p on $e_{p,q}$ can be captured by the ratio of abundances $e_p/e_q \propto \alpha$ (see SM Fig. D.1). For both localization strategies, J_p is larger than that for a well-mixed system across the full range of α, as shown in Fig. 6.2a.

The magnitude of this increase tends to grow with the number of enzymes in each cluster, irrespective of whether $c_{p,q}$ or $r_{p,q}$ is increased (see SM Fig. D.2). The two strategies differ in the shape of J_p: J_p^{cocl} tends to a (c_p, r_p-dependent) finite value for $\alpha \ll 1$; in contrast $J_p^{\mathrm{seq}} \to 0$ as $\alpha \to 0$ but the turning point of the sigmoidal J_p^{seq} shifts to smaller α values with increasing c_q or r_q. Interestingly, these differences result in a crossover, where for smaller α a coclustering strategy tends to provide a larger J_p, while the opposite is true for large α.

The expressions for J_p computed from Eq. 6.5 are not instructive for elucidating the origin and essential physics of the behavior shown in Fig. 6.2a. We therefore considered a simple approximation of the reaction-diffusion dynamics, summarized in Fig. 6.1c, that allows us to understand the system in detail. We consider the cluster and the bulk region outside the cluster as two separate domains. Within each domain the intermediates are taken to be well-mixed. Transport between the two domains is replaced by exchange with an effective permeability, D/Δ, where the effective exchange length scale Δ can be taken as a fit parameter to match the pathway fluxes to Eq. 6.5. For the sequestration scenario the density of intermediate in the cluster and bulk, ρ_c and ρ_b, can be calculated from the conservation equations

$$\frac{\partial \rho_c}{\partial t} = A_q \frac{D}{\Delta} (\rho_b - \rho_c) - V_q k_q c_q \rho_c \tag{6.6}$$

$$\frac{\partial \rho_b}{\partial t} = A_q \frac{D}{\Delta} (\rho_c - \rho_b) + [V(e_q) - V_q] \left[j_i - k_p e'_p \rho_b \right], \tag{6.7}$$

where $V(e_{p,q}) = \frac{4}{3}\pi R(e_{p,q})^3$, $V_{p,q} = \frac{4}{3}\pi r_{p,q}^3$ and $A_{p,q} = 4\pi r_{p,q}^2$ are the system volume, cluster volume and cluster surface area, respectively. This leads to a steady-state branching fraction

$$J_p^{\mathrm{seq}} = \frac{\alpha (1 + \beta_q \delta_q)}{1 + \alpha (1 + \beta_q \delta_q)}, \tag{6.8}$$

where $\beta_q = k_q c_q r_q^2 / (3D)$ and $\delta_q = \Delta / r_q$. For coclustering of E_I and E_P, the corresponding equations are

$$\frac{\partial \rho_c}{\partial t} = A_p \frac{D}{\Delta} (\rho_b - \rho_c) + V_p \left[j_i - k_p c_p \rho_c \right] \tag{6.9}$$

$$\frac{\partial \rho_b}{\partial t} = A_p \frac{D}{\Delta} (\rho_c - \rho_b) - [V(e_p) - V_p] \, k_q e'_q \rho_b, \tag{6.10}$$

with the resulting steady-state branching fraction,

$$J_p^{\mathrm{cocl}} = \frac{\alpha + \beta_p \delta_p}{1 + \alpha + \beta_p \delta_p}, \tag{6.11}$$

where $\beta_p = k_p c_p r_p^2 / (3D)$ and $\delta_p = \Delta / r_p$. Upon fitting $\delta_{p,q}$, Eqs. 6.8 and 6.11 are indistinguishable from the results of the full reaction-diffusion system, as shown in Fig. 6.2a. Notably, we find $\delta_{p,q}$ are slightly larger than 1 and vary little across the range of biologically realistic cluster radii (roughly 10 nm-1 µm, see Fig. 6.2b). In fact, simply setting $\delta_{p,q} = 1$ only marginally reduces the accuracy of the approximation (see SM Fig. D.3).

The dimensionless parameters $\beta_{p,q}$ are the ratio of timescales for I to diffusive through the cluster, $\tau_d \sim r_{p,q}^2 / D$, and reactions of I with the enzymes within the cluster, $\tau_r \sim (k_{p,q} c_{p,q})^{-1}$.

For both localization strategies the enhancement in branching towards P rather than Q is greatest when reactions within the cluster are faster than exchange between domains ($\beta \gg 1$), whereas there is little enhancement in the production of P when exchange is fast compared with reactions ($\beta \ll 1$). This appears counter intuitive in the case of sequestration of E_Qs, where we might expect a greater flux of Q as the efficiency of E_Q is increased. However, we must bear in mind that β_q is multiplied by α, which itself decreases with the total E_Q activity in the system.

Comparing Eqs. 6.8 and 6.11, we see that the crossover between regimes where $J_p^{\text{cocl}} > J_p^{\text{seq}}$ for smaller α and $J_p^{\text{cocl}} < J_p^{\text{seq}}$ for larger α occurs when $\alpha\beta_q\delta_q = \beta_p\delta_p$. Assuming $\delta_p \approx \delta_q$ (Fig. 6.2b), this condition can be written as

$$\frac{e_p}{e_q} \approx \frac{c_p}{c_q}\frac{r_p^2}{r_q^2}. \tag{6.12}$$

Thus, coclustering will produce a greater J_p when the ratio of abundances of E_P to E_Q in the system is less than the ratio of the density of the enzymes in their clustered state, multiplied roughly by the squared ratio of their radii. This quadratic dependence arises from a comparison of the diffusive timescales for intermediates within the two types of cluster, $\tau_d^{\text{cocl}}/\tau_d^{\text{seq}}$. Figure 6.2c confirms that Eq. 6.12 provides an excellent approximation for the crossover position in the full reaction-diffusion system.

6.3.2 Regulation of the branching fraction by sequestration and coclustering

We quantified the extent of upregulation of a desired product P for each enzyme localization strategy relative to uniformly-distributed enzymes by the enhancement factor $\epsilon = J_p/J_p^0$. Figure 6.3a shows that for both strategies the largest enhancement is achieved when the average enzymatic activity of E_Q is much larger than of E_P (small α) and the enzymatic activity in the cluster is large compared to the rate of diffusive exchange (large $\beta_{p,q}$). On the other hand, when the average enzymatic activity of E_P is much larger than that of E_Q ($\alpha \gtrsim 5$) the pathway will be strongly biased towards P already with uniformly-distributed enzymes, and hence any enhancement from enzyme clustering will be negligible.

A large enhancement factor alone does not guarantee that the resulting fraction of I directed towards P will be significant if the pathway was initially strongly biased towards Q. We therefore consider upregulation to be biologically relevant when the resulting branching fraction is $J_p \geq 0.2$ and the enhancement $\epsilon \geq 1.2$. These constraints delimit regions in parameter space, shown by solid lines in Fig. 6.3a (see also SM Fig. D.4 for J_p as a function of α and $\beta_{p,q}$). For the E_P-E_I coclustering strategy the minimum value of β_p is roughly constant, at least for $\alpha \lesssim 0.1$. This corresponds roughly to requiring that the coclusters are sufficiently large and dense that the required fraction of I can be converted to P before diffusing out of the cluster. For E_Q sequestration, on the other hand, as the enzymatic activity in the system is biased more towards E_Q (smaller α), the size and/or density of the E_Q clusters must be increased (larger β_q) in order to maintain large inter-cluster separations, such that the E_Ps distributed between the clusters are able to process a sufficient fraction of Is before they reach the E_Q clusters by diffusion.

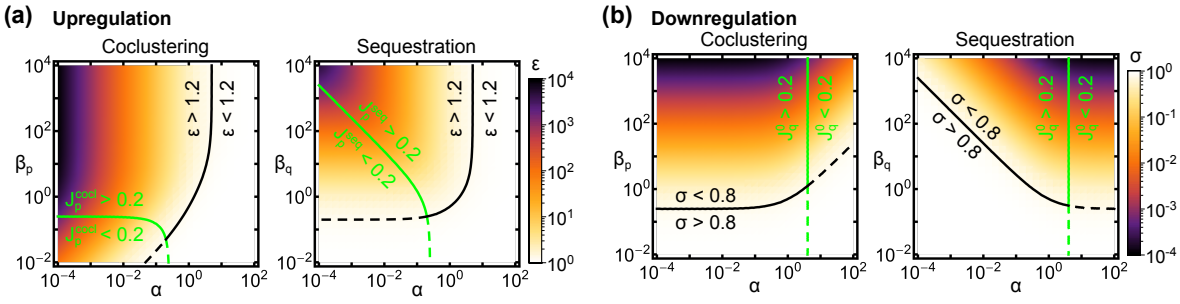

Figure 6.3: Capability of the spatial strategies to regulate the branching fraction. **(a)** Enhancement towards the P branch by E_I–E_P coclustering (left) and E_Q sequestration (right). Black lines divide regimes at the enhancement threshold $\epsilon = 1.2$, while green lines divide regimes based on the upregulated branching fraction $J_p^{\mathrm{cocl}}, J_p^{\mathrm{seq}} = 0.2$. **(b)** Downregulation of the Q branch. Black lines denote $\sigma = 0.8$, green lines indicate where the branching fraction in the absence of enzyme localization is $J_p^0 = 0.2$. For the rescaled exchange length we used $\delta_{p,q} = 1$

We can similarly consider the downregulation of an unwanted product Q, for which we define the suppression factor $\sigma = J_q/J_q^0$. In Fig. 6.3b we see that while the strongest suppression (smallest σ) can always be achieved for large $\beta_{p,q}$, this requires $\alpha \leq 1$ in the coclustering scenario but $\alpha \geq 1$ in the sequestration scenario. However, similarly to the upregulation of P above, strong downregulation may be of no significance if the basal branching fraction was already small. We therefore define the regime of biologically relevant downregulation as that in which the suppression factor $\sigma \leq 0.8$ and the basal branching fraction without enzyme clustering $J_q^0 \geq 0.2$. While the greatest suppression by E_Q sequestration occurs outside this region in the regime of low basal Q production (large α), sequestration can nevertheless suppress the production of Q by a large factor in regions where J_q^0 is significant. Indeed, similar parameter combinations allow for significant upregulation of P and for downregulation of Q, on the basis of the same interplay between the locations of I production and the timescales of reaction with E_P versus reaching E_Q enzymes by diffusion.

Interestingly, for both localization strategies and for either the upregulation of P or the downregulation of Q, the constraints combine to produce a lower bound in a similar range of $\beta_{p,q} \gtrsim 0.1 - 0.25$ for which a significant regulation can be achieved. Assuming a cluster packing density $c_{p,q} \sim 25\,\mathrm{mM}$ [142], a cluster radius of $r_{p,q} \sim 0.4\,\mu\mathrm{m}$ comparable to the size of clusters found in purine metabolism [83], and a diffusion coefficient of a small metabolite $D \sim 100\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$ [264], this lower bound translates into a catalytic efficiency of the clustered protein of $k_{p,q} \gtrsim 10^4\,\mathrm{M}^{-1}\,\mathrm{s}^{-1}$. This range encompasses approximately 90% of enzymes in carbohydrate metabolism and approximately 75% of other classes of metabolic enzymes [185], meaning that both forms of enzyme clustering may be viable metabolic regulation strategies in many systems.

6.4 Discussion

While the upregulation of fluxes and metabolic branching by enzyme colocalization have been studied extensively (see Chap. 4 and 5, and [202, 197, 142, 37]), the possibility for pathway downregulation through enzyme sequestration into non-membrane bound clusters

has received less attention. Using a reaction-diffusion model, we studied in this chapter how enzyme sequestration and coclustering can regulate the branching fraction at a metabolic branch point. Employing a piecewise well-mixed approximation allowed us to identify two non-dimensional parameters which govern the branching fraction in both scenarios. One parameter captures the ratio of the catalytic capacities of the two competing enzymes and the other parameter measures the relative timescales of diffusion through and reactions within a cluster. We found that both strategies can bias the fluxes at a metabolic branch point to a comparable degree, and that both strategies could achieve significant levels of regulation for realistic biological parameters and for the majority of metabolic enzymes. Thus our analysis shows that enzyme sequestration, in the same way as colocalization, represents a potential regulation strategy in metabolic pathways.

In this study we have neglected some details which may affect the flux distribution at the branch point. The enzymes in our model were treated as point-like particles rather than discrete reaction centers. As we have seen in Chap. 5, describing the enzymes as spatially extended molecules leads to steric blocking of metabolites. In the sequestration scenario, steric blocking hinders the intermediates from diffusing into the cluster and, therefore, the collective activity of the sequestered enzymes would be further repressed. In the coclustering scenario, in contrast, the same effect confines intermediates within the cluster, which leads to an enhancement of intermediate processing by the coclustered downstream enzyme. Furthermore, in our model we have neglected enzyme saturation. While we do not expect that including enzyme saturation will alter the qualitative finding that enzyme sequestration and coclustering can significantly regulate metabolic fluxes, it will be interesting to see how a non-linear reaction kinetics affects the detailed behavior of the fluxes distribution in the two scenarios.

Our analysis suggests that enzyme coclustering and sequestration observed in the *de novo* purine biosynthesis pathway [76, 78] present two strategies used to up- and downregulate metabolic fluxes without changing the inherent catalytic activities of the enzymes. The regulation of fluxes via cluster formation may allow the cells to quickly respond to changing environmental conditions. Compared to other regulation mechanisms which require control over enzyme abundances, enzyme aggregation can be achieved comparably quickly. Furthermore, when metabolic needs change, the dynamic clustering of enzymes is a more cost-efficient alternative for flux regulation than the costly change of enzyme expression levels. In particular, under conditions where a specific enzymatic activity is not needed, it may be more suitable for the cell to store the enzyme molecules and simultaneous downregulate their collective activity rather than degrading all enzymes. When the demand for the enzymatic activity rises again, enzymes can be gradually released from the cluster, thereby increasing the activity in the cytosol.

The theoretical findings presented in this chapter could be tested experimentally using different biotechnological techniques for spatial protein organization. For example, enzymes could be clustered using large nucleic acid or protein scaffolds. Tagging the enzymes with specific strands that are complementary to functionalized regions on the scaffolds, allows for binding of the enzymes to the scaffold. Additionally, enzyme clustering could be achieved by using short protein linkers that interconnect the enzymes or by fusing many copies of the same enzyme together. With a suitable technique for enzyme clustering and a metabolic branch point that allows to simultaneously measure the fluxes of the individual branches, it could be tested

whether enzyme sequestration and coclustering can bias the flux towards a specific product. Furthermore, if the experimental setup also allows to control the amount of enzymes in the cluster as well as the overall abundances of enzymes, it would be possible to experimentally validate the predicted behavior of the fluxes when these quantities are changed.

Supplementary Materials

A Optimal compartmentalization strategies for metabolic compartments

A.1 Carboxysome parameters

To demonstrate the behavior of the system we used parameter values representative of the α-carboxysome. For the catalytic efficiencies we used the measured values of carbonic anhydrase, $\kappa_1 \simeq 5\,(\mu\mathrm{M\,s})^{-1}$ [177], and RuBisCO, $\kappa_2 \simeq 0.06\,(\mu\mathrm{M\,s})^{-1}$ [178]. The diffusion coefficients of CO_2 and HCO_3^- were assumed to be equal and approximated by the diffusion coefficient of CO_2 in water, $D \simeq 10^3\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$ [176].
The upper bound of the total enzyme concentration $e_{\max}$ was estimated as the concentration at which the whole volume is filled with RuBisCO monomers. The RuBisCO octamere is approximately spherical with a radius $\sim 5\,\mathrm{nm}$ [181], thus we obtained for the maximal concentration $e_{\max} = 8/(4\pi(5\mathrm{nm})^3/3) \simeq 25\,\mathrm{mM}$.

The shell permeability has not yet been measured. However we can estimate the permeability for an uncharged CO_2 molecule using the measured crystal structure of the shell. We estimate the permeabilities of HCO_3^- and CO_2 for the α-carboxysome and β-carboxysome.
A α-carboxysome has $N_p \simeq 800$ pores, which are uniformly distributed over the carboxysome surface, while each pore has a radius of $r_p \sim 2\,\text{Å}$ [179]. The effective area of a pore that allows for the penetration of a CO_2 molecule is $\pi(r_p - r_w)^2$, with $r_w \simeq 1.7\text{Å}$ the van der Waals radius of carbon [265]. Thus, the probability a CO_2 molecule penetrates the shell is $\sim \frac{N_p\pi(r_p-r_w)^2}{A}$, while $A \simeq 0.036\,\mu\mathrm{m}^2$ the surface of a icosahedral carboxysome with diameter $\sim 0.123\,\mu\mathrm{m}$ [181]. The permeability is then given by the probability to penetrate the shell times the speed CO_2 diffuses through the shell,

$$p \simeq \frac{N_p\pi(r_p - r_w)^2}{A}\frac{D}{\delta} = 18\,\mu\mathrm{m\,s}^{-1}, \tag{A.1}$$

The thickness of the carboxysome shell is $\delta \simeq 3.5\,\mathrm{nm}$ [180]. However, due to the positive charge at the pore, the permeability of the negatively charged substrate HCO_3^- will be enhanced [53]. To account for this selective permeability we assumed that the permeability of HCO_3^- is 5 times higher than for CO_2. Thus, the permeabilities we used in our model are $p_s = 90\,\mu\mathrm{m\,s}^{-1}$ for HCO_3^- and $p_i = 18\,\mu\mathrm{m\,s}^{-1}$ for CO_2. Using the same derivation we estimate the permeabilities for the β-carboxysome. The number of pores per shell area is approximately the same as for the α-carboxysome, however the pores are larger with an average radius of $r_p \sim 2.75\,\text{Å}$ [57]. From this we obtain for the shell permeabilities, $p_s = 1080\,\mu\mathrm{m\,s}^{-1}$ and $p_i = 215\,\mu\mathrm{m\,s}^{-1}$ for HCO_3^- and CO_2, respectively. Hence, due to the larger pore, the permeabilities of the β-carboxysome shell are 12-fold higher than of the α-carboxysome shell.

A.2 Derivation of productivity for uniformly distributed enzymes

In the case of uniformly distributed enzymes in spherical compartments, the densities of substrate and intermediate also exhibit spherical symmetry. The productivity is therefore given by

$$\mathrm{P}(e_1, e_2, R) = \frac{\int_0^R 4\pi r^2 \kappa_2 e_2 i(r) dr}{4\pi R^3 (e_1 + e_2)/3}. \tag{A.2}$$

The integral in the numerator is the total rate of $\mathcal{P}$-production in the compartment, which can be computed analytically by solving the coupled reaction-diffusion equations, Eqs. 3.2-3.3. For ease of notation, we define four length scales as follows:

$$\begin{aligned} l_1 &= \sqrt{D/(\kappa_1 e_1)},\ l_s = D/p_s, \\ l_2 &= \sqrt{D/(\kappa_2 e_2)},\ l_i = D/p_i. \end{aligned} \tag{A.3}$$

Then the steady-state versions of the reaction-diffusion equations become

$$0 = \nabla^2 s(r) - s(r)/l_1^2, \tag{A.4}$$

$$0 = \nabla^2 i(r) + s(r)/l_1^2 - i(r)/l_2^2. \tag{A.5}$$

The boundary conditions at the permeable compartment boundary become

$$s'(R) = \left[s_0 - s(R)\right]/l_s,\ i'(R) = -i(R)/l_i \tag{A.6}$$

where primes indicate differentiation with respect to the radial coordinate r. Furthermore, due to the spherical symmetry of the compartments we impose no-flux boundary conditions at the center of the compartments, $s'(0) = i'(0) = 0$.

