Spatial Aspects of Enzymology

Optimal Arrangement and Conformational Dynamics

Alexander Buchner

München 2013

Spatial Aspects of Enzymology
Optimal Arrangement and Conformational Dynamics

Alexander Buchner

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

vorgelegt von
Alexander Buchner
aus Simbach am Inn

München, den 25. Oktober 2013

Erstgutachter: Prof. Dr. Ulrich Gerland
Zweitgutachter: Prof. Dr. Friedrich C. Simmel
Tag der mündlichen Prüfung: 18. Dezember 2013

Contents

List of Figures

Zusammenfassung

Die Zelle ist der Grundbaustein des Lebens, vom Einzeller bis hin zu hochkomplexen Organismen. Lebenswichtige Aufgaben wie DNA Transkription, Signaltransduktion und Stoffwechselregulierung werden in allen Organismen von Enzymen gleichermaßen erledigt. Die natürliche Selektion von Enzymen, bezüglich der Struktur wie auch der Substratspezifikation, führte zu einer Erhöhung der Reaktionsgeschwindigkeit. Neben der individuellen katalytischen Aktivität haben mehrstufige enzymatische Reaktionen eine synergistische Funktion entwickelt. Die verschiedenen Stoffwechselwege sind hintereinander geschaltete Enzymsequenzen, welche die Lebensadern der Zelle bilden. Das Zusammenspiel der Stoffwechselwege Glykolyse und Citratzyklus ist für die Erzeugung von ATP aus der Nahrung verantwortlich. Obwohl die Abfolge der Enzymsequenz in den meisten Stoffwechselwegen bekannt ist, wissen wir noch wenig über die exakte räumliche Anordnung. Die Entwicklungen der letzten Jahre im Bereich der Einzelmolekülfluoreszenzspektroskopie eröffneten neue Wege auf dem Gebiet der Enzymmessung. Diese neuartigen Methoden ermöglichten es einzelne Enzyme mit Nanometerpräzesion anzuordnen.
Wir sind überzeugt, dass die genaue Kenntnis über die räumliche Anordnung von Enzymen einen Leitfaden liefern wird, um das Zusammenspiel von biosynthetischen Prozessen zu verstehen. Um ein vollständigeres Bild über die Funktion der Enzyme zu erhalten, erachten wir es als sinnvoll sowohl die Aktivität der einzelnen Enzyme zu untersuchen, sowie auch die Gesamtaktivität einer Enzymkaskade in einer physiologischen Umgebung. Der erste Teil der Arbeit befasst sich mit der Aktivität der Enzyme, abhängig von deren strukturellem Aufbau. Dazu werden wir knapp die kürzlich entwickelten Methoden auf dem Gebiet der Einzelmolekülfluoreszenzspektroskopie erörtern. Dabei werden wir ganz besonders auf die Anwendungen im Bereich der Kraftspektroskopie eingehen und das Verhalten der Enzyme unter dem Einfluss einer äußeren Kraft studieren. Der dabei auftretende Faltungsprozess des Enzyms wird mit einem Random Walker Models beschrieben. Die Übergangsraten und die freie Energielandschaft eines bestimmten Enzyms werden durch Anpassung der Theorie an die experimentellen Daten ermittelt. Im zweiten Teil der Arbeit untersuchen wir den Einfluss der räumlichen Anordnung von Enzymen auf ihre Gesamteffizienz. Hierfür betrachten wir eine eindimensionale Anordnung, bestehend aus zwei unterschiedlichen Enzymarten, welche mit Hilfe eines linearen Reaktions-Diffusions-Modell beschrieben wird. Eine optimale Anordnung von Enzymen besteht nicht nur aus einer lokalen Anhäufung der Enzyme, sondern auch aus sog. "Back up" Enzymen. Welche Anordnung letztendlich eintritt, hängt von einem dimensionslosen Kontrollparameter ab. Der Grund für dieses

überraschende Verhalten liegt in der stochastischen Natur der molekularen Reaktion und Diffusion. Um dies zu quantifizieren, wird ein neues Konzept eingeführt, das sog. "Enzyme Exposure". Ebenfalls zeigen wir, dass dieses Verhalten kein Spezialfall ist, sondern vielmehr ein universell auftretender Effekt. Dazu erweitern wir unser Modell auf höhere Dimensionen und andere geometrische Anordnungen, weiterhin führen wir konkurrierende Nebenreaktionen sowie die Erweiterung auf nichtlineare Reaktionen ein. In jedem einzelnen Fall wird, abhänging von den Systemparametern, eine Delokalisierung des Enzymclusters gefunden. Wir glauben, dass ein solches Verhalten mit Hilfe künstlicher Enzymkaskaden beobachtet werden kann.
Im ersten Kapitel geben wir eine kurze Einführung in die Grundprinzipien der Enzymreaktion. Insbesondere die Reaktionskinetik und der entsprechenden strukturellen Veränderung der Reaktanten während der Substrat Enzym Komplex Bildung. Das zweite Kapitel fasst einige Messmethoden der Einzelmoleülfluoreszenzspektroskopie und stellt unser Random Walker Modell vor. Nach einer Auflistung über die Vorteile eines Enzymclusters in Kapitel 3, präsentieren wir unser mathematisches Modell zur Beschreibung der Enzymkaskaden im vierten Kapitel. Das letzte Kapitel gibt einen kurzen Ausblick auf zukünftige Forschungsvorhaben in diesem Bereich.

Abstract

The cell is the core building block of life, from an unicellular organism through to highly complex organisms such as mammals. All these organisms have in common that enzymes perform the various tasks sustaining life, DNA transcription, signal transduction and regulation of our metabolism. Natural selection of enzymes provoked an acceleration of the reaction velocity caused by a specification in the enzyme's geometrical organization and conformation. Besides their individual catalytic activity, multistep enzymatic reactions have developed a function synergistically. Different enzyme sequences form the diverse metabolic pathways representing the lifelines of cells. Metabolic pathways such as glycolysis and citric acid cycle are, amongst others, responsible for the conversion of nutrition to chemical energy in form of ATP. Although the enzyme sequence in many metabolic pathways is well known a deeper insight into the spatial arrangement remains widely elusive. Recent developments in single molecule spectroscopy gave rise to new insights on a single molecule level. Notably, the cut and paste method permits an accurate arrangement of enzymes with nanometer precision.
We are convinced that exact knowledge about the spatial arrangement of enzymes will provide a guide to understand the orchestration of biosynthetic processes. This thesis focuses on the activity of enzymes, both on single molecules and when incorporated into a greater pathway. These two different angles are taken to draw a more complete picture of the function of enzymes when embedded in a physiological context. The first part deals with the influence of the enzyme structure on its activity. We will acquaint the reader with the recently developed single molecule microscopy methods. Their application to the response of the enzyme towards external force will be emphasized. We introduce a random walker model describing the folding process of an enzyme. The kinetic rates and the free energy landscape of a particular enzyme is obtained by fitting the theory to the experimental data. In the second part we investigate the dependence of the efficiency on the spatial arrangement of enzymes within a particular pathway. Deliberately, we focus on a pathway containing two different types of enzymes described by a one dimensional linear reaction diffusion model. The optimal arrangement of enzymes shows, besides the formation of an enzyme cluster, the emergence of "back up" enzymes, as a single control parameter is adjusted. The reason for this surprising behavior lies in the stochastic nature of the molecular reaction and diffusion. To this end we introduce a new concept called "enzyme exposure". Consequently, we demonstrate the universality of those back up enzymes. This transition is robust against varying geometry and dimension as well as

introducing a competing pathway and allowing for intermediate densities of the order of the Michaelis-Menten constant K_M. We believe that our results can be tested by designing such an artificial enzyme pathway.
In the first chapter we give a brief introduction to the basic principles of enzyme reaction. In particular, the reaction kinetics and the corresponding structural changes during the process of substrate-enzyme complex formation. The second chapter summarizes some single molecule measurement methods and introduces our random walker model. After stating the advantages of clustering enzymes in chapter 3, we present our mathematical model describing the enzymatic pathway in chapter four. The final chapter gives a brief outlook on future research projects in that field.

.

Chapter 1

Introduction

Enzymes appear in nature in various different structures, each of them show a high specificity with respect to their biological function. But all of them have in common to lower the activation energy for a reaction. This makes enzymes a necessary part of reaction in living cells, such as in plants, animals and human beings. Enzymes are involved in various biochemical reactions for example in the muscle contraction myosin (ATPase) [1, 2], integrase in HIV (integrating its DNA into the host cell) [3], digestive enzymes for instance lipases [4] and protease[5], DNA transcription by RNA polymerase ribozymes [6] and Signaling pathway [7] just to name a few. There are 20 different amino acids forming the various proteins [8] and a single protein-based enzyme may consist of 62-2.500 amino acids [9, 10]. But protein based enzymes are not the only catalysts for biochemical reactions. In the late 80's, early 90's, it was found that RNA can form complicated structures catalyzing the generation of proteins in the cell, such as Ribozymes [11]. Another example is RNA splicing, which connects the exons by cutting out the introns [12]. Generally, an enzyme forms a complex with its matching substrate, which lowers the free activation energy by usually using non covalent interactions between the enzyme and the substrate [13]. Protein-based enzymes show an extremely high specification and therefore are only catalyzing a selective range of substrates. The grade of specification limits the number of different substrates an enzyme is able to catalyze, but in return increases the reaction velocity [14, 15]. The enzymatic activity is not distributed over the entire enzyme surface. Only locally confined parts of the enzymes are active, forming the so called "active center".

1.1 Kinetics of Enzymatic Reactions

In the following we will briefly explain the main ideas of enzyme kinetics. To this end we will introduce the basic notions needed in the main part of the thesis. These include the activation energy and the enzyme-substrate complex. Moreover, we will give a short introduction to the Michaelis-Menten kinetics and its features, which is the most prevalent reaction scheme used in this thesis.

1.1.1 General Principle

Protein based enzymes are one of nature's most effective catalysts. Like most catalyst enzymes, they are lowering the activation energy E_a in a reaction. A reaction happens spontaneously when the overall free energy difference ΔG, between the reactant and the product, is negative, Fig.(1.1). Note that in enzymology the reactants are usually called substrates. The height of the activation energy forms an energy barrier stabilizing the molecule. The higher this barrier the more stable the substrate and the longer it takes until the reaction takes place. Enzymes typically lower that energy barrier by binding the substrate and forming a so called enzyme-substrate complex, which has not necessarily the same stereochemistry or charge configuration as the substrate. A lower energy barrier leads to an increase in the reaction rate and thus accelerates the reaction [16, 13]. The stabilization of the enzyme-substrate complex is one of the keys to understand the concept of enzyme catalysis. Certain degrees of freedom for the substrate are frozen in the complex such as translation or rotation increasing the reaction velocity. For example in the esterification of hydro-carboxylic acids a lower relative freedom of movement of the carboxyl-group and hydroxyl-group accelerates the reaction [17]. From a mechanistic point of view the change in the energy barrier of the substrate is caused by a stabilization of the transition state, the enzyme-substrate complex, through chemical bonds such as electrostatic interactions and hydrogen bonds. Another feature of the transition state is the facilitation of proton movement during the reaction [18, 19, 20]. However, enzymes do not shift the chemical equilibrium in favor of the substrates or the product since both reaction rates, the forward and the backward rate, are changed by the same factor. The enzymes merely catalyze the reaction, so that the reaction runs more quickly than it would without the enzyme.

1.1.2 The Michaelis-Menten Kinetics

In 1913 Leonor Michaelis and Maud Menten (MM) suggested a model describing the kinetic features of an irreversible enzymatic reaction forming an intermediate complex

$$\mathrm{E} + \mathrm{S} \underset{\mathrm{k_r}}{\overset{\mathrm{k_f}}{\rightleftharpoons}} ES \xrightarrow{\mathrm{k_{cat}}} \mathrm{E} + \mathrm{P}, \tag{1.1}$$

where k_f is the association rate, k_r the dissociation rate and k_{cat} the catalytic rate. Upon contact the substrate S and the enzyme E form a complex ES, with rate k_f, which either generates a product P or it dissociates back k_r into its chemical precursors. In either case the enzyme regains its original state. The reaction velocity (how many products P are generated per unit time) is given by

$$v = k_{\mathrm{cat}}[E_T]\frac{[S]}{K_M + [S]}. \tag{1.2}$$

For that matter it is assumed that the enzyme-substrate complex reaches the steady sate faster than the substrate or product concentration. The so called Michaelis-Menten constant K_M primarily stands for the substrate concentration at which the reaction reaches

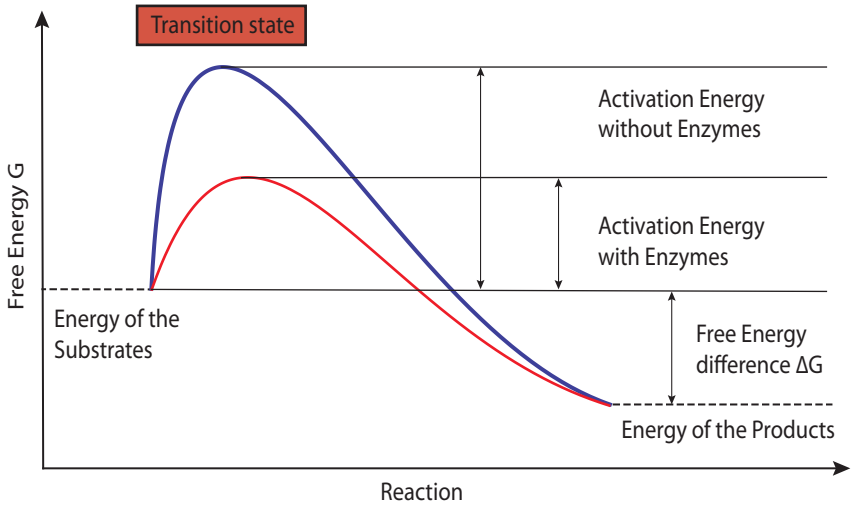

Figure 1.1: A schematic representation of the activation energy. Considering a chemical reaction with substrate S and Product P via an intermediate state called transition state. The reduction of the activation energy E_a (red curve) due to the enzymes compared to the activation energy (blue curve) without the enzymes as catalyst.

half of its maximal velocity. As it can be seen in Eq.(1.2) the maximal velocity is obtained by large substrate concentrations $[S] \gg K_M$. For that scenario the maximal velocity is determined by the catalytic rate k_{cat} and the total enzymes concentration $[E_T]$. For a full derivation of the reaction velocity from the kinetic rate equations see [21, 16]. Although the MM-reaction scheme is the commonly used one, other mechanism are needed, such as in the case of cooperative binding. The binding of one substrate changes the active site of the enzymes for a second substrate. A well known example for this reaction kinetics is hemoglobin. A protein responsible for the binding, respectively, transportation of oxygens in the red blood cells. The kinetics of the multiple bindings sites of hemoglobin are expressed by the Adair equation [21]. Further mechanisms are the so called ping-pong mechanism also known as the substituted enzyme mechanism or the reversible MM-kinetics [21]. For the sake of this thesis we will consider the widely used MM-reaction scheme.

1.1.3 The Catalytic Acceleration of Enzymes

The catalytic acceleration of enzymes is usually observed by comparing of the rate constant of the catalyzed reaction k_{cat} with the reaction rate in aqueous solution k_{aq} [22]. One of the most proficient enzymes is orotidine 5'-monophosphate decarboxylase that catalyzes orotidine 5'-monophosphate (OMP) to uridine 5'-monophosphate (UMP) which shows an acceleration $k_{\text{cat}}/k_{\text{aq}}$ of 17 orders of magnitude [18, 23]. Although enzymes do not change the chemical equilibrium, their acceleration of the reaction velocity make biochemical reactions possible on a time scale relevant for living organism. This demonstrates the enormous

importance of enzymes in catalyzing biochemical reaction.Another example is the carbonic anhydrase, which are enzymes catalyzing carbon dioxide and water to bicarbonate and protons, maintaining the acid-base balance in the blood of mammals by transporting the carbon dioxide out of the tissue. The increase in reaction rate can be 10^7 times faster than in the absence of those enzymes, catalyzing the astonishing number of up to 10^6 reactions per second [24]. In the case of superefficient enzymes k_{cat}/K_M can reach values of $10^{10} M^{-1} s^{-1}$ [25].

1.2 A Structural View on Enzymatic Reaction

The MM-equation is a very useful tool to describe the kinetics of an enzyme reaction. We have seen how the reaction rates enter into the kinetic description of the enzyme reaction, yet the exact value is of the rates must be determined experimentally. It is necessary to understand the structure and the chemical compounds of an enzymes in order explain its affinity and specificity to the substrates as well as the chemical origin of the kinetic rates. In the following we will briefly discuss a more mechanistic view of the enzyme reaction and the composition of an enzyme.

1.2.1 Active Site

The active site of the enzyme is the part that catalyzes the reaction. It is only a small part of the complex 3D amino acid structure of the enzyme. The structural composition of the active site of the enzyme matches its substrate counterpart. The transition state, however, demands at least a temporal rearrangement of its compounds. After the release of the product the active center regains its original state. Although the individual chemical composition of the active site varies form enzyme to enzyme, the general functionality is similar. Of the previously mentioned 20 different amino acids only 11 polar and charged residues contribute to the active centers [26]. This raises the question what is the function of the remaining part of the enzyme? One main aspect is the shielding of the substrate from the surrounding solvent, which creates an adequate environment for the biochemical reaction. For a reaction to happen the substrate and the enzymes must find each other, which is commonly a random diffusion process. The actual reaction is based on the collision between those two chemicals. It is speculated that the active center evolved in such a way that it attracts the substrate. Such an attraction can be imposed by localizing the polar groups of the amino acid ends towards the active center creating an overall electrostatic field [8, 16, 23]. The slightest change in geometry or the arrangement of the amino acid ends of the active center may prevent the reaction from occurring. Biochemically this is achieved by inhibitors, deactivating the enzyme by destabilizing the conformation and hence allosterically altering the active site [27]. Some of the enzymes even need so called cofactors, which are molecules binding to the enzymes regulating the active site. An example, we will discuss later in more detail, is the pyruvate dehydrogenase complex [28] which has up to five cofactors, and is one of the largest and best known multi-enzyme

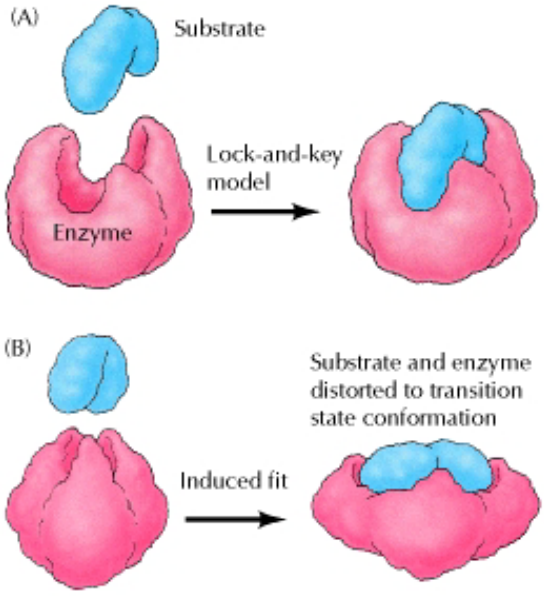

Figure 1.2: The basic reaction mechanism are depicted. In (A) the Lock-an-key mechanism, in which the substrate and the enzymes are rigid highly specified chemicals. In (B) the induced fit model is shown. During the reaction between the substrate and the enzyme their configuration gets altered. This picture is taken form [30].

complexes.

1.2.2 Lock Key

By means of evolution enzymes are highly specified. This fact requires a nearly perfectly matching substrate. One of the first enzyme models directly based on that statement is the so called "lock-key" - mechanism suggested by Fischer [29]. The enzyme pocket should be the exactly matching counterpart of the substrate. Despite its simplistic view, for instance neglecting environmental influences on the conformation of enzyme, the basic idea survives up to this day. However, it became apparent, that in several cases the substrate binding changes not only the configuration of the enzyme but also of the substrate itself. This model is called induced fit model, see Fig.(1.2), where the enzymes are not assumed as rigid object. Their active site constantly reconverts the active center while interacting with the substrate [30]. Many enzymes have flexible loops, floppy tails or mobile lids which during the reaction with the substrate try to hinder the substrate to escape into the solvent again. This sequestration prevents the structurally changed substrate to be

decomposed by the solvent while it is converted into a product [31]. More recent single molecule investigations revealed the fact that the catalytic rate of the enzymes is subjected to a fluctuation in time, called dynamic disorder [32]. Induced by the thermal fluctuation of the 3D enzyme configuration the overall reaction rate is actually a time averaged rate. This again changed the picture of an enzyme reaction, away form a purely static conformation of the enzymes towards a dynamic conformational change before and during the reaction.

1.3 RNA Based Enzymes

Not all enzymes consist of proteins, some of them are made entirely of RNA. For example, the production of proteins in a living cell is driven by an RNA based enzymatic reaction. The information encoded in the DNA-sequence gets transcribed into the messenger RNA (mRNA). In the translation process, where ribosome catalyzes proteins from the mRNA, the Ribozyme is the main participant, a RNA-molecule with an enzymatic action. The so called "RNA world hypothesis" [33] suggests that the origin of life and the self-replication mechanism is entirely based on RNA-molecules. Self-replication contains two key components, information storage and enzymatic activities, which is nowadays conducted by the DNA and the proteins [34]. It is believed that in the early stage of life both tasks were done by RNA-molecules. The nucleotides of the RNA, similar to the modern DNA, encode the information for gene expression and the three dimensional structure of the RNA ensures its enzymatic activity. As a first step to prove such an hypothesis it was necessary to show that RNA-molecules can reproduce simple RNA-strands. Arthur Zaug and Thomas Cech [12] showed that an intron in the RNA of *Tetrahymena thermophila* could catalyze a reaction with a reasonable rate enhancement ($k_{\text{cat}}/K_M \approx 10^3 M^{-1}s^{-1}$) . This started the investigation on RNA as a biomolecule with enzymatic properties. Later it was shown that RNA can even perform its own synthesis providing a central part in a laboratory based synthesis of a basic life form [35]. Although the catalytic velocity of RNA-molecules does not reach the one of protein based enzymes, in a world without those RNA-enzymes they are still fast enough. It was suggested that in order to improve the enzymatic activity evolution produced protein based enzymes. Their structure is essentially different from the RNA based catalysts. Proteins have many different functional groups forming a complex stable 3D structure, whereas ribozymes only consist of four different compounds and it is not clear whether their 3D structure can be controlled by nature as accurately as for proteins [11] . Despite the slower reaction rate, RNA-based enzymes still play a crucial role in the self-replication cycle forming the foundation of living organism.

1.4 Pyruvate Dehydrogenase Complex

As we have seen in the previous sections, multi-enzyme complexes can be found on many occasions in the cell. The pyruvate dehydrogenase complex (PDC) is one of the largest known multi-enzyme complex, it can be found in prokaryotes and eukaryotes alike [36, 37].

Apart form the enormous size on a molecular level with an approximate diameter of 500 $\mathring{A}$ [38]. The way it works, the complexity of its composition as well as the intricacy of its reactions made it not only an interesting but also a challenging enzyme-complex to study. The PDC serves as a link between the glycolysis and the citric acid cycle, which is loosely speaking responsible for the conversion of nutrition to ATP. The PDC itself converts pyruvate to acetyl-CoA according to the following reaction

$$CH_3COCOOH + NAD^+ + CoA\text{-}SH \longrightarrow CH_3CO\text{-}SCoA + FADH + H^+ + CO_2, \qquad (1.3)$$

where the pyruvate ($CH_3COCOOH$) with the Coenzyme A (CoA-SH) essentially produces carbon dioxide (CO_2) and the acetyl-CoA (CH_3CO-SCoA).

1.4.1 The Composition of the Pyruvate Dehydrogenase Complex

Independent of a particular organism the PDC consists of three different types of enzymes the pyruvate decarboxylase (E1), the dihydrolipoyl acetyltransferase (E2) and the dihydrolipoyl dehydrogenase (E3). However, in mammalian the complex is slightly more evolved, it has an additional binding site for E3 but more importantly the PDC can be regulatory controlled by a kinase and a phosphatases [39]. In eucaryotes those PDC's are located in the mitochondria at the inner membrane matrix. The inner core of the complex is built of E2 enzymes forming a highly symmetric complex, for example in the bovine kidney pyruvate dehydrogenase the 60 subunits of E2 are placed in sets of three at the 20 vertices of pentagonal dodecahedral core [38], see Fig.(1.3 A)). But for example in *E.coli* the core consists of only 24 E2 forming a cube with an E2 trimer at each of the eight vertices. The morphological unit of three E2 appears to be an essential feature of the core arrangement in both cases. Around the core the remaining E1/E3 can bind mutually exclusive to the inner linkers, consisting of polypeptides, of the E2 enzymes Fig.(1.3 B)). The self organizing E2 trimers play an essential role in the formation of the complex, since it acts as an scaffold for the other enzymes E1 and E3. The trimers themselves are jointly connected by a flexible polypeptide bound. Those linkers, keeping together the E2-core of the pyruvate dehydrogenase complex, are necessary for the so called breathing effect. The linkers stretch and contract again influencing the size of the core. A change in the diameter up to 40 $\mathring{A}$ was observed [40]. The reason for that breathing effect is speculated to augment the movement of the lipoamide swinging arm between the catalytic sites of E1, E2 and E3.

The inner linker connects the catalytic domain of the E2 with the E1/E3 binding domain and the outer linker connects the E1/E3 binding domain with the lipoyl domain. Depending on the particular complex there are one, two or even three such lipoyl domains. The enzymes E1 and E3 bind to the inner linkage, forming the outer region of the pyruvate dehydrogenase complex. In the aforementioned example, the bovine kidney pyruvate dehydrogenase complex, up to 22 E1 molecules and 6 E3 molecules bind to the inner scaffold of E2 enzymes. The lipoyl domain has to migrate between the enzymes E1 and E2 and subsequently between E2 and E3. Further, the lipoyl domains of E2 are attached to the

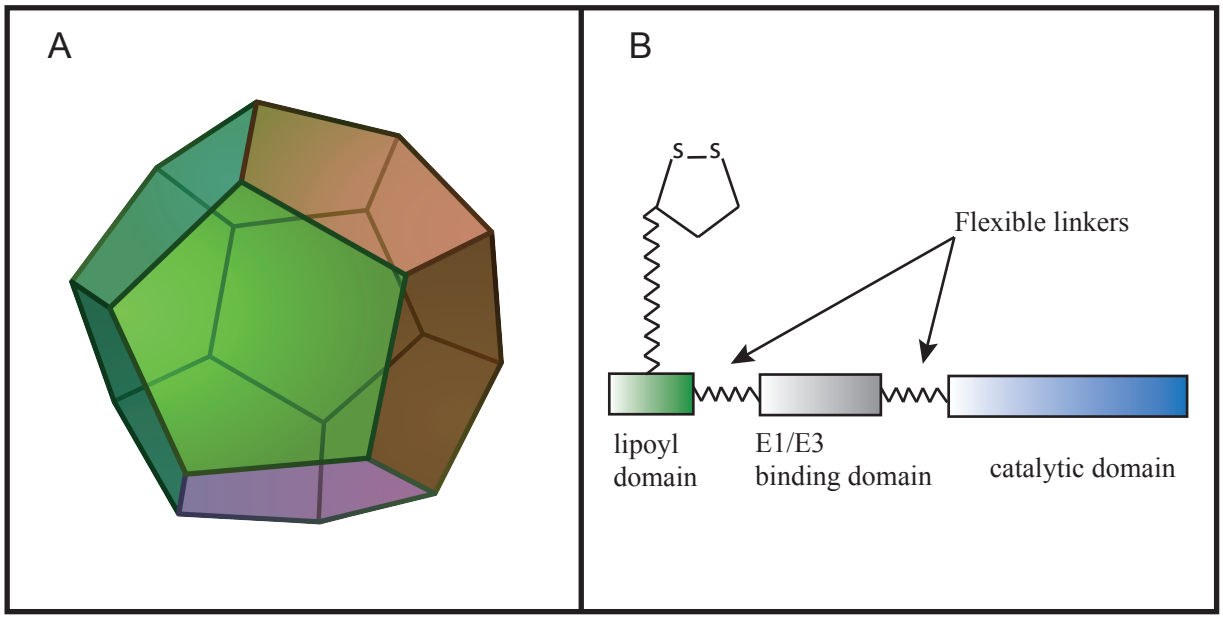

Figure 1.3: The pentagonal dodecahedral scaffold of E2 is illustrated in A), which is the core building block of the pyruvate dehydrogenase complex (PDC). B) depicts the composition of the enzyme E2 with one lipoyl domain with the swinging arm attached to it. The catalytic domain the E1/E3 binding domain and lipoyl domain are connected by flexible linkers.

lysine residues in the E2 polypeptide chains. Note that this bond is covalent, hence the lipoyl domain is a prosthetic group. If the lipoyl group is not attached to the domain the ring does not serve as an substrate for E1 [41]. The swinging arm provides a flexible essentially free rotating arm, representing a channeling linkage between the active centers of E1, E2 and E3 [42]. The tip of the outer linker of the lipoyl group is the dithiolane ring Fig.(1.3) and is the essential part for the reductive acetylation by E1, making the lipoyl domain the specified substrate of the reaction[43].

1.4.2 The Reaction Scheme of the Pyruvate Dehydrogenase Complex

In the following we discuss the individual chemical reaction of the pyruvate dehydrognase as depicted in Fig(1.4 A) and B)). The enzyme E1 converts the pyruvate, using the cofactor thiamine diphosphate (ThDP)[1] to carbondioxide (CO_2).The metastable enamine ThPD reacts with the dithiolane ring of the lipoyl domain, transferring the remaining acetyl-group onto the ring. The reaction mechanism for the enzyme E1 is best described by the "ping-pong"-mechanism, where two successive substrates react with the same enzyme at

[1]This sometimes also called thiamine pyrophosphate (TPP)

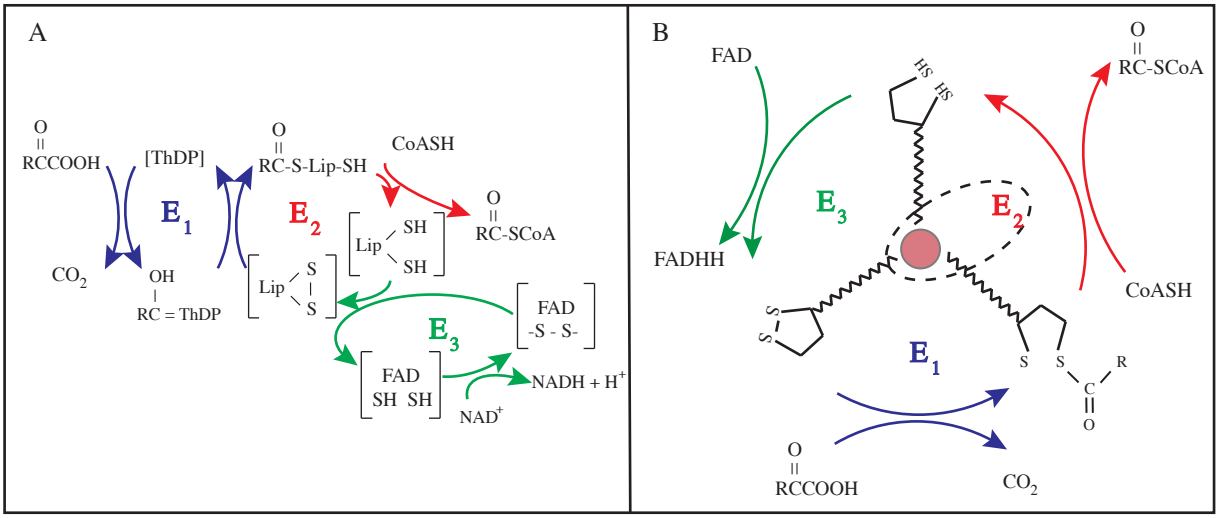

Figure 1.4: The full reaction scheme of the pyruvate dehydrogenase and the reaction of the swinging arm are presented. In A) the full chemical reaction scheme is shown including the enzymes (E1, E2 and E3) and the cofactors thiamine diphosphate (ThDP) of E1, the cofactor A (CoASH) of E2 and NAD^+ of E3. Picture B) shows the reaction scheme of the swinging arm, with the arm itself being covalently attached to the enzyme E2. The dashed ellipse indicates that the swinging arm is a prosthetic group of the enzyme E2 and does not exist independently of the enzyme. (See text for a detailed explanation.)

the active site [44]. In this case the pyruvate and the dithiolane ring of the lipoyl domain are these two different substrates. The arm with the acetyl-group, from the pyruvate, attached to it then "swings" towards the enzyme E2 and reacts with the coenzyme A to form the acetyl-CoA. Note that every time a reaction takes place the coenzyme A must be provided by the surrounding solution. This is the product of the pyruvate dehydrogenase, which subsequently enters the citric acid cycle. As the last step it is necessary to recover the disulfide bond of the lysine residue of the complex to enter the pyruvate dehydrogenase cycle again. To this end the arm moves towards the active site of E3 and a flavin-mediated oxidation takes place. This again is a two step ping pong reaction, at first the FAD reacts with dihydrolipoate leading to the $FADH_2$ and the original lipoate state of the swinging arm. In the last step the cofactor NAD^+ oxidizes $FADH_2$ back to its original state FAD producing a NADH and a H^+. Note that the NAD^+ the Cofactor A have to be provided by the solution whereas ThDP is a cofactor of E1, the lipoyl domain and the FAD are prosethic groups of E2 and E3 respectively [45]. The three main advantage of the complex are the stabilization of the participating proteins, an effective regulation by pyruvate dehydrogenase kinase (PDK) and pyruvate dehydrogenase phosphatase (PDP) activating respectively deactivating the complex, and an increase in the efficiency of the reaction cascade by substrate channeling, due to the swinging arm.

1.5 Outline of this Thesis

This thesis is structured into two main parts. In the first part we consider the consequences of manipulating the configuration of a single enzyme via external mechanical force for its activity. Based on an one dimensional random walker model we extract the kinetic and thermodynamic parameters of the enzyme's folding states from an experimentally obtained time series protocol. However, the main focus of the thesis lies on the second part. Here, we investigate a representative enzymatic pathway consisting of two different type of enzymes. We optimize the arrangement of the second enzymes E_2 relative to the first enzymes E_1 by means of maximizing the generation of the product from the second enzyme. To this end a novel concept called "enzyme exposure" is introduced to explain the optimal enzyme profile. In this work we are trying to make a further step towards a better theoretical understanding of spatially arranged enzyme cascades.

Chapter 2

Kinetics and Structure of a Single Enzyme

Enzymes have been investigated now for centuries, their behavior, their specificity and kinetic rates. Only until recently assay methods have been used to determine the kinetic parameters of enzymes. The advent of single molecule enzymology drastically changed the way enzymes are measured. Not only that their transient time is measured on a single molecule level to extremely high time resolution, also the exposure to external force has not been possible before. The mechanical exertion of force by an atomic force microscope (AFM) or an optical tweezer changes the conformation of the enzyme in a controlled way. Thus the activity depending on the conformation of the enzyme can be investigated, as well as the enzyme's folding path. While folding, the enzyme takes several different conformation, every single confromation is associated with a specific point in the corresponding energy landscape. The path through the energy landscape ultimately ends up in the global minimum, hence the enzyme reaches its ground state. Thermal fluctuations push the enzyme out of its ground state causing conformational changes in its composition. By using methods like Förster resonance energy transfer (FRET) or total internal reflection fluorescence microscopy (TIRF) the reaction can be made visible. One could ask the question why do we even want to perform single molecule measurements? What is the benefit beyond determining the kinetic rates? On the one hand single molecule microscopy methods resolve the the properties of enzymes on a single molecule level as opposed to enzyme assay measurements. This leads to the observation of an unexpected effect called dynamic disorder, in short it means fluctuating reaction rates. On the other hand revealing the composition and the structure of the enzymes helps to improve their activity, in terms of faster rates or greater/weaker specification to the substrate, by designed evolution.
In the following we begin with a brief overview of the enzyme assay measurement methods. Then we will give a survey of the single molecule experimental methods needed to explain the experimental setups, accompanied with new insights into the kinetics of enzymes. Note that there are many more microcopy methods and our small survey is far from being complete. Subsequently, we will introduce the experiment performed by the Gaub group to which a one dimensional random walker model describing the folding pathway is tested.