Now, we can compute the intermediate concentration $i(r)$ that is needed to evaluate Eq. A.2. Solving first for $s(r)$ yields

$$s(r) = \frac{s_0 R}{l_s} \frac{\mathrm{sinhc}(r/l_1)}{\left[\cosh(R/l_1) + (R/l_s - 1)\,\mathrm{sinhc}(R/l_1)\right]}, \tag{A.7}$$

where $\mathrm{sinhc}(x) = \sinh(x)/x$. Substituting into Eq. A.5, we solve for the concentration profile of intermediate,

$$i(r) = \frac{s(r)}{(l_1/l_2)^2 - 1}\left[1 - \left(\frac{\cosh(R/l_1) + (R/l_i - 1)\,\mathrm{sinhc}(R/l_1)}{\cosh(R/l_2) + (R/l_i - 1)\,\mathrm{sinhc}(R/l_2)}\right)\frac{\mathrm{sinhc}(r/l_2)}{\mathrm{sinhc}(r/l_1)}\right]. \tag{A.8}$$

Finally, we obtain the productivity by evaluating the integral in Eq. A.2,

$$\begin{aligned} &\mathrm{P}(e_1, e_2, R) = \\ &\frac{3 p_s s_0}{R\,(e_1 + e_2)\left[\cosh(R/l_1) + (R/l_s - 1)\,\mathrm{sinhc}(R/l_1)\right]\left[\cosh(R/l_2) + (R/l_i - 1)\,\mathrm{sinhc}(R/l_2)\right]} \\ &\times \Big\{\left[\cosh(R/l_1) - \mathrm{sinhc}(R/l_1)\right]\left[\cosh(R/l_2) - \mathrm{sinhc}(R/l_2)\right] \\ &+ (R/2l_i)\left[\mathrm{sinhc}\,(R/l_1 + R/l_2) + \mathrm{sinhc}\,(R/l_1 - R/l_2) - 2\,\mathrm{sinhc}(R/l_1)\,\mathrm{sinhc}(R/l_2)\right]\Big\}. \end{aligned} \tag{A.9}$$

A.3 Derivation of productivity for enzyme distribution observed in carboxysomes

We model the enzyme distribution in carboxysomes as a shell of $\mathcal{E}_1$ enzymes at the compartment boundary, which we take to have a thickness δ, and a sphere of $\mathcal{E}_2$ enzymes confined to the remaining volume, the region $r < R - \delta = \rho$,

$$e_1(r) = \begin{cases} 0 & r \leq \rho \\ e_1 & r > \rho \end{cases}, \quad e_2(r) = \begin{cases} e_2 & r \leq \rho \\ 0 & r > \rho \end{cases} \tag{A.10}$$

For this scenario, the productivity is given by,

$$\mathrm{P}(e_1, e_2, R) = \frac{\int_0^\rho 4\pi r^2 \kappa_2 e_2 i(r) dr}{4/3\pi \left[\rho^3 e_2 + (R^3 - \rho^3) e_1\right]}. \tag{A.11}$$

This expression can be computed analytically by solving the coupled steady state reaction-diffusion equations for the enzyme distribution in Eq. A.10. For this, we divide the system into two domains. Domain A ($\mathcal{E}_1$ shell) where $r > \rho$ and domain B ($\mathcal{E}_2$ sphere) where $r < \rho$. The reaction-diffusion system in each domain is then,

$$\left.\begin{aligned} 0 &= D\nabla^2 s_A(r) - \kappa_1 e_1 s_A(r), \\ 0 &= D\nabla^2 i_A(r) + \kappa_1 e_1 s_A(r), \end{aligned}\right\} \text{ Domain A } (r > \rho) \tag{A.12}$$

$$\left.\begin{aligned} 0 &= D\nabla^2 s_B(r), \\ 0 &= D\nabla^2 i_B(r) - \kappa_2 e_2 i_B(r). \end{aligned}\right\} \text{ Domain B } (r < \rho) \tag{A.13}$$

At the compartment shell we impose the boundary conditions,

$$D s'_A(R) = p_s \left[s_0 - s_A(R)\right], \; D i'_A(R) = -p_i i_A(R). \tag{A.14}$$

Moreover, because of the spherical symmetry of the compartment, we require no-flux boundary conditions at the center of the compartment $s'_B(0) = i'_B(0) = 0$. Finally, we impose four conditions at the interface between the two domains, to ensure that the concentration profiles are continuously differentiable,

$$s_A(\rho) = s_B(\rho), \; s'_A(\rho) = s'_B(\rho), \; i_A(\rho) = i_B(\rho), \; i'_A(\rho) = i'_B(\rho). \tag{A.15}$$

With these conditions, we obtain for the substrate concentration profile,

$$s_A(r) = \frac{s_0 \left[\cosh((r-\rho)/l_1)\rho/l_1 + \sinh((r-\rho)/l_1)\right] l_1/l_s}{\left[(\delta/R + \rho/l_s)\cosh(\delta/l_1) + (l_1(1/l_s - 1/R) + \rho/l_1)\sinh(\delta/l_1)\right] r/R}, \tag{A.16}$$

$$s_B(r) = \frac{s_0 R/l_s}{(\delta/R + \rho/l_s)\cosh(\delta/l_1) + (l_1(1/l_s - 1/R) + \rho/l_1)\sinh(\delta/l_1)}, \tag{A.17}$$

where we used the length scale notation defined in Eq. A.3. Using this solution, we can determine the intermediate concentration, which is needed to determine the productivity. Since the integral in Eq. A.11 runs only over domain b (from 0 to ρ), we only need to compute the intermediate concentration in domain b, which is given by,

$$i_B(r) = s_B(r)\,\mathrm{sinhc}\,(r/l_2)\,\frac{(\delta/R + \rho/l_i)\cosh(\delta/l_1) + (l_1(1/l_i - 1/R) + \rho/l_1)\sinh(\delta/l_1) - R/l_i}{(\delta/l_i + \rho/R)\cosh(\rho/l_2) + l_2\,(1/l_i - 1/R)\sinh(\rho/l_2)}. \tag{A.18}$$

Finally, we obtain the productivity by solving the integral in Eq. A.11,

$$\begin{aligned}\mathrm{P}(e_1,e_2,R) = &\frac{3p_s s_0}{\frac{R^2}{l_2}\left(e_1+(e_2-e_1)(\rho/R)^3\right)}\\ &\times\frac{-R/l_i+(\delta/R+\rho/l_i)\cosh(\delta/l_1)+(l_1(1/l_i-1/R)+\rho/l_1)\sinh(\delta/l_1)}{(\delta/l_i+\rho/R)\cosh(\rho/l_2)+l_2(1/l_i-1/R)\sinh(\rho/l_2)}\\ &\times\frac{\frac{\rho}{l_2}\cosh(\rho/l_2)-\sinh(\rho/l_2)}{(\rho/l_s+\delta/R)\cosh(\delta/l_1)+(l_1(1/l_s-1/R)+\rho/l_1)\sinh(\delta/l_1)}\end{aligned} \tag{A.19}$$

A.4 Optimization of intra-compartment enzyme distribution

The numerical optimization of enzyme profiles was performed using a modified version of the stochastic optimization algorithm described previously [202].

The compartment was discretized into $N = 200$ concentric spherical shells, $(n-1)\delta r \leq r \leq n\delta r$ with lattice size $\delta r = R/N$ and $n = 1..N$. We used as optimization variables the fractional occupancy of each site, $e_{T,n}/e_{\max}$, and the fraction of these enzymes that are of type $\mathcal{E}_1$, $\varphi_n = e_{1,n}/e_{T,n} = \phi_n/(1+\phi_n)$. The advantage of these variables is that both are restricted to the interval $[0,1]$. In these coordinates, the discrete version of the reaction-diffusion system (Eqs. 3.2-3.5) becomes

$$0 = \frac{D}{\delta r^2}\left[-4s_1+4s_2\right]-\kappa_1 e_{T,1}\varphi_1 s_1, \tag{A.20a}$$

$$\begin{aligned}0 = &\frac{D}{\delta r^2}\left[\left(\frac{2n-2}{2n-1}\right)^2 s_{n-1}-4\frac{2n^2-2n+1}{(2n-1)^2}s_n+\left(\frac{2n}{2n-1}\right)^2 s_{n+1}\right]\\ &-\kappa_1 e_{T,n}\varphi_n s_n,\quad 2\leq n<N\end{aligned} \tag{A.20b}$$

$$\begin{aligned}-\frac{p_s}{\delta r}\left(\frac{2N}{2N-1}\right)^2 s_0 = &\frac{D}{\delta r^2}\left[\left(\frac{2N-2}{2N-1}\right)^2 s_{N-1}-\left(\frac{2N-2}{2N-1}\right)^2 s_N\right]\\ &-\frac{p_s}{\delta r}\left(\frac{2N}{2N-1}\right)^2 s_N-\kappa_1 e_{T,N}\varphi_N s_N,\end{aligned} \tag{A.20c}$$

$$-\kappa_1 e_{T,1}\varphi_1 s_1 = \frac{D}{\delta r^2}\left[-4i_1+4i_2\right]-\kappa_2 e_{T,1}(1-\varphi_1)i_1, \tag{A.20d}$$

$$\begin{aligned}-\kappa_1 e_{T,n}\varphi_n s_n = &\frac{D}{\delta r^2}\left[\left(\frac{2n-2}{2n-1}\right)^2 i_{n-1}-4\frac{2n^2-2n+1}{(2n-1)^2}i_n+\left(\frac{2n}{2n-1}\right)^2 i_{n+1}\right]\\ &-\kappa_2 e_{T,n}(1-\varphi_n)i_n,\quad 2\leq n<N\end{aligned} \tag{A.20e}$$

$$-\kappa_1 e_{T,N}\varphi_N s_N = \frac{D}{\delta r^2}\left[\left(\frac{2N-2}{2N-1}\right)^2 i_{N-1} - \left(\frac{2N-2}{2N-1}\right)^2 i_N\right] - \frac{p_i}{\delta r}\left(\frac{2N}{2N-1}\right)^2 i_N - \kappa_2 e_{T,N}(1-\varphi_N) i_N. \tag{A.20f}$$

Each optimization run was initialized with a uniform enzyme distribution, $e_{T,n}/e_{\max} = 10^{-3}$ and $\varphi_n = 0.5$ for each of the $n = 1..N$ lattice sites. We verified that using random initial distributions as well as smaller lattice sizes did not alter the resulting optimal profiles. At each iteration of the optimization process, a set of 50 new test profiles were generated by selecting one site m at random and multiplying either $e_{T,m}/e_{\max}$ or φ_m by a factor ξ drawn from a log-normal distribution with location parameter $\mu = 0$ and scale parameter $\sigma = 0.01$. Thus changes that increase or decrease the corresponding enzyme variable are equally likely. For each of these modified enzyme configurations, the steady-state metabolite densities were calculated by first solving the linear system Eqs. A.20a-c for $\{s_n\}$, followed by Eqs. A.20d-f for $\{i_n\}$. The productivity was then evaluated as

$$\mathrm{P} = \frac{\sum_{n=1}^{N} V_n \kappa_2 e_{T,n}(1-\varphi_n) i_n}{\sum_{n=1}^{N} V_n e_{T,n}}, \tag{A.21}$$

where $V_n = 4\pi\delta r^3(1-3n+3n^2)/3$ is the volume of the n-th spherical shell. The initial enzyme distribution for the next round of modifications is constructed by taking the mean of the 10 enzyme profiles with the highest P values.

The results presented in Figs. 3.2 and 3.4 and Figs. A.1 and A.2 show the mean of the best profiles produced by each of 10 independent optimizations runs.

A.5 Optimal enzyme distributions for α-carboxysome parameters

Figure A.1 shows examples of the optimized enzyme profiles corresponding to the blue squares in Fig. 3.2a. As noted in Sec. 3.3.2, the total amounts of $\mathcal{E}_1$ and $\mathcal{E}_2$ in the compartment are in close agreement with the uniform enzyme scenario and the well-mixed model. These amounts are determined by the same trade-offs as described in the section "Well-mixed approximation". Specifically, increasing the number of enzymes will increase the turnover of either substrate or intermediate, but at the same time will lead to depletion of the corresponding metabolite and thereby reduce the efficiency with which each enzyme operates. The optimal amounts represent a compromise between these two effects for both reactions. Since diffusive mixing of substrate and intermediate in the compartment are fast for the depicted radii, the total amounts of enzymes play the largest role in determining the productivity, with their precise arrangement being of lesser importance.

In Fig. A.1, compartment sizes spanning the critical radius R_c are shown. At all radii we observe that $\mathcal{E}_1$ enzymes localize in a boundary layer at the outer edge of the compartment. The colocalization of $\mathcal{E}_1$ at the source of substrate in this way is a strategy to maximize the conversion of substrate into intermediate by maximizing the probability for a substrate molecule to encounter an $\mathcal{E}_1$ enzyme before diffusing out of the compartment again, consistent

with prior theoretical results [202]. In the absence of the constraint $e_1(r) + e_2(r) \leq e_{\max}$, the width of this shell becomes vanishingly small. However the maximal density constraint forces the available $\mathcal{E}_1$ enzymes to be arranged instead in a shell of finite width (such finite bands of enzymes subject to a maximal density constraint have also been observed previously [142]). This width w is approximately given by $e_{\max}(4\pi R^2 w) \approx e_1^*(R)(4\pi R^3/3)$, where $e_1^*(R)$ is the average density of e_1^* in the compartment as a whole as predicted by the well-mixed model. Thus for $R > R_c$, when $e_1^* \propto e_T^*(R) \sim R^{-1}$, the band width is approximately independent of R.

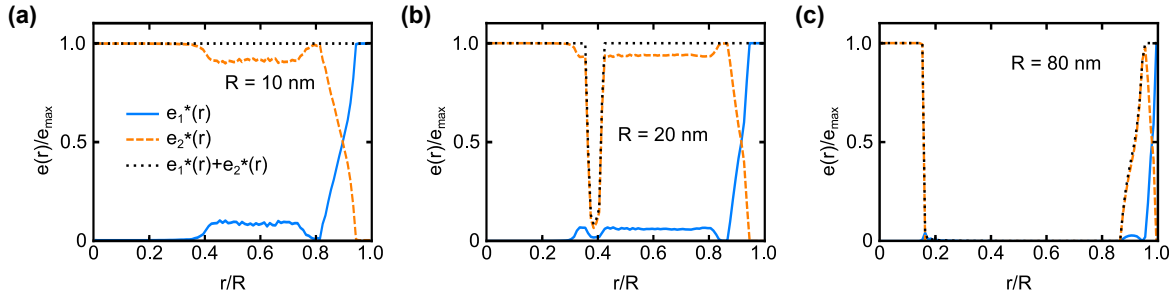

Figure A.1: Optimal intra-compartment enzyme distributions. Examples of the globally-optimized intra-compartment enzyme arrangement for different compartment sizes spanning the critical radius R_c, corresponding to Fig. 3.2. **(a)** $R = 10\,\text{nm}$, **(b)** $R = 20\,\text{nm}$, and **(c)** $R = 80\,\text{nm}$.

At $R > R_c$ (R =80 nm, Fig. A.1 bottom) we see that the majority of $\mathcal{E}_2$ enzymes localize in a band at the inner edge of the $\mathcal{E}_1$ domain . This reflects the same strategy for maximizing the conversion of $\mathcal{I}$ to $\mathcal{P}$ as described above for $\mathcal{E}_1$. The width of the band of enzymes is again determined by the maximal density constraint, in the same way. In addition to the band near the outer shell, the $e_2^*(r)$ distribution at large $R > R_c$ has a cluster of enzymes at the center of the compartment, although this cluster contains only a very small fraction of the total number of $\mathcal{E}_2$ enzymes (since the number of enzymes in a shell of width Δr is $\approx e_2(r) r^2 \Delta r$). The benefit provided by such a cluster is not immediately apparent, and we again emphasize that the productivity is largely insensitive to such fine details of the intracompartment enzyme distributions, suggesting that this feature originates in higher-order effects. (Indeed, this central cluster is absent when the dynamics is dominated by the decay of intermediates inside the compartment, see Fig. A.2c.) The center of the compartment stands out as the position where intermediate molecules are furthest from the compartment boundary and therefore have the highest expected time until they escape from the compartment. We therefore speculate that the diffusive trajectories for intermediate that spend the longest time in the compartment will be those that spend a disproportionate amount of time close to the center of the compartment. The central enzyme cluster may be used to exploit this class of intermediate molecules that spend a particularly large amount of time near the compartment center, thereby achieving some additional production of $\mathcal{P}$ with a very minor expenditure of enzymes.

Finally, as R is decreased below R_c (R =10 nm, Fig. A.1 top), the extension of the outer and inner enzyme bands becomes such that they span the entire compartment, and the optimal compartmentalization strategy consists of maximally filled compartments.

A.6 Optimal compartmentalization strategies with intermediate decay

We focused our analysis on the situation where intermediates are stable within compartments. In general, however, intermediates can be consumed by undesired side reactions or be intrinsically unstable. Both mechanisms lead to a decay of intermediates which has an effect on the optimal compartmentalization strategies. The addition of decay provides a second loss mechanism for intermediates, in addition to leakage across the compartment boundary. We therefore studied how the optimal compartmentalization strategies changed when decay of intermediate was added via a term $-\sigma i(r)$ in Eq. 3.3.

First, we examined a well-mixed system with decay. Mathematically, the effect of decay is simply to change the length scale parameter $\lambda_2 = 3p_i/(\kappa_2 e_{\max}) \to 3p_i(1 + R/R_\sigma)/(\kappa_2 e_{\max})$, with $R_\sigma = 3p_i/\sigma$. We can immediately see that λ_2 will be dominated by loss at the membrane when $R \ll R_\sigma$, and conversely by decay when $R \gg R_\sigma$. The contribution of decay to the total loss becomes more important as the compartment radius R is increased since the typical time for diffusion of intermediate to the boundary is longer, thereby increasing the probability of each intermediate molecule to decay before reaching the boundary. As R approaches R_σ, the optimal compartmentalization strategy deviates from the no-decay case (Fig. A.2a, solid lines). Specifically, the productivity $\mathrm{P_{wm}}$ begins to decrease from the plateau value. Simultaneously, the average enzyme density e_T^* deviates from the R^{-1} power law behavior and the enzyme ratio ϕ^* shifts towards a larger fraction of $\mathcal{E}_2$ enzymes. Increasing the number of available $\mathcal{E}_2$ enzymes in this way is an attempt to counteract the reduction in flux due to the loss of intermediates to decay by boosting the potential for conversion reactions into $\mathcal{P}$.

We next solved the reaction-diffusion system for the case of uniform enzymes, using the same procedure as above. The resulting productivity is given by

$$\mathrm{P}(e_1, e_2, R) = \frac{3p_s s_0}{R\,(e_1+e_2)\,(l_2/l_3)^2\big[\cosh(R/l_1) + (R/l_s - 1)\,\mathrm{sinhc}(R/l_1)\big]\big[\cosh(R/l_3) + (R/l_i - 1)\,\mathrm{sinhc}(R/l_3)\big]} \times \Big\{\big[\cosh(R/l_1) - \mathrm{sinhc}(R/l_1)\big]\big[\cosh(R/l_3) - \mathrm{sinhc}(R/l_3)\big] + (R/2l_i)\left[\mathrm{sinhc}\,(R/l_1 + R/l_3) + \mathrm{sinhc}\,(R/l_1 - R/l_3) - 2\,\mathrm{sinhc}(R/l_1)\,\mathrm{sinhc}(R/l_3)\right]\Big\}, \quad \text{(A.22)}$$

where the modified length scale $l_3 = \sqrt{D/(\kappa_2 e_2 + \sigma)}$. We find that the corresponding optimal compartmentalization strategy again matches the well-mixed approximation provided that R and σ are not too large (Fig. A.2a).

We nevertheless expect decay to push compartments out of the well-mixed regime by contributing to the establishment of gradients in intermediate concentration on length scales of $\lambda_\sigma \sim (D/\sigma)^{1/2}$. This effect is not observed in a system with uniform enzymes, however, because in this scenario the production of $\mathcal{I}$ is approximately uniform throughout the compartment (since $\mathcal{S}$ does not decay, $s(r)$ is approximately constant). While decay restricts the diffusion of intermediates throughout the compartment, it does so equally at all positions.

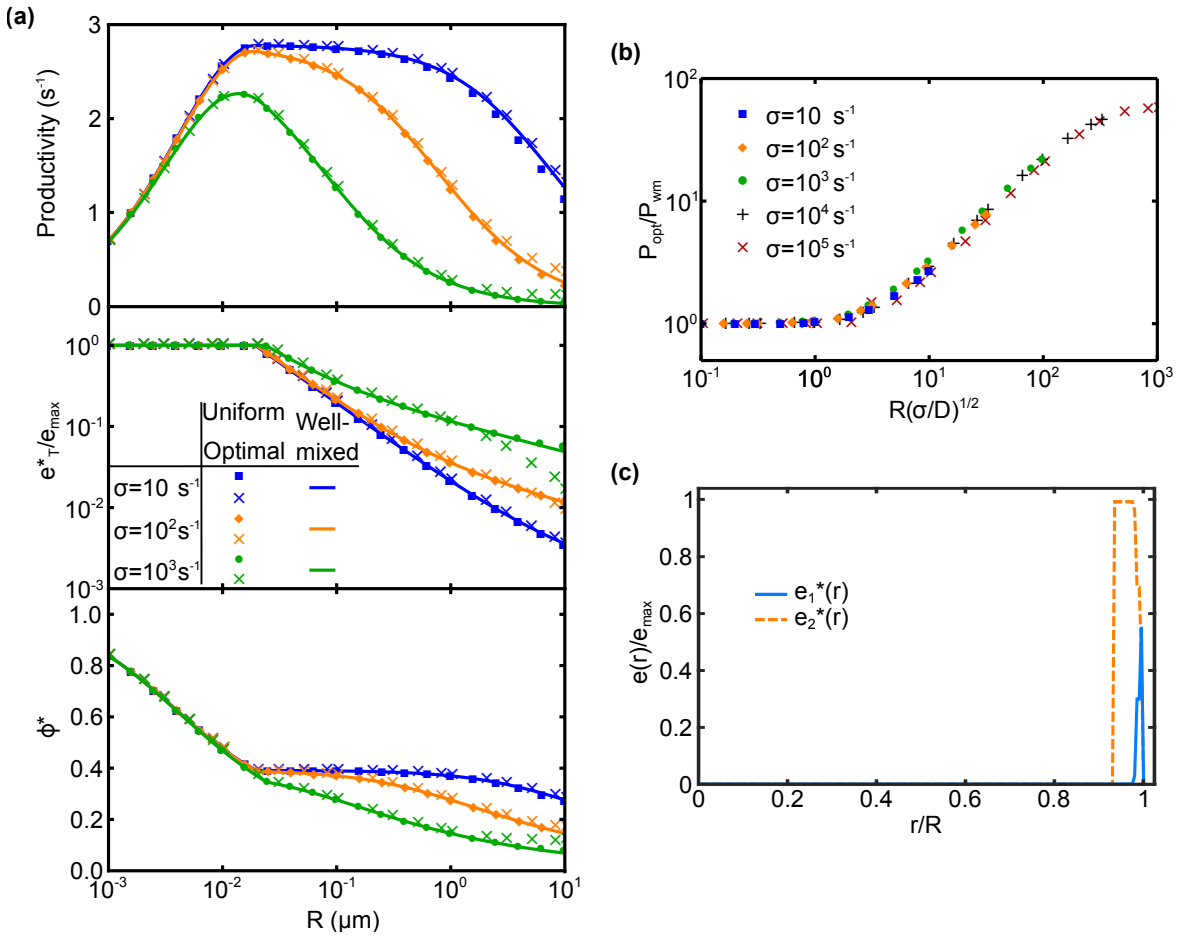

Figure A.2: Optimal compartmentalization strategies for α-carboxysome parameters with added decay of intermediate. (a) (Top) Optimal productivity, (middle) optimal total enzyme density, e^*_T, and (bottom) optimal ratio of $\mathcal{E}_1$ to $\mathcal{E}_2$ enzymes, ϕ^*. Lines are the solutions of the well-mixed model with decay. Solid symbols show the reaction diffusion-model with uniform enzymes. Crosses show results for numerically optimized enzyme distributions. **(b)** Deviation of the optimal productivity of the optimized enzyme distributions from the well-mixed productivity collapse when plotted as a function of R/λ_σ. **(c)** A typical example of the optimized enzyme distributions $e^*_1(r)$ and $e^*_2(r)$ for $R = 1\,\mu\text{m}$ and $\sigma = 10^4\,\text{s}^{-1}$.

However, this will not generally be the case if the distribution of $\mathcal{E}_1$ enzymes, and therefore the production of $\mathcal{I}$ is non-uniform.

We therefore performed numerical optimization of the enzyme distributions in the full reaction-diffusion system, as described above but with the addition of a term $-\sigma i_n$ to Eqs. A.20d-f. We found that indeed for large R values, the productivity of the optimized enzyme distribution was greater than that of the well-mixed model (Fig. A.2a, crosses). In particular, for large values of σ, this deviation occurs at radii for which the $\sigma = 0$ system is still in the well-mixed regime ($R \lesssim 3\,\mu\text{m}$). Plotting the deviation of the productivity against R/λ_σ we find that the data collapse to a single curve (Fig. A.2b), confirming that the deviation is indeed a result of the decay gradients that come to dominate the intracompartment dynamics.

The corresponding optimal enzyme profiles are strongly clustered at the outer boundary of

the compartment (see Fig. A.2c for an example). Enzymes of type $\mathcal{E}_1$ localize at the outer boundary so as to maximize the reaction with incoming substrate. Enzymes of the second type $\mathcal{E}_2$ form a shell on the inner side of the $\mathcal{E}_1$ region. By colocalizing $\mathcal{E}_2$ with $\mathcal{E}_1$, the system is able to increase the local enzyme density compared to a well-mixed or uniform arrangement, thereby increasing the probability of converting each $\mathcal{I}$ to $\mathcal{P}$. Simultaneously, decay means that intermediates will rarely reach the central region of the compartment at $r \lesssim (R - \lambda_\sigma)$. By emptying this region of enzymes, the system avoids wasting $\mathcal{E}_2$ enzymes in regions of the compartment where there is no $\mathcal{I}$ to act on.

While we have seen that decay can lead to breakdown of the well-mixed approximation, extremely fast decay rates are necessary for such an effect to be significant for biologically realistic compartment sizes ($\sigma \gtrsim 10^5 \mathrm{s}^{-1}$ for our α-carboxysome parameters). Such short lifetimes seem unrealistic for small-molecule metabolites. Furthermore, the environment inside microcompartments is typically highly controlled, making spurious side reactions with competing enzymes unlikely. Therefore, we do not expect such effects to be important in most biological contexts.

A.7 Corrections to the well-mixed approximation

To study the deviation of the optimal productivity from the well-mixed approximation, we performed a series expansion of Eq. A.9 in powers of D^{-1}, while substituting for the parameters describing the enzyme distribution $e_T(R) = e^*_{T,\mathrm{wm}}(R)\left[1 + \delta e_T^{(1)}(R) D^{-1} + ...\right]$ and $\phi(R) = \phi^*_{\mathrm{wm}}\left[1 + \delta\phi^{(1)}(R) D^{-1} + ...\right]$. In the regime $R > R_c$, the resulting expression for the productivity is

$$\mathrm{P}\{\delta e_T^{(1)}(R), \delta\phi^{(1)}(R); R\} = \frac{\kappa_1\kappa_2 p_s s_0}{\left[(\kappa_2 p_s)^{1/3} + (\kappa_1 p_i)^{1/3}\right]^3} + \frac{\kappa_1^{4/3}\kappa_2 p_s p_i^{1/3} R(p_i - p_s)s_0}{5D\left[(\kappa_2 p_s)^{1/3} + (\kappa_1 p_i)^{1/3}\right]^4} + O(D^{-2}) \tag{A.23}$$

$$= \mathrm{P}^*_{\mathrm{wm}}\left[1 + \frac{(p_i - p_s)R}{5\,(1 + \phi^*_{\mathrm{wm}})\,D} + O(D^{-2})\right]. \tag{A.24}$$

Note that the term of order D^{-1} is independent of the corrections to the enzyme distributions, $\delta e_T^{(1)}(R)$ and $\delta\phi^{(1)}(R)$. Thus to first order $\mathrm{P}\{\delta e_T^{(1)}(R), \delta\phi^{(1)}(R); R\} = \mathrm{P}^*(R)$.

The first-order correction term vanishes if $p_s = p_i$. Therefore, in order to understand how the optimal productivity behaves in this case, we computed also correction terms of the order $\mathcal{O}(D^{-2})$ for $p_s = p_i = p$, including also for completeness second-order corrections to the optimal enzyme concentrations, $e^*_T(R) = e^*_{T,\mathrm{wm}}(R)\left[1 + \delta e_T^{(1)}(R) D^{-1} + \delta e_T^{(2)}(R) D^{-2} + ...\right]$ and $\phi^*(R) = \phi^*_{\mathrm{wm}}\left[1 + \delta\phi^{(1)}(R) D^{-1} + \delta\phi^{(2)}(R) D^{-2} + ...\right]$. We find that the productivity takes the form

$$\mathrm{P}\{\delta e_T^{(1)}(R), \delta e_T^{(2)}(R), \delta\phi^{(1)}(R), \delta\phi^{(2)}(R); R\} = \mathrm{P}^*_{\mathrm{wm}}(R)\left[1 + f\{\delta e_T^{(1)}(R), \delta\phi^{(1)}(R), R\} D^{-2} + \mathcal{O}(D^{-3})\right]. \tag{A.25}$$

The leading correction term $f\{\delta e_T^{(1)}(R), \delta\phi^{(1)}(R), R\}D^{-2}$ only depends on the first-order corrections to e_T^* and ϕ^*, denoted by $\delta e_T^{(1)}(R)$ and $\delta\phi^{(1)}(R)$ respectively. Since we are interested in the optimal productivity, we then maximize f over $\delta e_T^{(1)}(R)$ and $\delta\phi^{(1)}(R)$, which gives $\delta e_T^{(1)*}(R) = -pR/5$ and $\delta\phi^{(1)*}(R) = 0$. The resulting correction to the productivity then reads

$$f^* = f\{\delta e_T^{(1)*}(R), \delta\phi^{(1)*}(R), R\} = -\frac{3p^2R^2}{175}. \tag{A.26}$$

Consequently, in the limit of equal permeabilities, the true optimal productivity $\mathrm{P}^*(R)$ is expected to deviate significantly from the well-mixed approximation $\mathrm{P}^*_{\mathrm{wm}}(R)$ for a compartment radius $R_x \sim \sqrt{175/3}\frac{D}{p}$. Akin to the case where $p_s > p_i$, $\mathrm{P}^*(R)$ initially plateaus for $R > R_c$ and then decreases again for very large radii.

A.8 High-permeability limit, $p_s, p_i \to \infty$

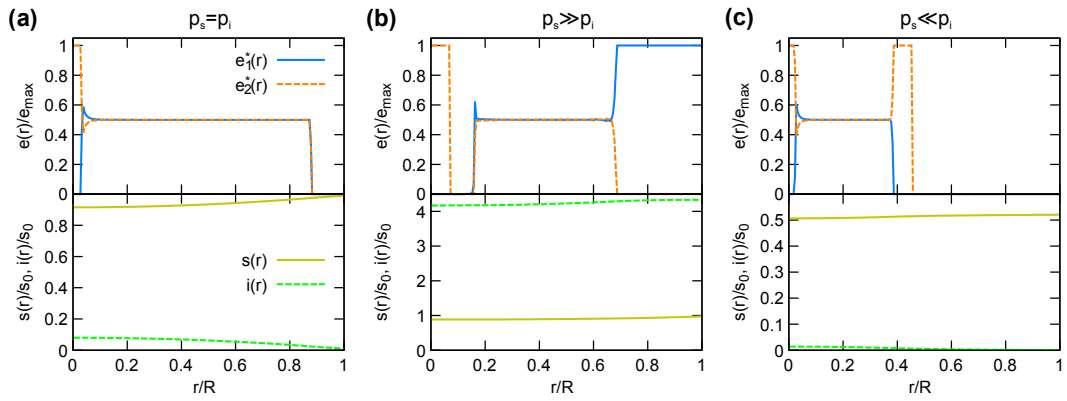

Figure A.3: Optimal intra-compartment enzyme distribution for high permeabilities. Optimal enzyme distributions for R =100 nm, Fig. 3.4(d-f), together with the corresponding metabolite density profiles. Parameters are as in Fig. 4: **(a)** $p_s = p_i = 10^4\,\mu\mathrm{m\,s^{-1}}$; **(b)** $p_s = 10^4\,\mu\mathrm{m\,s^{-1}}$, $p_i = 10\,\mu\mathrm{m\,s^{-1}}$; **(c)** $p_s = 10\,\mu\mathrm{m\,s^{-1}}$, $p_i = 10^4\,\mu\mathrm{m\,s^{-1}}$. Other parameters are fixed at $\kappa_1 = \kappa_2 = 0.4\,(\mu\mathrm{M\,s})^{-1}$, $e_{\max} = 25\,\mathrm{mM}$ and $D = 100\,\mu\mathrm{m^2\,s^{-1}}$.

In the $p_s, p_i \to \infty$ limit, the Robin boundary conditions Eqs. 4-5 reduce to Dirichlet boundary conditions,

$$s(R) = s_0,\ i(R) = 0. \tag{A.27}$$

Assuming uniformly distributed enzymes, $e_1(r) = e_1$ and $e_2(r) = e_2$, solving the reaction-diffusion equations Eqs. A.4-A.5 with these modified boundary conditions leads to the productivity

$$\mathrm{P}(e_T, \phi; R) = \frac{3s_0D}{e_TR^2(\kappa_2 - \kappa_1\phi)}\left\{\kappa_1\phi\left[1 - \sqrt{\frac{e_TR^2\kappa_2}{D(1+\phi)}}\coth\sqrt{\frac{e_TR^2\kappa_2}{D(1+\phi)}}\right]\right.$$
$$\left. - \kappa_2\left[1 - \sqrt{\frac{e_TR^2\kappa_1\phi}{D(1+\phi)}}\coth\sqrt{\frac{e_TR^2\kappa_1\phi}{D(1+\phi)}}\right]\right\}. \tag{A.28}$$

It can be seen that the enzyme density e_T and compartment radius R appear only in the combination e_TR^2.

A.9 Michaelis-Menten kinetics

We determined the optimal compartmentalization strategies with full Michaelis-Menten reaction kinetics for the reaction-diffusion system (Eqs. 3.2-3.5) with uniform enzyme profiles and for the well-mixed approximation. Because of the non-linear reaction kinetics in Eqs. 3.2-3.3 an analytical solution of the productivity is not possible. Therefore, we used a numerical method to solve the steady state reaction-diffusion system. This was done using in Matlab ODE solver for boundary value problems, `bvp5c`, together with the `fmincon` function for constrained optimization.

In the well-mixed approximation, the productivity can be determined analytically. The well-mixed steady state concentration of $\mathcal{S}$ and $\mathcal{I}$ with full Michaelis-Menten kinetics can be determined by solving the following system of equations,

$$0 = \frac{3p_s}{R}(s_0 - s) - \frac{k_{\mathrm{cat}}^{(1)} e_1 s}{K_{\mathrm{M}}^{(1)} + s}, \tag{A.29}$$

$$0 = \frac{3p_s}{R}(-i) + \frac{k_{\mathrm{cat}}^{(1)} e_1 s}{K_{\mathrm{M}}^{(1)} + s} - \frac{k_{\mathrm{cat}}^{(2)} e_2 i}{K_{\mathrm{M}}^{(2)} + i}. \tag{A.30}$$

The resulting substrate and intermediate concentration are given by,

$$s = \frac{\frac{3p_s}{k_{\mathrm{cat}}^{(1)}}\left(s_0 - K_{\mathrm{M}}^{(1)}\right) - e_1 R + \left[36\frac{k_{\mathrm{cat}}^{(1)}}{K_{\mathrm{M}}^{(1)}} p_s^2 s_0 + \left(\frac{3p_s}{k_{\mathrm{cat}}^{(1)}}\left(s_0 - K_{\mathrm{M}}^{(1)}\right) - e_1 R\right)^2\right]^{1/2}}{6p_s/k_{\mathrm{cat}}^{(1)}}, \tag{A.31}$$

$$i = \frac{R\frac{e_1 k_{\mathrm{cat}}^{(1)} s}{s + K_{\mathrm{M}}^{(1)}} - 3K_{\mathrm{M}}^{(2)} p_i - e_2 k_{\mathrm{cat}}^{(2)} R + \left[12RK_{\mathrm{M}}^{(2)} p_i \frac{e_1 k_{\mathrm{cat}}^{(1)} s}{s + K_{\mathrm{M}}^{(1)}} + \left(3K_{\mathrm{M}}^{(2)} p_i + e_2 k_{\mathrm{cat}}^{(2)} R - R\frac{e_1 k_{\mathrm{cat}}^{(1)} s}{s + K_{\mathrm{M}}^{(1)}}\right)^2\right]^{1/2}}{6p_i} \tag{A.32}$$

Finally, we can compute the productivity, by inserting Eq. A.32 in

$$\mathrm{P}(e_1, e_2, R) = \frac{i e_2 k_{\mathrm{cat}}^{(2)}}{(e_1 + e_2)\left(i + K_{\mathrm{M}}^{(2)}\right)}. \tag{A.33}$$

B Enzyme clustering and optimal enzyme arrangements

B.1 Equivalence of the expressions for the reaction efficiency

The reaction efficiency is defined via the reaction-diffusion equation (Eq. 4.2) as,

$$\frac{J_2}{J_1} = \int \alpha e(\mathbf{r}) \frac{\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \mathrm{d}\mathbf{r}. \tag{B.1}$$

Here, we demonstrate that in the regime of linear reaction kinetics ($\gamma = 0$), expression Eq. B.1 is equivalent to the definition of the reaction efficiency (Eq. 4.5),

$$\frac{J_2}{J_1} = \int_0^\infty P(E) p_\mathrm{r}(E) \mathrm{d}E, \tag{B.2}$$

which is based on considering individual intermediate trajectories.