Additionally, we present a manuscript demonstrating the validity of the model depending on the accuracy of the experiment. Moreover, we provide a rigorous derivation of the kinetic and thermodynamic parameters and extend the model to allow for multiple minima in the energy landscape.

2.1 Enzyme Measurement

The way how kinetic rates of enzymes are measured is divided into two classes. The first one is based on ensemble measurement and is exclusively restricted to measure the rates of the kinetic reaction. The second class, which has been developed more recently, is the measurement and manipulation of single biomolecules. With these single molecule measurement methods not only the rates of the enzymes can be measured to an astonishing accuracy but also the structure of the enzymes can be resolved.

2.1.1 The Enzyme Assay Method

The measurement of catalytic activity based on enzyme assays was for a long time the only way to obtain the kinetic parameters of the enzymes. In the case of the continuous assay a particular reaction is monitored over time [21]. This particular enzyme assay typically measures either the consumption of substrate particles or the generation of product particles over time. Many of these enzymatic reactions include the interconversion of NAD from the oxidized form to the reduced form or vice versa. The so called spectrophotometer measures the absorption of light at a particular wavelength by one form of the NAD. From the amount of these products the kinetic rates of the reaction can be determined. Certainly the main disadvantage is that the enzyme assay can only resolve bulk activities and therefore can not be used to resolve single enzyme properties [46].

2.1.2 Single Molecule Measurements

Over five decades ago Rotman performed the first single enzyme experiment [47]. To this end enzymes β-glactosidase and substrates 6-hydroxyfluoran-β-D-galactopyranoside (6HFG) were encapsulated in water droplets which themselves are placed in a oily emulsion. The enzyme cleaves the substrate and releases a fluorescent dye gathering within the droplet causing it to be fluorescent. Droplets which do not contain any enzymes will stay dark. This demonstrates, that within a seemingly "uniform" emulsion only parts of it are active. However, it took another 35 years to detect the first single enzyme reaction [48]. By improving the state of the art TIRF, fluorescently labeled myosin molecules were imaged to detect individual ATP turnover reactions. The ATP molecules themselves were labeled and therefore were detected when they entered the resolution region of the TIRF in which the enzyme was placed. In order to measure features of enzymes on a single molecule level a reporter system is required. Conceptually there are two different ways how to analyze single enzyme activity [49]. In the first category the substrate is modified such that it is

fluorescent or fluorogenic. Once the enzyme converts the substrate into a product it can be directly observed by the microscope. The second category however is somehow different, fluorescent dyes are coupled to the enzymes reporting a change in activity by a change in the enzyme's conformation. Conformational changes are usually detected by FRET [50], where the brightness of the acceptor dye depends on the distance to the donor. One of the first application of the FRET principle on the measurement of enzyme activity was done by Hu *et.al.* [51]. They labeled the enzyme staphylococcal nuclease with two different dyes. Upon the fluctuation of the intensity of the donor-acceptor emission they argued that the necessary reason for that effect was caused by a fluctuation in the enzyme's conformation. Note, this is only a conceptual excerpt of the vast field of high resolution single molecule microscopy, more intensive reviews can be found in [52, 53, 54]. Single molecule fluorescence microscopy showed that the kinetic rates of enzymes are not constant but fluctuate in time. The real time monitoring of the flavoenzyme cholesterol oxidase reducing fluorescent flavin adenine dinucleotide (FAD) to nonfluorescent $FADH_2$ exhibits that the turnovers are not independent of the previous turnovers [32]. Single molecule experiments record stochastic time traces of several consecutive reactions of the same enzyme. All these waiting times τ, which is the time for an reaction to occur, form a waiting time distribution, which according to MM kinetics should decay exponentially [32, 55]. The experiments [56] showed the expected exponential decay for small substrate concentrations, however for large substrate concentration they observed a deviation. Large substrate concentration basically means that the catalytic rate is the limiting one. Therefore the experiment suggests that the catalytic rate has not a single value but several. Such an effect can be modeled by assuming a probability distribution for the catalytic rate, leading to a fluctuating rate in time. This is called dynamic disorder. In contrast to those single molecule experiments, ensemble average experiments cannot distinguish between static [1] and dynamic disorder.
The second type of necessarily tools developed over the last decades are the atomic force microscope (AFM) [57, 58] and optical tweezer [59]. Although, for instance the AFM has a wider range of application, we restrict ourselves only to mechanical pulling experiments. Typically the AFM can apply a force between pN and nN [60]. In such experiments one end of the enzyme is immobilized to a prepared surface while the other end is attached to the cantilever of the AFM. When lifting the cantilever the AFM exerts a force on the enzyme until one of the ends or the enzyme itself ruptures. This force induces a conformational change in the enzyme or tests the strength of ligand bindings. We will discuss such an experimental setup later in more detail. The combination of AFM and single molecule microscopy proves particularly useful once the response of enzyme activity to external stress is investigated. While single molecule microscopy is an independent readout of the enzymatic turnovers the AFM is an optimal tool exerting external forces on the enzymes altering the activity of the enzyme.

[1] A static heterogeneity in the rates among the individual enzymes

2.2 Folding of Biomolecules

Recent developments in single molecule microscopy as well as single molecule force spectroscopy established new insights in the structure and folding of proteins. These two subjects are closely connected, while on a folding pathway the protein passes through several different conformations the final enzyme structure denotes the stable state. Single molecule measurements made it possible to monitor such a biomolecule on the folding path towards the stable state. Conventional enzyme assay methods can only probe the average features of enzymes. The exact way how proteins fold from the encoded 1D amino acid sequence into their 3D structure remains elusive even after more than 50 years of research [61, 62]. The exact structure of the enzyme is crucial for its biological activity[63]. The 3D structure of enzymes is not rigid, the non covalent folding provides a flexible frame, thus the enzyme does not exist in one definite state rather it fluctuates between several different. Binding of the substrate rearranges the active center and hence the conformation of the enzyme. The dynamic process of binding leads to breaking bonds and forming new ones along the reaction coordinate. For instance, not necessarily is the conformation of the enzyme, to which the substrate binds, the same as the one that performs the catalytic reaction[26]. This idea is closely related to the induced fit model we have previously discussed. A static enzyme structure only exists in its stable state or ground state while thermal fluctuations influence the conformation of the enzyme altering its activity. It was shown in [64] that even fluctuations on an atomic scale influence the catalytic activity of the enzymes. With the single molecule force spectroscopy we have a very powerful tool at hand to investigate the different structural conformations of an enzyme and link the folding path to a trajectory through the underlying multidimensional energy landscape. The energy landscape is a concept trying to unite biochemical and biophysical concept to have one framework. The number of reaction coordinates determines the dimensionality of the surface of the energy landscape. Unlike most purely physical situations, such as the time constant shape of the free energy functional of the Ising model [65], the energy landscape of biomolecules is not constant. Evolutionary entailed mutations erratically rearrange the surface of the energy landscape changing the path of the chemical reaction [66]. But not only the folding structure can be investigated using AFM. A further application is the testing of ligand receptor bonds [67, 68, 69]. In [70] for the first time it was achieved to measure the binding force of a single molecule. To this end, a single biotin molecule was bound covalently to the AFM via chemical linkers. Additionally, streptavidin was immobilized on a surface. After a streptavidin-biotin complex was formed, the AFM exerts force on the complex until it ruptures. This procedure was repeated several times to obtain a rupture force histogram, where they could read off the binding strength. A force histogram associated with several distinct peaks indicates several different bonds. In the following we will discuss recent results undermining the importance of force pulling experiments.

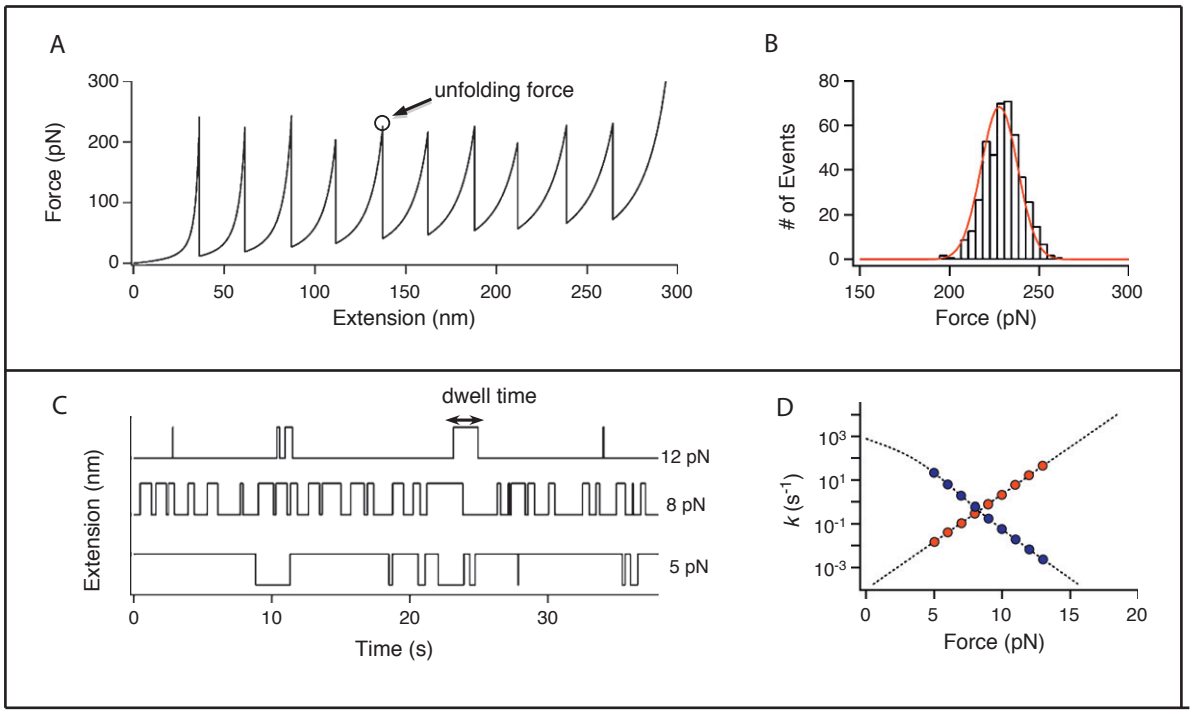

Figure 2.1: The two different ways of pulling experiments are shown, constant velocity and constant force respectively. A) shows the force-extension curve of a constant velocity experiment. The sudden drop of the force signalizes an unfolding of a bound of the polyprotein. Each of this sudden declines of the force symbolizes one event. Gathering all this events together and repeating the same experiment several times leads to a force histogram B). The red line is the fitted distribution of the force histogram. C) for a constant force pulling experiment the folding and unfolding of a bound over time at a fixed force is monitored. Different applied forces lead to different dwelling time in the folded respectively unfolded state of the bond. D) Using this dwelling times (red unfolding and blue folding) and the corresponding forces the kinetic rates can be estimated. Reprinted with permission from [60].

2.3 The Energy Landscape

In principle two different protocols are used in force pulling experiments to receive the kinetic rate parameters, the constant force method also called force clamp and the constant velocity method also known as force ramp [60]. Although conceptually these two methods are different they reveal similar results. In a constant velocity experiment, for example the cantilever of the AFM pulls at a constant velocity on the biomolecule. The force exerted along the pulling direction causes an unfolding of a subdomain leading to an increase in distance of the cantilever and a drastic decrease of the force. A sawtooth like pattern emerges Fig.(2.1 A), B)). The combination of all these peaks leads to a force distribution

$p(F)$. The modified Bell's expression [71] for time dependent force $k(t) = k_0 \exp[\beta F(t) x^\dagger]$ proved to be insufficient to model the pulling experiments. The distance form the free-energy minimum to the barrier is $x^\dagger$ and k_0 is the intrinsic force free rate of interest, $\beta = (k_B T)^{-1}$ and $F(t)$ the time dependent force. However, Hummer and Szabo [72] proposed a novel model using Kramers diffusion theory on an harmonic free energy surface with a sharp barrier. Testing their microscopic theory against computer simulation of a multimodule titin model they could extract information about the intrinsic kinetic parameter k_0, the position of the transition state $x^\dagger$ and the free energy barrier $\Delta G^\dagger$.
The second type of experiments is performed by using constant force loads. A constant force is applied to the biomolecule and at the force where both states, the folded and the unfolded state, coexists a transition occurs Fig.(2.1 C), D)). This transition between the two states leads to an abrupt distance change. Form the time-extension protocol the dwell time in the folded or unfolded state can be extracted. These dwell times depend on the value of the applied force and the kinetic rates can be red off, for example using Bell's framework. The two different approaches should give the same force free kinetic rates. A theoretical framework based on Kramers diffusion theory was evolved to explain both different pulling experiments with a single underlying theory [73]. The intrinsic rate coefficient, the location of the transition and the free energy activation can be obtained and were experimentally tested by applying the theory on DNA unzipping through a nano pore [74]. Later that concept was extended to higher dimensional energy landscapes [75, 76]. At this point it should be mentioned that pulling experiments are not restricted to proteins or enzymes in particular, they also helped to reveal the 3D RNA structure by testing the strength of reversible RNA hairpin folding [77, 78]. However, the crucial point in most of these folding experiments was, that for a distinct biomolecule different initial conformations, that correspond to different starting points on the energy surface, always led to the same final ground state. Pulling experiments suggest that evolution favors a funnel like shape of biomolecules [79, 80]. Finally, the pulling experiments are not exclusively *in vitro* studies, there are many physiological situation where biomolecules are exerted to external stress for instance the muscle contraction of titin, intra-cellular dynamics or blood flow.

2.4 The Experimental Setup

The powerful methods of single molecule microscopy and force spectroscopy, gave amongst others, new insight in structure, folding path and kinetic rates of enzymes. In the recent years there have been several attempts to combine those two methods. One of the first experiments of this kind probed the connection between force generating myosin heads and the ligand binding events induced by ATPase [82]. Further experiments [83, 84, 85, 86] have been performed, they clearly demonstrate the power of these experimental techniques. Here we will present the experiment that gave rise to our random walker model. But before we introduce our model we will discuss the experiment performed by Gumpp *et. al.* [81] in some greater detail. The enzyme lipase B from *Candida antarctica* (CalB) was monitored since it converts the fluorogenic substrate 5-carboxyfluorescein diacetate (CFDA) into the

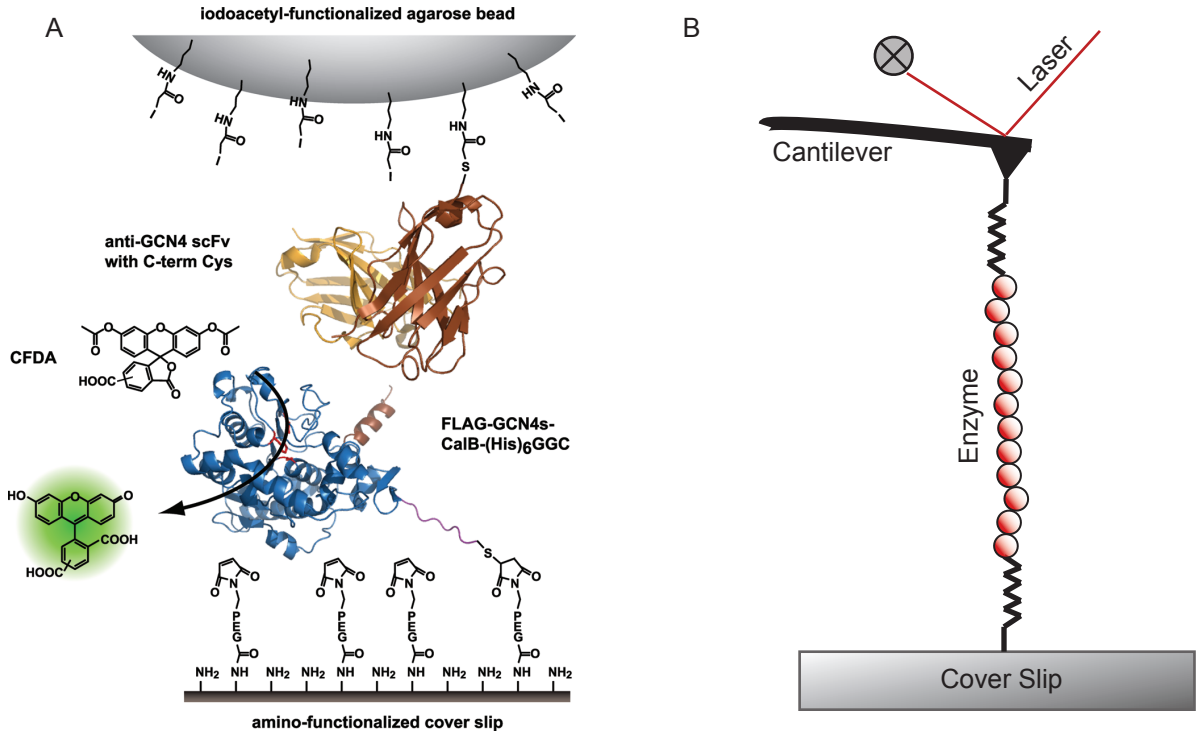

Figure 2.2: A) shows the biochemical details of the experimental setup. The enzyme CalB was modified, to covalently bind it to the cover slip. To this end cysteine was introduced at the C-terminus of CalB, which itself couples to the heterobifunctional polyethylene glycol spacer with the amino functionalized coverslip. Additionally, the GCN4s peptide sequence was linked to the N-terminus of CalB which binds to the antibody (anti-GCN4 scFv). The cysteine at the C-terminal of the antibody-antigen complex couples to an iodoacetly activated agarose bead fixed to the cantilever. Reprinted with permission from [81]. B) shows a schematic but essential representation of the experiment depicted in A).

fluorescent product 5-carboxyfluorescein. The different turnovers were recorded by TIRF. One end of the enzyme was covalently bound to a cover slip the other end was attached to the cantilever of the AFM through an antibody-antigen interaction. The enzymes unfolding induced by pulling with the AFM has a certain upper force limit. The rupture of the antibody-antigen interaction ensures a repeatable initial condition for the enzyme. After the rupture the enzyme refolds, driven by minimizing the free energy, into the initial stable configuration, where the enzyme's activity is regained. Likewise this time event can be measured by monitoring the generation of fluorescent products while in the unfolded state the enzyme's activity is suppressed. Therefore, the time period of the refolding process can be measured, and due to the well defined antibody-antigen interaction strength the experiment is repeatable under the same conditions. Hence an entire time series protocol and the resulting time histogram can be extracted. Consequently, the mean folding time,

the variance and all higher moments can be obtained.

2.5 A Stochastic Inverse Problem in Single-Molecule Enzymology

In the experimental paper [81] a one dimensional random walker was proposed to model the experimental data and extract the kinetic rates. In the following manuscript we will establish a fit routine for a more rigorous way of obtaining the kinetic and thermodynamic parameters of the folding pathway. From the random walker model the mean folding time, the variance and the skewness can be calculated. On the other hand, the mean, variance and skewness can also be calculated by the experimentally obtained times series of the pulling protocol. We equalize both sets of moments and can then inversely determine the kinetic parameters of the model. Gaining the kinetic parameters from a stochastic quantity is also known as a stochastic inverse problem [87] . Since the moments from the experiments depend strongly on the number of sequentially performed experiments, the kinetic parameters will show an error depending on the number of experiments. The manuscript will present the quality of the kinetic and thermodynamic parameters depending on the number of data points in the time series protocol under the assumption of a single minimum of the underlying free energy landscape. Further, we will test the validity of this assumption once we allow the energy landscape to display several minima. We will use the randomness parameter to characterize the "bumpiness" of the energy landscape.

A Stochastic Inverse Problem in Single-Molecule Enzymology

Alexander Buchner and Ulrich Gerland
Arnold Sommerfeld Center for Theoretical Physics,
Ludwig-Maximilians-Universität München, Germany
(Dated: January 14, 2014)

Enzymes are ubiquitous in nature, and in physiological situations they are often exposed to mechanical force changing their native conformation. Recent combination of atomic force microscopy and total internal reflection microscopy enabled the investigation of the change of enzyme activity under external stress in vitro. We provide a simple random walker model to determine the number of intermediate conformations and the kinetic transition rates between those states of the enzyme, while it is refolding into its native state. We generate a set of artificial time series, using a Gillespie algorithm, which represents the refolding of an enzyme starting form an unfolded state till it reaches its active state again. From this time series the mean, variance and skewness of the distribution are calculated. Fitting the values of those quantities using our model the kinetic transition rates and the number of different configurations of the enzyme are extracted. The finite number of realization in the algorithm leads to an inaccuracy in the averaging quantities, thus in determining the kinetic rates, a so called stochastic inverse problem. We give an estimate of the quality of the kinetic rates depending on the number of times in the time series protocol We also provide a lower bound to the free energy difference ΔG of the funnel in the energy landscape. Additionally, by means of the randomness parameter r a lower bound of the number of different configurations can be estimated.

INTRODUCTION

The developments of the last years in single molecule spectroscopy allow for a detailed measurements of biomolecules, that go beyond ensemble observations. The Atomic force microscope (AFM) [3, 9] and the optical tweezer [1] are widely used methods to stretch macromolecules by applying mechanical force. Such new experimental methods gave further insights into long outstanding problems like the folding pathway of proteins [24, 25] and the strength of receptor ligand bonds[11, 26]. It has been shown that a molecule passes through several different intermediate sates before it reaches the final native state. Moreover constant force experiments have been applied to investigate the kinetic and thermodynamic parameters determining the folding and unfolding of RNA hairpins [23, 32]. The three dimensional structure of specific RNA governs its catalytic activity. Force as a control parameter in single molecule experiments on proteins is an excellent tool to explore the underlying energy landscape of biomolecules. On the theoretical side it has been established that the shape of the underlying energy landscape conducts the conformational state of the biomolecules [13, 29, 33]. A funnel like shape of the energy landscape guides the protein through the different conformations to its native state [27]. However the energy landscape can be multidimensional often a single reaction coordinate, in the pulling direction, is assumed[10, 22]. Advances in molecule fluorescent techniques such as Förster resonance energy transfer (FRET) and total internal reflection microscopy (TIRF) improved time resolution of single molecules. The turnover rates of the enzymatic activity could be measured up to sub-millisecond time resolution [5]. Further the connection between conformal changes in the enzyme structure and its catalytic activity was revealed, the so called dynamic disorder [17, 19]. In the above described examples force and high resolution spectroscopy gave insights into the features of the energy landscape of polymers. But there are many examples where force acts on biomolecules in a cellular environment, such as muscle tissue, force signaling and cytoskeletal dynamics [33]. A more concrete physiological context is the force induced mechanical contraction of the muscle protein titin [7, 16, 30] . In this paper we focus on an experiment, which combines the AFM and the TIRF spectroscopy method. One end of the enzyme (lipase B) is fixed on an amino functionalized coverslip via a covalent bond, the other side is bound to the cantilever of the AFM via a reversible antibody antigen bound. With the AFM the enzyme is stretched until the rupture of the antibody antigen bound initializes the folding of the enzyme. The enzyme's final, hence refolded state, is monitored using TIRF which observes the enzymatic reaction converting a fluorogenic substrate into a fluorescent product. The combination of AFM and TIRF provides a distinct starting and final point, from which a times series protocol can be extracted[4]. In contrast to the pulling experiments where the rupture event leads to a force dependent survival probability of a particular conformation of the enzyme, in [4] the rupture event only initializes the refolding towards the native state.

In this paper we suggest a 1D random walk model describing the dynamics of the enzyme folding. Our goal is to determine the number of different intermediate states and the transition rates between them from a times series protocol, which is usually obtained by the experiment [4]. However we use an artificial time

series by performing a Gillespie algorithm with detailed knowledge about the microscopic rate parameters. From the time series protocol we get a first passage time distribution (FPTD) and the moments. By fitting our model to these moments the underlying kinetic and thermodynamic parameters are obtained. This is known to be a stochastic inverse problem [12]. Additionally, we receive the minimal depth of the funnel of the energy landscape ΔG for the case of equal forward and backward rates. Further, we allow for several minima in the energy landscape. Surprisingly, an energy landscape with only one minimum provides always a lower bound to the overall depth of the global minimum, when considered a general pathway of the same topology with several minima in the energy landscape. Additionally, the number of intermediates states appearing in the refolding are also bound from below. Thus we provide an estimate of the minimal number of intermediate conformation along its pathway.

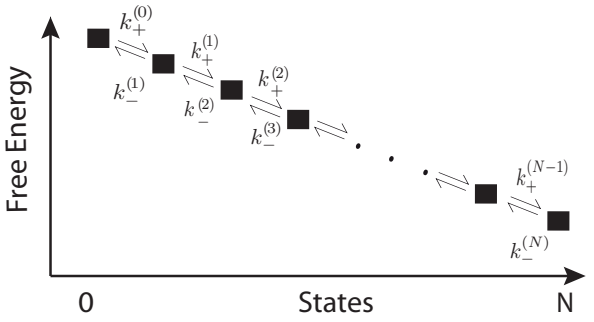

FIG. 1: A random walker with N different forward and backward rates modeling the folding pathway of an enzymes. The intermediate states represent the different 3D configurations of the enzyme in a specific folding state. The forward, respectively backward rates $k_+^{(i)}$ and $k_-^{(i)}$ represent the transition of the enzymes between those configurations. The downhill reaction is driven by a free energy difference induced by the pulling force.

METHODS AND MODEL

We model the dynamics of enzyme refolding as described in [4] by a 1D random walker, see Fig (1). The intermediate positions of the random walker represent the individual configurations of the enzyme and the back, respectively, forward rates the transitions between those configurations. The 0-th state of the of the random walker corresponds to the inactive state (unfolded state) of the enzyme and the N-th state to the active one (native folded state). It should be mentioned, that due to the random walker analogy enzyme always folds in the same way, hence the sequence of the folding configurations is fixed. As described in [4] the experiment is initialized by the rupture of the antibody-antigene interaction, representing the starting time of the experiment. The N-th state of our random walker, the final configuration, is recorded by a fluorescence signal emitted by the product of the enzyme reaction. We set $k_-^{(N)}$ equal to zero in order to make the N state absorbing, excluding spontaneous deactivation of the enzyme activity. For this model the Master equation reads

$$\frac{\partial}{\partial t}P(n,t|n',t') = -(k_+^{(n)} + k_-^{(n)})P(n,t|n',t') + k_+^{(n-1)}P(n-1,t|n',t') + k_-^{(n+1)}P(n+1,t|n',t') \quad (1)$$

with $k_-^{(n)}$, $k_+^{(n)}$ and $n = 0..N$, the backward and forward rates [14]. At time $t = 0$ the probability of finding the system in the inactive state is equal to one and the probability of finding the enzyme in any other state is zero.

Fitting the cumulative first passage time distribution from the time series protocol directly using the solution of the master equation to obtain the kinetic parameters is a highly degenerate problem, thus we will consider the moments. The first moment is given by

$$T_1(m) = \sum_{k=m}^{N-1}\sum_{l=0}^{k}\frac{1}{k_+^{(l)}}\prod_{j=l+1}^{k}\frac{k_-^{(j)}}{k_+^{(j)}}. \quad (2)$$

and all other moments can be constructed recursively according to

$$T_n(m) = \sum_{k=m}^{N-1}\sum_{l=0}^{k}\frac{(n-1)T_{n-1}(l)}{k_+^{(l)}}\prod_{j=l+1}^{k}\frac{k_-^{(j)}}{k_+^{(j)}}. \quad (3)$$

where n here is the order of the moment. For a complete derivation of the moments see Appendix. The specific value of the moments are obtained by evaluating the time series protocol from the experiment. For the rest of the article we focus on the following moments, the mean first passage time, the variance, and the skewness. While the skewness describes the uneven distribution of waiting times left and right of the maximum of the waiting time distribution. Having a finite number of experimental data and no underlying probability distribution the accuracy of the e.g. moments can only be provided to some extent. The sample mean, variance, and skewness can be calculated by the first passage times t_i (i=1..M) only. The sample mean is given by

$$\bar{t} = \frac{1}{M}\sum_{i=1}^{M}t_i \quad (4)$$

where n is the number of runs. In similar way the sample variance is given by

$$s^2 = \frac{1}{M-1}\sum_{i=1}^{M}(t_i - \bar{t})^2 \tag{5}$$

and the sample skewness is

$$sk_{sample} = \frac{\sqrt{M(M-1)}}{M-2}\frac{\frac{1}{M}\sum_{i=1}^{M}(t_i-\bar{t})^3}{(\frac{1}{M}\sum_{i=1}^{M}(t_i-\bar{t})^2)^{\frac{3}{2}}}. \tag{6}$$

An estimator is called unbaised, if the estimators expectation value is equal to the population value. For example in our case, if $\mathbb{E}(\bar{t}) = \langle t\rangle$, with $\langle t\rangle$ the mean value of a possible underlying distribution. The above defined sample mean and variance are unbiased, the sample skewness, though, is biased, yet the pre-factor assures an expectation value of zero for a symmetric distribution [28, 31]. One can easily calculate physical quantities using higher moments such as kurtosis. Although, the finite number of waiting times give rise to an inaccuracy in the value of the moments it turns out that the nature of our inverse stochastic problem allows for a more accurate determination of the kinetic parameters when fitting the moments rather than the cumulative waiting time distribution.

RESULTS

Determining the Free Parameters

As a first attempt we set all forward and all backward rates equal, $k_+^{(n)} = k_+$ and $k_-^{(n)} = k_-$ for all n. The parameters k_-, k_+, and N are free parameters of the theory. Since the fact of equal forward and backward rates, respectively does not hold true in general, we will relax this assumption later. They are determined by fitting the theoretical model to the experimental data. Form the time series protocol we can easily extract the sample mean value, the sample variance, and the sample skewness of the waiting time distribution. On the other hand with Eq (2,3), we can calculate these physical quantities depending on the free parameters k_-, k_+, and N. After some straightforward manipulations we arrive at the following expression for the mean

$$\langle T\rangle = \frac{1}{k_+}\frac{1}{1-K}\left(N - K\frac{1-K^N}{1-K}\right). \tag{7}$$

The variance $\sigma_T^2 = \langle T^2\rangle - \langle T\rangle^2$ and the skewness $sk_T = \frac{\langle T^3\rangle - 3\langle T^2\rangle\langle T\rangle + \langle T\rangle^3}{\sigma_T^3}$ are obtained in the very same fashion, yet the calculations become more and more extensive as one goes to higher moments. The variance reads

$$\sigma_T^2 = \frac{1}{k_+^2}\frac{1}{(1-K)^2}\left(N + \frac{K}{1-K}(4NK^N - 4 + 2N + 4K^N)\frac{K^2}{(1-K)^2}(4K^N + K^{2N} - 5)\right) \tag{8}$$

and similarly the skewness

$$\begin{aligned} sk_T = \frac{1}{k_+^3}\frac{1}{(1-K)^3}\Bigg(&2N + \frac{1}{1-K}(12N^2K^N + 30NK^N + 12N - 18 + 18K^N) \\ &+ \frac{K^2}{(1-K)^2}(12NK^{2N} + 12N + 12K^{2N} + 48K^N + 36NK^N - 60) \\ &+ \frac{K^3}{(1-K)^3}(2K^{3N} + 30k^N - 44 + 12K^{2N})\Bigg)/(\sigma_T^2)^{\frac{3}{2}} \end{aligned} \tag{9}$$

where $K = \frac{k_-}{k_+}$ is the equilibrium constant. The numerical values for the moments for Eqs.(7-9) can be extracted from the experiment. These sample quantities serve as the input values for our fit routing yielding to k^+, k^-, and N. The mean $\langle T\rangle$, the variance σ_T^2, and the skewness sk_T are usually called population quantities. The population mean, variance, and skewness are calculated by means of the probability distribution. For

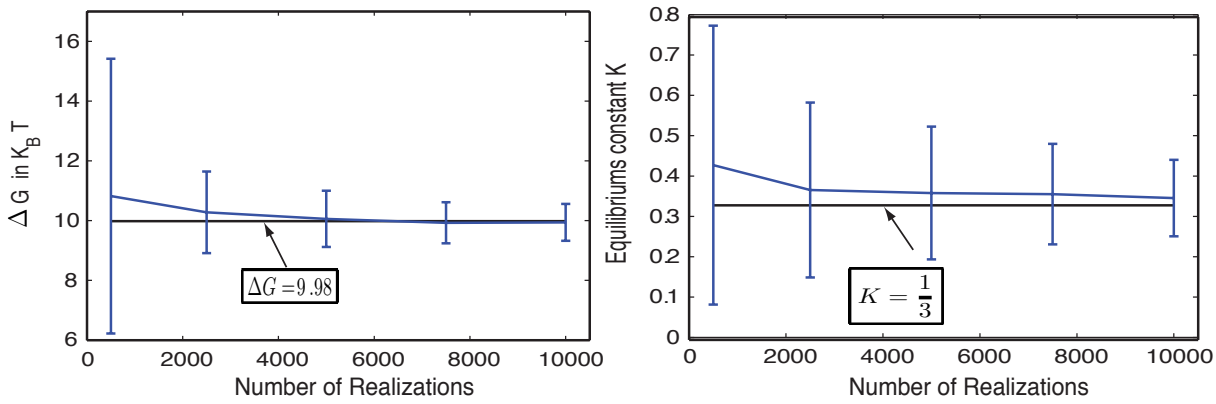

FIG. 2: In the left picture the free energy ΔG is plotted versus the number of Gillespie realizations. The black straight line denotes the original value $\Delta G = 9.98 k_B T$, the blue line with the error bars is the free energy difference obtained by the fit routine. The left picture depicts the equilibriums constant K versus the number of Gillespie realizations. The black straight line is the original value $K = \frac{1}{3}$ whereas the blue line the fitted value.

the rest of the article we refer to them as the mean value, variance, and skewness.

In the following we create a set of artificial time series by simulating the master equation (15) using the Gillespie algorithm [18] with $k_- = 1$, $k_+ = 3$, and $N = 9$. The equilibrium constant $K = \frac{1}{3}$ and the energy difference is $\Delta G = 9.98 k_B T$ according to

$$\left(\frac{k_-}{k_+}\right)^N = \exp\left(-\frac{\Delta G}{k_B T}\right). \tag{10}$$

. The essential dynamics of the system is captured by the model depicted in Figure(1) a simulation of suchlike produces the desired experimental data. The main advantage using Gillespie is, that we have controlled experiment, controlled in the sense of knowing the applied values for k_-, k_+, and N. This circumstance is not given in an actual experiment. In other words, by means of this artificial set of time series the sample mean, variance, and skewness are obtained. Equalizing the mean $\langle T \rangle$, the variance σ_T^2, and the skewness sk_T with the sample mean $\bar{t}$, variance s^2, and skewness sk_{sample}, the Eqs. (7-9) can be resolved in terms of k_-, k_+, and N. Considered the complexity of these equations one can not hope for an analytical expression of k_-, k_+, and N in terms of $\langle T \rangle$, σ_T^2, and sk_T. Therefore the values of the free parameters of the system have to be determined numerically, using a fit routine, established in Matlab, see Appendix. Plugging the numerically received sample values into Eqs. (7-9) the inaccuracy of those sample values, due to finite number of sampling, leads to an error in determining the values for k_-, k_+, and N. Assuming the underlying physics, that all forward rates, respectively, backward rates are equal, holds true, our model captures the essential dynamics and we provide a validity for the number of runs (the number of performed experiments) to determine the microscopic parameters appropriately. To this end we perform for each specific number of Gillespie realizations M several sets of artificial data J. In other words, a specific experiment with a certain number of iterations is carried out several times in a row. Now we want to examine the behavior of the convergence of the fitted parameter here denoted be k_-^f, k_+^f, and N^f to the real values k_-^r, k_+^r, and N^r. Recall, in a proper experiment these real values are not known. We calculate $\left\langle \mathrm{Log}(\frac{|k_-^f - k_-^r|}{k_-^r}) \right\rangle$, $\left\langle \mathrm{Log}(\frac{|k_+^f - k_+^r|}{k_+^r}) \right\rangle$, and $\left\langle \mathrm{Log}(\frac{|N^f - N^r|}{N^r}) \right\rangle$ depending on the logarithm of the number of runs $\mathrm{Log}(M)$, the average in this case is taken over all different experiments J for a specific number of runs. Consequently, the M dependence of the convergence is obtained, $k_-^f \sim \frac{1}{M^{0.67}}$, $k_+^f \sim \frac{1}{M^{0.65}}$, and $N^f \sim \frac{1}{M^{0.56}}$.