We start by defining the steady-state intermediate concentration, $\rho(\mathbf{r})$, in terms of trajectories of diffusing intermediate molecules. We denote the diffusive trajectory of a single intermediate molecule, in the absence of any E_2 enzymes, as $\{\mathbf{r}(t)\}$. Every such trajectory has a time τ_esc associated with it at which the trajectory ends, either because the intermediate decayed or escaped through the system boundary. When we now reintroduce the E_2 enzymes according to the distribution $e(\mathbf{r})$, we can define a reaction propensity at each point of the trajectory, $\alpha e(\mathbf{r}(t))$. Thus, the survival probability that an intermediate, following the trajectory $\{\mathbf{r}(t)\}$, has not reacted with E_2 before the time t follows $\dot{S}(t|\{\mathbf{r}(t)\}) = -\alpha e(\mathbf{r}(t)) S(t|\{\mathbf{r}(t)\})$. Integrating this equation leads to the survival probability,

$$S(t|\{\mathbf{r}(t)\}) = \exp\left[-\alpha \int_0^t \mathrm{d}t' \; e(\mathbf{r}(t'))\right], \quad t \leq \tau_\mathrm{esc}. \tag{B.3}$$

The steady-state intermediate concentration at point $\mathbf{r}$ can be interpreted as the sum of trajectories that spend a certain time at $\mathbf{r}$, weighted by the probability that the intermediate molecule has not yet reacted prior to each return to $\mathbf{r}$. This weighting factor is simply the survival probability $S(t|\{\mathbf{r}(t)\})$. With this, we can write the intermediate concentration profile as,

$$\rho(\mathbf{r}) = \int \mathrm{d}\{\mathbf{r}(t)\} \; p(\{\mathbf{r}(t)\}) \int_0^{\tau_\mathrm{esc}} \mathrm{d}t \; S(t|\{\mathbf{r}(t)\}) \delta\left[\mathbf{r} - \mathbf{r}(t)\right], \tag{B.4}$$

where the inner integral is the weighted time of a single trajectory spent at $\mathbf{r}$, and the outer integral sums over the contributions of all possible trajectories weighted by the probability

$p(\{\mathbf{r}(t)\})$ that a specific trajectory $\{\mathbf{r}(t)\}$ occurs. Substituting Eqs. B.4 and B.3 into Eq. B.1 and changing the order of integration, we obtain

$$\begin{aligned}\frac{J_2}{J_1} &= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{\tau_{\mathrm{esc}}} \mathrm{d}t \int \mathrm{d}\mathbf{r}\ \alpha e(\mathbf{r}) \exp\left[-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))\right] \delta\left[\mathbf{r}-\mathbf{r}(t)\right] \\ &= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{\tau_{\mathrm{esc}}} \mathrm{d}t\ \alpha e(\mathbf{r}(t)) \exp\left[-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))\right] \\ &= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{\tau_{\mathrm{esc}}} \mathrm{d}t\ \frac{\mathrm{d}}{\mathrm{d}t}\left\{-\exp\left[-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))\right]\right\} \\ &= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \left\{1-\exp\left[-\alpha \int_0^{\tau_{\mathrm{esc}}} \mathrm{d}t'\ e(\mathbf{r}(t'))\right]\right\}. \end{aligned} \tag{B.5}$$

Defining the enzyme exposure as $E = \int_0^{\tau_{\mathrm{esc}}} \mathrm{d}t\ e(\mathbf{r}(t))$, and changing the variable of integration from $\{\mathbf{r}(t)\}$ to E, we recover

$$\frac{J_2}{J_1} = \int_0^{\infty} P(E)(1-e^{-\alpha E})\mathrm{d}E. \tag{B.6}$$

B.2 One-dimensional system with unstable intermediates

Here, we analytically solve the reaction-diffusion equation (Eq. 4.7) which is given by,

$$0 = \partial x^2 \rho(x) - \alpha e(x)\rho(x) - \beta\rho(x), \tag{B.7}$$

with the boundary conditions $\partial_x \rho(x)|_{x=0} = -1$ and $\partial_x \rho(x)|_{x=1} = 0$.

B.2.1 Clustered enzyme profile

The clustered enzyme profile is given by $e_{\mathrm{c}}(x) = \delta(x - x_0)$, where x_0 denotes the point where the E_2 enzymes are clustered. Note that we take x_0 to zero at the end of the calculation to assure colocalization of E_2 with E_1 enzymes. We start by dividing the system into two domains, domain I where $x < x_0$ and domain II where $x > x_0$. In each domain Eq. B.7 reduces to

$$0 = \partial x^2 \rho(x) - \beta\rho(x), \tag{B.8}$$

which has the solution

$$\rho_i(x) = A_i e^{\sqrt{\beta}x} + B_i e^{-\sqrt{\beta}x} \tag{B.9}$$

with $i = \{I, II\}$. Applying the boundary conditions at $x = 0$ and $x = 1$ yields

$$A_I - B_I = (\sqrt{\beta})^{-1}, \qquad A_{II} e^{\sqrt{\beta}} - B_{II} e^{-\sqrt{\beta}} = 0. \tag{B.10}$$

To guarantee that the intermediate concentration profile is continuous at x_0 we require, $\rho_I(x_0) = \rho_{II}(x_0)$ leading to

$$A_I e^{\sqrt{\beta}x_0} + B_I e^{-\sqrt{\beta}x_0} = A_{II} e^{\sqrt{\beta}x_0} + B_{II} e^{-\sqrt{\beta}x_0}. \tag{B.11}$$

Furthermore, to assure particle conservation in the system we integrate Eq. B.7 from $x_0 - \epsilon$ to $x_0 + \epsilon$ and take the limit of small ϵ, leading to

$$\lim_{\epsilon \to 0} \left((\partial_x \rho_{II}(x))_{x_0+\epsilon} - (\partial_x \rho_I(x))_{x_0-\epsilon} - \alpha \rho(x_0) - \beta \int_{x_0-\epsilon}^{x_0+\epsilon} \rho(x) dx \right) = 0. \tag{B.12}$$

The last term on the left hand side vanishes in the limit $\epsilon \to 0$, leading to the condition

$$\left[A_{II} e^{\sqrt{\beta} x_0} - B_{II} e^{-\sqrt{\beta} x_0} - \left(A_I e^{\sqrt{\beta} x_0} + B_I e^{-\sqrt{\beta} x_0} \right) \right] - \frac{\alpha}{\sqrt{\beta}} \left(A_I e^{\sqrt{\beta} x_0} + B_I e^{-\sqrt{\beta} x_0} \right) = 0. \tag{B.13}$$

To calculate the reaction efficiency, we integrate over the reaction term of Eq. B.7, leading to

$$\frac{J_2}{J_1} = \alpha \int_0^1 \delta(x - x_0) \rho(x) dx = \alpha (A_{II} e^{\sqrt{\beta} x_0} + B_{II} e^{-\sqrt{\beta} x_0}) \tag{B.14}$$

After some straightforward algebra we arrive at expressions for all four constants $\{A_I, A_{II}, B_I, B_{II}\}$. Substituting the integration variables and taking the limit $x_0 \to 0$, we obtain

$$\left(\frac{J_2}{J_1} \right)_{\mathrm{c}} = \frac{\alpha}{\alpha + \beta^{\frac{1}{2}} \tanh(\beta^{\frac{1}{2}})}. \tag{B.15}$$

B.2.2 Uniform enzyme profile

Eq. B.7 with uniformly distributed enzymes $e_{\mathrm{u}}(x) = 1$ reads

$$0 = \partial x^2 \rho(x) - (\alpha + \beta) \rho(x). \tag{B.16}$$

The solution of this equation is given by

$$\rho(x) = A e^{\sqrt{\alpha+\beta} x} + B e^{-\sqrt{\alpha+\beta} x} \tag{B.17}$$

Using the boundary conditions at $x = 0$ and $x = 1$ leads to the conditions

$$A - B = (\sqrt{\alpha + \beta})^{-1}, \qquad A e^{\sqrt{\alpha+\beta}} - B e^{-\sqrt{\alpha+\beta}} = 0. \tag{B.18}$$

Similarly to above, the constants A and B can be obtained straightforwardly, and the reaction efficiency reads,

$$\left(\frac{J_2}{J_1} \right)_{\mathrm{u}} = \alpha \int_0^1 \rho(x) dx = \frac{\alpha}{\alpha + \beta} \tag{B.19}$$

B.3 Enzyme exposure probability distribution in one dimension

The reaction efficiency defined in terms of the enzyme exposure probability distribution $P(E)$ reads,

$$\frac{J_2}{J_1} = 1 - \int_0^\infty P(E) e^{-\alpha E} dE = 1 - \frac{J_{\mathrm{loss}}}{J_1}. \tag{B.20}$$

To compute an exact expression of $P(E)$ for a given enzyme configuration, it is convenient to calculate the inverse Laplace transformation of J_{loss}/J_1 with respect to α.

We previously have seen that the expression for the loss term often takes the form $f(\beta)/(\alpha + f(\beta))$, where $f(x)$ is any function. For such expressions, the inverse Laplace transformation can easily be computed and takes the form,

$$P(E) = f(\beta)e^{-f(\beta)E}. \tag{B.21}$$

For the uniform enzyme profile we have, $f(\beta) = \beta$ and thus $P_{\text{u}}(E) = \beta e^{-E\beta}$; for the clustered enzyme profile we have, $f(\beta) = \beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}}$ and thus $P_{\text{c}}(E) = \beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}} e^{-E\beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}}}$.

B.4 Two- and three-dimensional systems

In two and three dimensions we impose rotational symmetry, and thus the reaction-diffusion equation depends only on the radial coordinate,

$$r^{1-d}\partial_r \left(r^{d-1} \partial_r \rho(r) \right) - \alpha e(r)\rho(r) = 0, \tag{B.22}$$

where d is the spatial dimension. At the center of the system we localize the source leading to the boundary condition, $(2\pi r \partial_r \rho(r))|_{r=0} = -1$ in tow dimensions and $(4\pi r^2 \partial_r \rho(r))|_{r=0} = -1$ in three dimensions. At the outer boundary we use the absorbing boundary condition, $\rho(1) = 0$.

B.4.1 Clustered enzyme profile

We again start by deriving the expression for the efficiency generated by a clustered enzyme profile. The rescaled clustered configuration is given by $e_{\text{c}}(r) = \frac{\delta(r-r_0)}{2r_0}$ in two dimensions and $e_{\text{c}}(r) = \frac{\delta(r-r_0)}{3r_0^2}$ in three dimensions. As in Sec. B.2.1 we divide the system into two domains, domain I for $r < r_0$ and domain II for $r > r_0$. In each domain no enzymes are present and thus the solution of Eq. B.22 is

$$\rho_i(r) = \begin{cases} A_i \log r + B_i & d = 2 \\ \frac{A_i}{r} + B_i & d = 3, \end{cases} \tag{B.23}$$

where $i = \{I, II\}$. Applying the boundary conditions leads to the following two conditions in the respective dimension

$$B_{II} = 0, \qquad A_I = -\frac{1}{2\pi} \qquad \text{for } d = 2 \tag{B.24}$$

$$A_{II} = -B_{II}, \qquad A_I = \frac{1}{4\pi} \qquad \text{for } d = 3. \tag{B.25}$$

The remaining two conditions come again from matching the concentration at $r = r_0$, $\rho_I(r_0) = \rho_{II}(r_0)$, and the discontinuity of the derivative of $\rho(r)$ at $r = r_0$. With these four conditions we get after some straightforward algebra the expressions for the four constant $\{A_I, A_{II}, B_I, B_{II}\}$.

Since loss of intermediates occurs only at the outer boundary, we can determine the efficiency by computing the leakage flux at the outer boundary. This then leads to the expressions of the reaction efficiency in two and three dimensions.

$$\left(\frac{J_2}{J_1}\right) = \begin{cases} 1 - 2\pi r \partial_r \rho(r)|_{r=1} = \frac{\alpha \log r_0}{\alpha \log r 0 - 2} & d = 2 \\ 1 - 4\pi r^2 \partial_r \rho(r)\big|_{r=1} = \frac{\frac{\alpha}{3}(1-r_0)}{r_0 + \frac{\alpha}{3}(1-r_0)} & d = 3. \end{cases} \tag{B.26}$$

B.4.2 Uniform enzyme profile

In the case of uniformly distributed enzymes ($e_{\mathrm{u}}(r) = 1$), the reaction-diffusion equation Eq. B.22 reads

$$r^{1-d}\partial_r(r^{d-1}\partial_r\rho(r)) - \alpha\rho(r) = 0. \tag{B.27}$$

For $d = 2$ and $d = 3$ this is solved by,

$$\rho(r) = \begin{cases} A\mathrm{I}_0(\alpha r) + B\mathrm{K}_0(\alpha r) & d = 2 \\ \frac{1}{r}\left(Ae^{\sqrt{\alpha}r} + Be^{-\sqrt{\alpha}r}\right) & d = 3. \end{cases} \tag{B.28}$$

Together with the source and absorbing boundary condition this leads to the following conditions,

$$B = -\frac{1}{2\pi}, \qquad A\mathrm{I}_0(\alpha) + B\mathrm{K}_0(\alpha) = 0 \qquad \text{for } d = 2 \tag{B.29}$$

$$A + B = \frac{1}{4\pi}, \qquad Ae^{\sqrt{\alpha}} + Be^{-\sqrt{\alpha}} = 0 \qquad \text{for } d = 3. \tag{B.30}$$

The reaction efficiency is then given by,

$$\begin{cases} 1 - 2\pi r \partial_r \rho(r)|_{r=1} = 1 - \mathrm{I}(\sqrt{\alpha})^{-1} & d = 2 \\ 1 - 4\pi r^2 \partial_r \rho(r)\big|_{r=1} = 1 - \sqrt{\alpha}\mathrm{csch}(\sqrt{\alpha}) & d = 3. \end{cases} \tag{B.31}$$

B.5 Enzyme exposure probability distribution in 2D and 3D

As described above, the enzyme exposure probability distribution is given as the inverse Laplace transformation of the loss flux through the boundary. The loss flux for the clustered configuration is given by,

$$\frac{J_{\mathrm{loss}}}{J_1} = \begin{cases} \frac{2}{2 - \alpha \log r_0} & d = 2 \\ \frac{r_0}{r_0 + \frac{\alpha}{3}(1-r_0)} & d = 3. \end{cases} \tag{B.32}$$

Both loss fluxes have the general form as discussed in Sec. B.3. Thus, the enzyme exposure distributions are given by,

$$P_{\mathrm{c}}(E) = \begin{cases} \frac{-2}{\log r_0} e^{\frac{2E}{\log r_0}} & d = 2 \\ \\ \frac{3r_0}{1-r_0} e^{-\frac{3r_0}{1-r_0}E} & d = 3. \end{cases} \tag{B.33}$$

For uniformly distributed enzymes, we first calculate the singularities of the loss flux, which are the zero points of, $\mathrm{I}_0(\sqrt{\alpha})$ in two dimensions and $\sinh(\sqrt{\alpha})$ in three dimensions. These are given in 2D by the n^{th} zero of the Bessel function via $j_{0,n}^2$ and in 3D by $\alpha_n = -(n\pi)^2$ with $n \in \mathbb{N}$. This then leads to the following residues

$$\mathrm{Res}\left[\left(\frac{J_{\mathrm{loss}}}{J_1}\right) e^{\alpha E}, \alpha_n\right] = \begin{cases} \frac{2e^{-j_{0,n}^2 E} j_{0,n}}{\mathrm{J}_1(j_{0,n})} & d = 2 \\ \\ (-1)^{n+1} 2(\pi n)^2 e^{-(n\pi)^2 E} & d = 3. \end{cases} \tag{B.34}$$

With this, the exposure distribution is determined by summing over all residues,

$$P_{\mathrm{u}}(E) = \begin{cases} 2\sum_{n=1}^{\infty} \frac{e^{-j_{0,n}^2 E} j_{0,n}}{\mathrm{J}_1(j_{0,n})} & d = 2 \\ \\ 2\sum_{n=1}^{\infty} (-1)^{n+1} (\pi n)^2 e^{-(n\pi)^2 E} & d = 3. \end{cases} \tag{B.35}$$

C Enhancing biocatalytic fluxes via spatial organization

C.1 Ensembles of biocatalyst arrangements

The distributions of pathway fluxes for different model parameters and localization strategies were determined in each case by sampling an ensemble of 3000 random biocatalyst configurations and computing for each configuration the steady-state flux.

For the delocalized scenario, these configurations were generated by distributing the biocatalysts uniformly over the system. For a two dimensional spherical symmetric system a uniform distribution is achieved by picking for the center of each biocatalyst a radial position $r = \sqrt{z}$, where z is uniformly distributed over the interval $z \in [0, (R - r_c)^2)$, and an angular position θ from the interval $\theta \in [0, 2\pi)$. Similarly, in three dimensions the radial position is $r = z^{1/3}$, and the angular coordinates $\theta \in [0, 2\pi)$ and $\phi = \arccos(2v - 1)$ where $v \in [0, 1]$. After distributing all biocatalysts in this way, we tested whether any two biocatalyst overlapped. If any pair of biocatalysts had a distance smaller than $2r_c$ between their centers, then we moved this pair away from each other, along the line connecting the centers, until their separation becomes larger than $2r_c$. After all overlapping pairs had relocated, the procedure was repeated to avoid overlaps created by the repositioning.

For the different studied biocatalyst organizations, we also considered ensembles of configurations generated by a similar procedure. In the case of $\mathcal{A}$-$\mathcal{B}$ pairs and the complex configurations in Fig. 5.3e, their centers were distributed randomly over the system as described above. The center-center distance between biocatalysts within a pair or complex was fixed at r_c, while their orientation was chosen randomly. In the case of clustered biocatalysts, a center position for the cluster within the system was chosen randomly. The biocatalysts were then randomly positioned within a circular (in 2d) or spherical (in 3d) region so as to achieve a packing density of 60% in 2d or 50% in 3d. In all cases, cycles of rearrangements were made in order to avoid biocatalyst overlaps.

C.2 Optimization of the complex arrangement

To determine the optimal $\mathcal{B}$ configuration around a single $\mathcal{A}$ localized at the system center, we used a Monte Carlo optimization algorithm to iteratively explore the biocatalyst configuration space.

The optimization algorithm was initialized with a random configuration of $\mathcal{B}$s. From this configuration a new trial configuration was sampled by selecting one $\mathcal{B}$ at random and moving

it a distance l in a random direction to a new position. If this trial configuration led to an increase in reaction flux $j_{\mathcal{P}}$ it was accepted and used as the starting configuration for the next trial step, otherwise it was rejected and a new trial was generated from the previous best configuration. This procedure was repeated until a termination criterion of either a defined total number of iterations (set to 10^4), or a number of successive non-improving iterations (300–400), was reached. We made two further modifications to this basic algorithm that were found to speed up convergence of the optimization. First, after a trial step was accepted the subsequent trial step was taken in the same direction. Second, the step length was decreased during the course of the optimization process.

In general, this procedure does not guarantee a convergence to the global optimum. We therefore performed 30 realizations of the optimization procedure, with different initial configurations, for each set of model parameters. Most realizations resulted in the same final configuration, which we are confident to be the global optimum.

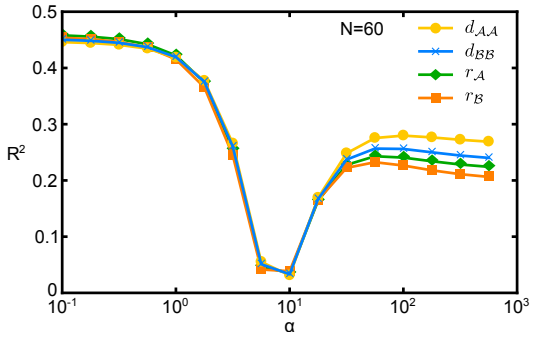

Figure C.1: Coefficient of determination. R^2 of the linear regression of $j_{\mathcal{P}}$ against $d_{\mathcal{AA}} = \langle|\mathbf{r}_i - \mathbf{r}_j|\rangle_{i,j\in\mathcal{A}}$, $d_{\mathcal{BB}} = \langle|\mathbf{r}_i - \mathbf{r}_j|\rangle_{i,j\in\mathcal{B}}$, $r_{\mathcal{A}} = \langle|\mathbf{r}_i|\rangle_{i\in\mathcal{A}}$, and $r_{\mathcal{B}} = \langle|\mathbf{r}_i|\rangle_{i\in\mathcal{B}}$ in systems of delocalized biocatalysts.

C.3 Permeable biocatalysts

To identify the impact of steric effects due to the discreteness of biocatalysts, we considered how the reaction flux is altered when the biocatalysts are permeable. The behavior of such a system is defined by reaction-diffusion equations where we distinguish between the spaces within and outside biocatalysts,

$$\left.\begin{aligned} D\nabla^2 C_{\mathcal{S}}(\mathbf{r}) - \kappa_{\mathcal{A}} C_{\mathcal{S}}(\mathbf{r}) &= 0 \\ D\nabla^2 C_{\mathcal{I}}(\mathbf{r}) + \kappa_{\mathcal{A}} C_{\mathcal{S}}(\mathbf{r}) &= 0 \end{aligned}\right\} \forall \mathbf{r} \in \mathcal{A}, \tag{C.1}$$

$$\left.\begin{aligned} D\nabla^2 C_{\mathcal{S}}(\mathbf{r}) &= 0 \\ D\nabla^2 C_{\mathcal{I}}(\mathbf{r}) - \kappa_{\mathcal{B}} C_{\mathcal{I}}(\mathbf{r}) &= 0 \end{aligned}\right\} \forall \mathbf{r} \in \mathcal{B}, \tag{C.2}$$

$$\left.\begin{aligned} D\nabla^2 C_{\mathcal{S}}(\mathbf{r}) &= 0 \\ D\nabla^2 C_{\mathcal{I}}(\mathbf{r}) &= 0 \end{aligned}\right\} \text{elsewhere}, \tag{C.3}$$

together with boundary conditions that require $C_{\mathcal{S},\mathcal{I}}$ to be continuous and smooth at the surfaces of all biocatalysts. However, it is not necessarily clear how the catalytic activity on

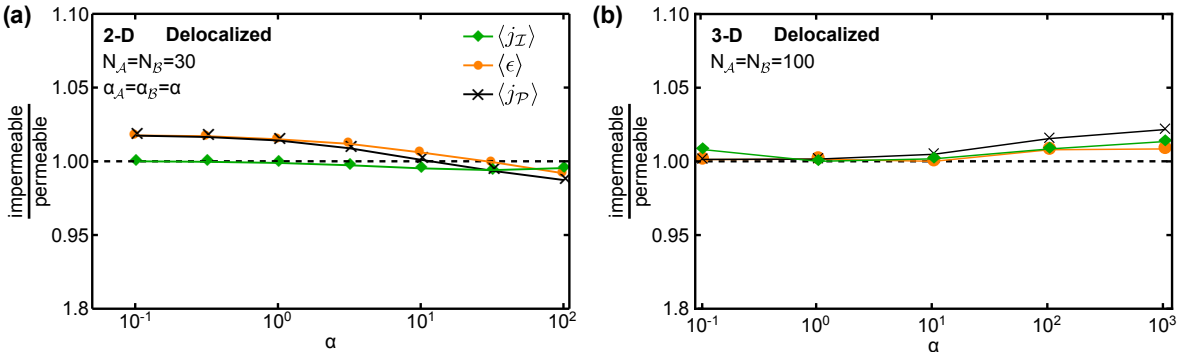

Figure C.2: Impermeable vs. permeable for delocalized arrangements. Comparison of $\langle J_\mathcal{I}\rangle$, $\langle\epsilon\rangle$, and $\langle J_\mathcal{P}\rangle$ for impermeable versus permeable biocatalysts for delocalized arrangements in two and three dimensions.

the surface of biocatalysts, $k_{\mathcal{A},\mathcal{B}}$, should be converted into a catalytic activity throughout the volume of such permeable biocatalysts, $\kappa_{\mathcal{A},\mathcal{B}}$. We used a mapping based on the equivalence of reaction fluxes in a simplified system with a single biocatalyst.

To this end we begin by considering a single $\mathcal{A}$ biocatalyst positioned at the system center. To determine its reaction flux, we solve the steady state diffusion equation, $D\nabla^2 C_\mathcal{S}(r) = 0$, with the boundary conditions used in the main text,

$$DC'_\mathcal{S}(r_c) = \frac{k_\mathcal{A}}{A_\mathcal{A}} C_\mathcal{S}(r_c), \tag{C.4}$$

$$C_\mathcal{S}(R) = C_0, \tag{C.5}$$

where $R = 100r_c$ is the outer boundary of the system. By solving this system, we obtain for the fluxes in 2 and 3 dimensions,

$$J_{\text{imp}} = \frac{k_\mathcal{A} C_0}{1 + \frac{k_\mathcal{A}}{2\pi D}\log\frac{R}{r_c}} \qquad d = 2, \tag{C.6}$$

$$J_{\text{imp}} = \frac{k_\mathcal{A} C_0}{1 + \frac{k_\mathcal{A}}{4\pi r_c D}\left(1 - \frac{r_c}{R}\right)} \qquad d = 3. \tag{C.7}$$

We next considered the scenario where the single central $\mathcal{A}$ biocatalyst is permeable to substrate. For this system, the reaction flux can be computed by dividing the system into two domains, inside and outside of the biocatalyst, and solving in each domain the corresponding reaction-diffusion equations,

$$D\nabla^2 C_\mathcal{S}(r) = 0 \qquad \text{if } r > r_c \tag{C.8}$$

$$D\nabla^2 C_\mathcal{S}(r) - \kappa_\mathcal{A} C_\mathcal{S}(r) = 0 \qquad \text{if } r < r_c \tag{C.9}$$

By requiring that the substrate concentration profile is continuous and differentiable at $r = r_c$, together with the boundary conditions $C_\mathcal{S}(R) = C_0$ and $C'_\mathcal{S}(0) = 0$, we can determine $C_\mathcal{S}(r)$. With this we compute the reaction flux by integrating the last term in Eq. C.9 over the domain of the biocatalyst. This yields the following expressions for the reaction flux in 2 and

3 dimensions,

$$J_{\text{perm}} = \frac{2\pi D C_0 \beta \, \mathrm{I}_1(\beta)}{\mathrm{I}_0(\beta) + \beta \, \mathrm{I}_1(\beta) \log \frac{R}{r_c}} \qquad d = 2, \tag{C.10}$$

$$J_{\text{perm}} = \frac{4\pi D C_0 \frac{R}{r_c} [\beta \coth(\beta) - 1]}{1 + (\frac{R}{r_c} - 1)\beta \coth(\beta)} \quad d = 3, \tag{C.11}$$

where $\beta = r_c \sqrt{\kappa_{\mathcal{A}}/D}$ and $\mathrm{I}_\alpha(x)$ is the modified Bessel function of the first kind.

To achieve an equivalence between the reaction rates κ_l and the rates k_l of the impermeable biocatalysts, we chose $\kappa_{\mathcal{A}}(k_{\mathcal{A}})$ such that the fluxes J_{perm} and J_{imp} match. Comparing Eqs. C.6-C.7 with C.10-C.11, we find that the fluxes will be equal when

$$\alpha_{\mathcal{A}} = \beta \mathrm{I}_1(\beta)/\mathrm{I}_0(\beta) \quad d = 2, \tag{C.12}$$

$$\alpha_{\mathcal{A}} = \beta \coth(\beta) - 1 \quad d = 3, \tag{C.13}$$

These mappings were then used for the rates $\kappa_{\mathcal{A}}(\alpha_{\mathcal{A}})$ and $\kappa_{\mathcal{B}}(\alpha_{\mathcal{B}})$ to solve Eqs. C.1-C.3 for general biocatalyst configurations. In systems for which steric effects are expected to be negligible, for example in the case of delocalized biocatalyst arrangements, using this mapping for the rates led to reaction fluxes which are approximately the same for impermeable and permeable biocatalysts (see Fig. C.2).

D Flux regulation via enzyme coclustering and sequestration

D.1 Supplementary figures

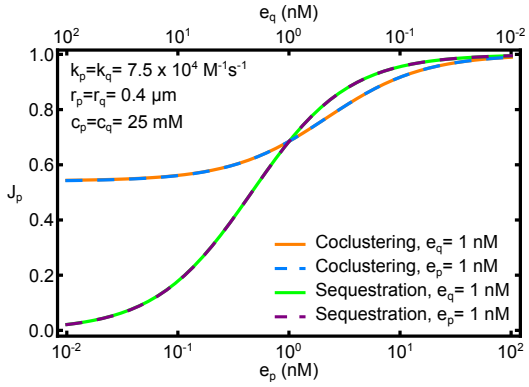

Figure D.1: Branching fraction only depends on the ratio $\mathbf{e_p/e_q}$. Branching fraction J_p as e_p and e_q are varied separately, with the other held fixed. The two curves for each scenario coincide, indicating that J_p depends only on the ratio e_p/e_q. Other parameter values are as shown.

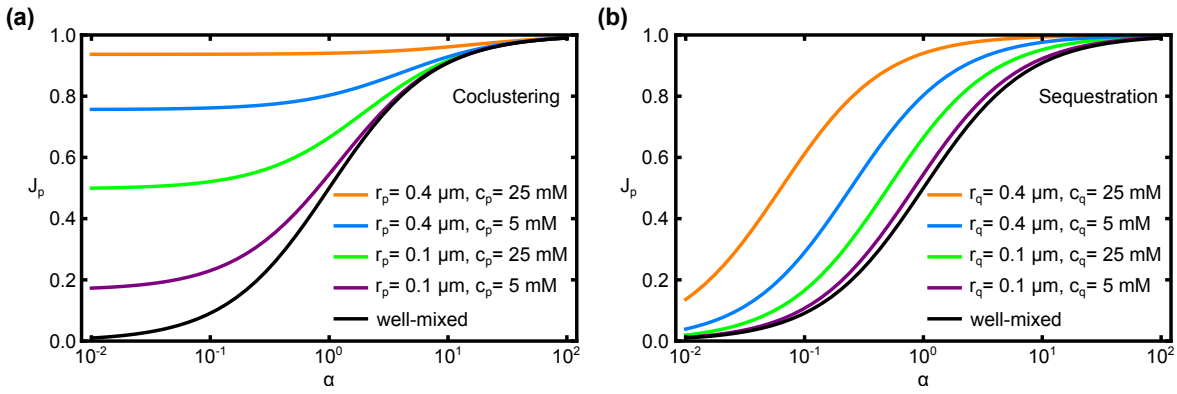

Figure D.2: Branching fraction for different cluster parameters. Branching fraction J_p for different values of the clustering parameters $r_{p,q}$ and $c_{p,q}$. Other parameters were fixed to $k_p = k_q = 10^6\,\mathrm{M^{-1}\,s^{-1}}$, and $D = 100\,\mu\mathrm{m^2\,s^{-1}}$.

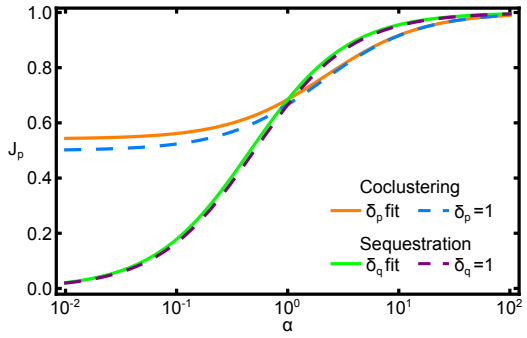

Figure D.3: Approximation with $\delta_{\mathbf{p,q}} = \mathbf{1}$. Comparison of the approximate model, Eqs. 8 and 11 of the main text, with $\delta_{p,q}$ fit to the full reaction-diffusion model (solid lines), and with $\delta_{p,q} = 1$ (dashed lines).

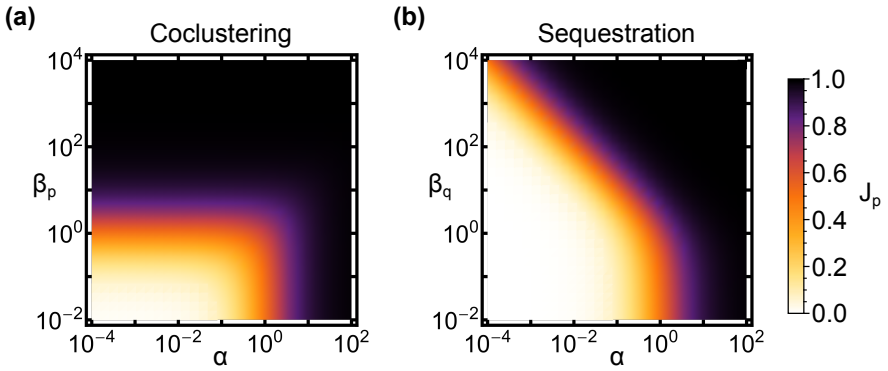

Figure D.4: Branching fractions. J_p for the **(a)** coclustering and **(b)** sequestration scenarios as a function of α and $\beta_{p,q}$.

D.2 Colocalization of consecutive enzymes into pairs

Instead of colocalizing several of the upstream and downstream enzymes into large coclusters, it has been suggested that flux regulation might also be achieved by colocalizing enzyme molecules into pairs. In this scenario, E_I-E_P pairs are well mixed with the competing E_Q enzymes. To study the ability of this strategy to direct the flux towards P, we develop a model that allows us to analytically determine the flux distribution at the branch point.

Let us imagine the system to be partitioned into identical reaction volumes where each volume contains only a single enzyme pair. Since we are interested in the branching fraction of the fluxes rather than absolute fluxes, we again only focus on one representative volume in isolation. The concentration of enzyme pairs in the system, and thus the concentration of E_P, e_p, is given by the radius R of the volume around the pair, $e_p = (4/3\pi R^3)^{-1}$. Inside the reaction volume, the enzyme pair is positioned such that E_P is localized at the center and E_I a distance r_0 away from the center. Furthermore, we assume that E_P is a spherical reaction center with radius r_E which is reactive at the entire surface. Therefore, the exact orientation of E_I does not matter, and we can assume the production of intermediates to occur uniformly on a spherical shell of radius r_0 around E_P. The transport of intermediates is again modeled by simple diffusion and intermediates also can react with the competing enzyme E_Q which is

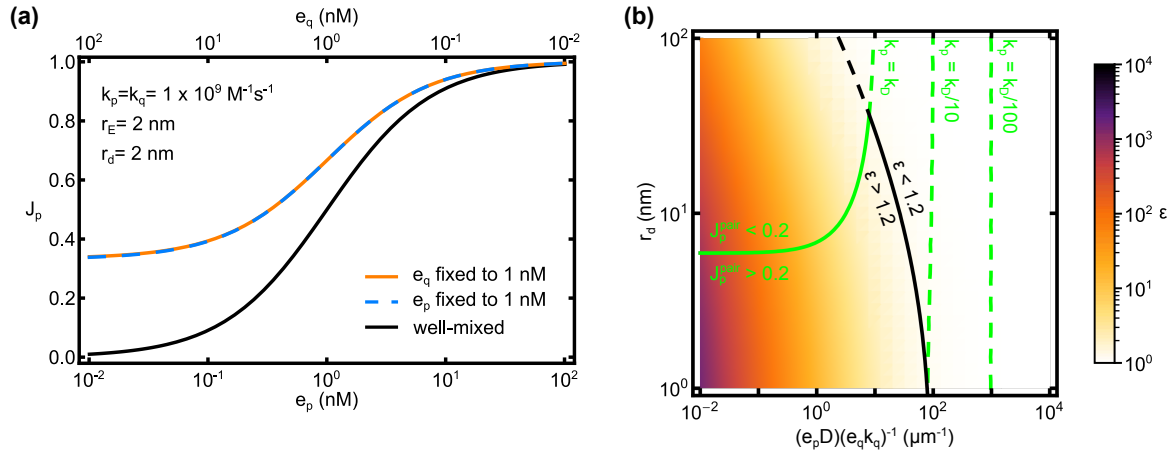

Figure D.5: Branching fraction and enhancement for enzyme pairs. (a) The branching fraction depends only on the ratio $\frac{e_p}{e_q}$. **(b)** Enhancement achieved by the pair strategy. The black curve is the contour line where $\epsilon = 1.2$. The green lines are the contour lines where $J_p^{pair} = 0.2$ for different catalytic efficiencies, $k_p = k_D$, $k_p = k_D/10$, and $k_p = k_D/100$.

assumed to be uniformly distributed throughout the reaction volume with the concentration, e_q. The corresponding reaction-diffusion equation governing the dynamics of intermediates is given by,

$$\frac{\partial \rho(r,t)}{\partial t} = D\nabla^2\rho(r,t) + j_i\delta(r - r_0) - k_q e_q \rho(r,t) \tag{D.1}$$

with the boundary conditions,

$$4\pi r_E^2 D \left.\frac{\partial \rho(r,t)}{\partial r}\right|_{r=r_E} = k_p^{int}\rho(r_E,t), \qquad \left.\frac{\partial \rho(r,t)}{\partial r}\right|_R = 0. \tag{D.2}$$

where $\rho(r,t)$ is the density profile of the intermediate I, D is the diffusion coefficient of I, k_p^{int} is the intrinsic reaction rate at the surface of E_P. In steady-state we can analytically solve this reaction-diffusion system and use the solution of the concentration profile to determine the branching fraction,

$$J_p^{pair} = \frac{k_p^{int}\rho(r_E)}{4\pi \int_0^{R(e_p)} r^2 j_i \delta(r - r_0)}. \tag{D.3}$$

To relate the intrinsic reaction rate, k_p^{int} to the catalytic efficiency, k_p, of E_P we assume the Smoluchowski-Collins-Kimball relation (see Sec. 1.1.5),

$$k_p = \frac{k_p^{int} k_D}{k_p^{int} + k_D}, \tag{D.4}$$

where k_D is the diffusion limited rate which for spherical reaction centers of radius r_E is given by, $k_D = 4\pi D r_E$. Using this relation for the intrinsic reaction rate and expressing the compartment radius in terms of the enzyme concentration e_p, we can determine J_p as a function of the catalytic efficiency of E_P, k_p, the concentration of E_P in pairs, e_p, and the rate of Q production catalyzed by E_Q, $k_q e_q$.

Unlike the clustering strategies discussed in Chap. 6, it is not possible to approximate the fluxes or extract from the analytical solutions the parameter combinations that govern the

behavior of the system. However, we find that the fluxes depend again only on the ratio between the average enzyme concentrations, e_p/e_q (see Fig. D.5a) but crucially not on the ratio of the overall activities. Instead, we identify as the parameter combinations that govern the fluxes, the ratio between average enzyme concentration compared to the ratio between the catalytic efficiency of E_Q and the diffusivity, $\frac{e_p}{e_q}\frac{D}{k_q}$, the ratio between the catalytic efficiency of E_P and the diffusivity, k_p/D, and the distance between E_I and E_P in the pair, $r_d = r_0 - 2r_E$. Notably, we find that the enhancement, $\epsilon = J_p/J_p^0$, is independent of k_p/D. This behavior can be understood by splitting the flux towards P into two contributions: the probability that an I molecule produced by E_I reaches E_P by diffusion times the probability that an I molecule is processed to P given it reached E_P. Only the latter probability depends on the catalytic efficiency of E_P and since it does not depend on the spatial arrangement of the enzymes the enhancement becomes independent of k_p.

We find that the condition, $\epsilon > 1.2$, can only be met if the overall catalytic activity of E_Q ($e_q k_q$) is large such that the distance an intermediate diffuses before it reacts with E_Q is small compared to the average distance between E_I and E_P in the well-mixed scenario (see Fig. D.5b). When we additionally require that, $J_p > 0.2$, E_P has to be diffusion limited ($k_p = k_D$) and in close proximity to E_I, $r_d < 6\,\mathrm{nm}$. For catalytic efficiencies, k_p, that are just an order of magnitude smaller than the diffusion limit, both requirements cannot be met simultaneously (see Fig. D.5b).

Bibliography

[1] Victor Henri. *Lois générales de l'action des diastases*. Librairie Scientifique A. Hermann, 1903.

[2] M. T. Michaelis and L. Menten. Die kinetik der invertinwirkung. *Biochem. Z.*, 49:333369, 1913.

[3] L. Stryer. *Biochemie*. Spektrum Akad. Verlag, Heidelberg, 1991.

[4] Alberts, Johnson, Lewis, Raff, Roberts, and Walter. *Molecular Biology of the cell*. Gerland Science, New York, 2002.

[5] A. Cornish-Bowden. *Fundamentals of enzyme kinetics*. Wiley-VCH, Weinheim, 2012.

[6] M. Eigen, W. Kruse, G. Maass, and L. DeMaeyer. Rate constants of protolytic reactions in aqueous solution. *Progress in reaction kinetics*, 2:287–318, 1964.