Using k_-^f, k_+^f, and N^f we can calculate the thermodynamical quantities such as the free energy ΔG and the equilibriums constant K. For the rest of the article we drop the subscript f. In Fig.(2) the mean value and the error bars (standard deviation) of ΔG and K are displayed depending on the number of runs M. It can be observed that the fitted values are in good agreement with the underlying values, even for a low number of iterations. However the comparably huge value for the error bars, especially in the case of the free energy, shows the ambiguity of such stochastic inverse problems. It is advisable to exceed a certain number of iterations for the first passage times and additionally checking the validity of the found values by performing the same experiment several times, as it has been outlined in the previous section.

Multiple Minima of the Free Energy Landscape

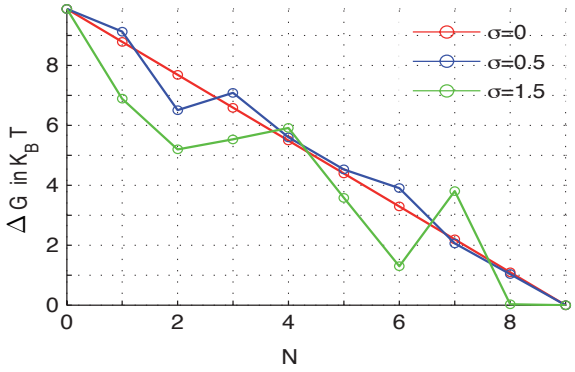

FIG. 3: The free energy landscape for different values of the bumpiness parameter σ. The red graph shows just one minimum at $N = 9$, this is the case when all forward. respectively, backward rates are equal $\sigma = 0$. Increasing the bumpiness $\sigma = 0.5$ (blue) to $\sigma = 1.5$ (green) local minima arise and the respective forward, backward rates become unequal according to Eq.(11) and Eq.(12).

As a next step we will show that the model with equal forward and backward rate, respectively provides a lower bound for the depth of the funnel in the energy landscape. In terms of the energy landscape equal rates lead to a smooth landscape with only one minimum, which is global. In oder to justify the usefulness of this assumption we show that the equal rates model provides a lower bound for the free energy ΔG and the number of states N, in the case where the underlying physical situation does not obey our previous assumption. To this end we generate a free energy landscape with several local minima. Assuming $\Delta G = 9.98 k_B T$ to be the thermodynamical constraint and additionally keeping the number of state fixed, namely $N = 9$, we create a bumpy energy landscape. Note, the energy value of initial the initial(stretched) state is $G = 9.98 k_B T$ and the final state (fully relaxed) state is zero. In between we divide the energy up into N bits with equal energy spacing between two sequential states. In our case the magnitude of energy between two sequential states is $\Delta G_i = 1.10 k_B T$ with $i = 1..N$. The energy of each intermediate state is shifted by a random number element of a normal distribution $N(0, \sigma)$ with zero mean and standard deviation of σ. The standard deviation is the control parameter of the bumpiness of the energy landscape is, see Figure(3). With these altered energies we calculated the shifted state depending rates

$$k_+(n) = k_0 \exp\left(\frac{-(G(n+1) - G(n))}{2k_B T}\right) \qquad (11)$$

and

$$k_-(n) = k_0 \exp\left(\frac{-(G(n-1) - G(n))}{2k_B T}\right) \qquad (12)$$

with k_0 setting the energy scale. Form here it immediately follows

$$\frac{k_-(n+1)}{k_+(n)} = \exp\left(-\frac{\Delta G_n}{k_B T}\right). \qquad (13)$$

as the equilibrium condition.

Assuming the enzyme refolding is determined by rates depending on the intermediate state, according to Eq(11) and Eq(12), but we still use your model of equal rates, clearly our fit routine will not reveal the true values of the kinetic parameters. Thus we will estimate the deviation form the fitted free energy and equilibriums constant to the real values . Plugging Eq.(11) and Eq.(12) into Eq.(3) and Eq.(2), respectively, we get the values for the first three moments and hence the mean value, variance, and skewness, but we keep the free energy $\Delta G = 9.98 k_B T$ fixed. The usual fit routine gives us the new values for state independent parameters k_-^{new}, k_+^{new}, and N^{new}. From Eq.(10), we receive the new free energy difference ΔG_{New}. The new free energy ΔG_{New} is always less than the actual free energy $\Delta G = 9.98 k_B T$. In Fig.(4) the lower bound for the number of states N and the free energy ΔG for equal forward and backward rates can be seen. The blue dashed line indicates the the free energy for a certain σ and the red dots represents the free energy ΔG_{New} obtained by the fit routine for equal rates. Same holds true for the number of sates N, where the fitted number of states (red dots) are always less than the initial number of states (blue dashed line). The χ^2-function is defined in the Appendix Eq.(27) and is the gauge for the validity of the fit. For σ values higher than 1 the fit routine becomes quite inefficient.

Furthermore it is informative to consider the randomness of our models. The regularity of the refolding enzyme is characterized by the randomness parameter

$$r = \frac{\sigma_T^2}{\langle T \rangle^2} \qquad (14)$$

[6]. In the case of time independent kinetic rates the value of the randomness parameter range between zero and one. For small values of r close to zero, the spread of the completion is small compared to its mean, therefore the enzyme (the reaction sequence) behaves very regularly almost deterministically. On the other hand values of r close to one point to a highly irregular behavior of the enzyme. Moreover, the randomness parameter is

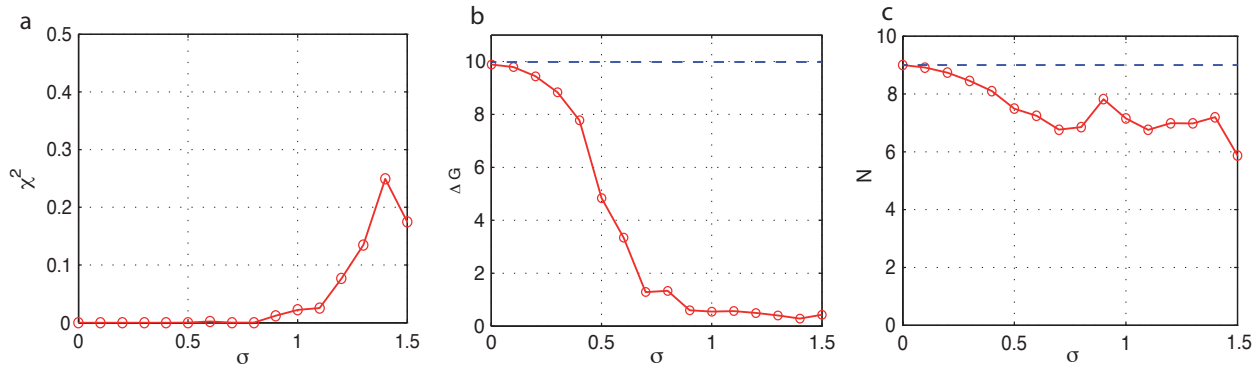

FIG. 4: In (a) the χ^2-function, denoting the validity of the fit routine calculated using Eq(27), shows a deviation form zero for values of *sigma* greater than one. However the fitted values of ΔG and N (red dots) for still provide a lower bound to the original values (dashed line), (b) and (c), independently of σ.

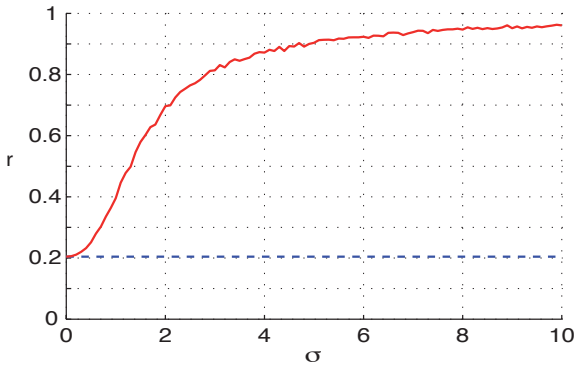

FIG. 5: The randomness parameter for unequal rates (red line) is always greater than the randomness parameter for equal rates (blue line). the overall free energy in both cases is kept fixed and equal independent of the value of σ.

measure of how many intermediate are present in our reactions sequence. In the case of negligible back reactions $r = \frac{1}{n}$ with n number of rate limiting steps, but not necessarily the overall number of steps involved. In other words $\frac{1}{r}$ is the minimal number of intermediate states involved in the reaction sequence. Including one back reaction int the reaction scheme, then r is equal to the number of all intermediate steps. However it does not matter where in the reaction sequence the back reaction occurs, thus several possibilities to arrange the reactions sequence exists, all produce the same value for r. In the following we compare the randomness of the model with equal rates to the one with rates depending on the bumpiness parameter σ. Keeping the free energy fixed $\Delta G = \Delta G(\sigma) = 9.98 k_B T$ we get the randomness parameter for equal rates $k_- = 1$, $k_+ = 3$, and $N = 9$ according to Eq.(14). The randomness parameter for the bumpy landscape obtained by the same equation however its value its different due to the different variance and mean value in that case. The randomness parameter for equal rates, blue dashed line in Figure(5), serves as a lower bound to randomness parameter for unequal rates, red solid line. By steadily increasing the value for σ the randomness approaches one, its highest possible value. Therefore, by including back reactions, a high value of r indicates a high bumpiness of the energy landscape. The bumpiness of the energy landscape leads to a higher value for the randomness parameter than in the case of a smooth energy landscape. With the measurement of the randomness parameter we can, firstly, roughly determine the number of intermediate states and, secondly, provide an estimate of the bumpiness of the underlying energy landscape.

DISCUSSION

Recent progress on single molecule experiments led to a combination of two different methods the AFM and the TIRF microscopy. The AFM clearly sets the staring time point of the stretched, inactive enzyme and the TIRF measures the final event once the enzyme is its relaxed, active state. We propose a random walker model describing the dynamic emerging from such a pulling experiment. The different conformational states are modeled by different positions the walker can take and the transitions between them are the free reaction parameter of our theory. We established a fit routine to estimate those parameters in the case of equal forward, respectively, backward rates. This fit routine was tested by generating an artificial set of time series using the Gillespie algorithm. The advantage of the artificial data set was, that its underlying parameters were known, therefore we could give an estimate on the validity of our fit

routine depending on the number of runs. Furthermore, we tested our fit routine against a situation where the transition rates are depending on the specific conformation of the enzyme. Still assuming the transition parameters to be equal, our fit routine provides a lower bound for the free energy ΔG and the number of states N. The so called randomness parameter was introduced giving us a measure of how regular the enzyme behaves. We think with our fit routine a helpful tool has been established, helping experimentalists to categories different types of enzymes via pulling experiments. We also saw the limitation of our model combined with the fit routine once the energy landscape becomes to bumpy, more and more local minima emerge, the inverse stochastic problem becomes ill-posted and merely quantitative conclusions can be drawn. In [25] it has been suggested that there could be several distinct pathways the enzyme follows until it reaches its native structure. This could be included into our model by incorporating different reaction branches into the model leading to new types of conformations. We believe that our fit routine is general enough to be applied to a random walker with branched pathways, that will give us further insight into the shape of the free energy landscape. We will leave such an extension of the model to future investigations.

This research was supported by the German Excellence Initiative via the program Nanosystems Initiative Munich and the German Research Foundation via the SFB 1032 Nanoagents for Spatiotemporal Control of Molecular and Cellular Reactions.

APPENDIX

Solving the Master Equation

The forward Master equation reads

$$\frac{\partial}{\partial t}P(n,t|n',t') = -(k_+^{(n)} + k_-^{(n)})P(n,t|n',t') + k_+^{(n-1)}P(n-1,t|n',t') + k_-^{(n+1)}P(n+1,t|n',t') \quad (15)$$

accounting for your problem, with $k_-^{(n)}$, $k_+^{(n)}$ and $n = 0..N$, are the backward and forward rates, respectively [14]. At time $t = 0$ the probability of finding the system in the inactive state is equal to one and the probability of finding the enzyme in any other state is zero. The corresponding backward Master equation reads

$$\frac{\partial}{\partial t'}P(n,t|n',t') = k_+^{(n')}(P(n,t|n',t') - P(n,t|n'+1,t')) + k_-^{(n')}P(n,t|n',t') - k_-^{(n')}P(n,t|n'-1,t') \quad (16)$$

In order to calculate the first passage time distribution we define

$$G_m(t) = \sum_{n=0}^{N-1} P(n,t|m,0) = Prob(T \geq t), \quad (17)$$

where T is the time interval in which the enzymes goes form the stretched state into the fully relaxed, hence reactive state, the first passage time. Note, that m is the starting position of the random walker. This means, the actually occupied state is still within the considered time interval. The enzyme has not yet reached its final state. Plugging this into the backward master equation and making use of the time homogeneity of the system (no time dependence of the rates) leads to $P(n,t|m,0) = P(n,0|m,-t)$ we receive

$$-\frac{\partial}{\partial t'}G_m(t) = k_+^{(m)}(G_m(t) - G_{m+1}(t)) + k_-^{(m)}(G_m(t) - G_{m-1}(t)), \quad (18)$$

obeying the boundary conditions $G_m(0) = 1$ for $0 \leq m < N$ and $G_N(0) = 0$. Since m runs from 0 up to N we can rewrite this as a vectorial differential equation

$$-\frac{\partial}{\partial t}\vec{G}(t) = \hat{M}\vec{G}(t). \quad (19)$$

The Matrix $\hat{M}$ reads

$$\begin{pmatrix} k_+^{(0)} & -k_+^{(0)} & 0 & \cdots & 0 \\ -k_-^{(1)} & k_+^{(1)} + k_-^{(1)} & -k_+^{(1)} & 0 & \cdots \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & -k_-^{(N-1)} & -k_+^{(N-1)} + k_-^{(N-1)} & -k_+^{(N-1)} \\ 0 & \cdots & & \cdots & 0 \end{pmatrix}$$

with the initial condition.

$$\vec{G}(0) = (1, 1...., 0)^T. \tag{20}$$

Equation (19) is solved by the following ansatz

$$\vec{G}(t) = e^{-\hat{M}t}\vec{G}(0). \tag{21}$$

Since we are just interested in the first passage time distribution starting from $m = 0$, we merely look at the first component of the solution

$$f(T) = -\frac{\partial}{\partial t}G_0(t)|_{t=T}. \tag{22}$$

The cumulative distribution of the first passage time reads $Prob(T < t) = 1 - Prob(T \geq t) = 1 - G_m(t)$.

In terms of $G_m(t)$ the n-th moment of the first passage time is

$$\begin{aligned} \langle T^n \rangle_m &= -\int_0^\infty t^n \partial_t G_m(t) \\ &= n\int_0^\infty t^{n-1} G_m(t) := nT_n(m) \end{aligned} \tag{23}$$

whereas the second equality is obtained by integration by parts. Now we plug Eq.(23) into Eq.(18) and get

$$\begin{aligned} nT_n(m) &= k_+^{(m)}(T_{n+1}(m) - T_{n+1}(m+1)) \\ &\quad + k_-^{(m)}((T_{n+1}(m) - T_{n+1}(m-1)). \end{aligned} \tag{24}$$

This can be solved quite straightforwardly by a recursive procedure and yields to

$$T_n(m) = \sum_{k=m}^{N-1}\sum_{l=0}^{k} \frac{(n-1)T_{n-1}(l)}{k_+^{(l)}} \prod_{j=l+1}^{k} \frac{k_-^{(j)}}{k_+^{(j)}}. \tag{25}$$

with $n \geq 2$. For the first moment we receive

$$T_n(m) = \sum_{k=m}^{N-1}\sum_{l=0}^{k} \frac{1}{k_+^{(l)}} \prod_{j=l+1}^{k} \frac{k_-^{(j)}}{k_+^{(j)}}. \tag{26}$$

Using this we are, in principal, able to present an analytic expression for any moment. Additionally, we can calculate the full analytic expression of the cumulative distribution of the waiting time, at least up to $N = 4$.

The Fit Routine

By using Matlab we establish a fit routine to extract the values for the three free parameters. The basic idea is as follows, we start with a randomly chosen value and use the built in function fsolve to find a minimum of the parameters space. To ensure the global minimum of the landscape is found we use the so called χ^2 function

$$\chi^2 = \left(\frac{\bar{t} - \langle T \rangle}{\bar{t}}\right)^2 + \left(\frac{s^2 - \sigma_T^2}{s^2}\right)^2 + \left(\frac{sk_{sample} - sk_T}{sk_{sample}}\right)^2. \tag{27}$$

Note, the χ^2 function depends on the three parameters k_-,k_+, and N. Such a procedure can be repeated several times with different starting points. Different starting points for the free parameters provide a broader scanning of the landscape of the parameter space. Keeping in mind the non linearity of the equations there is no a priori guarantee to run into the global minimum with certainty. Therefore we use the χ^2 function to find the global minimum, which is achieved by minimizing χ^2 with respect to k_-, k_+, and N.

[1] J. E. Bjorkholm A. Ashkin, J. M. Dziedzic and Steven Chu. Observation of a single-beam gradient force optical trap for dielectric particles.

[2] V. Barsebov and D. Thirumalai. Probing protein-protein interactions by dynamic force correlation spectroscopy. *Phys. Rev. Lett.*, 95(16):168302, October 2005.

[3] Quate C.F. Binning, G. and C.H. Gerber. Atomic force microscopy.

[4] H. Gumpp; E. M. Puchner; J. L. Zimmermann; U. Gerland; H. Gaub; K. Blank. Triggering enzymatic activity with force. *Nano Lett.*, 9(9):3290–3295, 2009.

[5] Hofkens J. Blank K, De Cremer G. Fluorescence-based analysis of enzymes at the single-molecule level.

[6] M.J. Schnitzer; S.M. Block. Statistical kinetics of processive enzymes. *Cold Spring Harb. Symp. Quant.Biol*, 60:793–802, 1995.

[7] Miklos S. Z. Kellermayer; Steven B. Smith; Henk L. Granzier; Carlos Bustamante. Folding-unfolding transitions in single titin molecules characterized with laser tweezers. *Science*, 276:1112–1116, 1997.

[8] Hendrik Dietz and Matthias Rief. Exploring the energy landscape of gfp by single-molecule mechanical experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 101(46):1619216197, 2004.

[9] Weisenhorn AL Gould SA Albrecht TR Quate CF Cannell DS Hansma HG Hansma PK. Drake B, Prater CB. Imaging crystals, polymers, and processes in water with the atomic force microscope.

[10] Olga K. Dudko, Gerhard Hummer, and Attila Szabo. Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96(10):108101, Mar 2006.

[11] Torsten Strunz Dario Anselmetti Hans-Joachim Güntherodt Annemarie Honegger Lutz Jermutus Louis Tiefenauer Falk Schwesinger, Robert Ros and Andreas Plückthun. Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates.

[12] Pak-Wing Fok and Tom Chou. Reconstruction of bond energy profiles from multiple first passage time distributions. *Proc. Royal Soc. A*, 466:13479–3499, 2010.

[13] Wolynes PG. Frauenfelder H, Sligar SG. The energy landscapes and motions of proteins.

[14] C.W. Gardiner. *Handbook of Stochastic Processess for physics, chemistry and natural science, 4th ed.* Springer Verlag, 2009.

[15] Matthias Rief; Julio M. Fernandez; Hermann E. Gaub. Elastically coupled two-level systems as a model for biopolymer extensibility. *Physical Review Letters*, 81(31):47644767, 1998.

[16] Matthias Rief; Mathias Gautel; Filipp Oesterhelt; Julio M. Fernandez; Hermann E. Gaub. Reversible unfolding of individual titin immunoglobulin domains by afm. *Science*, 276:1109–1112, 1997.

[17] Mukulesh Baruah Michel Sliwa Bert F. Sels Johan Hofkens Gert De Cremer, Maarten B. J. Roeffaers and Dirk E. De Vos. Dynamic disorder and stepwise deactivation in a chymotrypsin catalyzed hydrolysis reaction.

[18] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry B*, 81(25):2340–2361, 1977.

[19] X. Sunney Xie H. Peter Lu, Luying Xun. Single-moleclue enzymatic dynamics.

[20] Manfred Radmacher; Monika Fritz; Helen G. Hansma; Paul K. Hansma. Direct observation of enzyme activity with the atomic force microscope. *Science*, 265:1577–1579, 1994.

[21] Berthold Heymann and Helmut Grubmü ller. Dynamic force spectroscopy of molecular adhesion bonds. *Phys. Rev. Lett.*, 84(26):6126, June 2010.

[22] Olga K. Dudko; Gerhard Hummer and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *Proc. Natl. Acad. Sci. U.S.A.*, 105(41):1575515760, 2008.

[23] Steven B. Smith Ignacio Tinoco Jr. Carlos Bustamante Jan Liphardt, Bibiana Onoa. Reversible unfolding of single rna molecules by mechanical force.

[24] Matthias Rief Jan Philipp Junker, Fabian Ziegler. Ligand-dependent equilibrium fluctuations of single calmodulin molecules.

[25] Justin L. MacCallum Ken A. Dill. The protein-folding problem, 50 years on.

[26] Leung A Ritchie K Evans E. Merkel R, Nassoy P. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy.

[27] Michael Schlierf and Matthias Rief. Single-molecule unfolding force distributions reveal a funnel-shaped energy landscape. *Biophysical Journal: Biophysical Letters*, 90(4):L33–L35, 2006.

[28] A.; Stuart and J. K Ord. *Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory, 6th ed.* New York: Oxford University Press, 1998.

[29] Yohichi Suzuki and Olga K. Dudko. Single-molecule rupture dynamics on multidimensional landscapes. *Phys. Rev. Lett.*, 104(4):048101, Jan 2010.

[30] Hui Lu; Barry Isralewitz; Andre Krammer; Viola Vogel and Klaus Schulten. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophysical Journal*, 75:662671, 1998.

[31] Samuel Wilks. *Mathematical Statistics.* John Wiley and Sons, Inc. New York London, 1962.

[32] Hazen P. Babcock Rick Russell Taekjip Ha Daniel Herschlag Steven Chu Xiaowei Zhuang, Laura E. Bartley. A single-molecule study of rna catalysis and folding.

[33] Rief M. Zoldk G. Force as a single molecule probe of multidimensional protein energy landscapes.

2.6 Conclusion

Here, we have shown that the activity of enzymes strongly depends on the specific conformation of the enzyme. We considered the type of pulling experiment performed in [81]. In such experiments one end of the enzyme is immobilized to a surface, while the other end is bound via a controlled antibody-antigen interaction to the AFM. The AFM exerts a certain force on the enzyme changing its configuration until the antibody-antigen interaction ruptures initializing the starting time. Out of equilibrium the enzyme folds back towards its natural stable state regaining its activity, which is measured by monitoring the fluorescent reaction. This procedure is easily repeated satisfying the same conditions, which leads to a set of time series, from which the moments can be calculated. Motivated by the pulling experiment we refined the proposed random walker model and gave it a more rigorous footing. To this end we calculated the first passage time moments of our model depending on the kinetic parameters and fitted them to the experiments. Thus, the kinetic and thermodynamic parameters could be obtained by solving an inverse stochastic problem. The number of performed experiments determines the accuracy of the moments and hence leads to an error for the free parameters of the model. We calculated the validity of the free parameters of the model depending on the quality of the experiments preconditioned that the energy landscape showed only one minimum. Increasing the bumpiness of the energy landscape gradually, causes the fit routine to produce kinetic parameters with large errors. However, it was still possible to provide an upper bound for the maximal number of foldings states along the pathway and more importantly for the total free energy difference between the folded and unfolded state

Chapter 3

Spatial Arrangement of Enzymes

The cell was long consider as a structureless "bag", however, gradually this picture changed. Nowadays we know that the cell is highly structured. Separate phases such as organelles in particular the mitochondrion play an important role for the spatially confined functionality of the cell [91], see Fig.(3.2). Compartmentation of enzyme activities into multi-enzyme complexes [88, 89, 90] localize the active centers of the cell to very specific positions. A deeper understanding of the highly complex, intertwined metabolic pathways of the cell is still needed.

In the previous sections we have seen the main features of a single enzyme substrate kinetics. In living organisms, however, such a single enzyme substrate reaction is not an isolated system. The enzymes are embedded into a larger reaction scheme involving several enzymes. In a physiological context the arrangement of enzymes into a cascade is called metabolism. Each individual metabolic path is a life sustaining arrangement of enzymes in order to divide, grow and reproduce the cell. The product of one enzymes is the substrate of another enzyme, forming a cascade with multiple different enzymes arranged in series. In a sequence of steps, catalyzed by several different enzymes, the substrate of the first enzyme gets transformed into the product of the last enzyme of the cascade, where the ration of the initial substrate and the final product defines the efficiency of the pathway. In principle two essentially different metabolism are distinguished, the anabolic metabolism and the catabolic metabolism. In the anabolic metabolism larger chemical molecules are constructed from smaller units consuming energy, as it is seen in muscle tissues. The catabolic metabolism, however, breaks down lager molecules into subunits, while releasing energy, such as in glycolysis [6]. Metabolic paths consist of smaller enzymatic subunits. There are two essentially different enzyme reaction cascades: The static complex, which is often refereed to as metabolite channeling, and the dynamic complex. In the first case the enzymes form a fixed complex where the intermediates are directly transferred between the active site of the different enzymes. The static complex is a heterogenous enzyme complex, well known examples are the fatty acid complex, a multi-enzyme catalyzing acetyl-CoA to malonyl-CoA [93] and celluosome. The latter is a multi-enzyme complex with 11 different enzymes arranged on a scaffold breaking cellulose into sugar [94]. While the fatty acid is a intracellular compound the celluosome complex is assembled on the outer cell membrane

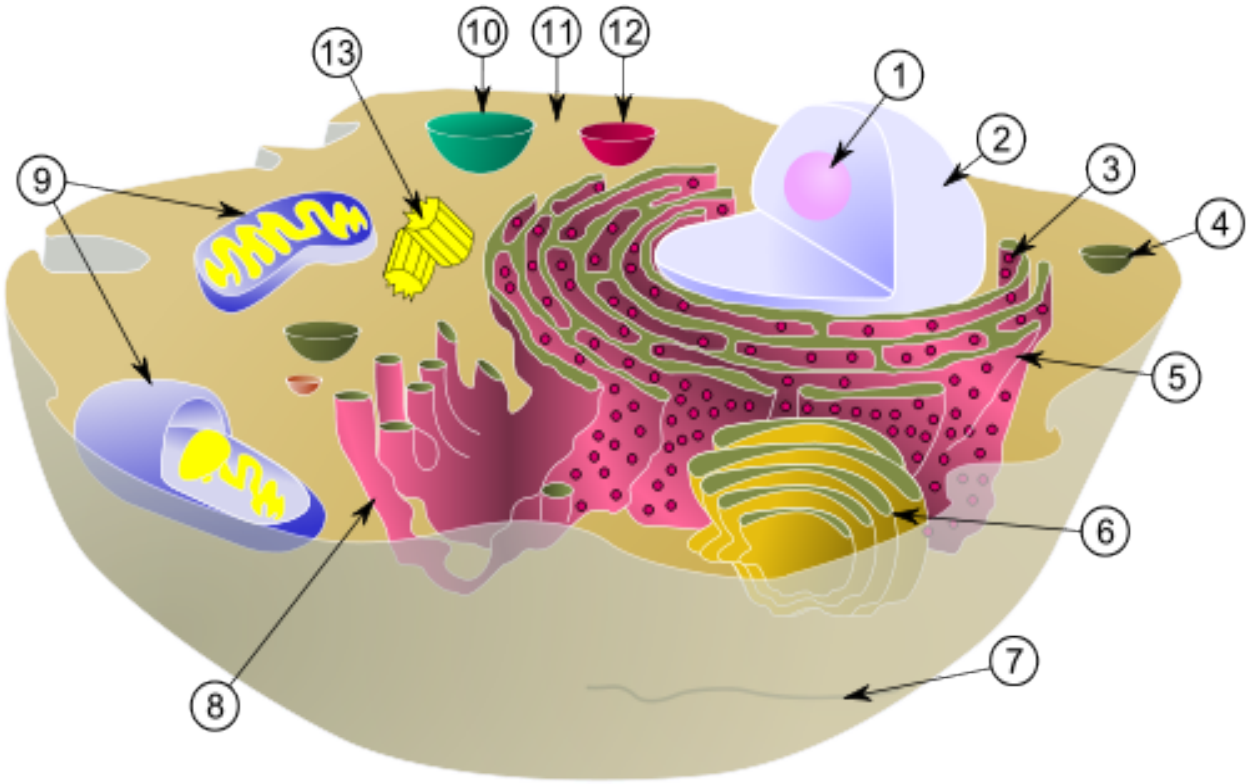

Figure 3.1: A schematic picture of cell is shown demonstrating the different micro-compartments. This is only a small selection of the cells interior and far from complete. 1=Nucleolus , 2=Nucleus, 3=Ribosome, 4=Vesicle, 5=Rough endoplasmic reticulum, 6=Golgi apparatus, 7=Cytoskeleton, 8=Smooth endoplasmic reticulum, 9=Mitochondrion, 10=Vacuole, 11=Cytosol, 12=Lysosome and 13=Centriole. This picture is taken form [92]

of anaerobic bacteria. Dynamic enzyme complexes exhibit a reversible complex formation. The first step of the glucose metabolism shows such a dynamic translocation, which phosphorylates glucose into glucose-6-P [95, 96, 97]. The isoform HKII of the hexokinase reversibly binds to the membrane of the mitochondria and the cytosol. The ratio of clustered and free enzymes is regulated by glucose-6-P and GSK3 [98]. In the next section we discuss the advantage of enzyme clustering.

3.1 Advantage of Enzyme Clustering

In the last section we have seen a few examples of enzyme clustering. Here, we ask the question, whether forming an enzyme-complex comprises any advantage over randomly distributed enzymes of the same type. So why investing energy or evolutionary effort in

forming an enzyme structure? In the following we want to investigate the advantage not only of substrate channeling but also the advantage of clustering enzymes when the intermediates are allowed to freely diffuse between the active sites. Since in this thesis we are particularly interested in the scenario of diffusive intermediates we will stress the advantage for that scenario when we itemize the advantages stated by Heinrich and coworkers[99] and Ovadi [100] the later considered the advantage from a more physiological perspective.

1. *Stabilization and preservation of intermediates.*
   The intermediates emerging from a reaction are stabilized against reaction with water when channeled from one active site of the enzyme complex to the next one. In the case of toxic intermediates the organism is protected against intoxicating itself. Additionally, diffusion between the active site may lead to a loss of intermediates, which do not participate in the reaction anymore. For the case of free diffusing intermediates no such stabilization exists since they are released into the bulk. One of the main goals is the reduction of the overall intermediate concentrations, preventing an undesired loss of these intermediates. This can be achieved by increasing the intermediate concentration locally, in the vicinity of the active site of the enzyme.

2. *Flux enhancement*
   In the channeling mechanism flux enhancement may be achieved by reducing the transient time of the intermediates between the active sites. With flux it is meant the production rate of product particles of the final enzyme in the cascade. However, the more likely fact is that the time for solvation and disovlation for the substrate, product and the active site of the enzyme is saved in each channelling step. The enhancement of the flux for the diffusive intermediates will be our major concern. We will ask the question, does the flux depend on the spatial arrangement of enzymes under the condition of freely diffusing intermediates?

3. *Stereospecific transfer*
   When the intermediates get channeled directly their steric conformation can be preserved which make them even more adjusted for the reaction with the next enzyme. We will not address this question, since the intermediates are released into the bulk and it is not clear whether a specific steric pre-arranged intermediate is sustained during the diffusion process

4. *Spatial organization*
   Due to spatial organization either in static complexes or when considering diffusive intermediates the products can be directed towards definite places in the cell. The micro-compartmentation of enzymes confines the product to a sub-part of the cell preventing overcrowding.

5. *Maintenance of low intermediate concentration*
   In the case of substrate channelling the concentration of intermediates is reduced to a minimum. Once the intermediate is allowed to diffuse freely in the surrounding solution such an optimal intermediate concentration is no longer be guaranteed.

Intermediates are able to diffuse away and thus are lost for further reaction. For the metabolic path to work in the most efficient way, in the sense that most initial substrates are converted into final product particles, it is important to minimize the loss of any intermediates. The existence of enzyme complexes in the cell reduces the overall intracellular concentration of metabolites. Thus, the cell is prevented form being overcrowded, since one type of intermediate has just one use for a particular pathway.

6. *Reduction of transient time*
   Substrate channeling in multi-enzyme complexes reduces the transient time of the intermediates between the active sites of the complex. This leads to an increase in the overall reaction velocity of the complex.

7. *Coordinate regulation*
   A static enzyme-complexe not only stabilizes the intermediates but the respective enzymes themselves. By forming a complex, the ligandation of an enzymes at one site of the complex prevents the other enzymes, from inhibitor binding, due to changes in its conformation.

Generally, these listed points can be divided into two classes, one considering the structural advantage of the bio-chemicals such as preservation of their conformation, and the other is their arrangement in space. These advantages are not free of debate [101, 102, 103, 104]. In this thesis we limit ourselves to discuss the case of freely diffusing intermediates. However, an astonishing example of direct substrate channeling is the Tryptophan synthase, which catalyzes the final two reactions in the synthesis of L-tryptophan. The α-subunit catalyzes indole-3-glycerol phosphate to indole and the β-subunit uses the indole and combined with L-serine it forms L-tryptophan. It was shown for *Salmonella typhimurium* that these two subunits form a multi-enzyme complex $(\alpha\beta)_2$, where the active sites of the two subunits are at a distance of $25\mathring{A}$. Remarkably, the α and β site are connected by a hydrophobic intramolecular tunnel channeling the indole without releasing it into bulk [105, 106, 107]. In [108] a modular protein-scaffold for a three enzyme cascade was designed, forming an artificial multi-enzyme complex. Depending on the architecture of the scaffold a flux enhancement of up to 77-fold was achieved. This demonstrates the advantage of channeling intermediates through an assembly line of consecutive enzyme activities. A related topic that should be mentioned at this point is biochemical signaling [109, 110, 7] as it can be found in mitogen-activated protein kinase [111, 112] or Ras-proteins [113, 114]. Proteins form clusters at the cell membrane to amplify the signal transduction compared to freely diffusive proteins on the cell membrane [115, 116]

Here, we will focus on the advantages of an enzyme cascade with diffusing substrates. To this end we will ask the question how the spatial arrangement of enzymes does influence the reaction flux and is there a way to organize enzymes such that the active center of the respective enzyme is placed where the local concentration of the corresponding intermediates is highest. In other words we will consider exclusively the second class of possible

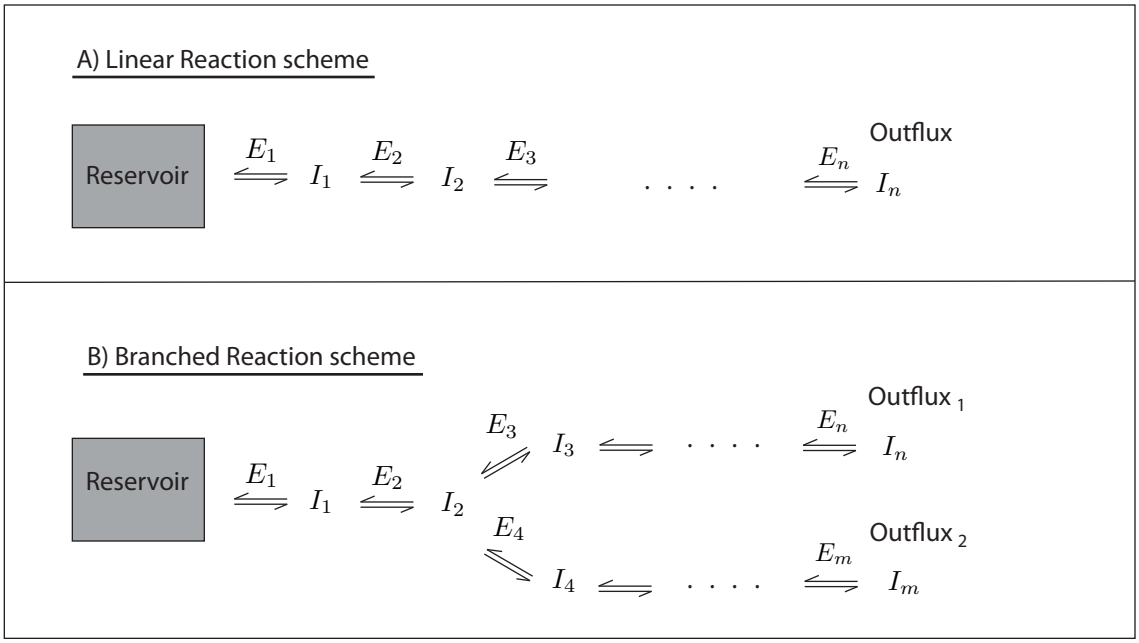

Figure 3.2: Two showcase examples of a reaction scheme are shown. A) displays a linear reaction scheme coupled to a reservoir supplying the pathway with an initial substrate. The reservoir can either contain a limited number of particles or it gives an infinite supply. B) the second pathway shows a single branched reaction pathway with two different outfluxes. They can either represent two different particle species or the same species in two different micro-compartments.

advantage of enzyme clustering. In our model we will neglect any spatial extension of the individual chemicals, both enzymes and intermediates. A step towards a more comprehensive model must include the individual structure of enzymes, but for a more detailed description of an extension of our model see the outlook. At the time the aforementioned advantages had been stated they were mostly driven by theoretical work [117, 118, 119, 120] concerning the transient time.

3.2 Metabolic Control Analysis

This section discusses the basic idea of the existing theory of enzyme kinetics in metabolic pathways. Since spatial arrangement were neglected in this description we add this section just for the sake of completeness.
We have seen that the MM-mechanism describes the kinetics of a single enzyme, but when we want to describe the kinetics of metabolism another of description is needed. We cannot obtain a full understanding of the physiological role of enzymes in a metabolism when the rest of the pathway is suppressed. The behavior of an enzyme can be entirely

different when it is disconnected from the physiological surrounding. The precise structure of the pathway, whether it is linear or branched, is not important, see Fig(3.2). The pathway can even be multiple times branched with feedback loops, positive or negative, or induced by inhibitors. In the past four decades a theory on the behavior of a metabolic pathway depending on the components has been developed. Metabolic control analysis is the theory of steady states of enzymes catalyzing several substrates, where at least one of the metabolites is connected to an external reservoir. In other words, metabolic control analysis describes control coefficients, which measures the relative change of steady state quantities, such as the flux (J) or the concentration of the intermediates (I) to a change in the kinetic parameters such as the enzyme activities. In contrast, we are interested in the underlying physical reasons governing the arrangement of enzymes to maximize the efficiency of the pathway, thus we will not make any use of the metabolic control analysis. For the interested reader, extensive reviews can be found in [121, 122, 123]. So far we have not specified any type of model. But before we consider such a model we will introduces the techniques with which our model and its prediction can be tested.

3.3 Artificial Enzyme Cascades

The importance of metabolism in the cell cannot be stressed enough. Understanding the basic principles of these mechanisms is one of the main goals of modern bio-science. A key step is to engineer an artificial enzymatic cascade and investigating its kinetic behavior in a controlled environment. The design of optimized bio-chemical enzyme cascades generating products such as therapeutical molecules (anti cancer drugs) is of great interest. Especially, when the availability of the initial substrates are limited and a high efficiency is desired, while traditional synthetic routes fail. As we have seen before, enzyme cascades bear a number of advantages, both concerning the structural composition of the enzymes and their arrangement in space. While the advantage in multi-enzyme complexes with direct substrate channeling are rather clear, the fact whether spatially arranged enzyme cascades with diffusing intermediates bear any advantage over well stirred cascades is far from being resolved. Altering the structure is the domain of classical biochemistry, by comparing the activity of wild type enzymes with engineered mutations. Creating mutations of biomolecules in a lab is an efficient way to imitate natural evolution. We have addressed that question on the structural features of enzymes in the second chapter of this thesis.

3.3.1 DNA-Scaffolds

A relatively young field is the spatial arrangement of enzymes. Since the advent of single molecule spectroscopy techniques it became possible to control and monitor enzyme reactions on a single molecule level. While several attempts of exact enzyme positioning have been made in the past [124, 125, 126], progress came with the combination of two major techniques. The combination of DNA-assembly and single molecule spectroscopy made exact positing of enzymes possible. The idea of using DNA to construct immobile

junction networks dates back to the 80's [127]. Thereby, it was made use of the double helix formation of complementary DNA strands proposed by Watson and Crick [128] . Instead of forming a linear strand, networks can be created by utilizing the specificity of the double helical base pairing two different DNA strands which have complementary "sticky ends" can be linked together. The basic idea is to put together smaller oligonucleotides to accomplish a greater structure. However, a major improvement in DNA-nanotechnology was the discovery of the DNA origami-method by Rothemund [129]. The main difference to the conventional method was instead of former larger structure from smaller compounds large DNA strands were formed into different structure. In the following we give a short description of the basic concepts of DNA-origami, for a deeper studying of that matter the reader is referred to the original paper by Rothemund. This method uses a single strand of viral DNA and several complementary oligonucleotides, the so called staple strands. These staple strands are used to form the long viral DNA strand, due to complementary base pair binding, into to the desired shape. Every shape has its own different set of single strands. The staple strands and the long DNA-origami strand are put together, heated, and then cooled down. While cooling the staple strands bind to the longer strand forming cross-linkers and shaping it into the desired form. This process is called self-assembly, because after adding the oligonucleotides and the long viral DNA strand the folding of the DNA-origami runs without external intervention. Using DNA-origami assembling techniques nanoscale objects with nanometer precision can be constructed. Almost any shape in one, two and three dimensions can be achieved by this technique [130], even an analog of the chinese map [131]. One of the main advantage is an explicit knowledge of the exact base sequence of the DNA-scaffold. This allows for an accurate binding of the complementary sequence at a precise locations. With the help of the "Cut and Paste" method functional units can be placed at specific positions on the DNA-origami scaffold. To this end the functional unit [1] is connected with an DNA-oligomer which is picked up by the AFM form a storage site. The cantilever of the AFM is armed with a matching DNA single strand complementary to the one at the functional unit. Thus the enzyme can be transported to its final position where its second "free" single DNA strand hybridizes with the DNA strand at the target site [132]. Thus, in principle any artificial enzyme cascade can be constructed. The combination of DNA-origami and AFM establishes a way of screening the surface of the scaffold verifying whether the functional unit has bound to its destined place [133]. Further it is possible to assemble several molecules to create a composite macromolecule with controlled spacing between them [134]. This leads to a possible realization of enzymatic assembly lines. So called nano-factories could combine several aspects of physiochemical process, to create a novel and efficient way to synthetically produce biofuels [135], as we have seen for the cellulosome.

[1] For our purposes we will consider enzymes, but there is no restriction on that.

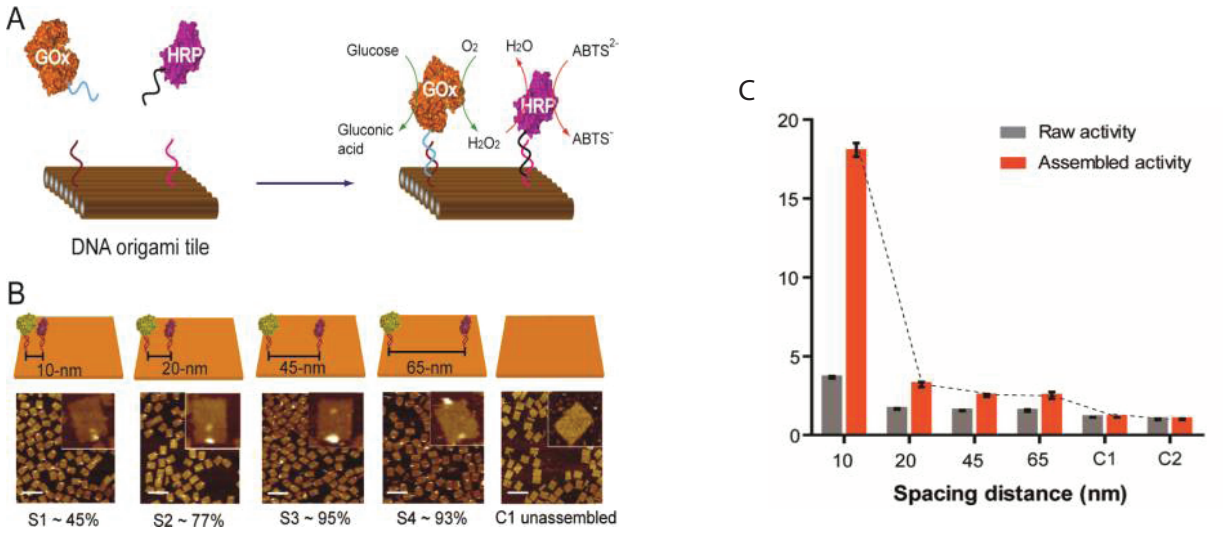

Figure 3.3: The experiment conducted by Fu and coworkers. A) illustrates the experiment with the two dimensional DNA-origami scaffold and the two enzymes, glucose oxidase (GOx) and horseradish peroxidase (HRP), making up the cascade. B) visualizes the different distances between the enzymes using AFM imaging. C) shows the enhancement of the cascaded activity induced by the distance between the enzymes compared to the raw activity. Note that C1 is the scaffold control and C2 the free enzymes. Reprinted with permission form [136]

3.3.2 The Design of Artificial Enzyme Cascades

In the following we give a few examples of spatially arranged enzymes and how the distance between them influences their reaction activity. Such an enzymatic arrangement is one possible way to have a post-translational alteration of the enzyme activities. One of the first attempts was the DNA-directed assembly of the two immobilized enzymes, NAD(P)H:FMN Oxidoreductase and Luciferase, catalyzing a sequential reaction. Bringing the two enzymes closer together enhances the cascade activity by more than twice, suggesting that spatial proximity of the two enzymes is beneficial for the bienzymic system [137]. Another experiment was performed by Willner *et.al.* [138] in which they gained greater control of the enzyme positioning of glucose oxidase (GOx) and horseradish peroxidase (HRP) by forming a hexagon DNA strip assembly including hinges to which two enzymes are tethered. The topology of the DNA scaffold pre-determines the position of the enzymes, facilitating a spatial control over the enzyme cascade. Again an enhancement of the cascade activity depending on the spatial separation was observed. In [136] an improvement over the spatial parameters due to the use of DNA origami was achieved. The cascade consisting of GOx and HRP was placed on a two dimensional origami scaffold, see Fig(3.3). A five fold increase of the activity at 20nm distance for the two enzymes compared to the free enzymes, was observed. This points towards an enhancement of the cascade activity by a

more effective transfer of the immediate H_2O_2. However an even greater enhancement, up to 20 fold was achieved at a distance of 10 nm. While three dimensional brownian diffusion does not allow for such an enhancement, the authors suggested that dimensionally limited diffusion caused the drastic increase of cascade activity. Similar to the linear diffusion for transcription factors along DNA [139]. However, their speculations are open for debate initiating further experimental research in that field. More extensive reviews on dependence of the cascade activity of spatially separated enzymes in a consecutive reaction based on a DNA-scaffold can be found in [140, 141, 135]. DNA-scaffolds are so far limited to *in vitro* assembling. However, it was shown in [142] that constructing *in vivo* RNA-scaffolds lead to a major increase of hydrogen production in cells, depending on the architecture of the scaffold, compared to the unscaffolded production.

Chapter 4

Optimal Arrangement of Enzymes

Up to now we have we have mainly discussed the physiological importance of metabolic pathways and its recent in vitro experimental investigations. With the dawn of single molecule microscopy and the aforementioned cut and paste method it became possible to design artificial pathways. Explicit manipulation of the distribution of the enzymes and the monitoring of their turnover rates is now possible.

4.1 A Two Enzyme Pathway

Here, we want to make a first step towards understanding spatially arranged enzymatic pathways. We focus entirely on the influence of the spatial arrangement and do not consider complex branched enzymatic pathways. To this end we take the simplest possible pathway consisting of two different types of enzymes E_1 and E_2. This system as we will show in this chapter already reveals a rich physical diversity. The overall reaction cascade is irreversible, thus the product of the first enzyme works exclusively as an substrate for the second one and cannot react back to the educt S. Both types of enzymes act according to the Michaelis-Menten kinetics leading to the following reaction scheme

$$\mathrm{E}_1 + \mathrm{S} \underset{\mathrm{k}_\mathrm{r}^{(1)}}{\overset{\mathrm{k}_\mathrm{f}^{(1)}}{\rightleftharpoons}} ES \xrightarrow{\mathrm{k}_\mathrm{cat}^{(1)}} \mathrm{E}_1 + \mathrm{I} \tag{4.1}$$

and

$$\mathrm{E}_2 + \mathrm{I} \underset{\mathrm{k}_\mathrm{r}^{(2)}}{\overset{\mathrm{k}_\mathrm{f}^{(2)}}{\rightleftharpoons}} EI \xrightarrow{\mathrm{k}_\mathrm{cat}^{(2)}} \mathrm{E}_2 + \mathrm{P}. \tag{4.2}$$

Let us assume all E_1 molecules together working as a source of I molecules, producing a constant influx J_1, independently of their individual positions. This is achieved for example by a constant supply of substrates. In this case only the dynamics, and thus the rates, of the second reaction are explicitly relevant and consequently we drop the superscript of the rates, hence $k_f^{(2)} = k_f$, $k_r^{(2)} = k_r$ and $k_{cat}^{(2)} = k_{cat}$. We ask the question, how to distribute the second enzyme E_2 relative to enzymes E_1 in order to generate as many products P

as possible? Both enzyme types have a fixed position in space, however the intermediates I move freely within a confined, finite system according to simple diffusion kinetics, with the coefficient D. Additionally, we allow the intermediates to form a waste product Q. We assume that this consumption of I molecules can happen anywhere in the system and is not influenced by the distribution of the second enzymes. Therefore this competing pathway can be modeled as a first order reaction with constant reaction rate σ

$$\mathrm{I} \xrightarrow{\sigma} \mathrm{Q}. \tag{4.3}$$

Enzymes are a complex assembly of molecules. Due to their molecular composition only parts of the entire enzymes are active. The reaction with the substrate is highly specific and works according to the so called key lock mechanism. For instance we have seen in chapter 2, that a change in the environment of an enzyme, in particular when it is exposed to external mechanical force, alters the conformation of the enzymes and as a result their activity. Thermal fluctuations also affect the enzymes and influences the evolutionary highly adapted substrate-enzyme reaction. Here we focus only on the impact of the spatial arrangement of a certain pathway and we will neglect the explicit structure of the participating chemicals, Fig(4.1). Although, the particle numbers of the enzymes and their corresponding intermediates are low, demanding a stochastic consideration, for reasons of simplicity we regard a mean field type approach. Accordingly, we assume the enzymes E_2 and the intermediates I to be point like with zero extension, meaning that we can write them in terms of densities. Our approach models the pathway as a reaction diffusion system

$$\frac{\partial \rho(\mathbf{r},t)}{\partial t} = D\nabla^2\rho(\mathbf{r},t) - \frac{k_{\mathrm{cat}} e(\mathbf{r})\rho(\mathbf{r},t)}{K_M + \rho(\mathbf{r},t)} - \sigma\rho(\mathbf{r},t), \tag{4.4}$$

where $\rho(\mathbf{r},t)$ is the density of the intermediates, $e(\mathbf{r})$ the distribution of E_2 enzymes and K_M is the MM-constant. Note that there is no time dependence in the enzyme distribution, since we presume that the enzymes equilibrate on a much faster time scale than the intermediates. Generally, there are several standard approaches to solve such models, see [143, 144]. Productions of intermediates by the first enzyme will be embedded into the boundary condition and will take an explicit form once we regard a particular model. We are interested in the efficiency of the system which is the ratio of the steady state creation of products relative to the incoming intermediates governed by J_1. We only consider the steady state distribution of the intermediates, hence $\rho(\mathbf{r},t) \to \rho(\mathbf{r})$. But before we consider any specific system we perform the following transformations

$$\mathbf{r} \to \frac{\mathbf{r}'}{R} \qquad \text{and} \qquad \rho(\mathbf{r}) \to \frac{\rho'(\mathbf{r}')J_1 R}{D} \tag{4.5}$$

where R is the system size and the transformation of the intermediate density assures that the production rate is normalized to 1. The reaction diffusion equation (4.4) can be rearranged to a dimensionless equation of the form

$$0 = \nabla^2\rho'(\mathbf{r}') - \frac{\alpha e'(\mathbf{r}')\rho'(\mathbf{r}')}{1+\gamma\rho'(\mathbf{r}')} - \beta\rho'(\mathbf{r}'). \tag{4.6}$$

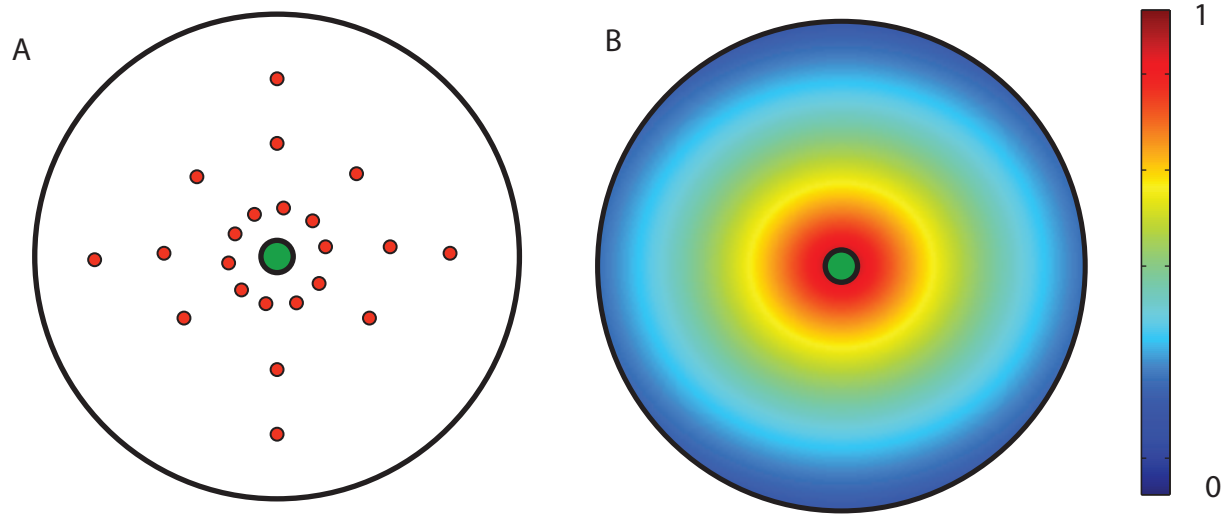

Figure 4.1: A two dimensional representation of the model is depicted. A) the enzymes E_1 (green) are placed in the center of the circle giving rise to a production of intermediates (not shown). The enzymes E_2 are placed around the center converting the intermediates to product particles generating the reaction flux. In B) a course grained version of the enzymes E_2 is shown. We do not regard the enzymes E_2 as discrete particles with a finite size, we rather approximate their distribution by the continuous function $e(\mathbf{r})$. The right legend shows the value of the enzymes density $e(\mathbf{r})$ at a specific point in space.

In which we have rescaled the enzyme distribution, $e(\mathbf{r}) = \frac{e'(\mathbf{r})}{\bar{e}}$, with the mean enzyme density $\bar{e} = V^{-1}\int_V d^3\mathbf{r}e(\mathbf{r})$, where V is the volume of the system. The first dimensionless parameter $\alpha = \frac{\tau_D}{\tau_R}$, with $\tau_D = \frac{R^2}{D}$ the timescale for a particle to diffuse to the end of the system and the reaction times scale $\tau_R = \frac{K_M}{k_{cat}\bar{e}}$ denoting the mean time it takes for an intermediate to react. We call the regime $\alpha < 1$, where the reaction time is larger than the diffusion time, diffusion dominated, likewise, the regime $\alpha > 1$ is called reaction dominated. However, in the presence of a competing pathway the system is also determined by the timescales set due to the rate of the second pathway. The second dimensionless parameter $\beta = \frac{\tau_D}{\tau_L}$, with $\tau_L = \sigma^{-1}$ the reaction timescale for the second undesired pathway. These two dimensionless parameters capture the relative timescales of reaction with E_2 or the competing pathway compared to the diffusion of the intermediates towards the end of the system. The third parameter $\gamma = \frac{J_1 R}{K_M D}$ controls the influx relatively to the saturation of the enzymes E_2. For the rest of the work we drop the prime notation and we think this will not cause any confusion. By integrating Eq.(4.6) we get

$$0 = \int_V d^3\mathbf{r}\nabla(\nabla\rho(\mathbf{r})) - \int_V d^3\mathbf{r}\frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} - \int_V d^3\mathbf{r}\beta\rho'(\mathbf{r}). \tag{4.7}$$

The first term can be rewritten as

$$\int d^3\mathbf{r}\nabla(\nabla\rho(\mathbf{r})) = -\left|\nabla\rho(\mathbf{r})|_{B_{inner}}\right| + \left|\nabla\rho(\mathbf{r})|_{B_{out}}\right| = -\frac{J_{out}}{J_1} + 1 \tag{4.8}$$

which is the difference of the normalized incoming flux due to the enzymes E_1 and the loss current through the boundary. The inner boundary B_{inner} is the boundary formed by the enzymes E_1, hence the current at the inner boundary is given by J_1. The outer boundary B_{out} denotes the boundary due to the finite system size. In the case of purely reflecting boundaries the corresponding loss current is equal to one. In the absorbing case the current lost through the boundary is given by J_{out}. We wrote the absolute values, since both currents are negative because of the decreasing density gradient form the production center to the outer absorbing boundary. The reaction current is

$$\frac{J_2}{J_1} = \int_V d^3\mathbf{r} \frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \tag{4.9}$$

and the second part of the loss current originated by the competing pathway reads

$$\frac{J_{\text{leak}}}{J_1} = \int_V d^3\mathbf{r}\beta\rho(\mathbf{r}). \tag{4.10}$$

By combining all the currents we receive the flux conservation

$$1 = \frac{J_2}{J_1} + \underbrace{\frac{J_{\text{leak}}}{J_1} + \frac{J_{out}}{J_1}}_{J_{\text{loss}}/J_1}, \tag{4.11}$$

with $\frac{J_2}{J_1}$ denoting the efficiency of the pathway, which is the fraction of the intermediate particles that are converted into product particles. Note this definition of the efficiency only holds true if the arrangement of the enzymes E_2 do not change the production rate of the E_1 cluster. A possible scenario could be that the enzymes E_2 cluster so tightly around the E_1 cluster that steric effects would hinder the substrate particles Eq.(4.1) to diffuse to the cluster. Therefore, it is not possible to generate intermediates in the first place, cutting of the supply for the enzymes E_2. The second and third part on the right hand side of Eq.(4.11) represents the overall loss current which consists of two parts the loss of intermediate particles through the boundary and the loss by the competing pathway. In a confined system, where e.g. the membrane is impermeable, the outer boundary becomes reflecting hence the current through that boundary is zero, thus the leakage current is the only remaining current. Obviously, in the absence of any loss current particle conservation demands the efficiency of that particular pathway to be one.

We will investigate different enzyme arrangements and their impact on the efficiency of the pathway. To this end, we will firstly ask the question whether clustering the enzymes E_2 around E_1 bears any advantage over distributing E_2 uniformly. As a second step, and this is at the moment of greater interest for synthetic biology, we will discuss the possibility of an optimal enzyme arrangement of E_2 that is neither a cluster nor an uniform distribution, which maximizes the pathway's efficiency. But before discussing a particular system we will say a few words about the features of the MM-reaction.

4.1.1 Limits of the MM-reaction

The MM-reaction as it is shown in Eq.(4.7) is a non-linear reaction and there is little hope to find an exact analytic solution to this differential equation. However, we can consider two different limits, the low density limit and the high density limit, respectively. To this end we take the limit $\gamma \gg 1$ of the reaction term

$$\frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \to \frac{\alpha}{\gamma} e(\mathbf{r}) \tag{4.12}$$

corresponding to the hight density limit. Recall, the parameter $\gamma = \frac{J_1 R}{K_M D}$, thus the limit can essentially be rewritten as $J_1 \gg K_M \frac{D}{R}$. This means that the incoming flux generated by the enzymes E_1 is larger than the consumption of these intermediates by the second enzymes E_2. As a consequence the system fills up with intermediates, hence every enzyme of the second type works at full capacity, regardless of its actual position. The reaction current becomes independent of the enzyme arrangement

$$\frac{J_2}{J_1} = \int_V d^3\mathbf{r} \frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \to_{\gamma \gg 1} \frac{\alpha}{\gamma}, \tag{4.13}$$

since we normalized the total number of enzymes to be one. Note, that this statement only holds true once the strength of the competing pathway is low enough not to significantly influence the intermediate density. The second limit $\gamma \ll 1$ leads to a diluted system, in which the consumption of intermediates is several times higher than their production by the first enzyme E_1. The reaction term becomes linear in that limit

$$\frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \to \alpha\rho(\mathbf{r})e(\mathbf{r}). \tag{4.14}$$

In the manuscript on optimization of collective enzyme activity via spatial localization we will talk about the full-non-linear case and it will be shown that the relevant phenomenology for the optimal enzyme profile is mainly covered by the linear case.

4.1.2 Dimensionality Reduction

Here we consider a rotationally symmetric enzyme distribution $e(\mathbf{r}) = e(|\mathbf{r}| = r)$ and the enzymes E_1 are clustered in the center of our system, setting the inner boundary of the system. Further we assume that the outer boundary of the system shows no irregularities, so that the radial symmetry is not broken. The outer boundary is either absorbing, for example a semi permeable membrane where the intermediates can leave the system but once outside they can not reenter or reflecting, represented by an impermeable membrane. Due to the symmetry of the system it is advisable to express Eq.(4.6) in terms of spherical coordinates $\rho(\mathbf{r}) \to \rho(r, \theta, \phi)$. Since the system is spherically symmetric and the boundary conditions do not break this symmetry, only the $l = 0$ eigenvalue contributes. Therefore the m eigenvalue is restricted to be zero and the spherical harmonics reduce to one,

$Y_{l=0,m=0}(\theta,\phi) = 1$. We are left with a reaction diffusion equation only depending on the radial coordinate r. The enzymes in the center is encoded in the boundary condition at $r = 0$ for the corresponding 1D system. The reaction diffusion equation reads

$$0 = \frac{1}{r^2}\partial_r(r^2\partial_r\rho(r) - \frac{\alpha e(r)\rho(r)}{1+\gamma\rho(r)} - \beta\rho(r). \tag{4.15}$$

The inner boundary is $\big(4\pi r^2\partial_r\rho(r)\big)_{r=0} = 1$ and the outer boundary condition is either absorbing $\big(\rho(r)\big)_{r=1} = 0$ or reflecting $\big(4\pi r^2\partial_r\rho(r)\big)_{r=1} = 0$. Similarly, the 1D reduction of a 2D spherical system can be obtained by reducing the two dimensional polar coordinates. In the one dimensional case the nabla operator is simply the second derivative of the intermediate density. In the next section we will consider a minimal two site model. It will be shown that the phenomenology of the optimization of the enzyme distribution is qualitatively similar to the continuous system in the various dimensions. We demonstrate that the physical reason for a particular enzyme arrangement bears a certain universality, which can be extended to any higher dimensional system. We will use the minimal model as a pedagogical introduction for the concept of enzyme exposure. After that the consequences of the optimization procedure for an one dimensional system with linear reaction will be discussed. As a next step we will show the robustness of our results against changes in the geometry, respectively, dimensionality and extending the reactions to the full MM-non-linear reaction.

4.2 Minimal Model

As a starting point of our investigations on optimal spatial arrangements of enzymes we consider a simple two state model. We will show that such a model captures the essential physics behind the enzyme arrangement. However its applicability is rather limited, the results can be generalized to more realistic models. This will be done in the next sections. For the minimal model it is more instructive to think in terms of particles rather than concentrations. However, the parameters defined in the last section remain unchanged when moving to the the particle picture.

The minimal model consists of two sites. At each site the intermediate can potentially react with one of the enzymes sitting at these sites. For reasons of simplicity we exclusively restrict ourselves to the linear reaction discussed before. The incoming substrate particles enter the system by hopping onto site A, this incoming current is denoted by J_1. The intermediate particles can either jump to the next state B via the diffusion hopping rate $d = 1$ (remember this is because of the definition of α of the previous section) or react with the enzymes sitting at state A with the reduced reaction rate α. With e_A the number of particles at site A the overall consumption rate is αe_A. At site B the intermediate particle has several choices either it leaves the system with the diffusive rate d or it jumps back to state A with rate d or it reacts with the enzymes, and the associated enzyme number e_B, sitting on site B with the reaction rate α. This reaction scheme can be seen in Fig.

(4.2). The reaction current J_2 is the combination of the substrate particles reacting with

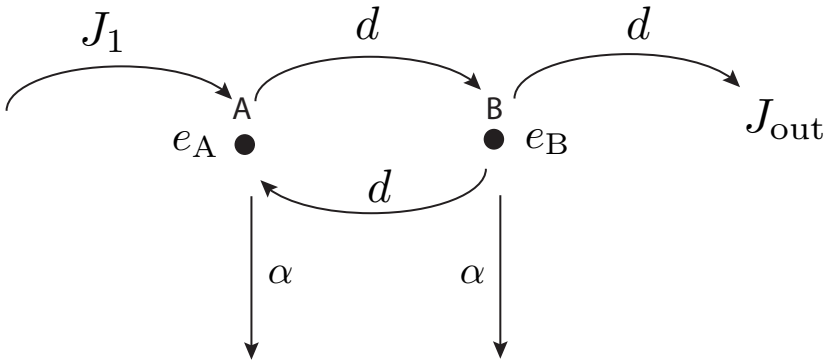

Figure 4.2: The two site minimal model is shown. The possible transitions for a substrate particle are reaction rate α at site A and site B contributing to the reaction current J_2, and the diffusion rate d which is equal to one. The fraction of enzymes at site A is e_A and at site B is e_B.

the enzymes on site A or on site B. Those particles which are not reacting with any of the enzymes will eventually leave the system, due to particle conservation, and thus contribute to the outgoing current J_{out}. Note, that in the previous section we fixed the overall number of enzymes to be constant E_T (using the reduced variables E_T is normalized to one). We rewrite $e_A = f$ and $e_B = (1 - f)$ with $f \in [0, 1]$, where f is the fraction of total enzymes sitting on site A. In order to determine the reaction current, we monitor a particle through the system. An intermediate sitting on site A has two possibilities either it reacts or it moves on to the next site. The reaction probability is given by

$$P_{R_A} = \frac{\alpha f}{1 + \alpha f} \tag{4.16}$$

respectively the probability that the substrate particle moves on to the next site is

$$P_F = \frac{1}{1 + \alpha f} \tag{4.17}$$

where, obviously, $P_{R_1} + P_F = 1$. A substrate particle sitting on site B has three different possibilities, leaving the system and thus contributing to the loss current, reacting with an enzyme, and jumping back to site A. All of them are associated with a certain probability. The probability for the intermediate particle to react reads

$$P_{R_B} = \frac{\alpha(1 - f)}{2 + \alpha(1 - f)}, \tag{4.18}$$

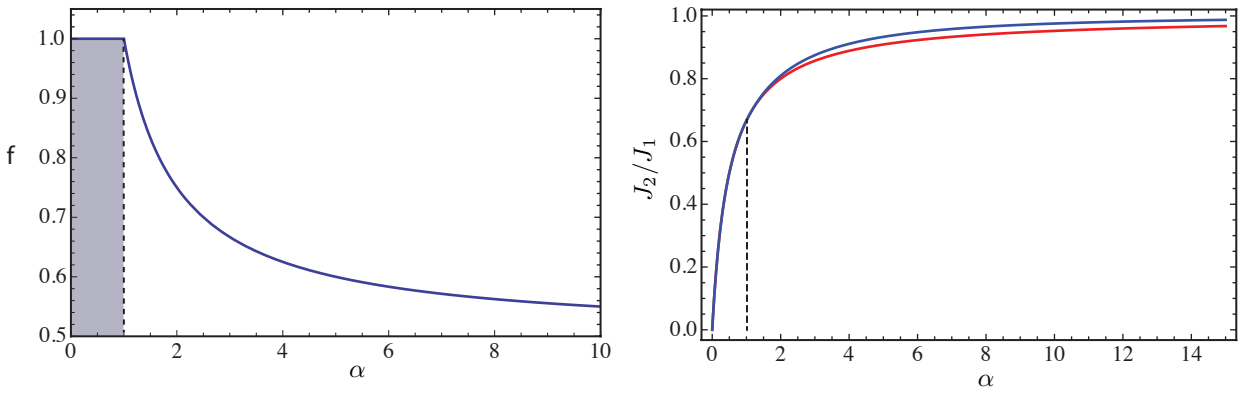

Figure 4.3: The left panel shows the optimal fraction f of enzymes sitting at site A, with respect to the product generation. For values of $\alpha \leq 1$ all of the enzymes cluster at the first site(gray shaded). After a sharp transition at $\alpha = 1$ the enzymes form the first site are transferred to site B, changing the shape of the optimal enzyme profile qualitatively. The fraction f that get transferred form site A to B scales with $f \approx \alpha^{-1}$. The panel on the right hand side shows the reaction current generated by the optimal enzyme distribution (blue line) versus the clustered enzyme distribution (red line), where all enzymes are placed at site A.

the probabilities to leave the system and hopping back to site A are identical, they are given by

$$P_L = \frac{1}{2+\alpha(1-f)} \qquad \text{and} \qquad P_B = \frac{1}{2+\alpha(1-f)}. \tag{4.19}$$

And again the probability conservation is satisfied $P_{R_1} + P_L + P_B = 1$. The normalized loss current is given by the sum of all the different paths an intermediate particle can take before leaving the system. The simplest path the particle can take, when starting at site A is the direct one given by the product of the probabilities of the substrate particle moving from site A to B and immediately leaving the system $P_F P_L$. For the next "higher order" path the particle instead of leaving the system, when it is sitting at site B it jumps back to site A and then forward to B again. In other words the particle makes an extra loop. By going to higher and higher order every time we add an extra loop, thus

$$\begin{aligned} \frac{J_{\text{out}}}{J_1} &= P_F P_L + P_F P_B P_F P_L + P_F (P_B P_F)^2 P_L + \cdots \\ &= P_F \sum_{n=0}^{\infty} \Big(P_B P_F \Big)^n P_L = \frac{P_F P_L}{1 - P_B P_F}, \end{aligned} \tag{4.20}$$

where in the last step we have applied the geometric sum. By using the definition of the corresponding probabilities Eq.(4.17) and Eq.(4.19) we obtain after some straightforward

manipulations

$$\frac{J_{\text{out}}}{J_1} = \big((1+\alpha f)(2+\alpha(1-f))-1\big)^{-1}. \tag{4.21}$$

To find the optimal enzyme distribution for the two site model, we require the out flux of the intermediates Eq.(4.21) to be minimal, hence

$$\frac{\partial J_{\text{out}}/J_1}{\partial f} = 0 \qquad \Longrightarrow \qquad f = \frac{1+\alpha}{2\alpha}. \tag{4.22}$$

Since f cannot be greater than one, due to the fixed total number of enzymes we find that the solution is given by

$$f(\alpha) = \begin{cases} 1 & \text{if } \alpha \leq 1 \\ \frac{1+\alpha}{2\alpha} & \text{if } \alpha > 1. \end{cases} \tag{4.23}$$

Surprisingly, the optimal distribution shows a sharp transition of the enzyme distribution at $\alpha = 1$. For $\alpha < 1$ all enzymes cluster at the first site A while the second is unoccupied. However, for $\alpha > 1$ a fraction of the enzymes is moved toward the second site. In the limit $\lim_{\alpha\to\infty} f(\alpha) = \frac{1}{2}$ the enzymes are equally distributed between the the two sites. Note that actually in the case of α goes to infinity the normalized loss current in Eq.(4.21) goes to zero independently of f, thus the distribution does not matter anymore. Every particle entering the system immediately reacts since the reaction probability is one in the limit. In Fig.(4.3) the left panel shows the fraction of enzymes at the first site. After the sharp transition the fraction decays as $f \approx \alpha^{-1}$. In the right panel the optimized normalized reaction flux J_2/J_1 (blue line) is depicted. Further, the normalized current produced by a clustered distribution (red line), where $f = 1$, thus all enzymes sitting at site A, coincides with the optimal distribution until $\alpha = 1$, as expected. However, we see a slight deviation for α much greater than one of around 5 percent, clearly favoring the enzyme distribution where both sites are occupied. We call these emerging enzymes at the site B "back up" enzymes. The effect of these back up enzymes becomes even more important when we consider a continuous system.