[7] M. Schurr. The role of diffusion in bimolecular solution kinetics. *Biophys. J.*, 10(8):700–716, 1970.

[8] D. Shoup and A. Szabo. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J.*, 40(1):33–39, 1982.

[9] M. Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Zeitschrift für physikalische Chemie*, 92(1):129–168, 1918.

[10] F. C. Collins and G. E. Kimball. Diffusion-controlled reaction rates. *Journal of colloid science*, 4(4):425–437, 1949.

[11] P. Debye. Reaction rates in ionic solutions. *Transactions of the Electrochemical Society*, 82(1):265–272, 1942.

[12] R. A. Alberty and G. G. Hammes. Application of the theory of diffusion-controlled reactions to enzyme kinetics. *J. Phys. Chem.*, 62(2):154–159, 1958.

[13] K.C. Chou. The kinetics of the combination reaction between enzyme and substrate. *Scientia Sinica*, 19(4):505–528, 1976.

[14] T. T. Li and K. C. Chou. The quantitative relations between diffusion-controlled reaction rate and characteristic parameters in enzyme-substrate reaction systems. i. neutral substrates. *Scientia Sinica*, 19(1):117–136, 1976.

[15] David Fell and Athel Cornish-Bowden. *Understanding the control of metabolism*, volume 2. Portland press London, 1997.

[16] H. Kacser and J. A. Burns. The control of flux. *Biochem. Soc. Trans.*, 23:341366, 1995.

[17] R. Heinrich and T. A. Rapoport. A linear steady state theory of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.*, 42(1):89–95, 1974.

[18] K. J. Kauffman, P. Prakash, and J. S. Edwards. Advances in flux balance analysis. *Curr. Opin. Biotech.*, 14(5):491–496, 2003.

[19] J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nat. Biotechnol.*, 28(3):245–248, 2010.

[20] C. H. Schilling, J. S. Edwards, D. Letscher, B. Ø. Palsson, et al. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnol. Bioeng.*, 71(4):286–306, 2000.

[21] R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, 2012.

[22] R. Ramakrishna, J. S. Edwards, A. McCulloch, and B. O. Palsson. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am. J. Physiol.*, 280(3):R695–R704, 2001.

[23] D. A. Beard, S. Liang, and H. Qian. Energy balance for analysis of complex metabolic networks. *Biophys. J.*, 83(1):79–86, 2002.

[24] Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A-L Barabási, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proc. Natl. Acad. Sci. U.S.A.*, 104(31):12663–12668, 2007.

[25] E. W. Miles, S. Rhee, and D. R. Davies. The molecular basis of substrate channeling. *J. Biol. Chem.*, 274(18):12193–12196, 1999.

[26] H. O. Spivey and J. Ovádi. Substrate channeling. *Methods*, 19(2):306–321, 1999.

[27] X. Huang, H. M. Holden, and F. M. Raushel. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Annu. Rev. Biochem.*, 70(1):149–180, 2001.

[28] R. N. Perham. Swinging arms and swinging domains in multifunctional enzymes: catalytic machines for multistep reactions. *Annu. Rev. Biochem.*, 69(1):961–1004, 2000.

[29] P. A. Srere. The metabolon. *Trends Biochem. Sci.*, 10(3):109–110, 1985.

[30] R. Heinrich, S. Schuster, and H.-G. Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *The FEBS Journal*, 201(1):1–21, 1991.

[31] F. M. Raushel, J. B. Thoden, and H. M. Holden. Enzymes with molecular tunnels. *Acc. Chem. Res.*, 36(7):539–548, 2003.

[32] C. C. Hyde, S. A. Ahmed, E. A. Padlan, E. W. Miles, and D. R. Davies. Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from salmonella typhimurium. *J. Biol. Chem.*, 263(33):17857–17871, 1988.

[33] B. A. Manjasetty, J. Powlowski, and A. Vrielink. Crystal structure of a bifunctional aldolase–dehydrogenase: Sequestering a reactive and volatile intermediate. *Proc. Natl. Acad. Sci. U.S.A.*, 100(12):6992–6997, 2003.

[34] A. Teplyakov, G. Obmolova, B. Badet, and M.-A. Badet-Denisot. Channeling of ammonia in glucosamine-6-phosphate synthase. *J. Mol. Biol.*, 313(5):1093–1102, 2001.

[35] Z. H. Zhou, D. B. McCarthy, C. M. O'Connor, L. J. Reed, and J. K. Stoops. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *Proc. Natl. Acad. Sci. U.S.A.*, 98(26):14802–14807, 2001.

[36] M. Smolle, A. E. Prior, A. E. Brown, A. Cooper, O. Byron, and J. G. Lindsay. A new level of architectural complexity in the human pyruvate dehydrogenase complex. *J. Biol. Chem.*, 281(28):19772–19780, 2006.

[37] I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov, and M. Sigman. Substrate channelling as an approach to cascade reactions. *Nat. Chem.*, 8(4):299–309, 2016.

[38] D. R. Knighton, C.-C. Kan, E. Howland, C. A. Janson, Z. Hostomska, K. M. Welsh, and D. A. Matthews. Structure of and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase. *Nat. Struct. Mol. Biol.*, 1(3):186–194, 1994.

[39] F. Wu and S. Minteer. Krebs cycle metabolon: structural evidence of substrate channeling revealed by cross-linking and mass spectrometry. *Angew. Chem. Int. Ed.*, 54(6):1851–1854, 2015.

[40] A. H. Elcock and J. A. McCammon. Evidence for electrostatic channeling in a fusion protein of malate dehydrogenase and citrate synthase. *Biochemistry*, 35(39):12652–12658, 1996.

[41] R. W. Guynn, H. J. Gelberg, and R. L. Veech. Equilibrium constants of the malate dehydrogenase, citrate synthase, citrate lyase, and acetyl coenzyme a hydrolysis reactions under physiological conditions. *J. Biol. Chem.*, 248(20):6957–6965, 1973.

[42] A. H. Elcock, M. J. Potter, D. A. Matthews, D. R. Knighton, and J. A. McCammon. Electrostatic channeling in the bifunctional enzyme dihydrofolate reductase-thymidylate synthase. *J. Mol. Biol.*, 262(3):370–374, 1996.

[43] D. F. Savage, B. Afonso, A. H. Chen, and P. A. Silver. Spatially ordered dynamics of the bacterial carbon fixation machinery. *Science*, 327(5970):1258–1261, 2010.

[44] W. Bonacci, P. K. Teng, B. Afonso, H. Niederholtmeyer, P. Grob, P. A. Silver, and D. F. Savage. Modularity of a carbon-fixing protein organelle. *Proc. Natl. Acad. Sci. U.S.A.*, 109(2):478–483, 2012.

[45] S. An, Y. Deng, J. W. Tomsho, M. Kyoung, and S. J. Benkovic. Microtubule-assisted mechanism for functional metabolic macromolecular complex formation. *Proc. Natl. Acad. Sci. U.S.A.*, 107(29):12872–12876, 2010.

[46] H. Zhao, J. B French, Y. Fang, and S. J. Benkovic. The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *ChemComm*, 49(40):4444–4452, 2013.

[47] C. M. Agapakis, P. M. Boyle, and P. A. Silver. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.*, 8(6):527–535, 2012.

[48] I. J. Van Der Klei, W. Harder, and M. Veenhuis. Biosynthesis and assembly of alcohol oxidase, a peroxisomal matrix protein in methylotrophic yeasts: a review. *Yeast*, 7(3):195–209, 1991.

[49] P. Ozimek, M. Veenhuis, and I. J. van der Klei. Alcohol oxidase: a complex peroxisomal, oligomeric flavoprotein. *FEMS yeast research*, 5(11):975–983, 2005.

[50] F. A. G. Reubsaet, J. H. Veerkamp, M. L. P. Brückwilder, J. M. F. Trijbels, and L. A. H. Monnens. Peroxisomal oxidases and catalase in liver and kidney homogenates of normal and di (ethylhexyl) phthalate-fed rats. *International journal of biochemistry*, 23(9):961–967, 1991.

[51] U. Schlattner, M. Tokarska-Schlattner, and T. Wallimann. Metabolite channeling: creatine kinase microcompartments. 2011.

[52] C. A. Kerfeld, S. Heinhorst, and G. C. Cannon. Bacterial microcompartments. *Annu. Rev. Microbiol.*, 64, 2010.

[53] S. Cheng, Y. Liu, C. S. Crowley, T. O. Yeates, and T. A. Bobik. Bacterial microcompartments: their properties and paradoxes. *Bioessays*, 30(11-12):1084–1095, 2008.

[54] T. O. Yeates, C. A. Kerfeld, S. Heinhorst, G. C. Cannon, and J. M. Shively. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.*, 6(9):681–691, 2008.

[55] T. A. Bobik, B. P. Lehman, and T. O. Yeates. Bacterial microcompartments: widespread prokaryotic organelles for isolation and optimization of metabolic pathways. *Mol. Microbiol.*, 98(2):193–207, 2015.

[56] T. A. Bobik. Polyhedral organelles compartmenting bacterial metabolic processes. *Appl. Microbiol. Biotechnol.*, 70(5):517–525, 2006.

[57] C. A. Kerfeld, M. R. Sawaya, S. Tanaka, C. V. Nguyen, M. Phillips, M. Beeby, and T. O. Yeates. Protein structures forming the shell of primitive bacterial organelles. *Science*, 309(5736):936–938, 2005.

[58] S. Tanaka, C. A. Kerfeld, M. R. Sawaya, F. Cai, S. Heinhorst, G. C. Cannon, and T. O. Yeates. Atomic-level models of the bacterial carboxysome shell. *science*, 319(5866):1083–1086, 2008.

[59] A. Kaplan, R. Schwarz, J. Lieman-Hurwitz, and L. Reinhold. Physiological and molecular aspects of the inorganic carbon-concentrating mechanism in cyanobacteria. *Plant Physiol.*, 97(3):851–855, 1991.

[60] M. R. Badger and G. D. Price. Co2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J. Exp. Bot.*, 54(383):609–622, 2003.

[61] A. Kaplan and L. Reinhold. Co2 concentrating mechanisms in photosynthetic microorganisms. *Annu. Rev. Plant Biol.*, 50(1):539–570, 1999.

[62] G. D. Price, M. R. Badger, F. J. Woodger, and B. M. Long. Advances in understanding the cyanobacterial co2-concentrating-mechanism (ccm): functional components, ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J. Exp. Bot.*, 59(7):1441–1461, 2007.

[63] M. Volokita, D. Zenvirth, A. Kaplan, and L. Reinhold. Nature of the inorganic carbon species actively taken up by the cyanobacterium anabaena variabilis. *Plant Physiol.*, 76(3):599–602, 1984.

[64] G. D. Price, F. J. Woodger, M. R. Badger, S. M. Howitt, and L. Tucker. Identification of a sulp-type bicarbonate transporter in marine cyanobacteria. *Proc. Natl. Acad. Sci. U.S.A.*, 101(52):18228–18233, 2004.

[65] L. Reinhold, M. Zviman, and A. Kaplan. A quantitative model for inorganic carbon fluxes and photosynthesis in cyanobacteria. *Plant Physiol. & Biochem.*, 27(6):945–954, 1989.

[66] L. Reinhold, R. Kosloff, and A. Kaplan. A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Canadian Journal of Botany*, 69(5):984–988, 1991.

[67] N. M. Mangan and M. P. Brenner. Systems analysis of the co2 concentrating mechanism in cyanobacteria. *Elife*, 3:e02043, 2014.

[68] J. M. Shively, C. E. Bradburne, H. C. Aldrich, T. A. Bobik, J. L. Mehlman, S. Jin, and S. H. Baker. Sequence homologs of the carboxysomal polypeptide csos1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes-polyhedral bodies. *Canadian Journal of Botany*, 76(6):906–916, 1998.

[69] I. Stojiljkovic, A. J. Bäumler, and F. Heffron. Ethanolamine utilization in salmonella typhimurium: nucleotide sequence, protein expression, and mutational analysis of the ccha cchb eute eutj eutg euth gene cluster. *J. Bacteriol.*, 177(5):1357–1366, 1995.

[70] E. M. Sampson and T. A. Bobik. Microcompartments for b12-dependent 1, 2-propanediol degradation provide protection from dna and cellular damage by a reactive metabolic intermediate. *J. Bacteriol.*, 190(8):2966–2971, 2008.

[71] D. M. Roof and J. R. Roth. Ethanolamine utilization in salmonella typhimurium. *J. Bacteriol.*, 170(9):3855–3863, 1988.

[72] S. R. Brinsmade, T. Paldon, and J. C. Escalante-Semerena. Minimal functions and physiological conditions required for growth of salmonella enterica on ethanolamine in the absence of the metabolosome. *J. Bacteriol.*, 187(23):8039–8046, 2005.

[73] A. H. Chen and P. A. Silver. Designing biological compartmentalization. *Trends Cell. Biol.*, 22(12):662–670, 2012.

[74] R. Narayanaswamy, M. Levy, M. Tsechansky, G. M. Stovall, J. D. O'Connell, J. Mirrielees, A. D. Ellington, and E. M. Marcotte. Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proc. Natl. Acad. Sci. U.S.A.*, 106(25):10147–10152, 2009.

[75] J. D. O'Connell, A. Zhao, A. D. Ellington, and E. M. Marcotte. Dynamic reorganization of metabolic enzymes into intracellular bodies. *Annu. Rev. Cell. Dev. Biol.*, 28:89–111, 2012.

[76] S. An, R. Kumar, E. D. Sheets, and S. J. Benkovic. Reversible compartmentalization of de novo purine biosynthetic complexes in living cells. *Science*, 320(5872):103–106, 2008.

[77] H. Zhao, C. R. Chiaro, L. Zhang, P. B. Smith, C. Y. Chan, A. M. Pedley, R. J. Pugh, J. B. French, A. D. Patterson, and S. J. Benkovic. Quantitative analysis of purine nucleotides indicates that purinosomes increase de novo purine biosynthesis. *J. Biol. Chem.*, 290(11):6705–6713, 2015.

[78] D. L. Schmitt and S. An. Spatial organization of metabolic enzyme complexes in cells. *Biochemistry*, 2017.

[79] A. M. Pedley and S. J. Benkovic. A new view into the regulation of purine metabolism: The purinosome. *Trends Biochem. Sci.*, 2016.

[80] Y. Deng, J. Gam, J. B. French, H. Zhao, S. An, and S. J. Benkovic. Mapping protein-protein proximity in the purinosome. *J. Biol. Chem.*, 287(43):36201–36207, 2012.

[81] M. Kyoung, S. J. Russell, C. L. Kohnhorst, N. N. Esemoto, and S. An. Dynamic architecture of the purinosome involved in human de novo purine biosynthesis. *Biochemistry*, 54(3):870–880, 2015.

[82] V. Baresova, M. Krijt, V. Skopova, O. Souckova, S. Kmoch, and M. Zikanova. Crispr-cas9 induced mutations along de novo purine synthesis in hela cells result in accumulation of individual enzyme substrates and affect purinosome formation. *Molecular genetics and metabolism*, 119(3):270–277, 2016.

[83] D. L. Schmitt, Y. Cheng, J. Park, and S. An. Sequestration-mediated downregulation of de novo purine biosynthesis by ampk. *ACS Chem. Biol.*, 11(7):1917–1924, 2016.

[84] I. Chitrakar, D. M. Kim-Holzapfel, W. Zhou, and J. B. French. Higher order structures in purine and pyrimidine metabolism. *J. Struct. Biol.*, 2017.

[85] A. M. Robitaille, S. Christen, M. Shimobayashi, M. Cornu, L. L. Fava, S. Moes, C. Prescianotto-Baschong, U. Sauer, P. Jenoe, and M. N. Hall. Quantitative phosphoproteomics reveal mtorc1 activates de novo pyrimidine synthesis. *Science*, 339(6125):1320–1323, 2013.

[86] I. Lieberman. Enzymatic amination of uridine triphosphate to cytidine triphosphate. *J. Biol. Chem.*, 222(2):765–775, 1956.

[87] C. W. Long and A. B. Pardee. Cytidine triphosphate synthetase of escherichia coli b i. purification and kinetics. *J. Biol. Chem.*, 242(20):4715–4721, 1967.

[88] K. P. Chakraborty and R. B. Hurlbert. Role of glutamine in the biosynthesis of cytidine nucleotides in escherichia coli. *Biochim. Biophys. Acta*, 47(3):607–609, 1961.

[89] M. Ingerson-Mahar, A. Briegel, J. N. Werner, G. J. Jensen, and Z. Gitai. The metabolic enzyme ctp synthase forms cytoskeletal filaments. *Nat. Cell Biol.*, 12(8):739–746, 2010.

[90] C. Noree, B. K. Sato, R. M. Broyer, and J. E. Wilhelm. Identification of novel filament-forming proteins in saccharomyces cerevisiae and drosophila melanogaster. *J. Cell Biol.*, pages jcb–201003001, 2010.

[91] J.-L. Liu. Intracellular compartmentation of ctp synthase in drosophila. *Journal of genetics and genomics*, 37(5):281–296, 2010.

[92] W. C. Carcamo, M. Satoh, H. Kasahara, N. Terada, T. Hamazaki, J. Y. F. Chan, B. Yao, S. Tamayo, G. Covini, C. A. von Mühlen, et al. Induction of cytoplasmic rods and rings structures by inhibition of the ctp and gtp synthetic pathway in mammalian cells. *PLoS one*, 6(12):e29690, 2011.

[93] J.-L. Liu. The enigmatic cytoophidium: compartmentation of ctp synthase via filament formation. *Bioessays*, 33(3):159–164, 2011.

[94] G. Azzam and J.-L. Liu. Only one isoform of drosophila melanogaster ctp synthase forms the cytoophidium. *PLoS Genet.*, 9(2):e1003256, 2013.

[95] K.-M. Gou, C.-C. Chang, Q.-J. Shen, L.-Y. Sung, and J.-L. Liu. Ctp synthase forms cytoophidia in the cytoplasm and nucleus. *Exp. Cell Res.*, 323(1):242–253, 2014.

[96] J.-L. Liu. The cytoophidium and its kind: filamentation and compartmentation of metabolic enzymes. *Annu. Rev. Cell. Dev. Biol.*, 32:349–372, 2016.

[97] G. N. Aughey, S. J. Grice, Q.-J. Shen, Y. Xu, C.-C. Chang, G. Azzam, P.-Y. Wang, L. Freeman-Mills, L.-M. Pai, L.-Y. Sung, et al. Nucleotide synthesis is regulated by cytoophidium formation during neurodevelopment and adaptive metabolism. *Biology open*, 3(11):1045–1056, 2014.

[98] K. Chen, J. Zhang, Ö. Y. Tastan, Z. A. Deussen, M. Y.-Y. Siswick, and J.-L. Liu. Glutamine analogs promote cytoophidium assembly in human and drosophila cells. *Journal of genetics and genomics*, 38(9):391–402, 2011.

[99] G. N. Aughey, S. J. Grice, and J.-L. Liu. The interplay between myc and ctp synthase in drosophila. *PLoS Genet.*, 12(2):e1005867, 2016.

[100] S. J. Calise, D. L. Purich, T. Nguyen, D. A. Saleem, C. Krueger, J. D. Yin, and E. K. L. Chan. rod and ringformation from imp dehydrogenase is regulated through the one-carbon metabolic pathway. *J Cell Sci*, 129(15):3042–3052, 2016.

[101] S. J. Calise, W. C. Carcamo, C. Krueger, J. D. Yin, D. L. Purich, and E. K. L. Chan. Glutamine deprivation initiates reversible assembly of mammalian rods and rings. *Cell. Mol. Life Sci.*, 71(15):2963–2973, 2014.

[102] R. M. Barry, A.-F. Bitbol, A. Lorestani, E. J. Charles, C. H. Habrian, J. M. Hansen, H.-J. Li, E. P. Baldwin, N. S. Wingreen, J. M. Kollman, et al. Large-scale filament formation inhibits the activity of ctp synthetase. *Elife*, 3, 2014.