How can we explain such a behavior? Once a particle reaches the second site its probability to leave the system without contributing to the reaction current is nonzero. Therefore one would have expected, in order to optimize the current, it is favorable to place all the enzymes at the first site. Clustering of the enzymes minimizes the chances for a substrate particle to reach the second site and thus reduces the chances of escape. However, with this apparently naive picture we get the wrong optimal enzyme distribution. In the next section we will introduce the notion of enzyme exposure to explain the origin of the smeared out enzyme distribution.

4.2.1 Enzyme Exposure

In order to explain a particular optimal enzyme distribution it is convenient to consider an alternative version describing the same model, see Fig.(4.5). An intermediate entering

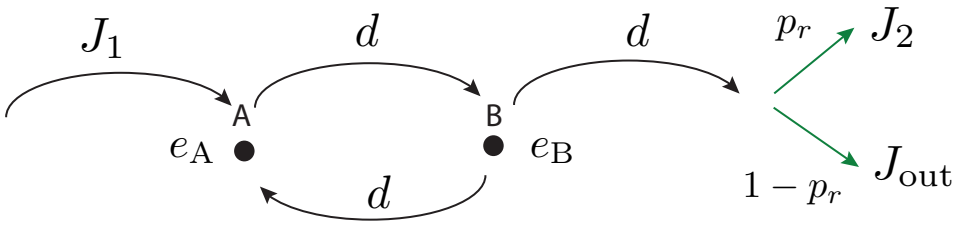

Figure 4.4: A schematic representation of an alternative minimal two state model. Showing the diffusive path through the system. After leaving the system it is decided whether the substrate contributes to the loss current J_{out} or the reaction current J_2 (green arrows).

the system arrives first at site A stays there for some time and moves on to site B. We have seen in the previous section that an intermediate particle may return to site A from site B several times. The "lifetime" τ_A is the total time the particle stays at site A on its trajectory till it escapes the system after τ_{escape}. Likewise the intermediate particle has a "lifetime" τ_B in state B, thus $\tau_A+\tau_B = \tau_{\mathrm{escape}}$. We now assume that the substrate does not react with the enzymes, but after the particle has left the system we decide whether or not it has reacted with one of the enzymes along its path through the system. In other words it is decided after the intermediate leaves the system whether it contributes to the reaction current or the loss current. The probability that this particle has not reacted at site A in the presence of the enzyme concentration e_A is given by $e^{-\alpha e_A \tau_A}$ and for the second site the probability of not reacting is $e^{-\alpha e_B \tau_B}$. From this we easily deduce the overall probability of a particle reacting along its trajectory is

$$p_r = 1 - e^{-\alpha(e_A\tau_A + e_B\tau_B)}. \tag{4.24}$$

Note, different paths lead to different lifetimes of an individual intermediate particle, therefore we need to consider an ensemble of paths through the system. To this end we define a new quantity called the "enzyme exposure" $E = e_A\tau_A + e_B\tau_B$ being the time an intermediate sees a certain amount of the enzymes on its path through the system. To clarify that definition we present two conceptually different examples. Assuming all enzymes are clustered at the first site the enzyme exposure is merely the remains of the intermediate at site A. In the other case where the enzymes are uniformly distributed over both sites $e_A = e_B = \frac{1}{2}$ the enzyme exposure is proportional to the escape time. In the second example the enzyme exposure is directly proportional to the diffusion time of an intermediate particle through the system. Once the enzymes are distributed unevenly over the two states the notion of enzyme exposure is not that simple anymore. The sum of all possible

paths through the system and their exposure to the enzymes concentration is then characterized by the enzyme exposure distribution $P(E)$. Note that this probability distribution not only depends on the paths but simultaneously on the enzyme distribution. At this point it should be mentioned that the enzyme exposure distribution does not depend on the reaction parameter α. The overall reaction flux is therefore the product of the enzyme exposure distribution and the reaction probability integrated over all possible paths

$$\frac{J_2}{J_1} = \int_0^\infty P(E) p_r(E) dE. \tag{4.25}$$

This expression decomposes the reaction-diffusion dynamics into two separate parts, the reaction part described by $p_r(E)$ and the diffusive part $P(E)$. The generalization of enzyme exposure concept of our two state model to continuous systems is straightforward and will be discussed in the following manuscript [1]. Plugging Eq.(4.24) into Eq.(4.25) leads immediately to

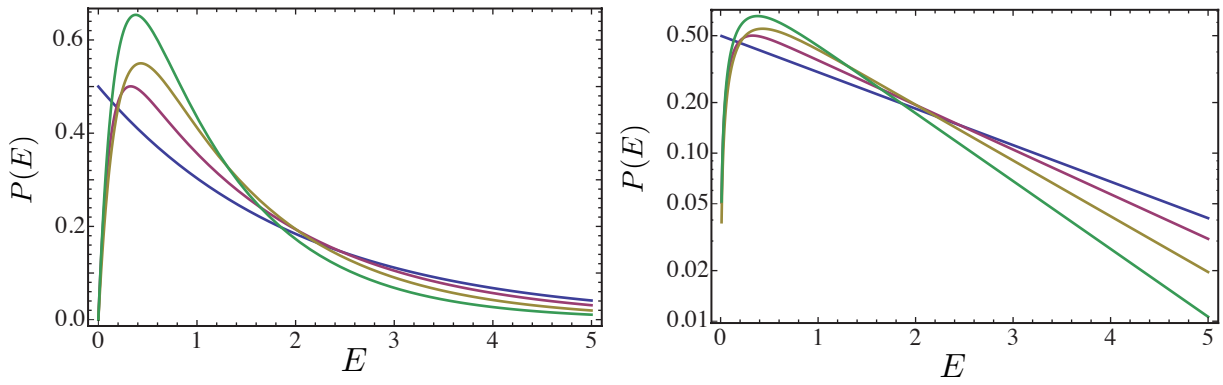

Figure 4.5: In both pictures the enzyme exposure distribution is shown the left one on a linear scale while the right one on a logarithmic scale. The different values for the fraction of enzymes sitting in site A is $f = 1$ (blue line), $f = 0.75$ (red line), $f = 0.5$ (yellow line) and $f = 0.25$ (green line).

$$\frac{J_2}{J_1} = 1 - \int_0^\infty P(E) e^{-\alpha E} dE = 1 - \frac{J_{\text{out}}}{J_1}. \tag{4.26}$$

where we used that $\int_0^\infty P(E)dE = 1$. Formally, we see that the probability of leaving the system without reacting is the Laplace transformation of $P(E)$. Hence, in order to calculate the enzyme exposure distribution we need to compute the inverse Laplace of the normalized loss current.

[1] Clustering and Optimal Arrangement of Enzymes in Reaction-Diffusion Systems.

Applying the inverse Laplace to Eq.(4.21) gives

$$P(E) = \frac{2e^{-E\left[\frac{1}{2f}+\frac{1}{1-f}\right]}}{[4f^2+(1-f)^2]} \sinh\left[E\sqrt{\frac{1}{4f^2}+\frac{1}{(1-f)^2}}\right]. \tag{4.27}$$

In the clustered case ($f = 1$) the enzyme exposure distribution reduces to a single exponential

$$P(E) = \frac{1}{2}e^{-\frac{E}{2}}. \tag{4.28}$$

For this situation the most probable paths are those with $E \approx 0$, since the intermediate can rapidly escape from the first site to the second where no enzyme is present and the contribution to the enzyme exposure is zero, see Fig(4.5). On the other hand once the enzymes are shared between both sides the distribution becomes narrower. Moreover the shape changes from a single exponential to a multi-exponential distribution with $P(E = 0) = 0$. Therefore, the probability for an intermediate particle to immediately escape without being exposed to the enzymes vanishes, since the second site is also occupied with a fraction of enzymes $(1 - f)$.

However, the fact an intermediate particle spends more time in site A than in site B becomes apparent once we consider the mean exposure $\langle E\rangle = 1 + f$, which decreases when f decreases. The mean value reflects our naive picture to cluster the enzyme independently of the value ofα. For the full description we need to focus on the product of the enzyme exposure distribution and the α dependent probability distribution $p_r(E)$. As shown in Eq.(4.25), the interplay between the change of $p_r(E)$ as α is varied and the shape of $P(E)$ depending on the fraction of enzymes in site A leads to an optimal f^* maximizing the current. To demonstrate this we consider two opposite case, the first one where the enzymes are all cluster at the first site $f = 1$ and the second one where the enzymes are equally distributed $f = \frac{1}{2}$.

To this end we consider at first the reaction probability $p_r(E)$ and the enzyme exposure $P(E)$ independently and then we take a closer look at their product. Just as a reminder, the area under the curve of the product is equal to the efficiency of a pathway with a certain enzyme arrangement, see Fig.(4.6). As already mentioned only the reaction probability p_r depends on α and in the limits $\alpha \ll 1$, $p_r(E)$ is almost linear in the relevant E region and for $\alpha \gg 1$ it becomes a step like function. Thus in the first case $\alpha \ll 1$ the trajectories with a larger E have a higher probability to react, therefore the enzyme exposure probability with the longer tail leads to a higher reaction current in the end. As depicted in Fig.(4.5) the enzyme exposure probability for $f = 1$ has the most probability weight in the tail. While the other case $\alpha \gg 1$ favors an absence of "short" $E \approx 0$ trajectories the uniformly distributed enzymes are preferred, in which case $P(E = 0) = 0$, see Fig.(4.6), exactly where the enzyme exposure distribution has most of its probability weight for the clustered enzymes. In these opposite cases it is easy to see the difference in the curve under the product $p_r(E)P(E)$. In Fig.(4.6) the left column the red shaded area of the clustered enzyme distribution clearly is lager than the blue area representing the uniformly distributed enzymes. This is only true for very small values of the reaction

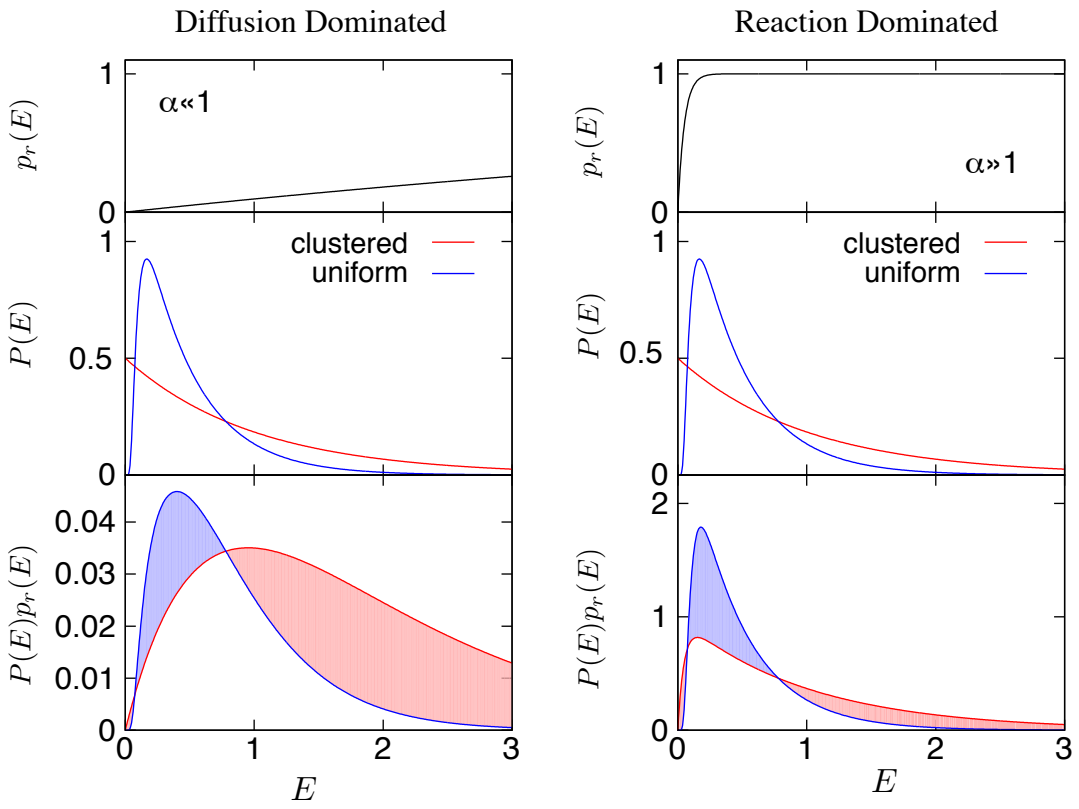

Figure 4.6: A description for the two different scenarios demonstrating the importance of the concept of "enzymes exposure". The transition from a clustered configuration ($\alpha \ll 1$, left) to the regime in which the uniform profile is preferable ($\alpha \gg 1$, right). (Middle) When enzymes are clustered at site A ($f = 1$) $P(E)$ has excess probability, compared to when enzymes are uniformly distributed, at small and large values of E. (Bottom) The reaction flux is given by the integral of $P(E)p_{\mathrm{r}}(E)$. For $\alpha \ll 1$ the extra probability in the large-E tail of $P(E)$ in the clustered configuration contributes more to J_2 than probability in the region $E \sim 1$. When $\alpha \gg 1$ only trajectories with $E \ll 1$ are subject to a low reaction probability, leading to a lower J_2 when enzymes are clustered.

strength $\alpha \ll 1$. For large reaction strength $\alpha \gg 1$ blue area is greater than the red area favoring the uniform distribution. Note that the transition between clustered and uniform enzyme distribution is gradually, for $\alpha \approx 1$ the optimal distribution is a superposition of both. It is straightforward to generalize the concept of enzyme exposure to continuous systems with and without competing pathways to any arbitrary dimensions. However, this approach is purely limited to linear systems. Once we consider for example the full MM-reaction the individual trajectories are no longer independent of each other.

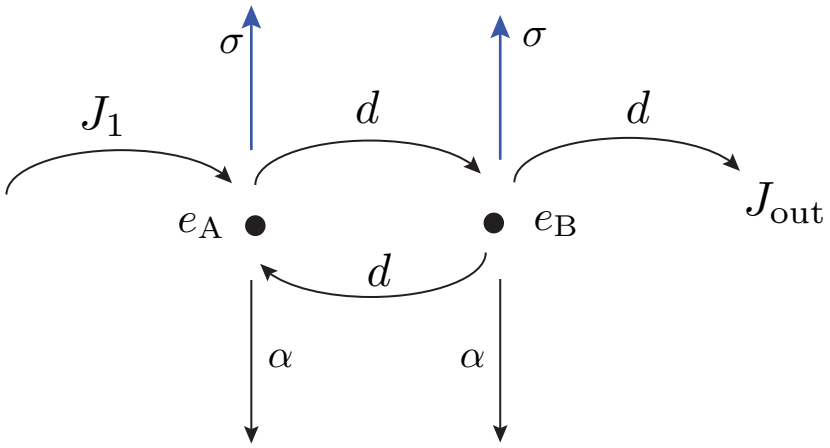

Figure 4.7: The two site minimal model from the previous section is extended by introducing a competing pathway. The strength of that pathway is σ (blue arrows).

4.2.2 Competing pathway

As mentioned in the introductory chapter 3. competing pathways have many different biochemical origins. Here we only deal with a simplified two state model, therefore we would like to focus only on the physical implication of such pathways and do not include any biochemical specifications for the competing pathway. Additionally, to the minimal model presented in Fig.(4.2) the intermediate particles can leak away on both sites A and B with the effective rates σ[2]. Our new two site model is depicted in Fig.(4.7). The leakage terms change the reaction probability for an intermediate particle arriving at site A as follows:

$$P_{R_A} = \frac{\alpha f}{1 + \alpha f + \sigma}, \tag{4.29}$$

for the hopping probability to site B

$$P_F = \frac{1}{1 + \alpha f + \sigma} \tag{4.30}$$

and for the leakage probability

$$P_{L_A} = \frac{\sigma}{1 + \alpha f + \sigma}. \tag{4.31}$$

Likewise, we find the probabilities for an intermediate particle sitting at site B to react

$$P_{R_B} = \frac{\alpha(1-f)}{2 + \alpha(1-f) + \sigma}, \tag{4.32}$$

[2]In the case of the minimal model $\sigma = \beta$.

to leak

$$P_{LB2} = \frac{\sigma}{2 + \alpha(1-f) + \sigma}, \tag{4.33}$$

to leave the system

$$P_{\text{out}} = \frac{1}{2 + \alpha(1-f) + \sigma}, \tag{4.34}$$

or to jump back to site A

$$P_B = \frac{1}{2 + \alpha(1-f) + \sigma}. \tag{4.35}$$

Similar, to the consideration without leakage we obtain the normalized loss current through the boundary by summing up all the possible paths, thus

$$\frac{J_{\text{out}}}{J_1} = \frac{P_F P_{\text{out}}}{1 - P_B P_F}. \tag{4.36}$$

Here, we have to be a bit careful, because of the two contributions to the overall loss current one stemming from the loss of substrate particles through the boundary and the second one from the loss due to leakage at both sites. The loss current due to leakage reads

$$\frac{J_{leakage}}{J_1} = \frac{P_{L_A} + P_F P_{L_B}}{1 - P_B P_F}, \tag{4.37}$$

which leads to the reaction current

$$\begin{aligned} \frac{J_2}{J_1} &= 1 - \frac{P_F P_{\text{out}} + P_{L_A} + P_F P_{L_B}}{1 - P_B P_F} \\ &= \frac{\alpha f(2 + \alpha(1-f) + \sigma) + \alpha(1-f)}{(2 + \alpha(1-f) + \sigma)(1 + \alpha f + \sigma) - 1}. \end{aligned} \tag{4.38}$$

As before we find the optimal fraction of enzymes f in the first site by setting the derivative with respect to f equal to zero, thus

$$f(\alpha, \sigma) = \begin{cases} 1 & \text{if } \alpha \leq 1 + 3\sigma + \sigma^2 \\ \frac{(1+\sigma(3+\alpha+\sigma)) - \sqrt{(1+\sigma(1+\sigma)(5+\alpha+2\sigma))}}{\alpha\sigma} & \text{if } \alpha > 1 + 3\sigma + \sigma^2. \end{cases} \tag{4.39}$$

Figure(4.8) shows that for fixed leakage rate σ increasing α leads to minimal fraction f of enzymes at site A, hence a maximal fraction at the site B. Up to that minimum the behavior of the optimal distribution is the same as for the leakage free case. For intermediate values of α it becomes favorable to locate some of the enzymes from site A to site B, but for a finite value of σ the uniform distribution $f = \frac{1}{2}$ is never reached. The relocation of enzymes towards site B for the optimal distributions has the same reason as in the case without a competing pathway, see explanation above. However, as σ is increased the enzymes tend to cluster again since the enzymes suppose to react as quickly as possible. A large value for σ only increases the probability for the substrate particle to contribute to the loss current. In the extreme case where σ is large enough, so that

on average an intermediate particle never hops onto the second site, it is reasonable that clustering the enzymes at site A achieves the best reaction flux. Surprisingly, when α increases more and more enzymes go back to the first site, in the optimal case. The need for distributed enzymes is reduced, because only those trajectories of intermediate particles that immediately leak away at the first site cause the reaction probability to be less than one. Therefore gathering the enzymes achieves a compensation of this effect. Again, the enzyme exposure distribution can be calculated using Eq.(4.26). In the general

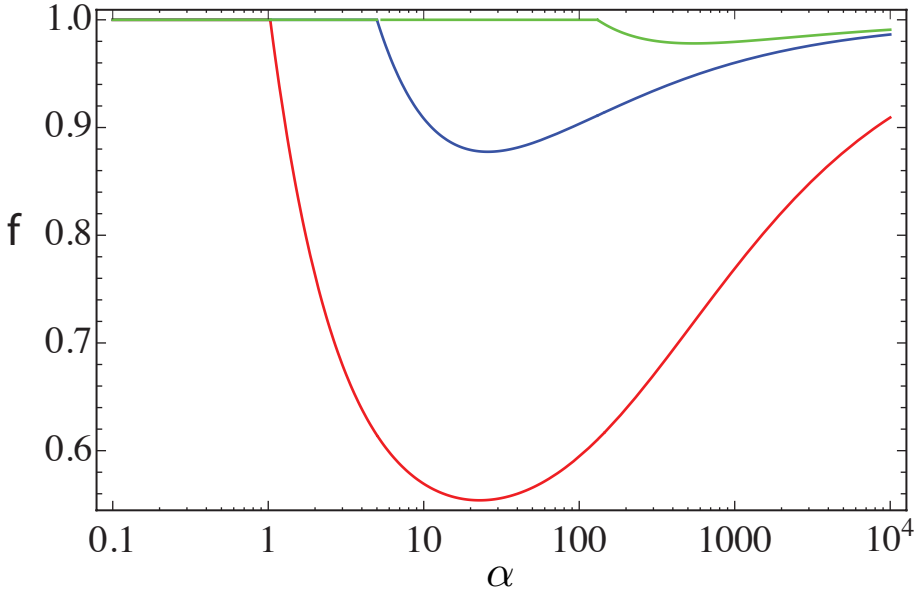

Figure 4.8: The optimal enzyme distributions in the presence of a competing pathway are depicted. The fraction f of enzymes in the first site shows an α-region in which clustering the enzymes is preferred. The extend of this region is determined by the value of the leakage parameter, red curve ($\sigma = 0.1$), blue curve ($\sigma = 1$) and green curve ($\sigma = 10$).

case, however, the enzyme exposure distribution could be obtained analytically and it is still multiexponetial but its concrete form is complicated and not very enlightening. In the special case of $f = 1$ the enzyme exposure reduces to a single exponential function

$$P(E)_{f=1} = \frac{(1+\sigma(3+\sigma))e^{-\frac{1+\sigma(3+\sigma)}{2+\sigma}E}}{2+\sigma}. \tag{4.40}$$

Thus from the perspective of enzyme exposure the same behavior as in the leakage-free case is recovered. The main differences, however, are that once $\sigma \neq 0$ the uniformly

distributed case is never optimal for large $\alpha \gg 1$. On the contrary, the enzymes tend to cluster again at the first site. The strength of the σ-value determines the maximal fraction $1 - f$ at the second site. In the following two manuscripts we will discuss implication of optimal arrangement when we consider continuous systems. Firstly, we will investigate a one dimensional system with length L. Again a clustered, an uniformly distributed and mixed optimal enzyme profile is observed. Further possible realizations of the system in terms of artificial reaction cascades are discussed. The second manuscript can be seen as a generalization of the idea of an optimal profile maximizing the efficiency of the system to higher dimensions with different geometries, also leakage will be included. Finally we will discuss the implications of non-linear reactions on the optimal profile.

4.3 Clustering and Optimal Arrangement of Enzymes in Reaction-Diffusion Systems

In the subsequent manuscript we investigate the optimal distribution of enzymes in space which maximizes the consumption of a locally-produced substrate. As in the minimal model we only consider linear reactions without a second pathway, thus $\sigma = 0$. Note, that we will relax both assumptions in the next manuscript. We demonstrate that this one-dimensional model system exhibits a transition as a function of the single control parameter α from a diffusion-dominated regime, in which the maximal out-flux is achieved by clustering enzymes near the source, to a reaction-dominated regime in which a distributed enzyme profile achieves a higher reaction flux. Remarkably, the origin of this transition lies in the fundamentally stochastic nature of the reactions and diffusion of single molecules, as we show through numerical results and analytic analysis of a minimal model, since varying the enzyme profile alters the shape of the distribution of enzyme exposure experienced by diffusing substrate molecules. Understanding systems with only a few different enzymes may create a toolbox for building larger multi-enzyme complexes with controlled functionality.

Clustering and Optimal Arrangement of Enzymes in Reaction-Diffusion Systems

Alexander Buchner, Filipe Tostevin, and Ulrich Gerland*

Arnold Sommerfeld Center for Theoretical Physics and Center for Nanoscience, Ludwig-Maximilians-Universität München, Theresienstraße 37, D-80333 München, Germany

(Received 5 February 2013; published 16 May 2013)

Enzymes within biochemical pathways are often colocalized, yet the consequences of specific spatial enzyme arrangements remain poorly understood. We study the impact of enzyme arrangement on reaction efficiency within a reaction-diffusion model. The optimal arrangement transitions from a cluster to a distributed profile as a single parameter, which controls the probability of reaction versus diffusive loss of pathway intermediates, is varied. We introduce the concept of *enzyme exposure* to explain how this transition arises from the stochastic nature of molecular reactions and diffusion.

DOI: 10.1103/PhysRevLett.110.208104 PACS numbers: 87.15.R−, 87.10.−e, 87.18.Nq

To efficiently catalyze multistep biochemical reactions, sets of enzymes have evolved to function synergistically. Cells not only keep concerted control over the concentrations and activities of enzymes in the same pathway, but often also arrange them in self-assembled multienzyme complexes [1]. Apart from the large molecular machines (polymerases, ribosomes, spliceosomes), one of the best-studied natural multienzyme complexes is the cellulosome, a complex where up to 11 different enzymes are arranged on a noncatalytic scaffolding protein [2]. This complex is assembled extracellularly by anaerobic bacteria to efficiently break down cellulose, the most abundant organic material on the planet. Similarly, enzyme complexes are used for intracellular metabolism [3]. However, neither the precise consequences of putting enzymes together into complexes are well understood, nor the degree to which complex formation confers a functional advantage in each case [4–7].

It has long been thought that physical association between collaborating enzymes might increase the effective reaction flux, minimize the pool of unwanted intermediate products, allow coordinate regulation by a single effector, and reduce transient time scales [8,9]. However, while enzymatic activity has been studied for over a century, suitable techniques to characterize such effects quantitatively have become available only recently. On the one hand, single-molecule enzymology allows us to monitor [10] and manipulate [11] the activity of individual enzyme molecules. On the other hand, enzyme molecules can be positioned with nanometer precision in artificial systems using "single-molecule cut-and-paste" [12] on two-dimensional surfaces or along one-dimensional channels, and with DNA origami structures even in three dimensions [13,14]. These experimental developments call for a theoretical analysis of the effects of spatial proximity and arrangement of enzymes, to uncover the principles for the design and optimization of multienzyme systems. Such principles could be applied to bioengineer systems that control biochemical reactions at will, such as for the production of drugs or biofuels [15,16]. Related issues also arise in the context of signaling proteins [17]; however, the functional criteria for the optimization of signaling systems are likely different [18,19].

Here, we ask under which conditions it is beneficial to localize enzymes rather than distribute them. Furthermore, what is the optimal arrangement and how does it depend on the system parameters? We base this study on simple-reaction diffusion models, which permit rigorous quantitative analysis, and assume the steady-state reaction flux is the single critical system property. Interestingly, this already leads to rich physical behavior, with a sharp transition from a regime in which it is optimal to cluster downstream enzymes in the vicinity of upstream enzymes, to a regime in which an extended enzyme profile generates a higher reaction flux. This behavior, which we explain by analyzing the "enzyme exposure" of molecules diffusing in the system, is a result of the stochastic nature of the reactions and diffusion of single molecules.

Clustered enzymes.—That colocalizing enzymes within the same pathway might indeed improve the efficiency of converting a substrate S into a final product P can be seen by considering a two-step reaction, $S \xrightarrow{E_1} I \xrightarrow{E_2} P$, as a minimal model where production of P via an intermediate I is catalyzed by the enzymes E_1 and E_2. Let us consider an E_1 molecule (or a small cluster thereof) as a local source of I molecules and describe the local arrangement of E_2 enzymes relative to E_1 by the distribution $e(\mathbf{r})$, normalized such that $E_T = \int e(\mathbf{r}) d\mathbf{r}$ is the total number of E_2 molecules per E_1 center. To determine the efficiency of an enzyme arrangement $e(\mathbf{r})$, we need to describe the reaction-diffusion dynamics of the density $\rho(\mathbf{r}, t)$ of intermediates. We assume simple diffusion, with coefficient D, and standard Michaelis-Menten kinetics [20] for the enzymatic reactions, with catalytic rate k_{cat} and Michaelis constant K_M for E_2. In the low-density regime, where the reaction term becomes linear, we then have

$$\partial_t \rho(\mathbf{r}, t) = D \nabla^2 \rho(\mathbf{r}, t) - \kappa e(\mathbf{r}) \rho(\mathbf{r}, t), \qquad (1)$$

with $\kappa = k_{\rm cat}/K_M$ measuring the enzyme efficiency. Intermediates will either react to form product or will be lost, either directly to the extracellular space (for extracellular enzymes) or across the cell membrane. We can implement this possible loss via an absorbing boundary condition, $\rho(r = R, t) = 0$, on a sphere with radius R that may be taken to infinity. On the other hand, intermediates are constantly generated by E_1 at the origin, with an average flux that we denote by J_1, yielding the source boundary condition $-D(4\pi r^2 \partial_r \rho)_{r=0} = J_1$. In the resulting nonequilibrium steady-state $\rho(\mathbf{r})$, product is generated at the rate

$$J_2 = \kappa \int_{r<R} e(\mathbf{r})\rho(\mathbf{r}) d\mathbf{r}. \tag{2}$$

Let us assume, for the moment, that enzyme E_2 is spread over a spherical shell with radius $r_0 < R$. We then find a total product flux of

$$J_2 = \frac{J_1}{1 + \frac{4\pi D R r_0}{E_T \kappa (R - r_0)}} \overset{R \gg r_0}{\longrightarrow} \frac{J_1}{1 + \frac{4\pi D r_0}{E_T \kappa}}. \tag{3}$$

This result indicates that reducing r_0—arranging the E_2 molecules close to the E_1 center—can dramatically increase the flux if loss of intermediate products is a concern. Whether this effect is biologically relevant crucially depends on the characteristic length scale $r_c = E_T \kappa / 4\pi D$, where J_2 begins to saturate. Enzyme efficiencies can be up to $\kappa \sim 10^8\ \mathrm{M^{-1}\,s^{-1}}$ (although superefficient enzymes can achieve $\kappa \sim 10^{10}\ \mathrm{M^{-1}\,s^{-1}}$ [21]), while biomolecular diffusion constants are typically larger than $D \sim 10\ \mu\mathrm{m^2\,s^{-1}}$, such that with $E_T \sim 10\, E_2$ molecules per E_1 center, r_c is at most of nanometer scale, comparable to the size of enzymes. Thus even our simplified model, which does not include interenzyme interactions such as direct channeling [22], suggests that in realistic biochemical settings, J_2 will be strongly dependent on the distance between enzymes down to the scale of their own size.

On a microscopic scale, the simple reaction-diffusion description we have used above will break down, since steric effects and the specific enzyme structure become important. Nevertheless, we can exploit the coarse-grained model to address more general questions on a mesoscopic scale. In particular, it is intriguing to ask whether colocalization is in fact the optimal enzyme arrangement, and whether the behavior will change qualitatively when the enzyme kinetics become nonlinear.

Clustered vs uniform arrangements.—Let us focus on the one-dimensional version of Eq. (1). This is not only a natural starting point for a theoretical study, but also relevant experimentally, e.g. for "molecular factories" in quasi-one-dimensional channels within future "lab-on-a-chip" devices. Specifically, we consider the one-dimensional steady-state $\rho(x)$ of a finite system, $x \in [0, L]$, with source/sink boundaries, $-D(\partial_x \rho)_{x=0} = J_1$ and $\rho(L) = 0$. We compare different E_2 enzyme distributions $e(x)$ with the same mean density $\bar{e} = L^{-1} \int_0^L e(x) dx = E_T/L$. The behavior of the system is determined by the dimensionless control parameter $\alpha = \kappa \bar{e} L^2 / D$, which measures the relative importance of reactions and diffusion in shaping $\rho(x)$. When $\alpha < 1$, the system is dominated by diffusion, as the typical reaction time scale $(\kappa \bar{e})^{-1}$ is longer than the typical diffusion time $\sim L^2/D$ to the absorbing boundary. Conversely, for large α, reactions are fast compared to diffusive escape. In the limit of $\alpha \to \infty$, J_2 approaches J_1 independent of the spatial arrangement of enzymes.

We first compare the reaction flux of clustered enzymes, $e_c(x) = \bar{e}\,\delta(x/L)$, and uniform enzymes, $e_u(x) = \bar{e}$. As shown in Fig. 1, the clustered configuration achieves a larger flux for $\alpha \lesssim 9$. Surprisingly, for larger α, the uniform configuration achieves a higher reaction flux. Thus it is not always preferable to simply localize enzymes where the concentration of intermediate is highest, which always occurs at $x = 0$. Rather, when reactions are fast compared to diffusion, the intermediates can be consumed more efficiently if E_2 is uniformly distributed throughout the system.

Enzyme exposure.—To examine the origin of this transition, we consider the fate of a single I molecule introduced into the system at $t = 0$. Whether it will have reacted by time T depends on the concentration of E_2 enzymes, $e(x(t))$, to which it has been exposed along its trajectory $x(t)$: the probability that it has not reacted is $\exp[-\kappa \int_0^T e(x(t)) dt]$. Therefore, the probability of escaping the system can be decomposed into the likelihood of particular trajectories through the system, and the probability of no reaction occurring along each trajectory. Indeed, the relative likelihoods of escape and reaction can be recaptured if, rather than assuming that I is consumed by the enzyme, we instead propagate a diffusive trajectory until it hits the absorbing boundary at time τ, and subsequently determine whether or not a reaction would have occurred based on the rescaled total enzyme exposure $E = D(L^2 \bar{e})^{-1} \int_0^\tau e(x(t)) dt$ and reaction probability $p_r(E) = 1 - \exp(-\alpha E)$.

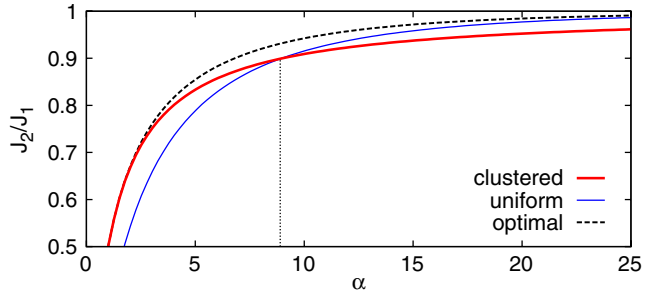

FIG. 1 (color online). Comparison of the reaction flux achieved by different enzyme profiles. A transition occurs at $\alpha \sim 9$ between regimes in which clustered and uniform enzyme profiles achieve a higher reaction flux. The optimal mixed enzyme distribution [Eq. (5) with $f = \alpha^{-1/2}$, dashed black line] achieves a still higher J_2 for intermediate values of α.

Given the stochasticity of diffusion, a given enzyme arrangement $e(x)$ will lead to a characteristic distribution of enzyme exposure, $P(E)$. For uniformly distributed enzymes, E is simply proportional to the time spent in the system, and $P(E)$ is therefore set by the distribution of escape times at the absorbing boundary $x = 1$ for a diffusing particle (see the Supplemental Material [23]),

$$P_u(E) = \sum_{n=0}^{\infty} \pi(-1)^n(2n+1)e^{-\pi^2(n+1/2)^2E}. \quad (4)$$

For a clustered configuration the appropriate distribution is found to be $P_c(E) = \exp(-E)$ (see the Supplemental Material [23]). Importantly, these distributions are independent of the reaction rate α, which enters into the reaction flux only via the reaction probability $p_r(E)$, which is in turn independent of the spatial arrangement of enzymes. Specifically, the reaction flux is given by $J_2 = J_1 \int_0^\infty P(E)p_r(E)dE$. Thus it is the interaction of these two distributions that determines which enzyme profile is preferable for a given value of α.

Figure 2 rationalizes the transition observed in Fig. 1. When $\alpha \ll 1$, such that $p_r(E \lesssim 1)$ is small, the majority of reaction events correspond to trajectories with large values of E. Compared to the uniform configuration, for which $P_u(E) \sim \exp(-\pi^2 E/4)$ for large E, the clustered configuration places more probability weight in the large-E tail of $P_c(E)$, and thus achieves a higher reaction flux when α is small. In the opposite limit of large $\alpha \gg 10$, only those

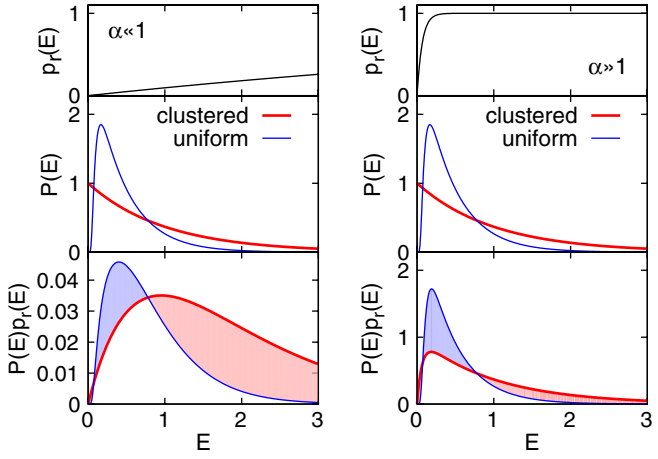

FIG. 2 (color online). Schematic depiction of the transition from a favorable clustered configuration ($\alpha \ll 1$, left) to the regime in which the uniform profile is preferable ($\alpha \gg 1$, right). Middle: when enzymes are clustered at $x = 0$ $P(E)$ has excess probability, compared to when enzymes are uniformly distributed, at small and large values of E. Bottom: the reaction flux is given by the integral of $P(E)p_r(E)$. For $\alpha \ll 1$ the extra probability in the large-E tail of $P(E)$ in the clustered configuration contributes more to J_2 than probability in the region $E < 1$. When $\alpha \gg 1$ only trajectories with $E \ll 1$ are subject to a low reaction probability, leading to a lower J_2 when enzymes are clustered.

trajectories with extremely small values of $E \ll 1$ have a significant probability of not reacting. Thus the uniform enzyme profile, for which $P_u(E \to 0) \to 0$, becomes preferable. The critical value of the transition, $\alpha \approx 9$, marks the point at which the reaction probability becomes large in the vicinity of the peak of $P_u(E)$.

Optimal profiles.—We have thus far compared only uniformly distributed and clustered configurations. However, it may be that another enzyme profile is able to achieve a reaction flux which is higher still. We therefore investigated what is the optimal enzyme distribution $e(x)$, for fixed $\bar{e}$, that maximizes the reaction flux J_2 (or alternatively, minimizes leakage $J_1 - J_2$). A direct analytic optimization of J_2 over $e(x)$ is not possible because of the nontrivial dependence of $\rho(x)$ on $e(x)$. We therefore studied the optimization of J_2 numerically on a discretized interval (see the Supplemental Material [23]).

These data show that for small $\alpha < 1$ the clustered configuration, with all enzymes colocalized with the source, is the optimal arrangement. Interestingly, the optimal profile undergoes a transition, distinct from that discussed above, at the critical value $\alpha = 1$. For $\alpha > 1$, in the optimal profile only a fraction of the available enzymes were clustered; the remaining enzymes were distributed approximately uniformly over an extended region with the enzyme density in this region equal to $\bar{e}$, as shown in Fig. 3.

Motivated by these numerical results, we studied enzyme profiles of the form

$$e(x) = \bar{e}\left\{f\delta\left[\frac{x}{L}\right] + \Theta\left[1 - f - \frac{x}{L}\right]\right\}, \quad (5)$$

where $\Theta(x)$ is the Heaviside function, and f is the fraction of enzymes that are clustered. We found that for this restricted class of profiles, the optimal profile indeed undergoes a transition from $f = 1$ for $\alpha \leq 1$ to $f = \alpha^{-1/2}$ for $\alpha > 1$. Examining the scaling of the fraction of enzymes that are clustered in the numerically optimized profiles, we find excellent agreement with this α scaling

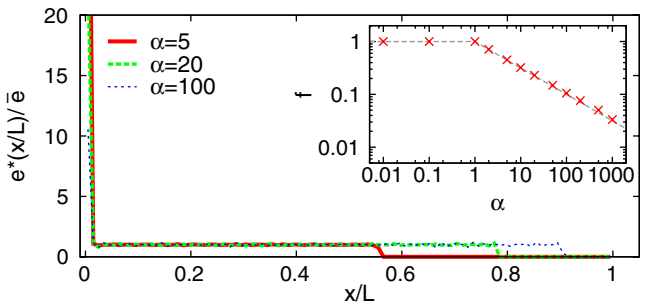

FIG. 3 (color online). Optimal enzyme density distribution for different values of α. Plotted profiles are the result of numerical optimization (see the Supplemental Material [23]) after 4×10^4 iterations with a lattice of 100 sites. Inset: the fraction of enzymes f located at the first lattice site in the numerically-optimized enzyme profile scales as $\alpha^{-1/2}$ for $\alpha > 1$.

(see Fig. 3 inset). The corresponding reaction flux tracks the envelope of the curves for the clustered and uniform configurations as α is varied (Fig. 1, dashed line).

The two distinct qualitative features of the optimal profile—the peak at $x = 0$ and the sharp decrease at $x = L(1 - \alpha^{-1/2})$—can be related to geometry of the system: enzymes cluster in the vicinity of the source and are excluded from the region nearest to the absorbing boundary. The distance from the end of the uniform enzyme domain to the boundary at $x = L$ scales with the typical diffusion length of substrate molecules in an enzyme density $\bar{e}$, which is $\sim L\alpha^{-1/2}$. If the enzyme concentration were to be uniform, $e(x) = \bar{e}$, substrate molecules that approach within this distance of the absorbing boundary have a high probability of diffusing out of the system rather than reacting. Any enzymes placed in this area contribute little to the reaction flux and can be used more effectively if relocated closer to the source.

We characterized $P(E)$ for mixed enzyme profiles of the form Eq. (5) by numerically sampling the enzyme exposure of continuous-time random walk trajectories on a lattice until their escape at $x = L$. The resulting distributions for different values of f are shown in Fig. 4. In the extreme cases of $f = 1$ and $f = 0$, the numerical results reproduce the analytic results of $P_c(E)$ and $P_u(E)$ above. At intermediate values of f, $P(E)$ retains a more pronounced large-E tail than $P_u(E)$, while still reducing the probability of extremely small E values relative to $P_c(E)$. As α is increased, the relative importance of these two features is reduced and increased, respectively. Thus the optimal $P(E)$ becomes more sharply peaked, corresponding to a smaller f.

So far we have considered only the case of linear reaction kinetics. In the nonlinear regime of the Michaelis-Menten kinetics, it is no longer possible to consider individual substrate trajectories independently since the reaction probability of a particular molecule depends on not only the local enzyme concentration but also the substrate density. Nevertheless, a qualitatively similar transition of the optimal enzyme distribution from clustered to distributed will occur provided the enzyme concentration

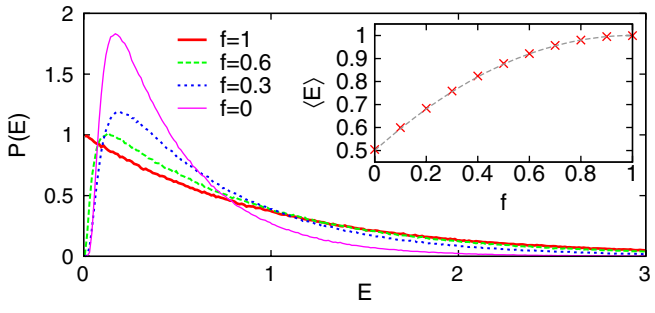

FIG. 4 (color online). Distributions $P(E)$ estimated from 2×10^6 simulated substrate trajectories subjected to an enzyme distribution $e(x_i) = \bar{e}\{fN\delta_{i,1} + \Theta[1 - f - (i/N)]\}$, with $N = 100$.

is not so low as to be saturated throughout the entire system, in which case the reaction current becomes independent of enzyme positioning.

Discussion.—In our model enzymatic pathway, the ultimate fate of each intermediate (I) molecule is either to react to product or to escape. For a given enzyme arrangement, the dimensionless parameter α controls the relative likelihood of these outcomes. Conversely, for each value of α there is an optimal enzyme arrangement that minimizes the loss of intermediates. In the small-α regime, where the reaction is slow and escape is likely, the best enzyme arrangement is a tightly clustered one. As α is increased, the system moves into the reaction-dominated regime and it becomes preferable to relocate some of the available E_2 enzymes away from the source. The transition of the optimal profile takes place at $\alpha \sim 1$. With a system size of $L \approx 100$ nm, α values in the range of 0.01–100 should be achievable in synthetic systems [12–14]. Thus it should be possible to directly test our results experimentally.

Intuitively, for large α, the more distant E_2 molecules may be interpreted as "backup enzymes" intended to catch the fraction of I molecules that were able to diffuse away from the cluster. The cost of removing some enzymes from the cluster is small since there remains a high probability of reaction for intermediates which spend a long time in the vicinity. Spreading these enzymes provides a larger benefit by recouping some of the escaped I molecules. Indeed, the optimal enzyme arrangement for $\alpha > 1$ is akin to a bet-hedging strategy, in the sense that the optimal placement for multiple E_2 molecules is not to cluster them all at the position where a single E_2 would do best, but instead hedge bets on the stochastic motion of their substrate by carefully distributing them.

Similar effects will also occur in systems with different geometries, including in higher dimensions. We have introduced the integrated "enzyme exposure" as a quantitative tool to characterize the effects of different enzyme arrangements. Importantly, the optimal enzyme profile does not necessarily maximize the average enzyme exposure (see Fig. 4 inset). Rather, it is the matching between the shape of the enzyme exposure distribution and the reaction probability that is key. While properties of diffusion such as recurrence change with dimension, the qualitative picture that clustered and distributed enzymes lead, respectively, to monotonically decaying and sharply peaked enzyme exposure distributions remains unchanged. Therefore the underlying physics of the transitions described is generic, although the magnitude of the effects will vary with the specific system. The concept of enzyme exposure provides a general framework for understanding the behavior of many different scenarios.

We have seen that the optimal enzyme distribution is determined by the distributions of timing of reaction and diffusion events. These are intrinsic single-molecule properties. Thus, we expect that the optimal enzyme profile

would remain unchanged if we considered instead discrete substrate and enzyme molecules. The only difference is that for finite numbers of enzyme molecules, $e(x)$ cannot be chosen arbitrarily but instead only certain discrete values are permitted. Thus $P(E)$ cannot be varied continuously, but rather one of a specific ensemble of allowed distributions must be chosen. While this will not change the qualitative behavior of the optimal profile as the system parameters are varied, it may quantitatively alter its shape for given parameter values. We leave this as a topic of future studies.

A. B. and F. T. contributed equally to this work. This research was supported by the German Excellence Initiative via the program "Nanosystems Initiative Munich" and the German Research Foundation via the SFB 1032 "Nanoagents for Spatiotemporal Control of Molecular and Cellular Reactions."

*gerland@lmu.de

[1] P. A. Srere. Annu. Rev. Biochem. **56**, 89 (1987).
[2] E. A. Bayer, H. Chanzy, R. Lamed, and Y. Shoham, Curr. Opin. Struct. Biol. **8**, 548 (1998).
[3] M. E. Campanella, H. Chu, and P. S. Low. Proc. Natl. Acad. Sci. U.S.A. **102**, 2402 (2005).
[4] A. Cornish-Bowden, Eur. J. Biochem. **195**, 103 (1991).
[5] P. Mendes, D. B. Kell, and H. V. Westerhoff, Eur. J. Biochem. **204**, 257 (1992).
[6] A. Cornish-Bowden and M. L. Cardenas, Eur. J. Biochem. **213**, 87 (1993).
[7] P. Mendes, D. B. Kell, and H. V. Westerhoff. Biochim. Biophys. Acta **1289**, 175 (1996).
[8] F. H. Gaertner, Trends Biochem. Sci. **3**, 63 (1978).
[9] R. Heinrich, S. Schuster, and H. Holzütter, Eur. J. Biochem. **201**, 1 (1991).
[10] S. Xie, Single Mol. **2**, 229 (2001).
[11] H. Gumpp, E. M. Puchner, J. L. Zimmermann, U. Gerland, H. E. Gaub, and K. Blank, Nano Lett. **9**, 3290 (2009).
[12] S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, and H. E. Gaub, Science **319**, 594 (2008).
[13] J. Müller and C. M. Niemeyer. Biochem. Biophys. Res. Commun. **377**, 62 (2008).
[14] J. Fu, M. Liu, Y. Liu, N. W. Woodbury, and H. Yan, J. Am. Chem. Soc. **134**, 5516 (2012).
[15] R. J. Conrado, J. D. Varner, and M. P. De Lisa, Curr. Opin. Biotechnol. **19**, 492 (2008).
[16] P. P. Peralta-Yahya, F. Zhang, S. B. del Cardayre, and J. D. Keasling, Nature (London) **488**, 320 (2012).
[17] D. Bray, Annu. Rev. Biophys. Biomol. Struct. **27**, 59 (1998).
[18] S. B. van Albada and P. R. ten Wolde, PLoS Comput. Biol. **3**, e195 (2007).
[19] A. Mugler, A. Gotway Bailey, K. Takahashi, and P. R. ten Wolde, Biophys. J. **102**, 1069 (2012).
[20] A. Cornish-Bowden, *Fundamentals of Enzyme Kinetics* (Wiley-Blackwell, Weinheim, Germany, 2012), 4th ed.
[21] M. E. Stroppolo, M. Falconi, A. M. Caccuri, and A. Desideri, Cell Mol. Life Sci. **58**, 1451 (2001).
[22] X. Huang, H. M. Holden, and F. M. Raushel, Annu. Rev. Biochem. **70**, 149 (2001).
[23] See Supplemental Material http://link.aps.org/supplemental/10.1103/PhysRevLett.110.208104 for details of the calculation of $P(E)$ and of the numerical optimization procedure.

4.4 Optimization of Collective Enzyme Activity via Spatial Localization

The following manuscript extends the previous model towards higher dimensions and different geometries. Even, if a competing pathways is present, a transition between clustered and uniformly distributed enzymes occurs, but only in the case of an absorbing boundary. Applying a reflecting boundary condition always prefers the clustered enzymes over the uniform distribution, but it is still valid that for certain parameters values (α, β) a transition between a clustered and an extended enzyme profile appears. We also consider the influence of a non-linear-MM-reaction on the optimal enzyme profile. Unfortunately, the concept of enzymes exposure fails since the individual paths of the intermediate are not independent of each other anymore. Yet, we see very similar results to the linear reaction case suggesting a screening effect of the enzymes due to their partial occupation by the intermediates.

Optimization of collective enzyme activity via spatial localization

Alexander Buchner, Filipe Tostevin, Florian Hinzpeter, and Ulrich Gerland[a)]
Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, 80333 München, Germany

(Received 8 July 2013; accepted 13 September 2013; published online 1 October 2013)

The spatial organization of enzymes often plays a crucial role in the functionality and efficiency of enzymatic pathways. To fully understand the design and operation of enzymatic pathways, it is therefore crucial to understand how the relative arrangement of enzymes affects pathway function. Here we investigate the effect of enzyme localization on the flux of a minimal two-enzyme pathway within a reaction-diffusion model. We consider different reaction kinetics, spatial dimensions, and loss mechanisms for intermediate substrate molecules. Our systematic analysis of the different regimes of this model reveals both universal features and distinct characteristics in the phenomenology of these different systems. In particular, the distribution of the second pathway enzyme that maximizes the reaction flux undergoes a generic transition from co-localization with the first enzyme when the catalytic efficiency of the second enzyme is low, to an extended profile when the catalytic efficiency is high. However, the critical transition point and the shape of the extended optimal profile is significantly affected by specific features of the model. We explain the behavior of these different systems in terms of the underlying stochastic reaction and diffusion processes of single substrate molecules. © *2013 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4823504]

I. INTRODUCTION

The action of enzymes is essential for nearly all processes in living cells. Often these enzymes are organized into large multi-molecular complexes associated with specific functional tasks,[1] and this organization can be crucial to the successful operation of the enzymatic system. These "molecular factory" assemblies, in which the product of one enzymatic reaction becomes the substrate for the next, are common in metabolic pathways of both prokaryotes and eukaryotes. Examples include the cellulosome,[2] the pyruvate dehydrogenase complex[3] and glycolytic enzymes.[4] In some cases, such as the cellulosome,[2] enzymes are arranged on an inert scaffold in a specific way. In others, such as tryptophan synthase complexes,[5] direct enzyme-enzyme interactions lead to self-assembly into a complex.

Despite the ubiquity of these multi-enzyme complexes, we still lack a deep understanding of the consequences of particular arrangements of enzymes for metabolic pathway operation. Many advantages of co-localization have been proposed,[6–8] particularly via the direct transfer or "channeling" of substrates from one enzyme to another. For example, reducing the transit time of pathway intermediates between enzymes can minimize the loss of unstable intermediates or the interference of competing pathways. Channeling could also potentially enhance the local density of substrates in the vicinity of the enzymes and reduce exposure to toxic intermediates; however, whether or not these effects can actually occur has been disputed.[9–12] On the other hand, compartmentalization of metabolic enzymes can also increase the flux of biosynthetic pathways,[13] indicating that the pathway kinetics can be influenced by localization even in the absence of direct channeling. Similar questions about the role of co-localization also arise in the context of protein signaling cascades. For example, there the clustering of enzymes can generate a greater amplification of the signal than distributing enzymes.[14,15] However, differing functional criteria between signaling scenarios, where discrimination between different inputs is crucial, and metabolic systems, where maintaining a specific flux may be more desirable, mean that these systems are likely subject to different design pressures. More generally, little is known about the effects of the placement of enzymes beyond simple co-localization or clustering scenarios.

Recently, there has been a growing focus on the experimental study of colocalized enzymes. Techniques have been developed that allow for the attachment of enzymes to a scaffold,[16] which was shown to significantly increase the yield of the mevalonate production pathway.[17] The "single-molecule cut-and-paste" technique[18] allows for the positioning of enzymes on a surface with nanometer precision. DNA origami permits the highly controlled production of three-dimensional structures,[19] enabling the quantitative study of the effects of more complex spatial arrangements of enzymes. Over the last few years much progress has been made in engineering of artificial enzymatic pathways on DNA[20–22] and RNA assemblies,[23] even *in vivo*. In particular, a distance-dependence of the activity of a pathway consisting of glucose oxidase (GOx) and the horseradish peroxidase (HRP) was demonstrated:[24] when the enzymes are brought closer together, the efficiency of the two enzyme complex increases.

Here we study theoretically the impact of enzyme positioning on the flux of pathways. It has been demonstrated previously[25] that in a simple linear reaction-diffusion model, in different parameter regimes co-localization can increase or

[a)]Electronic mail: gerland@lmu.de

decrease the pathway flux compared to the uniform distribution of enzymes. In this paper, we extend these results to a range of reaction-diffusion systems. In particular, we also consider nonlinear reactions and different spatial dimensions. We demonstrate that the qualitative features of these diverse models are similar. In general, a transition occurs as a function of the effective reaction rate between regimes in which clustering or distributing enzymes in space generates a higher pathway efficiency. We calculate the optimal enzyme distribution that maximizes the efficiency of the pathway. The universal nature of our results in these diverse systems shows that the observed transitions arise from general properties of reactions and diffusion and highlights the applicability of the observed behavior to diverse biochemical pathways.

II. MODEL

We consider a simple model reaction pathway consisting of two enzymatic reaction steps. In the first reaction step, an enzyme E_1 converts a substrate S into an intermediate I; subsequently, a second enzyme E_2 converts I into the final product of the pathway, P. We are interested in how the spatial organization of the enzymes affects the efficiency of the pathway $S \overset{E_1}{\rightarrow} I \overset{E_2}{\rightarrow} P$ in converting substrate S to product P. To this end we assume that the E_1 enzymes are fixed in position, and examine the impact of the location of the E_2 enzymes relative to E_1. In this scenario, E_1 enzymes act as a source of intermediate I, with a total production rate J_1. In order to additionally include possible undesirable non-specific competition for the intermediate by secondary pathways, or decay in the case that I is unstable, we also allow for the conversion of I into an alternative waste product Q. Under the assumption that these processes are independent of the spatial arrangement of E_2 enzymes, they are simply modeled as a first-order reaction with a constant, position-independent, rate σ. The density of intermediate I, $\rho(\mathbf{r}, t)$ can then be modeled by the reaction-diffusion equation

$$\frac{\partial \rho(\mathbf{r}, t)}{\partial t} = D\nabla^2 \rho(\mathbf{r}, t) - \frac{k_{\text{cat}} e(\mathbf{r}) \rho(\mathbf{r}, t)}{K_M + \rho(\mathbf{r}, t)} - \sigma \rho(\mathbf{r}, t), \quad (1)$$

where D is the diffusion constant of I and $e(\mathbf{r})$ is the (static) density of E_2 enzymes. The model is illustrated in Fig. 1. In writing Eq. (1) we have assumed that the conversion of I to P by the enzyme E_2 can be described by standard Michaelis-Menten kinetics with catalytic rate k_{cat} and Michaelis constant K_M. We implement the production of intermediate by E_1 enzymes through boundary conditions to Eq. (1). In this work we restrict ourselves to the case of a uniform source at the inner boundary, $-D\nabla\rho(\mathbf{r}_{\text{in}}) \cdot \mathbf{n}(\mathbf{r}_{\text{in}}) = J_1/A_{\text{in}}$, where $\mathbf{n}(\mathbf{r}_{\text{in}})$ is the unit vector normal to the boundary and A_{in} is the area of the inner boundary; thus the total influx integrated over the boundary equals the production rate J_1. For the outer boundary we limit ourselves to reflective ($\nabla\rho(\mathbf{r}_{\text{out}}) \cdot \mathbf{n}(\mathbf{r}_{\text{out}}) = 0$) or absorbing ($\rho(\mathbf{r}_{\text{out}}) = 0$) boundary conditions. The former could represent the confinement of the intermediate reaction product by the membranes of a cell or organelle, while the latter might describe an intermediate that can easily cross the membrane and be lost to the extracellular environment. We note that our treatment could easily be generalized to mixed boundary conditions representing partial confinement.

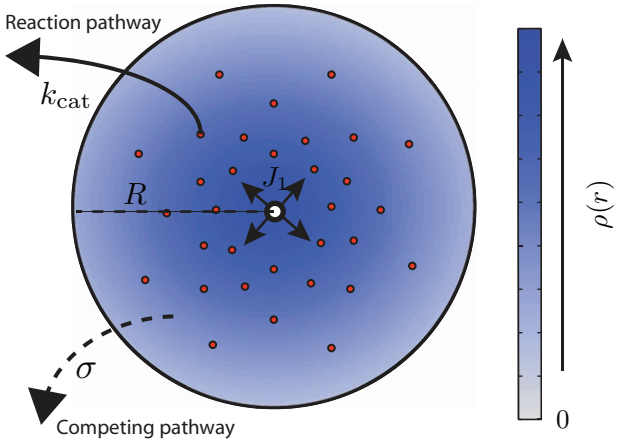

FIG. 1. Illustration of the type of reaction-diffusion systems considered in this paper. A cluster of E_1 enzymes at the center of the system acts as a source of intermediate I with total production rate J_1. Intermediates diffuse from this center and can either be converted to the desirable product by the enzymes E_2 (red circles), via a Michaelis-Menten reaction with catalytic rate k_{cat}, or can be lost to a competing pathway with the spatially uniform rate σ. We coarse-grain the positions of E_2 in space into the continuous distribution $e(\mathbf{r})$ and the local density of intermediates as $\rho(\mathbf{r})$. While a two-dimensional system with a hard-wall outer boundary is shown here for clarity, we consider systems of all dimension as well as different outer boundary conditions.

In the following we will be concerned only with the steady-state flux through the reaction pathway. At steady-state form, Eq. (1) can be recast into the dimensionless form

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

where $\mathbf{r}'$ denotes that the spatial coordinate has been rescaled by a characteristic length-scale of the system, R, which we will take to be the system size; and ρ' indicates that the density has further been rescaled such that the total production rate of intermediate is equal to 1. Additionally, we have defined the rescaled enzyme density $e'(\mathbf{r}) = e(\mathbf{r})/\bar{e}$ with $\bar{e} = V^{-1}\int_V e(\mathbf{r})\mathrm{d}\mathbf{r}$ the average enzyme density over the system volume V. The dimensionless parameters $\alpha = (k_{\text{cat}}\bar{e}/K_M)(R^2/D)$ and $\beta = \sigma(R^2/D)$ respectively capture the relative timescales of reactions with E_2 and with secondary pathway enzymes, compared to the typical time to diffuse a distance R. The 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 and includes the effect of varying the activity of E_1 enzymes via the intermediate production rate J_1. In the following we drop the prime notation and work exclusively with the dimensionless system; this should not be a source of confusion.

Integrating Eq. (2) and applying the boundary conditions leads to the flux-conservation equation

$$1 = \underbrace{\int_V \frac{\alpha e(\mathbf{r})\rho(\mathbf{r})}{1 + \gamma\rho(\mathbf{r})}\mathrm{d}\mathbf{r}}_{J_2/J_1} + \underbrace{\int_V \beta\rho(\mathbf{r})\mathrm{d}\mathbf{r} - \int_{\partial V} \nabla\rho(\mathbf{r}) \cdot \mathbf{n}(\mathbf{r})\mathrm{d}\mathbf{r}}_{J_{\text{loss}}/J_1}. \quad (3)$$

On the left-hand side we have the (rescaled) production of intermediate by E_1. This must be balanced by the flux of

reactions by E_2 enzymes, J_2/J_1, plus the loss of intermediate. This loss can occur to secondary pathways (the second term on the right-hand side of Eq. (2)) and via escape at the boundaries of the system (the third term, where ∂V is the outer boundary of the system that is not a source of intermediate). Assuming that the efficiency of the conversion of substrate to intermediate by E_1 is independent of the localization of E_2 enzymes, such that J_1 is constant, the efficiency of the system can be described by the ratio J_2/J_1, the fraction of intermediates that are converted into the correct product P. In this work we will examine how changing $e(\mathbf{r})$ affects the pathway efficiency J_2/J_1. To compare different enzyme profiles on an equal footing, the total amount of E_2 is held constant via the condition $V^{-1}\int_V e(\mathbf{r})\mathrm{d}\mathbf{r} = 1$.

In the remainder of this paper, we systematically characterize the effects of varying E_2 localization in different regimes of Eq. (2). In Sec. III A, we will focus on the low density limit of the intermediate product, in which the rate of reaction with E_2 becomes linear in $\rho(\mathbf{r})$. It was previously shown[25] that in an open one-dimensional system where intermediate is lost at an absorbing boundary, different parameter regimes exist in which the optimal enzyme profile consists of either co-localization of E_2 with E_1, or a configuration wherein only a fraction of E_2 enzymes are co-localized and the remainder are distributed over a finite region. Here we extend these results to consider closed systems where the loss of intermediate occurs only via position-independent secondary reactions, and to three- and two-dimensional systems. Finally, in Sec. III B, we also consider the full nonlinear reaction model. In all cases we observe a transition in the optimal profile from co-localized to distributed as a function of the system parameters, analogous to that reported in the specific minimal model of Buchner *et al.*,[25] demonstrating the generality of the underlying physics. However, we also highlight qualitative differences in the phenomenology of these different regimes.

III. RESULTS

A. Linear reaction models

1. Enzyme exposure

To understand the impact of different enzyme configurations on the overall pathway flux, the concept of integrated "enzyme exposure" has proven to be useful.[25] It allows for the decomposition of the reaction flux of linear systems into two factors, one that depends only on the enzyme distribution $e(\mathbf{r})$ and describes the diffusive dynamics of the system and another that is independent of $e(\mathbf{r})$ but captures the reaction dynamics. The concept is best explained with the help of a thought experiment where we first consider the dynamics of individual intermediate molecules in the absence of any E_2 enzymes, also shown schematically in Fig. 2. We introduce a single intermediate molecule at $t = 0$ at the source and track its stochastic path until it leaves the system, either through the boundary of the system or via a reaction with a competing pathway. We denote the time at which the trajectory ends, either by escaping at the system boundary or through a competing reaction, as t_{escape}. By repeatedly applying this procedure for many such molecules, we can generate an ensemble of trajectories $\mathbf{r}(t)$ through the system that is independent of the distribution of E_2 enzymes.

Next, we suppose that we were to re-introduce E_2 enzymes according to the distribution $e(\mathbf{r})$. For each of the diffusive intermediate trajectories generated above, the instantaneous propensity of reaction with an E_2 enzyme is given by $\alpha e(\mathbf{r}(t))$ (in the linear reaction regime). For each trajectory, the survival probability $S(t)$ that no reaction has occurred up to the time t follows the differential equation $\dot{S}(t) = -\alpha e(\mathbf{r}(t))S(t)$. We can, therefore, straightforwardly calculate the probability that a reaction would have occurred at some point along the trajectory as $1-\exp[-\alpha\int_0^{t_{\mathrm{escape}}} e(\mathbf{r}(t))\mathrm{d}t]$.

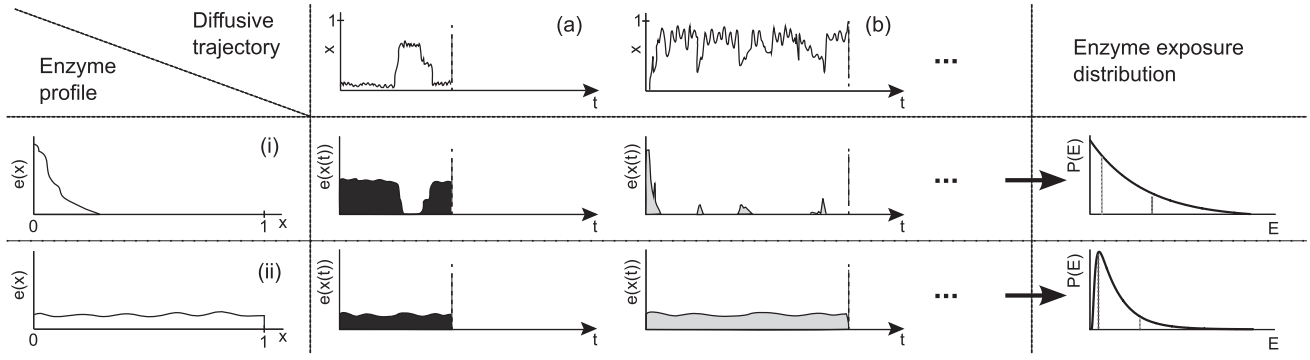

FIG. 2. A schematic illustration of the underlying concept of "enzyme exposure." Diffusion generates an ensemble of trajectories of intermediate molecules (top row), which have different enzyme exposure values in the presence of different enzyme distribution patterns (left column). For example, the diffusive path (a) spends a relatively long time in the vicinity of the origin. This leads to a higher enzyme exposure value $E = \int_0^{t_{\mathrm{escape}}} e(x(t))dt$ (shaded areas) for an enzyme distribution that is clustered near the origin [distribution (i), middle row]. Trajectory (b), which spends little time near the origin, leads to a very small exposure value in the presence of distribution (i). For a more uniformly distributed profile [distribution (ii), bottom row] the exposure value is determined primarily by how long the particle stays in the system. Enumerating the value of E for all possible diffusive trajectories leads to the $e(x)$-dependent distribution of enzyme exposures $P(E)$ (right column).

Finally, we define the enzyme exposure for each individual trajectory to be

$$E = \int_0^{t_{\text{escape}}} e(\mathbf{r}(t))\mathrm{d}t. \tag{4}$$

The ensemble of possible trajectories in the system $\mathbf{r}(t)$, each with a characteristic t_{escape}, therefore generates a distribution of enzyme exposure values, $P(E)$. This distribution is a function of the arrangement of E_2 enzymes $e(\mathbf{r})$ via Eq. (4), but importantly is independent of the reaction with E_2 enzymes, since the diffusive trajectories were generated in the absence of such reactions. The probability of reaction along a trajectory is then given by $p_r(E) = 1 - \exp[-\alpha E]$, which depends on the reaction parameter α but crucially not on the E_2 distribution itself. The overall probability of reaction is recovered by the expression

$$\frac{J_2}{J_1} = \int_0^{\infty} P(E)p_r(E)\mathrm{d}E, \tag{5}$$

which ensures a proper weighting of the likelihood of a particular trajectory occurring (see the supplementary material[26] for a derivation showing the equivalence of Eq. (5) with the expression for J_2/J_1 defined in Eq. (3)). Thus, as depicted schematically in Fig. 3, we have decomposed the reaction-diffusion dynamics of Eq. (2) into a diffusion- and $e(\mathbf{r})$-dependent component $P(E)$, and a reaction-dependent component $p_r(E)$, with the efficiency of the reaction pathway determined by the product of these two distributions.

Interestingly, using Eq. (3) the reaction efficiency can be rewritten as

$$\frac{J_2}{J_1} = 1 - \int_0^{\infty} P(E)e^{-\alpha E}\mathrm{d}E = 1 - \frac{J_{\text{loss}}}{J_1}, \tag{6}$$

wherein we see that the fraction of intermediate molecules lost via the reaction $I \to Q$ or through the boundary, J_{loss}/J_1, takes the form of the Laplace transform of $P(E)$, with transform variable α. It is generally more straightforward to calculate J_{loss} as a function of α for a given E_2 profile $e(\mathbf{r})$ and to compute $P(E)$ by performing an inverse Laplace transformation, than it is to calculate $P(E)$ directly from considering individual diffusive trajectories.

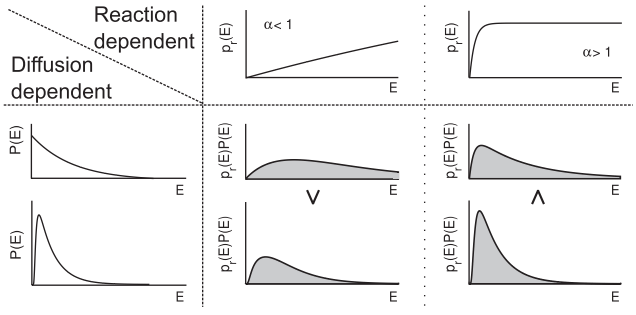

FIG. 3. The reaction efficiency can be calculated as the overlap integral of the $e(\boldsymbol{r})$-dependent enzyme exposure distribution $P(E)$, and the α-dependent reaction probability $p_r(E)$, according to $J_2/J_1 = \int_0^{\infty} P(E)p_r(E)\mathrm{d}E$. Broader exposure distributions, which maximize the likelihood of large-E trajectories, are preferable when α is small (middle column). Narrower $P(E)$ distributions, which minimize the likelihood of small values of E, are favored when α is large (right).

2. A competing pathway

We now consider the case where intermediates are unable to cross a cell membrane and therefore cannot escape via the boundaries of the system, but can be lost to a competing, spatially uniform, reaction pathway. That is, all intermediate molecules ultimately end up as either the desirable product P or undesirable product Q. For a one-dimensional system in the linear, low-density, regime of the Michaelis-Menten reaction, we have the reaction-diffusion equation

$$0 = \partial_x^2\rho(x) - \alpha e(x)\rho(x) - \beta\rho(x) \tag{7}$$

with a source boundary condition at the left edge, $\partial_x\rho|_{x=0} = -1$, and a reflecting boundary at the right edge, $\partial_x\rho|_{x=1} = 0$. The parameters α and β, defined above, reflect the relative reactivities of intermediate with E_2 and competing pathway enzymes, respectively, in units of the typical time to diffuse a distance of the system size R. Consequently, α and β can also be interpreted as describing the system size in units of the typical distance from the source at $x = 0$ that an intermediate molecule will diffuse before leaving the system via reaction with E_2 and via the competing pathway, respectively. When $\beta \ll 1$, intermediate molecules will typically be able to explore the entire system, and therefore we should expect that the spatial arrangement of enzymes will have little effect on the reaction flux, since the intermediate will be exposed to each enzyme regardless of where it is placed. In contrast, when $\beta \gg 1$ very few intermediate molecules will diffuse far from the source and we should expect that the amount of enzyme located close to the source will have a strong influence on the pathway efficiency.