[103] C. Noree, E. Monfort, A. K. Shiau, and J. E. Wilhelm. Common regulatory control of ctp synthase enzyme activity and filament formation. *Mol. Biol. Cell*, 25(15):2282–2290, 2014.

[104] G. N. Aughey and J.-L. Liu. Metabolic regulation via enzyme filamentation. *Crit. Rev. Biochem. Mol. Biol.*, 51(4):282–293, 2016.

[105] M. Pilhofer and G. J. Jensen. The bacterial cytoskeleton: more than twisted filaments. *Curr. Opin. Cell Biol*, 25(1):125–133, 2013.

[106] Y. S. Ji, J. J. Gu, A. M. Makhov, J. D. Griffith, and B. S. Mitchell. Regulation of the interaction of inosine monophosphate dehydrogenase with mycophenolic acid by gtp. *J. Biol. Chem.*, 281(1):206–212, 2006.

[107] G. Labesse, T. Alexandre, L. Vaupré, I. Salard-Arnaud, J. L. K. Him, B. Raynal, P. Bron, and H. Munier-Lehmann. Mgatp regulates allostery and fiber formation in impdhs. *Structure*, 21(6):975–985, 2013.

[108] G. D. Keppeke, S. J. Calise, E. K. L. Chan, and L. E. C. Andrade. Assembly of impdh2-based, ctps-based, and mixed rod/ring structures is dependent on cell type and conditions of induction. *Journal of Genetics and Genomics*, 42(6):287–299, 2015.

[109] C.-C. Chang, W.-C. Lin, L.-M. Pai, H.-S. Lee, S.-C. Wu, S.-T. Ding, J.-L. Liu, and L.-Y. Sung. Cytoophidium assembly reflects upregulation of impdh activity. *J Cell Sci*, 128(19):3550–3555, 2015.

[110] M. Jin, G. G. Fuller, T. Han, Y. Yao, A. F. Alessi, M. A. Freeberg, N. P. Roach, J. J. Moresco, A. Karnovsky, M. Baba, et al. Glycolytic enzymes coalesce in g bodies under hypoxic stress. *Cell reports*, 20(4):895–908, 2017.

[111] N. Miura, M. Shinohara, Y. Tatsukami, Y. Sato, H. Morisaka, K. Kuroda, and M. Ueda. Spatial reorganization of saccharomyces cerevisiae enolase to alter carbon metabolism under hypoxia. *Eukaryotic cell*, 12(8):1106–1119, 2013.

[112] S. Jang, J. C. Nelson, E. G. Bend, L. Rodríguez-Laureano, F. G. Tueros, L. Cartagenova, K. Underwood, E. M. Jorgensen, and D. A. Colón-Ramos. Glycolytic enzymes localize to synapses under energy stress to support synaptic function. *Neuron*, 90(2):278–291, 2016.

[113] E. Puchulu-Campanella, H. Chu, D. J. Anstee, J. A. Galan, W. A. Tao, and P. S. Low. Identification of the components of a glycolytic enzyme metabolon on the human red blood cell membrane. *J. Biol. Chem.*, 288(2):848–858, 2013.

[114] C. L. Kohnhorst, M. Kyoung, M. Jeon, D. L. Schmitt, E. L. Kennedy, J. Ramirez, S. M. Bracey, B. T. Luu, S. J. Russell, and S. An. Identification of a multienzyme complex for glucose metabolism in living cells. *J. Biol. Chem.*, 292(22):9191–9203, 2017.

[115] X. Zhao, H. Palacci, V. Yadav, M. M. Spiering, M. K. Gilson, P. J. Butler, H. Hess, S. J. Benkovic, and A. Sen. Substrate-driven chemotactic assembly in an enzyme cascade. *Nat. Chem.*, 2017.

[116] O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman, and I. Willner. Enzyme cascades activated on topologically programmed dna scaffolds. *Nat. Nanotechnol.*, 4(4):249, 2009.

[117] J. Fu, M. Liu, Y. Liu, N. W. Woodbury, and H. Yan. Interenzyme substrate diffusion

for an enzyme cascade organized on spatially addressable dna nanostructures. *J. Am. Chem. Soc.*, 134(12):5516–5519, 2012.

[118] J. E. Dueber, G. C. Wu, G. R. Malmirchegini, T. S. Moon, C. J. Petzold, A. V. Ullal, K. L. J. Prather, and J. D. Keasling. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.*, 27(8):753, 2009.

[119] A. D. Lawrence, S. Frank, S. Newnham, M. J. Lee, I. R. Brown, W.-F. Xue, M. L. Rowe, D. P. Mulvihill, M. B. Prentice, M. J. Howard, et al. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.*, 3(7):454–465, 2014.

[120] Y. Fu, D. Zeng, J. Chao, Y. Jin, Z. Zhang, H. Liu, D. Li, H. Ma, Q. Huang, K. V. Gothelf, et al. Single-step rapid assembly of dna origami nanostructures for addressable nanoscale bioreactors. *J. Am. Chem. Soc.*, 135(2):696–702, 2012.

[121] V. Linko, M. Eerikäinen, and M. A. Kostiainen. A modular dna origami-based enzyme cascade nanoreactor. *ChemComm*, 51(25):5351–5354, 2015.

[122] J. Fu, Y. R. Yang, A. Johnson-Buck, M. Liu, Y. Liu, N. G. Walter, N. W. Woodbury, and H. Yan. Multi-enzyme complexes on dna scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.*, 9(7):531, 2014.

[123] R. J. Conrado, G. C. Wu, J. T. Boock, H. Xu, S. Y. Chen, T. Lebar, J. Turnšek, N. Tomšič, M. Avbelj, R. Gaber, et al. Dna-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic acids research*, 40(4):1879–1889, 2011.

[124] C. J. Delebecque, A. B. Lindner, P. A. Silver, and F. A. Aldaye. Organization of intracellular reactions with rationally designed rna assemblies. *Science*, 333(6041):470–474, 2011.

[125] Y. Zhang, S. Tsitkov, and H. Hess. Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade. *Nat. Commun.*, 7:13982, 2016.

[126] Y. Zhang and H. Hess. Toward rational design of high-efficiency enzyme cascades, 2017.

[127] R. H. Doi and A. Kosugi. Cellulosomes: plant-cell-wall-degrading enzyme complexes. *Nat. Rev. Microbiol.*, 2(7):541, 2004.

[128] C. Fontes and H. J. Gilbert. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.*, 79:655–681, 2010.

[129] E. A. Bayer, J.-P. Belaich, Y. Shoham, and R Lamed. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.*, 58:521–554, 2004.

[130] E. A. Bayer, R. Kenig, and R. Lamed. Adherence of clostridium thermocellum to cellulose. *J. Bacteriol.*, 156(2):818–827, 1983.

[131] H.-P. Fierobe, A. Mechaly, C. Tardif, A. Belaich, R. Lamed, Y. Shoham, J.-P. Belaich, and E. A. Bayer. Design and production of active cellulosome chimeras selective in-

corporation of dockerin-containing enzymes into defined functional complexes. *J. Biol. Chem.*, 276(24):21257–21261, 2001.

[132] S. Moraïs, Y. Barak, J. Caspi, Y. Hadar, R. Lamed, Y. Shoham, D. B. Wilson, and E. A. Bayer. Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio*, 1(5):e00285–10, 2010.

[133] S.-L. Tsai, N. A. DaSilva, and W. Chen. Functional display of complex cellulosomes on the yeast surface via adaptive assembly. *ACS Synth. Biol.*, 2(1):14–21, 2012.

[134] S.-L. Tsai, G. Goyal, and W. Chen. Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. *Appl. Environ. Microbiol.*, 76(22):7514–7520, 2010.

[135] F. Liu, S. Banta, and W. Chen. Functional assembly of a multi-enzyme methanol oxidation cascade on a surface-displayed trifunctional scaffold for enhanced nadh production. *ChemComm*, 49(36):3766–3768, 2013.

[136] T. S. Moon, J. E. Dueber, E. Shiue, and K. L. J. Prather. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered e. coli. *Metab. Eng.*, 12(3):298–305, 2010.

[137] L. Bülow, P. Ljungcrantz, and K. Mosbach. Preparation of a soluble bifunctional enzyme by gene fusion. *Nat. Biotechnol.*, 3(9):821, 1985.

[138] P. Ljungcrantz, H. Carlsson, M. O. Mansson, P. Buckel, K. Mosbach, and L. Buelow. Construction of an artificial bifunctional enzyme,. beta.-galactosidase/galactose dehydrogenase, exhibiting efficient galactose channeling. *Biochemistry*, 28(22):8786–8792, 1989.

[139] R. J. Conrado, J. D. Varner, and M. P. DeLisa. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. *Curr. Opin. Biotech.*, 19(5):492–499, 2008.

[140] L. Albertsen, Y. Chen, L. S. Bach, S. Rattleff, J. Maury, S. Brix, J. Nielsen, and U. H. Mortensen. Diversion of flux toward sesquiterpene production in saccharomyces cerevisiae by fusion of host and heterologous enzymes. *Appl. Environ. Microbiol.*, 77(3):1033–1040, 2011.

[141] S. I. Lim, B. Yang, Y. Jung, J. Cha, J. Cho, E.-S. Choi, Y. H. Kim, and I. Kwon. Controlled orientation of active sites in a nanostructured multienzyme complex. *Sci. Rep.*, 6:39587, 2016.

[142] M. Castellana, M. Z. Wilson, Y. Xu, P. Joshi, I. M. Cristea, J. D. Rabinowitz, Z. Gitai, and N. S. Wingreen. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nat. Biotechnol.*, 32(10):1011, 2014.

[143] J. L. Avalos, G. R. Fink, and G. Stephanopoulos. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. *Nat. Biotechnol.*, 31(4):335, 2013.

[144] W. C. DeLoache, Z. N. Russ, and J. E. Dueber. Towards repurposing the yeast peroxisome for compartmentalizing heterologous metabolic pathways. *Nat. Commun.*, 7:11152, 2016.

[145] Y. J. Zhou, N. A. Buijs, Z. Zhu, D. O. Gómez, A. Boonsombuti, V. Siewers, and J. Nielsen. Harnessing yeast peroxisomes for biosynthesis of fatty-acid-derived biofuels and chemicals with relieved side-pathway competition. *J. Am. Chem. Soc.*, 138(47):15368–15377, 2016.

[146] J. Sheng, J. Stevens, and X. Feng. Pathway compartmentalization in peroxisome of saccharomyces cerevisiae to produce versatile medium chain fatty alcohols. *Sci. Rep.*, 6:26884, 2016.

[147] P. Xu, K. Qiao, W. S. Ahn, and G. Stephanopoulos. Engineering yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc. Natl. Acad. Sci. U.S.A.*, 113(39):10848–10853, 2016.

[148] S. Kumar, F. M. Hahn, E. Baidoo, T. S. Kahlon, D. F. Wood, C. M. McMahan, K. Cornish, J. D. Keasling, H. Daniell, and M. C. Whalen. Remodeling the isoprenoid pathway in tobacco by expressing the cytoplasmic mevalonate pathway in chloroplasts. *Metab. Eng.*, 14(1):19–28, 2012.

[149] H. M. Huttanus and X. Feng. Compartmentalized metabolic engineering for biochemical and biofuel production. *Biotechnol. J.*, 2017.

[150] J. B. Parsons, S. D. Dinesh, E. Deery, H. K. Leech, A. A. Brindley, D. Heldt, S. Frank, C. M. Smales, H. Lünsdorf, A. Rambach, et al. Biochemical and structural insights into bacterial organelle form and biogenesis. *J. Biol. Chem.*, 283(21):14366–14375, 2008.

[151] J. C. Cameron, S. C. Wilson, S. L. Bernstein, and C. A. Kerfeld. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell*, 155(5):1131–1140, 2013.

[152] J. S. Plegaria and C. A. Kerfeld. Engineering nanoreactors using bacterial microcompartment architectures. *Curr. Opin. Biotech.*, 51:1–7, 2018.

[153] C. Aussignargues, B. C. Paasch, R. Gonzalez-Esquer, O. Erbilgin, and C. A. Kerfeld. Bacterial microcompartment assembly: The key role of encapsulation peptides. *Commun. Integr. Biol.*, 8(3):e1039755, 2015.

[154] C. Fan, S. Cheng, S. Sinha, and T. A. Bobik. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. U.S.A.*, 109(37):14995–15000, 2012.

[155] C. Fan, S. Cheng, Y. Liu, C. M. Escobar, C. S. Crowley, R. E. Jefferson, T. O. Yeates, and T. A. Bobik. Short n-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. U.S.A.*, 107(16):7509–7514, 2010.

[156] M. B. Quin, S. A. Perdue, S.-Y. Hsu, and C. Schmidt-Dannert. Encapsulation of multiple cargo proteins within recombinant eut nanocompartments. *Appl. Microbiol. Biotechnol.*, 100(21):9187–9200, 2016.

[157] E. Y. Kim and D. Tullman-Ercek. A rapid flow cytometry assay for the relative quan-

tification of protein encapsulation into bacterial microcompartments. *Biotechnol. J.*, 9(3):348–354, 2014.

[158] H. J. Wagner, C. C. Capitain, K. Richter, M. Nessling, and J. Mampel. Engineering bacterial microcompartments with heterologous enzyme cargos. *Engineering in Life Sciences*, 17(1):36–46, 2017.

[159] M. C. Yung, F. A. Bourguet, T. S. Carpenter, and M. A. Coleman. Re-directing bacterial microcompartment systems to enhance recombinant expression of lysis protein e from bacteriophage ϕx174 in escherichia coli. *Microbial cell factories*, 16(1):71, 2017.

[160] M. F. Slininger Lee, C. M. Jakobson, and D. Tullman-Ercek. Evidence for improved encapsulated pathway behavior in a bacterial microcompartment through shell protein engineering. *ACS Synth. Biol.*, 6(10):1880–1891, 2017.

[161] S. Sinha, S. Cheng, Y. W. Sung, D. E. McNamara, M. R. Sawaya, T. O. Yeates, and T. A. Bobik. Alanine scanning mutagenesis identifies an asparagine–arginine–lysine triad essential to assembly of the shell of the pdu microcompartment. *J. Mol. Biol.*, 426(12):2328–2345, 2014.

[162] D. P. Patterson, B. Schwarz, R. S. Waters, T. Gedeon, and T. Douglas. Encapsulation of an enzyme cascade within the bacteriophage p22 virus-like particle. *ACS Chem. Biol.*, 9(2):359–365, 2013.

[163] P. C. Jordan, D. P. Patterson, K. N. Saboda, E. J. Edwards, H. M. Miettinen, G. Basu, M. C. Thielges, and T. Douglas. Self-assembling biomolecular catalysts for hydrogen production. *Nat. Chem.*, 8(2):179, 2016.

[164] F. Hinzpeter, U. Gerland, and F. Tostevin. Optimal compartmentalization strategies for metabolic microcompartments. *Biophys. J.*, 112(4):767–779, 2017.

[165] K. Hult and P. Berglund. Enzyme promiscuity: mechanism and applications. *Trends Biotech.*, 25(5):231–238, 2007.

[166] E. Gantt and S. F. Conti. Ultrastructure of blue-green algae. *J. Bacteriol.*, 97(3):1486–1493, 1969.

[167] J. M. Shively, G. L. Decker, and J. W. Greenawalt. Comparative ultrastructure of the thiobacilli. *J. Bacteriol.*, 101(2):618–627, 1970.

[168] B. D. Rae, B. M. Long, M. R. Badger, and G. D. Price. Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate co2 fixation in cyanobacteria and some proteobacteria. *Microbiol. Mol. Biol. Rev.*, 77(3):357–379, 2013.

[169] M. R. Badger, D. Hanson, and G. D. Price. Evolution and diversity of co2 concentrating mechanisms in cyanobacteria. *Functional Plant Biology*, 29(3):161–173, 2002.

[170] D. R. Ort, S. S. Merchant, J. Alric, A. Barkan, R. E. Blankenship, R. Bock, R. Croce, M. R. Hanson, J. M. Hibberd, S. P. Long, T. A. Moore, J. Moroney, K. K. Niyogi, M. A. J. Parry, P. P. Peralta-Yahya, R. C. Prince, K. E. Redding, M. H. Spalding, K. J.

van Wijk, W. F. J. Vermaas, S. von Caemmerer, A. P. M. Weber, T. O. Yeates, J. S. Yuan, and X. G. Zhu. Redesigning photosynthesis to sustainably meet global food and bioenergy demand. *Proc. Natl Acad. Sci. USA*, 112(28):8529–8536, 2015.

[171] J. Zarzycki, S. D. Axen, J. N. Kinney, and C. A. Kerfeld. Cyanobacterial-based approaches to improving photosynthesis in plants. *J. Exp. Bot.*, 64:787–798, 2013.

[172] G. D. Price, J. J. L. Pengelly, B. Forster, J. Du, S. M. Whitney, S. von Caemmerer, M. R. Badger, S. M. Howitt, and J. R. Evans. The cyanobacterial ccm as a source of genes for improving photosynthetic co2 fixation in crop species. *J. Exp. Bot.*, 64(3):753–768, 2013.

[173] M. T. Lin, A. Occhialini, P. J. Andralojc, M. A. J. Parry, and M. R. Hanson. A faster rubisco with potential to increase photosynthesis in crops. *Nature*, 513(7519):547–550, 2014.

[174] S. Frank, A. D. Lawrence, M. B. Prentice, and M. J. Warren. Bacterial microcompartments moving into a synthetic biological world. *J. Biotech.*, 163(2):273–279, 2013.

[175] M. Kafri, E. Metzl-Raz, G. Jona, and N. Barkai. The cost of protein production. *Cell Reports*, 14(1):22–31, 2016.

[176] L. Fridlyand, A. Kaplan, and L. Reinhold. Quantitative evaluation of the role of a putative co 2-scavenging entity in the cyanobacterial co 2-concentrating mechanism. *Biosystems*, 37(3):229–238, 1996.

[177] S. Heinhorst, E. B Williams, F. Cai, C. D. Murin, J. M Shively, and G. C Cannon. Characterization of the carboxysomal carbonic anhydrase csosca from *halothiobacillus neapolitanus*. *J. Bacteriol.*, 188:8087–8094, 2006.

[178] O. Mueller-Cajar and S. Whitney. Evolving improved *Synechococcus Rubisco* functional expression in *Escherichia coli*. *Biochem. J.*, 414:205–214, 2008.

[179] M. R. Tsai, Y. ans Sawaya, G. C. Cannon, F. Cai, E. B. Williams, S. Heinhorst, C. A. Kerfeld, and T. O. Yeates. Structural analysis of csos1a and the protein shell of the *halothiobacillus neapolitanus* carboxysome. *PLoS Biol.*, 5(6):e144, 05 2007.

[180] J. M. Shively, F. L. Ball, and B. W. Kline. Electron microscopy of the carboxysomes (polyhedral bodies) of thiobacillus neapolitanus. *J. Bacteriol.*, 116(3):1405–1411, 1973.

[181] C. V. Iancu, D. M. Ding, H. J.and Morris, D. P. Dias, A. D. Gonzales, A. Martino, and G. J. Jensen. The structure of isolated synechococcus strain wh8102 carboxysomes as revealed by electron cryotomography. *J. Mol. Biol.*, 372(3):764–773, 2007.

[182] J. M. Shively, F. Ball, D. H. Brown, and R. E. Saunders. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of thiobacillus neapolitanus. *Science*, 182(4112):584–586, 1973.

[183] A. K.-C. So, G. S. Espie, E. B. Williams, J. M. Shively, S. Heinhorst, and G. C. Cannon. A novel evolutionary lineage of carbonic anhydrase (ε class) is a component of the carboxysome shell. *J. Bacteriol.*, 186(3):623–630, 2004.