We begin by examining the case where E_2 enzymes have the uniform density $e_{\text{u}}(x) = 1$ throughout the domain $x \in [0, 1]$. Substituting into Eq. (7) leads to the straightforward solution $\rho_{\text{u}}(x) = \cosh[(1-x)\sqrt{\alpha+\beta}]/(\sqrt{\alpha+\beta}\sinh\sqrt{\alpha+\beta})$. From this expression, the reaction efficiency can be calculated using the definitions in Eq. (3) and is given by

$$\left(\frac{J_2}{J_1}\right)_{\text{u}} = \frac{\alpha}{\alpha+\beta}. \tag{8}$$

Next we suppose that all E_2 enzymes are clustered at $x = 0$, co-localized with E_1. This is represented by the distribution $e_{\text{c}}(x) = \delta(x)$, which leads to the solution $\rho_{\text{c}}(x) = \rho_0\cosh[(1-x)\sqrt{\beta}]/\cosh\sqrt{\beta}$, where ρ_0 can be found by imposing the flux conservation equation (3) with $\gamma = 0$. Ultimately, this leads to a reaction efficiency of

$$\left(\frac{J_2}{J_1}\right)_{\text{c}} = \frac{\alpha}{\alpha+\beta^{1/2}\tanh\beta^{1/2}}. \tag{9}$$

Comparing these two expressions, we can see that for given values of α and β the clustered E_2 configuration always generates a higher reaction flux than a uniform distribution of E_2 since $\tanh\beta^{1/2} \le \beta^{1/2}$. (Note that this situation changes considerably if we allow for escape of intermediate at the boundary in addition to loss via secondary reactions. For full details, see the Appendix.) Intuitively, this is because secondary reactions, parametrized by β, limit how far intermediate molecules diffuse away from the source at $x = 0$. Thus for a uniform profile, the effective enzyme density that

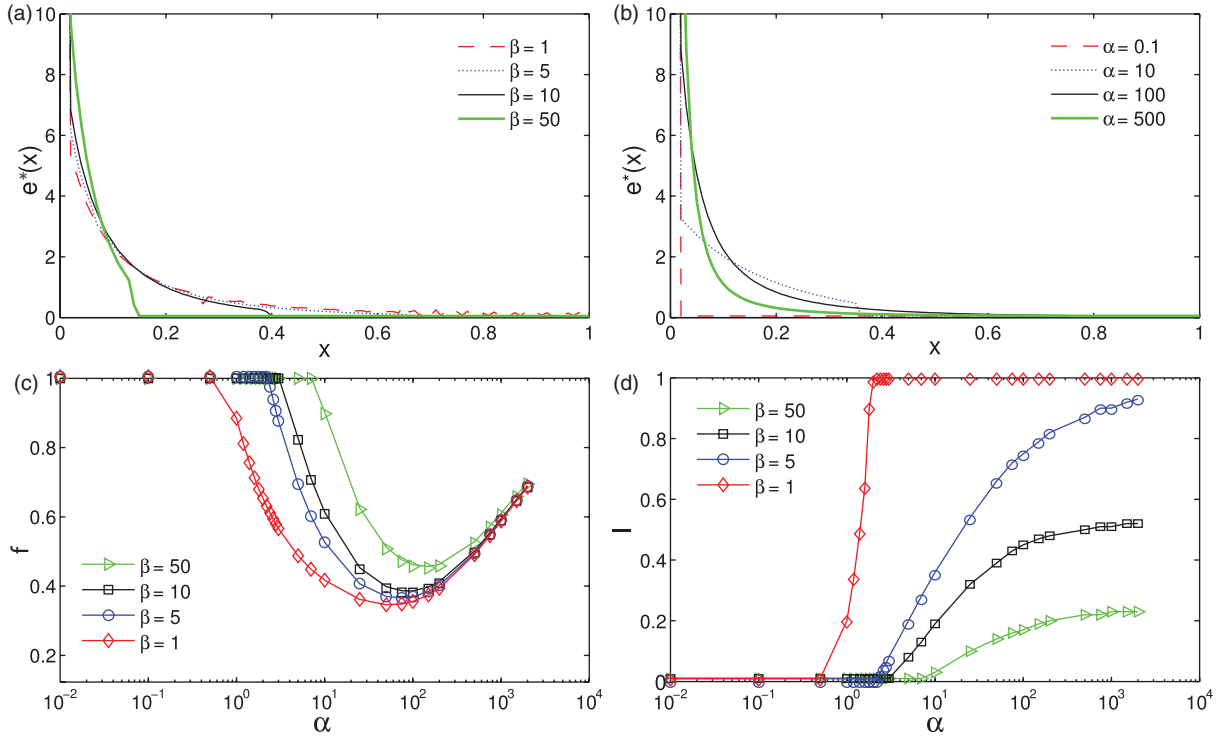

FIG. 4. Top: Optimal E_2 enzyme profiles as (a) β is varied at constant $\alpha = 50$ and (b) α is varied at constant $\beta = 5$. All the optimal profiles are obtained by numerical optimization on a lattice with $N = 100$ sites. Bottom: Quantifying the optimal profile. (c) The fraction f of E_2 enzymes that are clustered in the optimal profile shows a non-monotonic dependence on α. (d) The extent of distributed enzymes l increases monotonically with α until reaching a maximal extension that depends on β.

intermediate molecules will experience is reduced; for large β, this reduction is by a factor of order $\beta^{1/2}$. This effect can also be quantified by studying the enzyme exposure distributions corresponding to the two enzyme profiles, which are[26]

$$P_{\mathrm{u}}(E) = \beta e^{-E\beta}, \tag{10a}$$

$$P_{\mathrm{c}}(E) = \beta^{1/2} \tanh \beta^{1/2} e^{-E\beta^{1/2} \tanh \beta^{1/2}}. \tag{10b}$$

We see that $P_{\mathrm{u}}(E)$ is more concentrated near $E = 0$ than $P_{\mathrm{c}}(E)$; thus a higher proportion of trajectories rapidly react via secondary pathways before being exposed to a significant level of E_2 enzyme. These intermediate molecules, therefore, have a low probability of reacting with E_2 leading to a low pathway efficiency.

We now turn to the question of what is the E_2 profile that maximizes the reaction efficiency. We investigated this by performing a numerical optimization of $e(x)$ on a discrete lattice of N sites, as described previously.[25] Briefly, we use an evolutionary algorithm with mutation and mixing. We begin each optimization step with a trial enzyme profile $e(x)$. We generate 50 mutations of this profile, by selecting one lattice site at random and moving a random fraction of the E_2 enzymes at this site to another randomly chosen site. For each of these modified $e(x)$ profiles, the discrete reaction-diffusion system (an order-N linear system) is solved and J_2/J_1 calculated. As the initial profile for the next mutation round, we take the mean of the ten most-efficient mutant configurations in the previous round. We have found this procedure to achieve more rapid and robust convergence than a simple Monte Carlo exploration of the space of possible configurations. The optimal profiles reported below are the configurations with the highest reaction efficiency that occurred at any point during the optimization process.

The upper panels of Fig. 4 show the optimal enzyme profiles found numerically for different combinations of the parameters α and β. Importantly, we find that the fully clustered configuration is not always the optimal distribution; for different parameter values, the optimal E_2 profile can be either a fully clustered configuration at $x = 0$, or a mixed profile in which only a finite fraction of the available enzymes are clustered. This is reminiscent of the behavior of an open system with an absorbing boundary but without a competing pathway, reported previously,[25] although the shape of the enzyme profile differs.

We quantify the level of clustering in the optimal enzyme profiles (see Fig. 4, lower panels) by the fraction f of E_2 that are located at the lattice site $x = 0$, and the extent of the optimal profile via the distance l over which the optimal enzyme density is above a threshold of 10^{-3}. Examining first the behavior as β is varied (upper left) we see that for larger β the enzyme profile becomes more concentrated at smaller values of x; f increases and l decreases. This is simply because intermediate molecules typically diffuse less far from the source before reacting via the secondary pathway. Turning now to the behavior as a function of α, we see that there exists a sharp transition: below a (β-dependent) critical value of α, the optimal profile is the fully clustered configuration, $f = 1$. As the threshold is crossed, a fraction of enzymes are relocated

away from $x = 0$ and distributed over an extended region; f decreases and l increases. Interestingly, as α is increased further we find that the available enzymes tend to once again relocate towards $x = 0$; f passes though a minimum and begins to increase again. This is not accompanied by a decrease in l, but the density of enzymes is reduced at larger x and increased at smaller x. For large $\alpha \gtrsim 1000$, the optimal profiles for different values of β become more similar (with the exception of the position at which the enzyme profile cuts off sharply, which remains β-dependent).

The pattern of changes in the optimal profile can be understood as follows. When α is small the reaction efficiency is optimized by a clustered configuration since this is the enzyme distribution that maximizes the number of large-E trajectories. However, the clustered configuration also leads to a large population of trajectories, those corresponding to I molecules that rapidly diffuse away from the cluster and do not return, with extremely small values of E. At intermediate values of α, it becomes favorable to move some enzymes away from the cluster and to distribute more widely. This increases the probability of reaction for trajectories that rapidly leave the cluster, while not significantly reducing the probability of reaction for trajectories that spend a significant amount of time in the vicinity of the cluster. For large α, however, the need for these distributed enzymes is reduced, since only for those trajectories that rapidly escape via the secondary pathway is the reaction probability much less than 1. Thus, by again concentrating enzymes around $x = 0$ it is possible to maximize the probability of reaction for those trajectories that spend only a very short time in the system, and therefore do not diffuse far from $x = 0$. Here we see a significant difference from an open system where loss occurs only at the boundary, for which there is no impetus to cluster enzymes again.[25] If loss occurs only at $x = 1$, intermediate molecules must always diffuse past all E_2 molecules in order to escape from the system. However, if loss occurs in the vicinity of $x = 0$, then E_2 enzymes placed far from the source are essentially wasted.

3. Higher-dimensional systems

We now consider systems in more than one spatial dimension, beginning with a three-dimensional spherical geometry. We impose angular symmetry, such that position within the system can be parametrized by a single radial coordinate, r. We place E_1 enzymes at the center $r = 0$ of a spherical volume of radius 1. In the first instance, we neglect any secondary pathways ($\beta = 0$) and impose an absorbing boundary condition at $r = 1$. With these simplifications, Eq. (2) becomes

$$0 = r^{-2}\partial_r[r^2\partial_r\rho(r)] - \alpha e(r)\rho(r) \tag{11}$$

with the boundary conditions $[4\pi r^2\partial_r\rho(r)]_{r=0} = -1$, accounting for the production of intermediate, and $\rho(1) = 0$. This system is a three-dimensional analog of that discussed previously in Ref. 25.

We once again begin by exploring the configurations in which E_2 are either placed on a shell of radius r_0, $e_c(r) = \delta(r - r_0)/(3r_0^2)$, or uniformly distributed throughout the spherical volume, $e_u(r) = 1$ (recall that these distributions are scaled by the average density such that $V^{-1}\int_V e(\mathbf{r})d\mathbf{r} = 1$). The reaction efficiency is then given by

$$\left(\frac{J_2}{J_1}\right)_c = \frac{\frac{\alpha}{3}(1 - r_0)}{r_0 + \frac{\alpha}{3}(1 - r_0)} \tag{12}$$

for the enzyme shell and

$$\left(\frac{J_2}{J_1}\right)_u = 1 - \sqrt{\alpha}\,\mathrm{csch}\sqrt{\alpha} \tag{13}$$

for uniformly distributed enzymes.

As the shell radius r_0 approaches zero, the efficiency of reaction of intermediate with E_2 approaches one. However, this situation is not physically realistic, as enzymes have a finite size and there will be a maximum packing density which limits the potential shell radii. If r_0 is taken to be small but finite, for small α the clustering of enzymes into a tightly packed shell configuration still achieves a higher reaction flux (see Fig. 5(a)) than the uniformly distributed configuration. However, above a critical (r_0-dependent) α value, the uniform enzyme arrangement is able to achieve a higher reaction flux. This transition is analogous to that seen in the equivalent one-dimensional system.[25] However, as can be seen in Fig. 5(a) the region in which the uniform enzyme distribution is favored is shifted to much higher α values, from $\alpha \approx 9$ in the one-dimensional system up to $\alpha \approx 85$ for the three-dimensional system with $r_0 = 0.05$. Since the transition

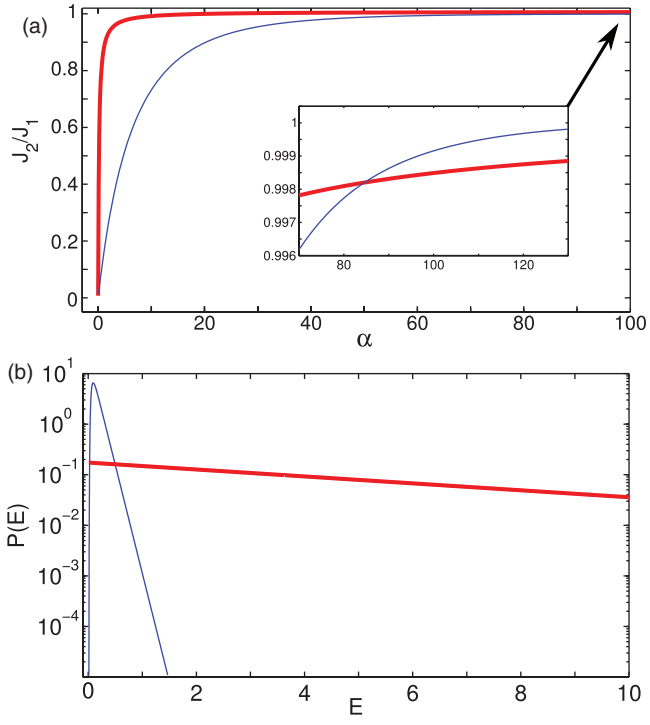

FIG. 5. (a) Comparison of the reaction flux for clustered (thick red line, $r_0 = 0.05$) and distributed (thin blue line) E_2 profiles. The clustered distribution achieves a significantly higher efficiency for small α. A transition is observed at $\alpha \approx 85$, above which the distributed profile reaches a marginally higher efficiency. (b) Enzyme exposure distributions for the clustered (thick red, $r_0 = 0.05$) and the distributed (thin blue) E_2 profiles. The enzyme exposure distribution for the distributed E_2 profile is sharply peaked at $E \ll 1$, whereas the distribution for the clustered profile extends to large values of E.

occurs at extremely large α values, for which almost all intermediate particles react, the difference between the efficiencies of the two profiles is extremely small. Furthermore, in the low-α domain the advantage provided by clustering of enzymes is much more significant than in the one-dimensional system, with an increase in the reaction flux by more than a factor of four. Thus clustering of enzymes is more strongly favored in three-dimensional than in one-dimensional systems.

Calculating the corresponding enzyme exposure distributions,

$$P_{\mathrm{u}}(E) = 2\sum_{n=1}^{\infty}(-1)^{n+1}(\pi n)^2 e^{-(\pi n)^2 E}, \tag{14a}$$

$$P_{\mathrm{c}}(E) = \frac{3r_0}{1-r_0} e^{-\frac{3r_0}{1-r_0}E}, \tag{14b}$$

we find that $P_{\mathrm{u}}(E)$ is sharply peaked around a small but finite value of E (see Fig. 5(b); the mean and variance of $P_{\mathrm{u}}(E)$ are 1/6 and 1/90, respectively). For the clustered profile, $P_{\mathrm{c}}(E)$ takes its customary exponential form. Thus, we can see that for small $\alpha \ll 1$, almost all of the weight of $P_{\mathrm{u}}(E)$ lies in the region $E \lesssim 1$ where $p_r(E)$ is small; the exponential tail of P_{c}, though, results in a larger mean value of E and more trajectories in regions where $p_r(E)$ is significant. Thus for $\alpha \ll 1$, the clustered configuration is more efficient. On the other hand when $\alpha \gg 1$, essentially all the probability weight of $P_{\mathrm{u}}(E)$ lies in the region $E \gtrsim \alpha^{-1}$, where $p_r(E) \approx 1$; however, $P_{\mathrm{c}}(E)$ is actually largest in the region $E \ll 1$ where $p_r(E)$ is small. These trajectories, which correspond to I molecules that rapidly diffuse away from the E_2 cluster and do not return, generally will not lead to reactions and thus reduce the relative efficiency of the clustered configuration.

How can the greater impact of clustering in a spherical geometry be understood intuitively? On a typical path to the boundary, a single intermediate particle originated from the center explores only a fraction of the whole sphere. If the same number of enzymes are distributed on a spherical shell further from the center, the effective "reaction cross-section" is smaller because the fraction of this shell that an intermediate will typically explore decreases, and with it the fraction of the enzymes in the system to which the intermediate molecule will be exposed. This is in contrast to the one-dimensional case, where the particle passes all enzymes before getting absorbed by the boundary. To derive a benefit from distributing enzymes, a larger α value is therefore required in three dimensions to compensate for this reduction in the effective level of E_2 enzymes to which intermediate molecules are exposed.

Next we investigated the optimal enzyme distribution as a function of the control parameter α. As noted above, if the clustering of all E_2 enzymes at $r = 0$ is permitted then $J_2/J_1 \to 1$ independent of α, yielding the maximal possible flux. However, we note again that this configuration with an infinite packing density is not physically realizable. Instead we impose a limit to the possible packing density through a minimal radius r_0 within which E_2 enzymes cannot be placed. We adapt the optimization procedure described above by solving Eq. (11) on a radial lattice where each lattice site represents a concentric shell of the system. This fixes a minimal value of $r_{0,\,\mathrm{min}} = (2N)^{-1}$ at the mid-point of the innermost shell; larger values of r_0 can be prescribed, but not smaller shells.

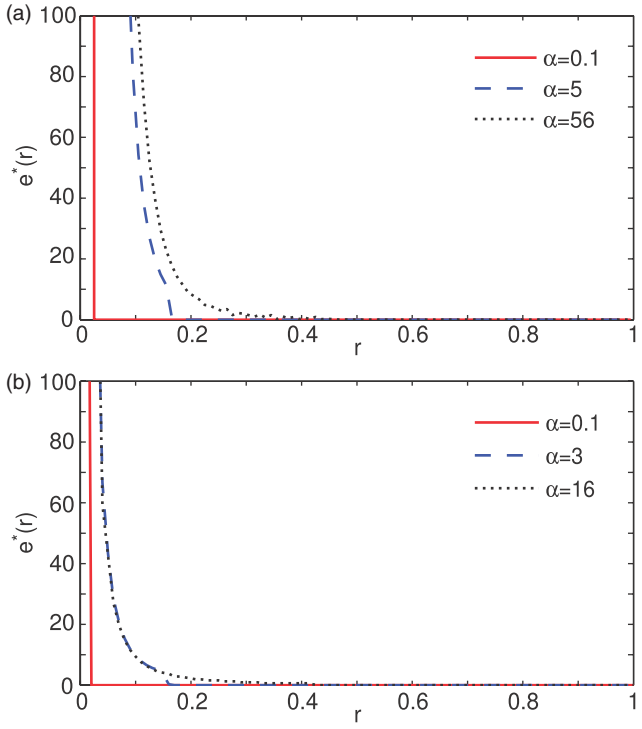

FIG. 6. Changes in the optimal enzyme profile as a function of α are qualitatively similar in two and three dimensions. (a) For the three-dimensional system described by Eq. (11), an extended region of distributed enzymes emerges for $\alpha \gtrsim 0.05$. (b) In two dimensions, the critical value for the emergence of a distributed enzyme fraction is $\alpha \approx 0.5$.

As shown in Fig. 6(a), when a finite minimal clustering radius is imposed we again find that a purely clustered configuration is optimal for small α while for α larger than a critical value an extended region of distributed enzymes emerges, in the same way as in one dimension.[25] The critical α value at which this transition occurs is strongly dependent on the minimal allowed radius, r_0. Interestingly, for $N = 100$ and $r_0 = r_{0,\,\mathrm{min}} = 0.005$ the transition point $\alpha \approx 0.05$ is significantly lower than in the one-dimensional system, despite the above arguments that distributing enzymes is less efficient in three dimensions than in one; only at $r_0 \sim 0.2$ does the transition point reach the $\alpha = 1$ observed in one dimension. The increased penalty, via the reduction in reaction cross-section, of moving enzymes to larger values of r is instead reflected in the shape of the extended enzyme "tail:" unlike in one dimension, the enzyme density in the extended region of the profile is not constant but decreases roughly as $e^*(r) \sim r^{-4}$.

We have seen that one- and three-dimensional systems have qualitatively similar phenomenology in terms of whether a clustered or uniform profile is preferable and in terms of the optimal enzyme profile, although the quantitative aspects of these transitions vary. The same also holds true for the two-dimensional case: we find that this system also displays a transition from a clustered to uniform E_2 configuration being preferable at an α value between those at which the transition occurs in one- and three-dimensional systems (data not shown). Figure 6(b) shows that once again an extended

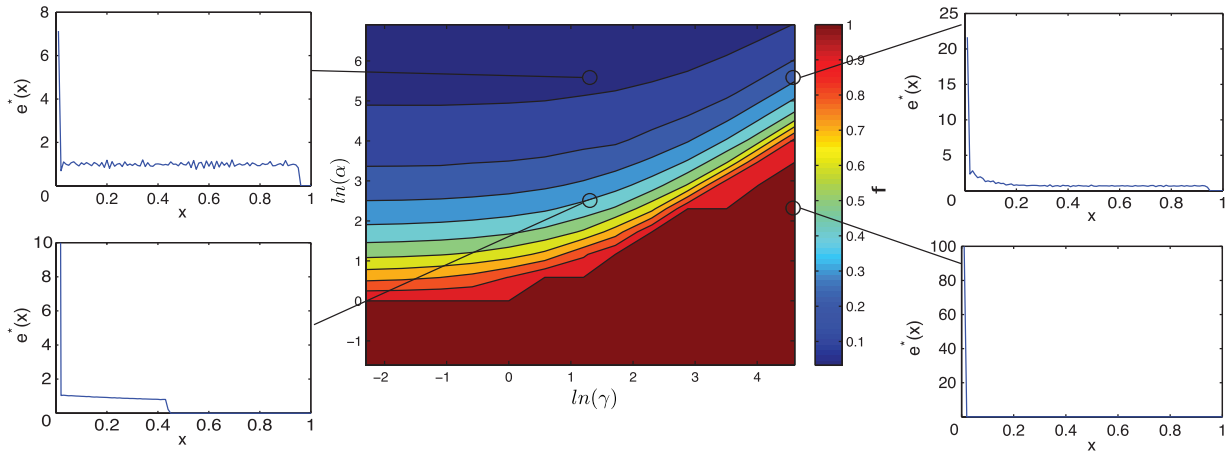

FIG. 7. Central panel: the fraction f of E_2 enzymes that are clustered in the numerically optimized enzyme distribution (on a lattice of 100 sites) as the parameters α and γ are varied. Outer panels depict the optimal profiles at specific parameter combinations, showing how the profile shape changes across the transition from clustered to fully distributed. Optimizations were run for 2×10^5 steps on a lattice of $N = 100$ sites with a solution error-tolerance of 10^{-7}.

optimal profile appears at $\alpha \approx 0.5$ (for $r_0 = 0.005$; the effect of varying r_0 is much weaker in two dimensions than in three). In the extended region, the optimal density is found to decrease as $e^*(r) \sim r^{-2}$. We can therefore see that the underlying physics of these transitions is generic, and is not dependent on the statistics of diffusion in particular dimensions.

B. Nonlinear reactions

We now turn to the case of fully nonlinear reactions. We begin by considering the case $\beta = 0$, in which there is no competition with secondary pathways for the intermediate I. We furthermore restrict ourselves to a one-dimensional system on the domain $x \in [0, 1]$, with E_1 enzymes located at $x = 0$ and an absorbing boundary condition at $x = 1$. That is, we consider a reaction-diffusion equation of the form

$$0 = \partial_x^2 \rho(x) - \frac{\alpha e(x) \rho(x)}{1 + \gamma \rho(x)}, \tag{15}$$

together with source-sink boundary conditions $\partial_x \rho|_{x=0} = -1$ and $\rho(1) = 0$, where γ is the effective saturation parameter defined in Sec. II.

We again seek to find the E_2 distribution $e(x)$ that maximizes the reaction efficiency J_2/J_1 via numerical optimization. The nonlinear nature of the reaction terms mean that the discretized reaction-diffusion equation for a given $e(x)$ no longer takes the form of a linear system that can be solved directly. Instead, we used a shooting approach to calculate $\rho(x)$ and thereby J_2. An initial trial solution for $\tilde{\rho}(x_N)$ at the rightmost lattice site is selected. This trial is then used to successively solve the nonlinear reaction-diffusion equation at the remaining lattice sites (the equation for site N depends on $\rho(x_N)$ and $\rho(x_{N-1})$, that for site $N - 1$ depends on $\rho(x_N)$, $\rho(x_{N-1})$, and $\rho(x_{N-2})$ and so on). Once a trial solution $\tilde{\rho}(x)$ has been calculated for all sites, this is tested against the reaction-diffusion equation at site x_1, which includes the source boundary condition. If the equation is satisfied to within a certain tolerance, then the solution is accepted. Otherwise, the trial value of $\tilde{\rho}(x_N)$ is refined and the process is repeated. The mutation, selection, and mixing steps of the optimization were unchanged.

Figure 7 shows results for the optimal enzyme profiles for different values of α and γ. We first verified that this solution technique accurately reproduces the results of the linear system in the limit of small γ, which should correspond the results for the linear-reaction case that have been described previously.[25] Indeed, we find that when $\alpha < 1$ all E_2 enzymes should be co-localized with the E_1 enzymes at $x = 0$. As $\alpha > 1$ is increased, the fraction f of E_2 enzymes that cluster at $x = 0$ decreases with the remaining enzymes being distributed uniformly over an extended region such that $e(x) = 1$ in this region. Figure 7 shows that the same qualitative behavior is also observed for larger values of γ. For all values of γ tested, the optimal profile undergoes a transition from fully clustered at small α to a mixed profile with a clustered fraction and extended, lower-density, region for larger α. The critical α value at which this transition occurs increases with γ, since in the fully clustered configuration a larger γ serves to reduce the effective reaction rate. Additionally, we find that the optimal profiles deviate in shape and extension away from the source. In the mixed-profile regime, the extended "tail" of enzymes need not have a constant density. This reflects the fact that for intermediate values of α the level of saturation of E_2 enzymes will vary with position. Finally, it appears that once the transition to a mixed profile has begun, the fraction of E_2 enzymes clustered at $x = 0$ decreases more quickly as α is increased if γ is large than if γ is small. We attribute this to the fact that large values of α tend to dramatically reduce the intermediate density within the system, thereby moving a system that was in the saturated regime for small α into the unsaturated regime for large α. Indeed, for extremely large values of α the optimal profile becomes independent of γ and approaches that expected in the linear reaction case.

The inclusion of nonlinear reaction terms complicates the analysis of systems of this type via enzyme exposure. This is because the probability of reaction of an individual I molecule

depends not only on the enzyme density, but also on other intermediate molecules in the system. One can define the effective enzyme activity at each position as $e(x)/[1 + \gamma\rho(x)]$ and thereby calculate an effective enzyme exposure for a trajectory as $\tilde{E} = \int_0^{t_{\text{escape}}} e(x(t))/[1 + \gamma\rho(x(t))]\mathrm{d}t$ taking $\rho(x)$ to be the solution to Eq. (15). However, it must be noted that this does not lead to a true decomposition of the reaction flux into diffusion- and reaction-dependent terms because $\rho(x)$ itself depends on the reaction parameter α.

Finally, we briefly consider the results of the full model described in Sec. II, including both a nonlinear reaction, $\gamma \neq 0$, and a competing pathway, $\beta \neq 0$, in a radially symmetric three-dimensional geometry. Figure 8(a) shows that the optimal enzyme distributions are qualitatively similar to those in Fig. 6(a) for a system without competition and with only linear reactions. However, examining the fraction of enzymes that are clustered at r_0 (Fig. 8(b)), we see that f is always larger than in the limiting case $\beta, \gamma \to 0$. This is consistent with our results above for one-dimensional systems with only competition or only nonlinearity in the reaction with E_2 enzymes, where we found that increasing the strength of either of these effects will increase the tendency for clustering of E_2 enzymes. The non-monotonic dependence of f on α that was previously found (Fig. 4(c)) when only competition is present in the model is also preserved in the case $\gamma \neq 0$. In a similar way to Fig. 7, Fig. 8(b) also suggests that the principal effect of varying γ is to alter the threshold value of α beyond which the purely clustered profile becomes sub-optimal. In summary, these results indicate that the qualitative effects of each of the modifications that we have previously considered individually are representative of the impact of the same elements in the full model.

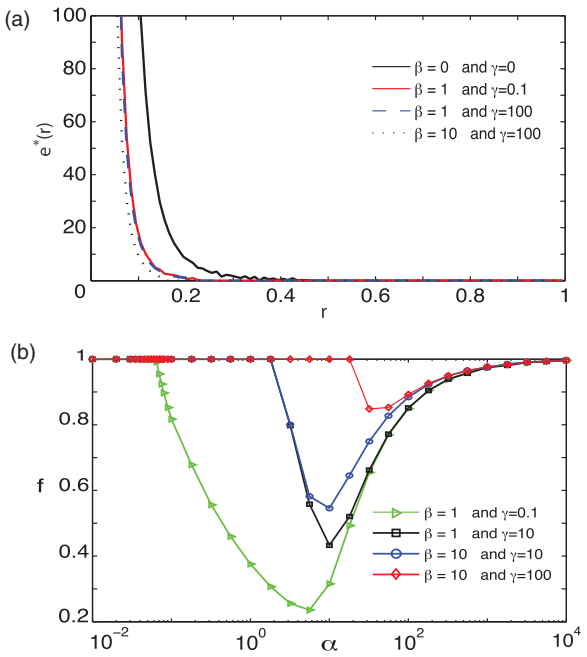

FIG. 8. (a) Optimal enzyme distributions for the full model, with parameter values as indicated, with $\alpha = 56$. (b) Clustered fraction of E_2 enzymes in the optimal profile f as a function of the catalytic activity of E_2. A non-monotonic dependence is seen when $\beta \neq 0$.

IV. DISCUSSION

In this work we have demonstrated that the existence of a transition between clustered and distributed optimal arrangements of enzymes is a generic feature of diverse reaction-diffusion systems. While the exact shape of the optimal enzyme profile and the parameter-dependence of this transition varies with the specific system, the fact that a non-trivial optimal profile exists is a general result of the interplay of reaction and diffusion in such systems. The transition ultimately emerges from the stochastic dynamics of individual intermediate molecules, as demonstrated by its dependence on the interplay between the distributions of enzyme exposure and reaction probability. By examining these distributions, we are led to an intuitive explanation for the transition. When reactions are slow, clustering of enzymes is beneficial because this provides the highest enzyme density in the region in which intermediate molecules are most likely to spend a significant amount of time. When reactions become fast, a limited density of enzymes will already ensure the rapid reaction of these molecules; in this scenario, it becomes preferable to distribute a fraction of enzymes more widely so as to provide an opportunity to react with those intermediate molecules that rapidly escape from the enzyme cluster.

For specific systems, our analysis has revealed some further notable features. In systems with competing pathways, the optimal enzyme profile tends to concentrate near the source again as the reaction rate is increased further, which results from the fact that intermediates can be lost from the region near the source as opposed to only at the boundary of the system. We have also seen that the benefit of clustering increases with the effective dimension of the system, as the increased space available to diffusive trajectories means that intermediates typically only have access to a small fraction of distributed enzymes. Finally, our results for nonlinear reactions suggest that clustering again enhances pathway efficiency if the availability of intermediates (determined by the activity of the first enzyme in the pathway) is increased. This observation suggests that it may be desirable to dynamically regulate the localization of enzymes, and specifically the formation of multi-enzyme complexes, in response to the availability of substrate or the flux of upstream reactions. A potential example of such regulation is provided by mammalian hexokinase isoform (HKII), which is thought to undergo reversible translocation between the outer membrane of mitochondria and a more diffuse cytoplasmic distribution depending on factors including glucose-6-P and GSK3,[27,28] thereby altering the relative flux of glucose through different metabolic pathways.

While there are several well-documented examples of enzyme clustering, including the pyruvate dehydrogenase and cellulosome complexes mentioned above[2,3] as well as glycolytic enzymes in various cell types,[4,29,30] enzyme clustering is not thought to be the default strategy in molecular biology. For example, while the cellulosome is a conglomeration of enzymes tethered to the outside of bacteria, other enzyme classes such as proteases[31] do not typically form tethered complexes but rather are simply secreted into the extracellular environment. We are not aware of specific enzyme systems

that display a combination of a cluster with a more diffuse arrangement near the cluster. Observation of such localization patterns will be difficult due to the relatively low density and dynamic nature of the distributed region in close proximity to the high-density cluster. Such enzyme distributions could be generated with the help of pre-existing cellular structures such as the cytoskeleton[32] or membrane sub-domains.[33] Simpler arrangements consisting of a localized and a uniform fraction would naturally arise from weak, transient interactions between the enzymes.

Our principal conclusions could be tested experimentally using, for example, the "single-molecule cut-and-paste" technique[18] or DNA origami[19,21,34] to construct specific arrangements of enzymes. Such constructs could also be incorporated into microfluidic chambers featuring localized sources and sinks of substrate. The reaction efficiency could be measured by observing the relative quantities of reacted and un-reacted substrate in the efflux channel, or using fluorogenic substrates. Such experiments would allow for quantification of the relative reaction efficiencies of different enzyme arrangements as the parameter α is varied by altering, for example, the number of enzymes in the system or the substrate diffusion constant.

While the continuous reaction-diffusion models that we have considered here present a useful mesoscopic description of the enzymatic systems, they make a number of approximations that will limit their validity at extremely short length scales. Foremost amongst these is that neither enzymes nor intermediate molecules occupy any finite volume. In reality, there will be an upper limit to the number of enzymes that can be clustered within a certain region. Furthermore, steric hindrance by enzymes will affect the trajectories of intermediate molecules. Thus, the tight packing of enzymes may strengthen the effect of clustering by physically blocking the escape of intermediates. However, by the same measure, a tight clustering of enzymes may prevent the access of initial substrates into the cluster. At such short length-scales, it is also not clear to what extent the motion of intermediates can be represented as normal diffusion. Additionally, enzymes are not reactive over their entire surfaces but only at specific catalytic binding sites. While rotational orientation can generally be neglected for freely diffusing enzymes, these effects may become significant if enzymes are attached to rigid scaffolds. A more complete understanding of these issues, together with more complex reaction schemes including cooperativity and allosteric regulation, will be crucial for a complete understanding of the design principles underlying enzyme arrangements in living cells as well as the effective engineering of synthetic biochemical systems.

ACKNOWLEDGMENTS

This research was supported by the German Excellence Initiative via the program "Nanosystems Initiative Munich" and the German Research Foundation via the SFB 1032 "Nanoagents for Spatiotemporal Control of Molecular and Cellular Reactions" and via the Focus area SPP 1617. F.T. is supported by a research fellowship from the Alexander von Humboldt Foundation.

APPENDIX: COMPETING PATHWAY WITH ABSORBING BOUNDARY CONDITIONS

Here we briefly consider a system described by Eq. (7), but with an absorbing boundary condition $\rho(1) = 0$ rather than the reflecting boundary considered in Sec. III A 2. Comparing the known results for the limit $\beta \to 0$,[25] and for the reflecting boundary condition discussed above, we can predict that there should be a qualitative difference in whether clustered or uniform profiles are preferred as β is varied. For small $\beta \ll 1$, loss to the competing pathway will be negligible compared to loss through the boundary at $x = 1$. We would, therefore, expect that as α is increased, the system should undergo a transition from a regime in which the clustered configuration is preferable to a regime in which the uniform profile provides a higher efficiency. On the other hand, for large β the length scale associated with the loss to the secondary pathway is short compared to the system size. In this case, the choice of boundary condition of $x = 1$ should have little influence on the dynamics, which should resemble that described in Sec. III A 2 where the clustered configuration is always preferable.