[184] E. Y. Kim and D. Tullman-Ercek. Engineering nanoscale protein compartments for synthetic organelles. *Curr. Opin. Biotech.*, 24(4):627–632, 2013.

[185] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry*, 50(21):4402–4410, 2011.

[186] F. Cai, Z. Dou, S. L. Bernstein, R. Leverenz, E. B. Williams, S. Heinhorst, J. Shively, G. C. Cannon, and C. A. Kerfeld. Advances in Understanding Carboxysome Assembly in Prochlorococcus and Synechococcus Implicate CsoS2 as a Critical Component. *Life*, 5(2):1141–1171, 2015.

[187] Y. Kaneko, R. Danev, K. Nagayama, and H. Nakamoto. Intact carboxysomes in a cyanobacterial cell visualized by hilbert differential contrast transmission electron microscopy. *J. Bacteriol.*, 188(2):805–808, 2006.

[188] C. V. Iancu, D. M. Morris, Z. Dou, S. Heinhorst, G. C. Cannon, and G. J. Jensen. Organization, Structure, and Assembly of -Carboxysomes Determined by Electron Cryotomography of Intact Cells. *J. Mol. Biol.*, 396:105–117, 2010.

[189] A. Mugler, F. Tostevin, and P. R. ten Wolde. Spatial partitioning improves the reliability of biochemical signaling. *Proc. Natl Acad. Sci. USA*, 110:5927–5932, 2013.

[190] J. Nunnari and P. Walter. Regulation of organelle biogenesis. *Cell*, 84(3):389–394, 1996.

[191] J. J. Smith, T. W. Brown, G. A. Eitzen, and R. A. Rachubinski. Regulation of peroxisome size and number by fatty acid β-oxidation in the yeast Yarrowia lipolytica. *J. Biol. Chem.*, 275(26):20168–20178, 2000.

[192] M. Schrader, N. A. Bonekamp, and M. Islinger. Fission and proliferation of peroxisomes. *Biochim. Biophys. Acta*, 1822(9):1343–1357, 2012.

[193] J. A. K. W. Kiel, I. J. van der Klei, M. A. van den Berg, R. A. L. Bovenberg, and M. Veenhuis. Overproduction of a single protein, pc-pex11p, results in 2-fold enhanced penicillin production by penicillium chrysogenum. *Fungal Genet. Biol.*, 42(2):154–164, 2005.

[194] L. C. Gomes, G. Di Benedetto, and L. Scorrano. During autophagy mitochondria elongate, are spared from degradation and sustain cell viability. *Nat. Cell Biol.*, 13(5):589–598, 2011.

[195] M. Liesa and O. S. Shirihai. Mitochondrial dynamics in the regulation of nutrient utilization and energy expenditure. *Cell Metab.*, 17(4):491–506, 2013.

[196] J. C. Cameron, S. C. Wilson, S. L. Bernstein, and C. A. Kerfeld. Biogenesis of a Bacterial Organelle: The Carboxysome Assembly Pathway. *Cell*, 155:1131–40, 2013.

[197] A. Buchner, F. Tostevin, F. Hinzpeter, and U. Gerland. Optimization of collective enzyme activity via spatial localization. *J. Chem. Phys.*, 139(13):135101, 2013.

[198] E. A. Bayer, H. Chanzy, R. Lamed, and Y. Shoham. Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol*, 8(5):548–557, 1998.

[199] S. B. Van Albada and P. R. Ten Wolde. Enzyme localization can drastically affect signal amplification in signal transduction pathways. *PLoS Comput. Biol.*, 3(10):e195, 2007.

[200] A. Mugler, A. G. Bailey, K. Takahashi, and P. R. ten Wolde. Membrane clustering and the role of rebinding in biochemical signaling. *Biophys. J.*, 102(5):1069–1078, 2012.

[201] S. Schoffelen and J. C. M. van Hest. Multi-enzyme systems: bringing enzymes together in vitro. *Soft Matter*, 8(6):1736–1746, 2012.

[202] A. Buchner, F. Tostevin, and U. Gerland. Clustering and optimal arrangement of enzymes in reaction-diffusion systems. *Phys. Rev. Lett.*, 110(20):208104, 2013.

[203] J. E. Wilson. Ambiquitous enzymes: variation in intracellular distribution as a regulatory mechanism. *Trends Biochem. Sci.*, 3(2):124–125, 1978.

[204] S. John, J. N. Weiss, and B. Ribalet. Subcellular localization of hexokinases i and ii directs the metabolic fate of glucose. *PLoS one*, 6(3):e17674, 2011.

[205] S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, and H. E. Gaub. Single-molecule cut-and-paste surface assembly. *Science*, 319(5863):594–596, 2008.

[206] P. Guo. The emerging field of rna nanotechnology. *Nat. Nanotechnol.*, 5(12):833, 2010.

[207] M. J. Lee, J. Mantell, L. Hodgson, D. Alibhai, J. M. Fletcher, I. R. Brown, S. Frank, W.-F. Xue, P. Verkade, D. N. Woolfson, et al. Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat. Chem. Biol.*, 14(2):142, 2018.

[208] P. A. Srere. Complexes of sequential metabolic enzymes. *Ann. Rev. Biochem.*, 56(1):89–124, 1987.

[209] K. Jørgensen, A. V. Rasmussen, M. Morant, A. H. Nielsen, N. Bjarnholt, M. Zagrobelny, S. Bak, and B. L. Møller. Metabolon formation and metabolic channeling in the biosynthesis of plant natural products. *Curr. Opin. Plant Biol.*, 8(3):280–291, 2005.

[210] H. Lee, W. C. DeLoache, and J. E. Dueber. Spatial organization of enzymes for metabolic engineering. *Metab. Eng.*, 14(3):242–251, 2012.

[211] A. Küchler, M. Yoshimoto, S. Luginbühl, F. Mavelli, and P. Walde. Enzymatic reactions in confined environments. *Nat. Nanotech.*, 11(5):409–420, 2016.

[212] J.-L. Lin, L. Palomec, and I. Wheeldon. Design and analysis of enhanced catalysis in scaffolded multienzyme cascade reactions. *ACS Catal.*, 4(2):505–511, 2014.

[213] Y. Yamada, C.-K. Tsung, W. Huang, Z. Huo, S. E. Habas, T. Soejima, C. E. Aliaga, G. A. Somorjai, and P. Yang. Nanocrystal bilayer for tandem catalysis. *Nat. Chem.*, 3(5):372–376, 2011.

[214] C. Xie, C. Chen, Y. Yu, J. Su, Y. Li, G. A. Somorjai, and P. Yang. Tandem catalysis for CO_2 hydrogenation to C_2-C_4 hydrocarbons. *Nano Lett.*, 17:3798–3802, 2017.

[215] J. Su, C. Xie, C. Chen, Y Yu, G. Kennedy, G. A. Somorjai, and P. Yang. Insights into the mechanism of tandem alkene hydroformylation over a nanostructured catalyst with multiple interfaces. *J. Am. Chem. Soc.*, 138(36):11568–11574, 2016.

[216] J. Müller and C. M. Niemeyer. DNA-directed assembly of artificial multienzyme complexes. *Biochem. Biophys. Res. Comm.*, 377(1):62–67, 2008.

[217] R. J. A. Wanders, H. R. Waterham, and S. Ferdinandusse. Metabolic interplay between peroxisomes and other subcellular organelles including mitochondria and the endoplasmic reticulum. *Front. Cell Dev. Biol.*, 3:83, 2015.

[218] M. Linka and A. P. M. Weber. Shuffling ammonia between mitochondria and plastids during photorespiration. *Trends Plant Sci.*, 10(10):461–465, 2005.

[219] J. B. French, S. A. Jones, H. Deng, A. M. Pedley, D. Kim, C. Y. Chan, H. Hu, R. J. Pugh, H. Zhao, Y. Zhang, et al. Spatial colocalization and functional link of purinosomes with mitochondria. *Science*, 351(6274):733–737, 2016.

[220] E. H. Ma and R. G. Jones. TORCing up purine biosynthesis. *Science*, 351(6274):670–671, 2016.

[221] A. Boetius, K. Ravenschlag, C. J. Schubert, D. Rickert, F. Widdel, A. Gieseke, R. Amann, B. B. Jørgensen, U. Witte, and O. Pfannkuche. A marine microbial consortium apparently mediating anaerobic oxidation of methane. *Nature*, 407(6804):623–626, 2000.

[222] J. Müller and J. Overmann. Close interspecies interactions between prokaryotes from sulfureous environments. *Front. Microbiol.*, 2:146, 2011.

[223] B. K. Mobarry, M. Wagner, V. Urbain, B. E. Rittmann, and D. A. Stahl. Phylogenetic probes for analyzing abundance and spatial organization of nitrifying bacteria. *App. Env. Microbiol.*, 62(6):2156–2162, 1996.

[224] E. Flores and A. Herrero. Compartmentalized function through cell differentiation in filamentous cyanobacteria. *Nat. Rev. Microbiol.*, 8(1):39–50, 2010.

[225] E. Costa, J. Pérez, and J.-U. Kreft. Why is metabolic labour divided in nitrification? *Trends Microbiol.*, 14(5):213–219, 2006.

[226] A. Schramm, L. H. Larsen, N. P. Revsbech, N. B. Ramsing, R. Amann, and K.-H. Schleifer. Structure and function of a nitrifying biofilm as determined by in situ hybridization and the use of microelectrodes. *App. Env. Microbiol.*, 62(12):4641–4647, 1996.

[227] A. P. Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *J. Biol. Chem.*, 276(14):10577–10580, 2001.

[228] H. Berry. Monte Carlo simulations of enzyme reactions in two dimensions: fractal kinetics and spatial segregation. *Biophys. J.*, 83(4):1891–1901, 2002.

[229] S. Schnell and T. E. Turner. Reaction kinetics in intracellular environments with macromolecular crowding: simulations and rate laws. *Prog. in Biophys. Mol. Biol.*, 85(2):235–260, 2004.

[230] Z. Hu, J. Jiang, and R. Rajagopalan. Effects of macromolecular crowding on biochemical

reaction equilibria: a molecular thermodynamic perspective. *Biophys. J.*, 93(5):1464–1473, 2007.

[231] J. A. Dix and A. S. Verkman. Crowding effects on diffusion in solutions and cells. *Annu. Rev. Biophys.*, 37:247–263, 2008.

[232] N. Agmon and A. Szabo. Theory of reversible diffusion-influenced reactions. *J. Chem. Phys.*, 92(9):5270, 1990.

[233] A. Vijaykumar, P. G. Bolhuis, and P. R. Ten Wolde. The intrinsic rate constants in diffusion-influenced reactions. *Farad. Discuss.*, 195:421–441, 2017.

[234] R. Samson and J. M. Deutch. Exact solution for the diffusion controlled rate into a pair of reacting sinks. *J. Chem. Phys.*, 67(2):847, 1999.

[235] H.-K. Tsao. Competitive diffusion into two reactive spheres of different reactivity and size. *Phys. Rev. E*, 66(1):011108, 2002.

[236] C. Eun, P. M. Kekenes-Huskey, and J. A. McCammon. Influence of neighboring reactive particles on diffusion-limited reactions. *J. Chem. Phys.*, 139(4):044117, 2013.

[237] L. J. Reed. Multienzyme complexes. *Acc. Chem. Res.*, 7(2):40–46, 1974.

[238] J. L. S. Milne, D. Shi, P. B. Rosenthal, J. S. Sunshine, G. J. Domingo, X. Wu, B. R. Brooks, R. N. Perham, R. Henderson, and S. Subramaniam. Molecular architecture and mechanism of an icosahedral pyruvate dehydrogenase complex: a multifunctional catalytic machine. *EMBO J.*, 21(21):5587–5598, 2002.

[239] J. J. Thomson. On the Structure of the Atom: an Investigation of the Stability and Periods of Oscillation of a number of Corpuscles arranged at equal intervals around the Circumference of a Circle; with Application of the Results to the Theory of Atomic Structure. *Philos. Mag.*, 7:237–265, 1904.

[240] J. I. Prosser and G. W. Nicol. Archaeal and bacterial ammonia-oxidisers in soil: the quest for niche specialisation and differentiation. *Trends Microbiol.*, 20(11):523–531, 2012.

[241] W. Martens-Habbena, P. M. Berube, H. Urakawa, R. José, and D. A. Stahl. Ammonia oxidation kinetics determine niche separation of nitrifying archaea and bacteria. *Nature*, 461(7266):976–979, 2009.

[242] J. N. Wilking, V. Zaburdaev, M. De Volder, R. Losick, M. P. Brenner, and D. A. Weitz. Liquid transport facilitated by channels in Bacillus subtilis biofilms. *Proc. Natl. Acad. Sci. U.S.A.*, 110(3):848–852, 2013.

[243] Y. Zhang, S.-Z. Li, J. Li, X. Pan, R. E. Cahoon, J. G. Jaworski, X. Wang, J. M. Jez, F. Chen, and O. Yu. Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and mammalian cells. *J. Am. Chem. Soc.*, 128(40):13030–13031, 2006.

[244] R. E. Schwartz. The five-electron case of Thomson's problem. *Exp. Math.*, 22(2):157–186, 2013.

[245] V. A. Yudin. The minimum of potential energy of a system of point charges. *Discrete Mathematics and Applications*, 3(1):75–82, 1993.

[246] N. N. Andreev. An extremal property of the icosahedron. *East J. Approx*, 2(4):459–462, 1996.

[247] S. Smale. Mathematical problems for the next century. *Math. Intell.*, 20(2):7–15, 1998.

[248] A. Maritan, C. Micheletti, A. Trovato, and J. R. Banavar. Optimal shapes of compact strings. *Nature*, 406:287–290, 2000.

[249] S. Torquato and F. H. Stillinger. Jammed hard-particle packings: From Kepler to Bernal and beyond. *Rev. Mod. Phys.*, 82(3):2633, 2010.

[250] D. L. D. Caspar and A. Klug. Physical principles in the construction of regular viruses. In *Cold Spring Harbor symposia on quantitative biology*, volume 27, pages 1–24. Cold Spring Harbor Laboratory Press, 1962.

[251] R. Zandi, D. Reguera, R. F. Bruinsma, W. M. Gelbart, and J. Rudnick. Origin of icosahedral symmetry in viruses. *Proc. Natl Acad. Sci. USA*, 101(44):15556–15560, 2004.

[252] T. Tarnai, Z. Gaspar, and L. Szalai. Pentagon packing models for "all-pentamer" virus structures. *Biophys. J.*, 69(2):612–618, 1995.

[253] C. Nisoli, N. M. Gabor, P. E. Lammert, J. D. Maynard, and V. H. Crespi. Annealing a magnetic cactus into phyllotaxis. *Phys. Rev. E*, 81(4):046107, 2010.

[254] D. Mackenzie. Proving the perfection of the honeycomb. *Science*, 285(5432):1338–1339, 1999.

[255] E. G. Krebs and J. A. Beavo. Phosphorylation-dephosphorylation of enzymes. *Ann. Rev. Biochem.*, 48(1):923–959, 1979.

[256] Q. Wang, Y. Zhang, C. Yang, H. Xiong, Y. Lin, J. Yao, H. Li, L. Xie, W. Zhao, Y. Yao, et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science*, 327(5968):1004–1007, 2010.

[257] L. Gerosa and U. Sauer. Regulation and control of metabolic fluxes in microbes. *Curr. Opin. Biotech.*, 22(4):566–575, 2011.

[258] F. H. Gaertner. Unique catalytic properties of enzyme clusters. *Trends Biochem. Sci.*, 3(1):63–65, 1978.

[259] M. Z. Wilson and Z. Gitai. Beyond the cytoskeleton: mesoscale assemblies and their function in spatial organization. *Curr. Opin. Microbiol.*, 16(2):177–183, 2013.

[260] Y. Zhang, K. F. M. Beard, C. Swart, S. Bergmann, I. Krahnert, Z. Nikoloski, A. Graf, R. G. Ratcliffe, L. J. Sweetlove, A. R. Fernie, et al. Protein-protein interactions and metabolite channelling in the plant tricarboxylic acid cycle. *Nat. Commun.*, 8:15212, 2017.

[261] T. Laursen, J. Borch, C. Knudsen, K. Bavishi, F. Torta, H. J. Martens, D. Silvestro, N. S. Hatzakis, M. R. Wenk, T. R. Dafforn, et al. Characterization of a dynamic metabolon producing the defense compound dhurrin in sorghum. *Science*, 354(6314):890–893, 2016.

[262] S. F. Banani, H. O. Lee, A. A. Hyman, and M. K. Rosen. Biomolecular condensates: organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.*, 18(5):285, 2017.

[263] A. A. Hyman, C. A. Weber, and F. Jülicher. Liquid-liquid phase separation in biology. *Annu. Rev. Cell. Dev. Biol.*, 30:39–58, 2014.

[264] R. Milo and R. Phillips. *Cell biology by the numbers.* Garland Science, 2015.

[265] A. Bondi. van der waals volumes and radii. *J. Phys. Chem.*, 68(3):441–451, 1964.

Acknowledgements

An dieser Stelle möchte ich mich bei einigen Leuten bedanken, die mich während der gesamten Zeit meiner Doktorarbeit unterstützt haben und zum Gelingen dieser Arbeit beigetragen haben. Zuallererst möchte ich mich bei meinem Doktorvater Uli Gerland für die Betreuung meiner Doktorarbeit bedanken. Vor allem bedanke ich mich für die vielen Freiheiten im wissenschaftlichen Arbeiten und die freundliche Arbeitsatmosphäre.

Als nächstes möchte ich mich bei Filipe Tostevin für seine stete Hilfsbereitschaft und gute wissenschaftliche Zusammenarbeit in einigen Projekten bedanken. Von deiner wissenschaftlichen Expertise konnte ich sehr viel lernen. Außerdem bedanke ich mich bei Nanni Giunta für die vielen wissenschaftlichen Diskussionen, die nette Zusammenarbeit im Enzym-Projekt und das Korrekturlesen von Teilen meiner Arbeit. Deine herzliche und offene Art hat den Arbeitsalltag immer sehr bereichert. Bedanken möchte ich mich auch bei Elena Biselli für die vielen lustigen Diskussionen über Essen und Kultur, Linda Martini für deine ruhige und angenehme Art als Bürokollegin, Severin Schink für die vielen entspannten Diskussionen und deine Expertise im Fleischgrillen, Bernhard Altaner und Johannes Nübler für das Korrekturlesen meiner Paper, Laura und Karin für die stete Hilfsbereitschaft und der gesamten Gerland Gruppe für die gute Zusammenarbeit und die vielen schönen Jahre.

Bedanken möchte ich mich auch bei den vielen Freunden die ich im Laufe meiner Studienzeit und Doktorarbeit gefunden habe. Philip, ich freue mich schon, wenn wir an der Copacabana oder im Mr. Mumble's auf die fertige Promotion anstoßen. Danke für die vielen lustigen Momente. Uli L., ohne unsere gemeinsamen Mittagessen in Garching wäre der Uni Alltag nur halb so schön gewesen. Ich freue mich schon auf unsere nächsten Museumsbesuche. Franzi und Linda (die girls), danke für die vielen gemeinsamen Mittagessen und die lustige Zeit in Venedig. Alexandra, danke für die schöne Reise nach San Francisco.

Vor allem aber möchte ich mich bei meiner Familie bedanken, die mich immer unterstützt und mir in allen Lebenslagen zur Seite steht. Eure bedingungslose Unterstützung hat mir den Weg zu dieser Promotion geebnet. Zu guter Letzt möchte ich mich ganz besonders bei meiner wundervollen Freundin Anna-Sophia bedanken. Danke, dass du immer für mich da bist.

Ich versichere, die Arbeit selbstständig angefertigt und dazu nur die im Literaturverzeichnis angegebenen Quellen benutzt zu haben.

München, den 15. Mai 2018