The reaction fluxes obtained by solving this system for uniform and clustered enzyme configurations are

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{u}} = \frac{\alpha}{\alpha+\beta}\left(1 - \operatorname{sech}\sqrt{\alpha+\beta}\right), \tag{A1}$$

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{c}} = \frac{\alpha}{\alpha + \sqrt{\beta}\coth\sqrt{\beta}}. \tag{A2}$$

Figures 9(a) and 9(b) confirm that there is a difference in whether the clustered or uniform profile is more efficient in different regimes of β, in keeping with our expectations. Figure 9(c) plots the critical value of α as a function of β and

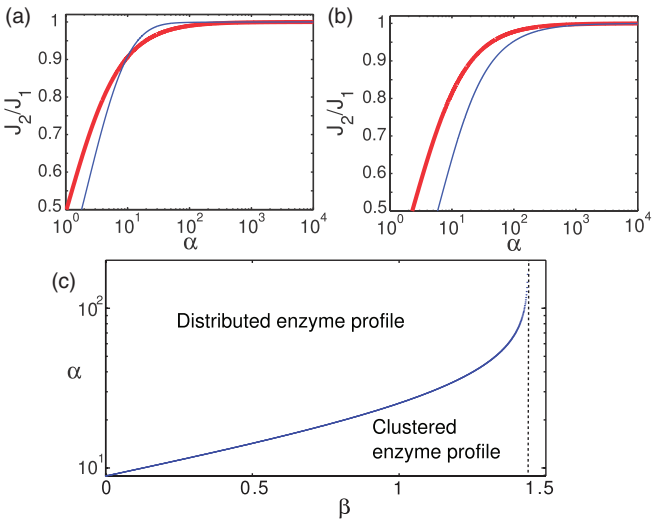

FIG. 9. ((a) and (b)) Comparison of the reaction efficiency for clustered (thick red line) and distributed (thin blue line) E_2 enzymes reveal a qualitatively different behavior in different β ranges. For $\beta = 0.1$ (a) a cross-over is observed at which the profile with the higher efficiency changes. For $\beta = 5$ (b) clustering is preferable for all values of α. (c) The transition value of α dividing the regimes where the clustered or uniform profiles achieve a higher reaction flux, found by solving $(J_2/J_1)_{\mathrm{u}} = (J_2/J_1)_{\mathrm{c}}$ numerically. Above the critical $\beta_c \approx 1.4$ (dashed line) the transition disappears and the clustering of enzymes is always more efficient.

demonstrates that the transition disappears at a finite value of $\beta_c \approx 1.4$. When $\beta > \beta_c$ the clustered configuration always achieves a higher reaction flux.

[1]P. A. Srere, Annu. Rev. Biochem. **56**, 89 (1987).
[2]E. A. Bayer, H. Chanzy, R. Lamad, and Y. Shoham, Curr. Opin. Struct. Biol. **8**, 548 (1998).
[3]A. de Kok, A. F. Hengeveld, A. Martin, and A. H. Westphal, Biochim. Biophys. Acta **1385**, 353 (1998).
[4]M. E. Campanella, H. Chu, and P. S. Low, Proc. Natl. Acad. Sci. U.S.A. **102**, 2402 (2005).
[5]M. F. Dunn, D. Niks, H. Ngo, T. R. Barends, and I. Schlichting, Trends Biochem. Sci. **33**, 254 (2008).
[6]F. H. Gaertner, Trends Biochem. Sci. **3**, 63 (1978).
[7]J. Ovádi, J. Theor. Biol. **152**, 1 (1991).
[8]R. Heinrich, S. Schuster, and H.-G. Holzütter, Eur. J. Biochem. **201**, 1 (1991).
[9]A. Cornish-Bowden, Eur. J. Biochem. **195**, 103 (1991).
[10]P. Mendes, D. B. Kell, and H. V. Westerhoff, Eur. J. Biochem. **204**, 257 (1992).
[11]A. Cornish-Bowden and M. L. Cárdenas, Eur. J. Biochem. **213**, 87 (1993).
[12]P. Mendes, D. B. Kell, and H. V. Westerhoff, Biochim. Biophys. Acta **1289**, 175 (1996).
[13]R. J. Conrado, T. J. Mansell, J. D. Varner, and M. P. DeLisa, Metab. Eng. **9**, 355 (2007).
[14]S. B. van Albada and P. R. ten Wolde, PLOS Comput. Biol. **3**, e195 (2007).
[15]A. Mugler, A. Gotway Bailey, K. Takahashi, and P. R. ten Wolde, Biophys. J. **102**, 1069 (2012).
[16]R. J. Conrado, J. D. Varner, and M. P. De Lisa, Curr. Opin. Biotechnol. **19**, 492 (2008).
[17]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, Nat. Biotechnol. **27**, 753 (2009).
[18]S. K. Kufer, E. M. Puchner, H. Gumpp, T. Liedl, and H. E. Gaub, Science **319**, 594 (2008).
[19]S. M. Douglas, H. Dietz, T. Liedl, B. Högberg, F. Graf, and W. M. Shih, Nature (London) **459**, 414 (2009).
[20]C. M. Niemeyer, J. Koehler, and C. Wuerdemann, ChemBioChem **3**, 242 (2002).
[21]O. I. Wilner, Y. Weizmann, R. Gill, O. Lioubashevski, R. Freeman, and I. Willner, Nat. Nanotechnol. **4**, 249 (2009).
[22]O. Idan and H. Hess, Curr. Opin. Biotechnol. **24**, 606 (2013).
[23]C. J. Delebecque, A. B. Lindner, P. A. Silver, and F. A. Aldaye, Science **333**, 470 (2011).
[24]J. Fu, M. Liu, Y. Liu, N. W. Woodbury, and H. Yan, J. Am. Chem. Soc. **134**, 5516 (2012).
[25]A. Buchner, F. Tostevin, and U. Gerland, Phys. Rev. Lett. **110**, 208104 (2013).
[26]See supplementary material at http://dx.doi.org/10.1063/1.4823504 for full details of this and following calculations.
[27]J. E. Wilson, Trends Biochem. Sci. **3**, 124 (1978).
[28]S. John, J. N. Weiss, and B. Ribalet, PLoS ONE **6**, e17674 (2011).
[29]D. T. Sullivan, R. MacIntyre, N. Fuda, J. Fiori, J. Barrilla, and L. Ramizel, J. Exp. Biol. **206**, 2031 (2003).
[30]L. E. Anderson, N. Gatla, and A. A. Carol, Photosynth. Res. **83**, 317 (2005).
[31]C. Wandersman, Mol. Microbiol. **3**, 1825 (1989).
[32]C. Masters, J. Cell Biol. **99**, 222s (1984).
[33]K. Simons and W. L. C. Vaz, Annu. Rev. Biophys. Biomol. Struct. **33**, 269 (2004).
[34]P. W. K. Rothemund, Nature (London) **440**, 297 (2006).

4.5 Conclusion

In this part of the thesis we considered an enzymatic pathway consisting of two different types of enzymes E_1 and E_2. Arranging enzymes E_2 relatively to E_1 to maximize the efficiency of a pathway leads to a surprising transition between clustered E_2 enzymes and distributed. At first we investigated a one dimensional system with one control parameter α. This parameter governs the ratio between the escape of an intermediate I generated by E_1 through the boundary or the reaction with one of the E_2 enzymes. To each value of α an optimal enzyme profile is associated. Small α favors a clustering of enzymes E_2 at E_1, but when we consider higher values $\alpha > 1$ so called back up enzymes emerge and eventually the enzymes E_2 distribute uniformly over the entire system. Intuitively, clustering E_2 at the source should always be preferred, since accumulating the enzymes at the maximum of the intermediate density should lead to maximal efficiency. To unravel this seemingly counterintuitive phenomena we introduce the concept of enzyme exposure, explaining the stochastic nature of the back up enzymes. Heuristically, once α is large enough the cost of relocating some of the enzymes away from the cluster is small since the probability for an intermediate to react at the cluster remains high. The gain, however, to move these enzymes towards the bulk is increased by creating the portability to catch intermediates that escaped the cluster. Although diffusive properties change throughout the dimension the general feature of emerging back up enzymes persists. We showed that the appearance of back up enzymes is an universal feature of such a pathway we have considered. They still emerge at higher dimensions, different geometries and even when a competing reaction for the intermediates, whose strength is regulated by β, is introduced. Additionally, when considering non-linear MM reaction of the intermediate with E_2, where the parameter γ governs the strength of the non-linearity, again the back up enzymes appear. However, the exact shape of the enzyme strongly depends on the individual values of α, β and γ.

Chapter 5

Outlook

In this thesis we have investigated the impact of forcing an enzyme out of its equilibrated ground state on the activity. A change in conformation causes a deactivation of the enzyme. It has been argued, that the structure of enzymes influences the specificity towards a substrate. We have assumed a linear one dimensional random walker model to describe the folding pathway of the enzyme. Thus we have assumed that the enzyme follows always the same folding trajectory through the energy landscape. By introducing a branched random walker model we believe to account for different folding pathways. Generally, we believe that the mean first passage time formalism is still suitable for describing a branched folding path. The kinetic and thermodynamic parameters can be obtained using the fit routine we have established in the manuscript. This will allow for a more precise investigation of the underlying landscape. On the other hand we have studied the collective enzyme activity depending on spatial arrangement of enzymes. Surprisingly, clustering the enzymes E_2 in the vicinity of the highest substrate concentration proofs not always to be the most effective way to maximize the efficiency. In fact extending the enzymes over a lager range, away from highest substrate concentration optimizes the efficiency of the pathway in the case of large reaction rate. We observe an emergence of back up enzymes. While the shape of the enzyme profile depends on type of reaction, geometry, dimensionality and the presence of competing pathways the back up enzymes however are an universal feature of that particular enzymes cascade. Our reaction diffusion description neglects the fact that enzymes and their substrates respectively their products have a finite size, consequently they can not both occupy the same space. We believe that combining the structural aspect of enzymes, as discussed in the first part of the thesis and their arrangement in space will lead to new effects. As a first step for further investigation we thus suggest to consider a system where the enzymes E_2 posse a finite extension. There will be several new features to concern with. Assuming a three dimensional spherical system with the source still in the center and neglecting competing pathways. For reasons of simplicity we merely considering a situation where the enzymes E_2 only can be arranged on a sphere. Obviously it is best to arrange the enzymes relative to the source such that the intermediates are hindered form reaching the boundary. Due to this steric effect all of the intermediates are eventually consumed by the enzymes E_2, which form a cage for the intermediates. In other

words, the enzymes confine the intermediates which leads to an efficiency of one for that particular system, since no intermediate particles are lost. However in a more realistic system we have to take into account that the enzymes E_1 representing the source also need substrate particles to provide intermediates for the enzymes E_2 forming the sphere. The enclosure of intermediates due to the enzyme sphere also leads to an exclusion of the substrate particle for the enzymes E_1 at the source, hence the efficiency drops to zero. A spherical arrangement maximizing the efficiency has to optimize the balance between enclosing intermediates and excluding substrates for E_1. This optimal arrangement will depend on the reaction rates and on the size of the different molecules, such as the substrates, the intermediates and enzymes E_2. For an unrestricted arrangement of the enzymes E_2 relative to E_1 we expect that the back up enzymes still occur however the exact arrangement of the enzymes will be influenced by the confinement effect due to their finite size. We have encountered such an effect when we contemplated the pyruvate dehydrogenase complex. The pyruvate decarboxylase form a cage only to leave the coenzymes for the inner enzymes, the dihydrolipoyl acetyltransferase, through. Further this shielding of larger molecules form the cellular milieu ensures the swinging arm to operate freely within the cage, mediating substrates between the active centers of the enzymes. As we have seen, a change in the structure of the enzymes causes a deactivation of the enzymes suggesting that our enzyme cascade model should be extended to incorporate such a deactivation. In a biological system this is usually achieved by inhibitors that change the conformation of the enzyme. The activity of the enzyme is solely confined to a small portion of the entire enzyme, it would be interesting to probe the implication of the orientation of the active center on the efficiency. For instance in the case where E_1 is still placed at the center of a sphere, intuitively we would expect the active centers should be oriented towards the source. Since the intermediate concentration is a decreasing function towards the outer boundary, the active center is exposed to a higher concentration of the intermediates once it faces to the center instead of to the outer boundary. The change of the intermediate density is determined by the size of the enzyme E_2. On the other hand we have already seen an optimal arrangement of enzymes at positions where the intermediate density does not show a maximum. We leave this question, whether and how the orientation of the active center influences the efficiency of the pathway to future investigations. Further we still lack a rigorous analytic solution for the optimal enzyme profile that maximizes the efficiency under the condition that the reaction diffusion equation has to be satisfied. The most promising approach is to write the optimization in terms of a variation of the reaction flux under the condition of satisfying the reaction diffusion equation and the conservation of the total number of enzymes E_2. We expect further insights into the physical origin of the effective enzyme activity for non-linear reactions. A more complete understanding of enzyme positioning will be obtained once we extend our model to more complicated reaction, such as including cooperativity, multiple binding sites or allosteric effects. These mechanism are indispensable for the regulation of cellular process and therefore of great interest for the creation of an artificial cell. We believe that extending the linear pathway to a higher number of different types of enzymes will lead to new yet undiscovered effects. Even a formation of a stripe-like pattern for the different enzymes might be possible. We

hope that with this work we could shed new light on the interesting but also highly complex topic of clustering of enzymes.

Appendix A

Supplementary Material

A.1 Clustering and Optimal Arrangement of Enzymes in Reaction-Diffusion Systems

Supplementary Material for "Clustering and optimal arrangement of enzymes in reaction-diffusion systems"

Alexander Buchner,* Filipe Tostevin,* and Ulrich Gerland†
Arnold Sommerfeld Center for Theoretical Physics and Center for Nanoscience, Ludwig-Maximilians-Universität München, Germany
(Dated: March 12, 2013)

Derivation of enzyme exposure distribution $P(E)$

Uniform configuration

In the case of a uniform enzyme profile, $e_{\mathrm{u}}(x) = \bar{e}$, the value of E for an individual substrate trajectory is simply proportional to the time taken to reach the absorbing boundary at $x = L$,

$$E = \frac{D}{L^2\bar{e}} \int_0^\tau \mathrm{d}t\, e(x(t)) = \frac{D}{L^2\bar{e}} \int_0^\tau \mathrm{d}t\, \bar{e} = \frac{D\tau}{L^2}. \tag{S1}$$

Thus the distribution $P(E)$ is determined by the distribution of escape times at the absorbing boundary, $f(\tau)$. The calculation of the first-passage time distribution for a diffusing particle [1] is included here for completeness. We begin from the renewal equation

$$p(L,t|0,0) = \int_0^t \mathrm{d}\tau\, f(\tau) p(L,t|L,\tau). \tag{S2}$$

Here $p(x,t|x',t')$ is the probability of a diffusing particle being found at position x at time t given that it was at position x' at time t', which is given by the solution to the diffusion equation on the semi-infinite domain $x \geq 0$ with a reflecting boundary at $x = 0$,

$$p(x,t|x',t') = \frac{1}{\sqrt{4\pi D(t-t')}} \left[e^{-\frac{(x-x')^2}{4D(t-t')}} + e^{-\frac{(x+x')^2}{4D(t-t')}} \right]. \tag{S3}$$

Taking the Laplace transform of Eq. S2 with respect to t we obtain

$$\tilde{f}(z) = \frac{\tilde{p}(L,z|0,0)}{\tilde{p}(L,z|L,0)}. \tag{S4}$$

Substituting in the Laplace transform of $p(x,t|x,t')$ with respect to $t - t'$,

$$\tilde{p}(x,z|x',0) = \frac{1}{\sqrt{4Dz}} \left[e^{-(x-x')\sqrt{\frac{z}{D}}} + e^{-(x+x')\sqrt{\frac{z}{D}}} \right], \tag{S5}$$

we find $\tilde{f}(z) = \operatorname{sech}\sqrt{zL^2/D}$. The escape time distribution $f(\tau)$ can be recovered by noting that $\tilde{f}(z)$ has an infinite series of poles at $z = -\frac{\pi^2 D}{L^2}(n+1/2)^2$ for $n = 0, 1, 2\ldots$, with associated residues $(-1)^n(2n+1)\pi D/L^2$, yielding

$$f(\tau) = \frac{\pi D}{L^2} \sum_{n=0}^{\infty} (-1)^n (2n+1) e^{-\pi^2 (n+1/2)^2 \tau D/L^2}. \tag{S6}$$

Equation 4 of the main text follows from applying Eq. S1.

Clustered configuration

To calculate $P(E)$ for the clustered enzyme configuration we consider the enzyme profile

$$e(x) = \begin{cases} \frac{\bar{e}L}{\delta x} & 0 \leq x < \delta x \\ 0 & \delta x \leq x < L \end{cases}. \tag{S7}$$

Thus for any given trajectory, E is related to the total time T spent in the region $0 \leq x < \delta x$ by $E = DT/(L\delta x)$.

Molecules are introduced into the system at $x = 0$. The distribution of times at which the intermediate leaves the region $0 \leq x < \delta x$ for the first time, $f_1(\tau)$, can be calculated as described in the previous section, and is given by Eq. S6 with L replaced by δx. Once the molecule has left the domain of enzymes, it can either diffuse to $x = L$ and escape from the system, or can diffuse back into the domain $x < \delta x$. The latter will occur with probability $p_{\rm ret} = 1 - \epsilon/(L - \delta x)$ if the molecule is initially located at a small displacement $+\epsilon$ from the boundary $x = \delta x$ [1]. For a molecule which re-enters the domain $0 \leq x < \delta x$, the distribution of times until it subsequently leaves again can be calculated via the procedure described above. Assuming once again a small displacement $-\epsilon$, the escape time distribution $f_2(\tau)$ satisfies the corresponding renewal equation

$$p(\delta x, t|\delta x - \epsilon, 0) = \int_0^t \mathrm{d}\tau\, f_2(\tau) p(\delta x, t|\delta x, \tau), \tag{S8}$$

and has the Laplace transform

$$\tilde{f}_2(z) = \frac{\cosh\left[(\delta x - \epsilon)\sqrt{z/D}\right]}{\cosh\left[\delta x\sqrt{z/D}\right]}. \tag{S9}$$

Since multiple rounds of return are possible, the overall distribution of times spent in the domain $0 \leq x < \delta x$ can be expressed in terms of a series of convolutions,

$$\begin{aligned} f(T) = (1 - p_{\rm ret}) f_1(T) + p_{\rm ret}(1 - p_{\rm ret}) \int_0^T \mathrm{d}\tau'\, f_1(\tau') f_2(T - \tau') \\ + p_{\rm ret}^2 (1 - p_{\rm ret}) \int_0^T \int_{\tau'}^T \mathrm{d}\tau'' \mathrm{d}\tau'\, f_1(\tau') f_2(\tau'' - \tau') f_2(T - \tau'') + \ldots \end{aligned} \tag{S10}$$

In Eq. S10, the first term represents the probability that the molecule spends a time T traversing the domain containing enzymes and then escapes from the system; the second term represents the probability that the molecule returns to the domain $0 \leq x < \delta x$ once after it initial leaves, and spends a total time T in the domain; the third term contains the probability that the molecule returns twice, and so on.

Equation S10 can be expressed concisely in the Laplace domain,

$$\tilde{f}(z) = (1 - p_{\rm ret}) \tilde{f}_1(z) \sum_{n=0}^{\infty} \left[p_{\rm ret} \tilde{f}_2(z)\right]^n = \frac{(1 - p_{\rm ret}) \tilde{f}_1(z)}{1 - p_{\rm ret} \tilde{f}_2(z)}. \tag{S11}$$

Substituting in the expressions for $p_{\rm ret}$ and $\tilde{f}_i(z)$, taking the limit $\epsilon \to 0$, we ultimately find that for small δx

$$\tilde{f}(z) \approx \frac{1}{1 + zL\delta x/D}, \tag{S12}$$

for which the inverse Laplace transform can be performed straightforwardly to yield

$$f(T) = \frac{D}{L\delta x} e^{-TD/L\delta x}. \tag{S13}$$

Transforming from T to E, we recover $P(E) = \exp(-E)$. Importantly, while $f(T)$ becomes ill-defined in the limit $\delta x \to 0$, $P(E)$ does not suffer this problem.

Numerical optimization of enzyme profiles

We studied the optimization of enzyme profiles using a stochastic algorithm consisting of multiple rounds of modification of the enzyme profile and mixing of the best-performing profiles, as described below. This procedure achieved a higher maximal flux, and required less computation time to converge to this optimal profile, than simulated annealing of the enzyme profile using the same mutation procedure at each iteration.

We discretized the domain $0 \leq x \leq L$ into N lattice sites with lattice spacing $dx = L/N$. Each optimization run was initialized with a uniform enzyme distribution, $e(x_i) = 1$ for each of the $i = 1..N$ lattice sites. At each iteration

of the optimization process, a set of 50 new test profiles were generated by selecting one site at random and moving a random fraction of the enzymes present to another randomly-selected site. For each of these modified enzyme configurations, the steady-state $\rho(x)$ was calculated by solving the system of N discrete reaction-diffusion equations,

$$-\frac{J_1}{dx} = \frac{D}{dx^2}\left[\rho(x_2) - \rho(x_1)\right] - \kappa e(x_1)\rho(x_1) \tag{S14a}$$

$$0 = \frac{D}{dx^2}\left[\rho(x_{i+1}) - 2\rho(x_i) + \rho(x_{i-1})\right] - \kappa e(x_i)\rho(x_i) \quad \text{for} \quad i = 2..N-1 \tag{S14b}$$

$$0 = \frac{D}{dx^2}\left[\rho(x_{N-1}) - 2\rho(x_N)\right] - \kappa e(x_N)\rho(x_N). \tag{S14c}$$

Equations S14a and S14c incorporate the source and sink boundary conditions at $x = 0$ and $x = L$ respectively. From each solution $\rho(x)$, the reaction flux is calculated as

$$J_2 = J_1 - \frac{D}{dx}\rho(x_N). \tag{S15}$$

The initial enzyme distribution for the next round of modifications is constructed by taking the mean of the 10 enzyme profiles with the highest J_2 values.

The results shown in Fig. 3 of the main text show the individual enzyme profiles which produced the highest J_2 throughout the entire optimization process. Multiple realizations of this optimization procedure produced the same optimal profile, suggesting that the observed profiles represent the global optimum. The optimal profiles generated were found to be highly robust to changes in the fineness of the discretization N, the number of trial profiles and the number of profiles contributing to the average at each iteration, as well as to the initial enzyme configuration at the start of the optimization.

* These authors contributed equally to this work.
† Electronic address: gerland@lmu.de

[1] S. Redner *A Guide to First-Passage Processes.* Cambridge University Press, 2001.

A.2 Optimization of collective enzyme activity via spatial localization

Supplementary Material for "Optimization of collective enzyme activity via spatial localization"

Alexander Buchner, Filipe Tostevin, Florian Hinzpeter, and Ulrich Gerland*
Arnold Sommerfeld Center for Theoretical Physics and Center for NanoScience, Ludwig-Maximilians-Universität, 80333 München, Germany

I. EQUIVALENCE OF EXPRESSIONS FOR THE REACTION EFFICIENCY

The reaction-diffusion equation leads us to define the reaction efficiency according to the Eq. 3 of the main text,

$$\frac{J_2}{J_1} = \int \alpha e(\mathbf{r}) \frac{\rho(\mathbf{r})}{1+\gamma\rho(\mathbf{r})} \mathrm{d}\mathbf{r}. \tag{S1}$$

Here we show that the alternative expression (Eq. 5 of the main text)

$$\frac{J_2}{J_1} = \int_0^\infty P(E) p_\mathrm{r}(E) \mathrm{d}E, \tag{S2}$$

which arises from the examination of individual intermediate trajectories, can be derived from Eq. S1 in the linear regime where $\gamma \to 0$.

We begin by reformulating the steady-state density $\rho(\mathbf{r})$ in terms of trajectories of diffusing molecules of intermediate. We denote the diffusive trajectory of a single intermediate molecule, in the absence of any E_2 enzymes, as $\{\mathbf{r}(t)\}$. Such a trajectory has associated with it a time t_escape after which the trajectory is terminated, either by escape across the system boundary or loss to a secondary pathway. The reintroduction of E_2 enzymes according to the distribution $e(\mathbf{r})$ leads to an instantaneous propensity for conversion to the correct product at each point along the trajectory of $\alpha e(\mathbf{r}(t))$. Thus the survival probability $S(t|\{\mathbf{r}(t)\})$ that an intermediate molecule on the trajectory $\{\mathbf{r}(t)\}$ has not undergone a reaction with E_2 before the time t follows $\dot{S}(t|\{\mathbf{r}(t)\}) = -\alpha e(\mathbf{r}(t)) S(t|\{\mathbf{r}(t)\})$. This equation can be integrated to yield

$$S(t|\{\mathbf{r}(t)\}) = \exp\left[-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))\right], \quad t \le t_\mathrm{escape}. \tag{S3}$$

At steady state, each trajectory $\{\mathbf{r}(t)\}$ of an intermediate molecule will make a contribution to the total intermediate density at point $\mathbf{r}$ that depends on the total time that the trajectory spends at $\mathbf{r}$, weighted by the probability that the intermediate molecule has not yet undergone a reaction prior to each return to $\mathbf{r}$. This later weighting factor is simply the survival probability $S(t|\{\mathbf{r}(t)\})$. Therefore, the local enzyme density can be rewritten as

$$\rho(\mathbf{r}) = \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{t_\mathrm{escape}} \mathrm{d}t\ S(t|\{\mathbf{r}(t)\}) \delta\left[\mathbf{r} - \mathbf{r}(t)\right], \tag{S4}$$

where the inner integral is the weighted time spent by a single trajectory at $\mathbf{r}$, and the outer integral sums over the contributions of all possible trajectories weighted by the probability $p(\{\mathbf{r}(t)\})$ of a specific trajectory $\{\mathbf{r}(t)\}$ occurring. Substituting Eqs. S4 and S3 into Eq. S1 and changing the order of integration, we find

$$\begin{aligned}
\frac{J_2}{J_1} &= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{t_\mathrm{escape}} \mathrm{d}t \int \mathrm{d}\mathbf{r}\ \alpha e(\mathbf{r}) e^{-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))} \delta\left[\mathbf{r} - \mathbf{r}(t)\right] \\
&= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{t_\mathrm{escape}} \mathrm{d}t\ \alpha e(\mathbf{r}(t)) e^{-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))} \\
&= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \int_0^{t_\mathrm{escape}} \mathrm{d}t\ \frac{\mathrm{d}}{\mathrm{d}t}\left[-e^{-\alpha \int_0^t \mathrm{d}t'\ e(\mathbf{r}(t'))}\right] \\
&= \int \mathrm{d}\{\mathbf{r}(t)\}\ p(\{\mathbf{r}(t)\}) \left[1 - e^{-\alpha \int_0^{t_\mathrm{escape}} \mathrm{d}t'\ e(\mathbf{r}(t'))}\right].
\end{aligned} \tag{S5}$$

*Electronic address: gerland@lmu.de

Finally, defining $E = \int_0^{t_{\text{escape}}} \mathrm{d}t\ e(\mathbf{r}(t))$ we can change the variable of integration from $\{\mathbf{r}(t)\}$ to E, recovering

$$\frac{J_2}{J_1} = \int_0^\infty P(E)(1 - e^{-\alpha E})\mathrm{d}E. \tag{S6}$$

II. ONE DIMENSION INCLUDING A COMPETING PATHWAY

We consider the rescaled reaction diffusion equation as stated in the main text

$$0 = \partial x^2 \rho(x) - \alpha e(x)\rho(x) - \beta\rho(x), \tag{S7}$$

with the source boundary conditions $\partial_x \rho(x)|_{x=0} = -1$. Below we consider the cases of reflecting ($\partial_x \rho(x)|_{x=1} = 0$) and absorbing ($(\rho(x)|_{x=1} = 0$) boundaries at $x = 1$.

A. Reflecting boundary, $\partial_x \rho(x)|_{x=1} = 0$

1. *Clustered enzyme profile*

The enzyme profile is taken to be clustered at some point x_0, $e_{\text{c}}(x) = \delta(x - x_0)$. In the end we take the limit x_0 goes to zero, leading to a clustering of E_2 at the origin. We divide the system into two parts, part I where $x < x_0$ and part II where $x > x_0$. In each part Eq. S7 reduces to

$$0 = \partial x^2 \rho(x) - \beta\rho(x), \tag{S8}$$

which has the solution

$$\rho_i(x) = A_i e^{\sqrt{\beta}x} + B_i e^{-\sqrt{\beta}x} \tag{S9}$$

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

$$A_I - B_I = (\sqrt{\beta})^{-1} \qquad A_{II} e^{\sqrt{\beta}} - B_{II} e^{-\sqrt{\beta}} = 0. \tag{S10}$$

In order to determine all constants we impose two additional conditions, firstly the matching condition of the concentration of intermediates at x_0, $\rho_I(x_0) = \rho_{II}(x_0)$ leading to

$$A_I e^{\sqrt{\beta}x_0} + B_I e^{-\sqrt{\beta}x_0} = A_{II} e^{\sqrt{\beta}x_0} + B_{II} e^{-\sqrt{\beta}x_0} \tag{S11}$$

The second condition, which captures particle conservation in the system, is found by integrating Eq. S7 from $x_0 - \epsilon$ to $x_0 + \epsilon$ and taking the limit of small ϵ,

$$\lim_{\epsilon \to 0} \left((\partial_x \rho_{II}(x))_{x_0+\epsilon} - (\partial_x \rho_I(x))_{x_0-\epsilon} - \alpha\rho(x_0) - \beta \int_{x_0-\epsilon}^{x_0+\epsilon} \rho(x)dx \right) = 0. \tag{S12}$$

The last term on the left hand side vanishes in the limit $\epsilon \to 0$. This leads to the expression

$$\left[A_{II} e^{\sqrt{\beta}x_0} - B_{II} e^{-\sqrt{\beta}x_0} - \left(A_I e^{\sqrt{\beta}x_0} + B_I e^{-\sqrt{\beta}x_0} \right) \right] - \frac{\alpha}{\sqrt{\beta}} \left(A_I e^{\sqrt{\beta}x_0} + B_I e^{-\sqrt{\beta}x_0} \right) = 0. \tag{S13}$$

Calculate the reaction current yields

$$\frac{J_2}{J_1} = \alpha \int_0^1 \delta(x - x_0)\rho(x)dx = \alpha(A_{II} e^{\sqrt{\beta}x_0} + B_{II} e^{-\sqrt{\beta}x_0}) \tag{S14}$$

After some straightforward algebra we arrive at expressions for all four constants $\{A_I, A_{II}, B_I, B_{II}\}$. Last we plug them into the above equation and take the limit $x_0 \to 0$,

$$\left(\frac{J_2}{J_1} \right)_{\text{c}} = \frac{\alpha}{\alpha + \beta^{\frac{1}{2}} \tanh(\beta^{\frac{1}{2}})}. \tag{S15}$$

2. Uniform enzyme profile

The reaction diffusion equation with a uniform enzyme profile $e_{\mathrm{u}}(x) = 1$ reads

$$0 = \partial x^2 \rho(x) - (\alpha + \beta)\rho(x). \tag{S16}$$

The solution is given by

$$\rho(x) = Ae^{\sqrt{\alpha+\beta}x} + Be^{-\sqrt{\alpha+\beta}x} \tag{S17}$$

Applying the boundary conditions at $x = 0$ and $x = L$ leads to the conditions

$$A - B = (\sqrt{\alpha+\beta})^{-1} \qquad Ae^{\sqrt{\alpha+\beta}} - Be^{-\sqrt{\alpha+\beta}} = 0. \tag{S18}$$

Similarly to above, the constants A and B can be obtained straightforwardly, and the efficiency of the pathway is

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{u}} = \alpha \int_0^1 \rho(x)dx = \frac{\alpha}{\alpha+\beta} \tag{S19}$$

B. Absorbing boundary, $\rho(x)|_{x=1} = 0$

The approach is very similar to the one the preceding section. The only difference, however, is that the boundary condition $\rho(x)|_{x=1} = 0$ leading to a slightly different second condition for the clustered enzymes

$$A_{II}e^{\sqrt{\beta}} + B_{II}e^{-\sqrt{\beta}} = 0. \tag{S20}$$

Likewise, we obtain

$$Ae^{\sqrt{\alpha+\beta}} - Be^{-\sqrt{\alpha+\beta}} = 0 \tag{S21}$$

as the second boundary condition for the uniformly distributed enzymes. Similarly, to the approach in section section II A 1 and II A 2, respectively the corresponding efficiency for the clustered case is given by

$$\left(\frac{J_2}{J_1}\right)_c = \frac{\alpha}{\alpha + \beta^{\frac{1}{2}} \coth(\beta^{\frac{1}{2}})} \tag{S22}$$

and for the uniform case

$$\left(\frac{J_2}{J_1}\right)_u = \frac{\alpha}{\alpha+\beta}(1 - \operatorname{sech}(\sqrt{\alpha+\beta})). \tag{S23}$$

III. ENZYME EXPOSURE PROBABILITY DISTRIBUTION IN 1D

As shown in the main text, the efficiency of the pathway in terms of the enzyme exposure probability distribution $P(E)$ reads

$$\frac{J_2}{J_1} = 1 - \int_0^\infty P(E)e^{-\alpha E}dE = 1 - \frac{J_{\mathrm{loss}}}{J_1}. \tag{S24}$$

To obtain an exact expression of $P(E)$ for the respective case it is convenient to calculate the inverse Laplace transformation of J_{loss}/J_1 with respect to α. We have seen above that the expression for the efficiency often has the form $\alpha/(\alpha + f(\beta))$; hence the loss term is $f(\beta)/(\alpha + f(\beta))$. The inverse Laplace transformation is easily obtained and has the form

$$P(E) = f(\beta)e^{-f(\beta)E}. \tag{S25}$$

For the case of a reflecting outer boundary we therefore have for the uniform enzyme profile, $f(\beta) = \beta$ and thus $P_{\mathrm{u}}(E) = \beta e^{-E\beta}$; for the clustered enzyme profile, $f(\beta) = \beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}}$ and $P_{\mathrm{c}}(E) = \beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}} e^{-E\beta^{\frac{1}{2}} \tanh \beta^{\frac{1}{2}}}$.

In the case of an absorbing boundary, the clustered distribution also gives rise to a similar expression for the efficiency, with $f(\beta) = \beta^{\frac{1}{2}} \coth \beta^{\frac{1}{2}}$ and thus $P_{\mathrm{c}}(E) = \beta^{\frac{1}{2}} \coth \beta^{\frac{1}{2}} e^{-E\beta^{\frac{1}{2}} \coth \beta^{\frac{1}{2}}}$. However, with a uniform distribution the loss flux does not have the form discussed above, and thus the calculation of the inverse Laplace transformation is more involved. We rewrite Eq. S23 and use Eq. S24 to obtain

$$\left(1 - \frac{J_2}{J_1}\right) = \frac{J_{\mathrm{loss}}}{J_1} = \frac{\alpha}{(\alpha+\beta)\cosh(\sqrt{\alpha+\beta})} + \frac{\beta}{\beta+\alpha}. \tag{S26}$$

Where the second term has the form as we have already discussed above, thus

$$P_{\mathrm{u}}(E) = \mathcal{L}^{-1}\left[\frac{\alpha}{(\alpha+\beta)\cosh(\sqrt{\alpha+\beta})}\right] + \beta e^{-\beta E}. \tag{S27}$$

The inverse Laplace transformation of the first term is calculated by determining the singularities in terms of α and than calculate their residues by using the Laurent expansion. Finding the poles is here equivalent to determining the roots of the denominator,

$$g(\alpha,\beta) = (\alpha+\beta)\cosh(\sqrt{\alpha+\beta}) = 0. \tag{S28}$$

This is satisfied for

$$\alpha = -\beta \qquad \text{and} \qquad \alpha_n = -\pi^2(n+\frac{1}{2})^2 - \beta \tag{S29}$$

for $n \in \mathbb{N}$. And the residues are given by

$$\begin{aligned}
\mathrm{Res}(\alpha g^{-1}(\alpha,\beta)e^{\alpha E}, \alpha = -\beta) &= -\beta e^{-\beta E} \\
\mathrm{Res}(\alpha g^{-1}(\alpha,\beta)e^{\alpha E}, \alpha_0 = -\frac{\pi^2}{4} - \beta) &= \frac{\pi^2+4\beta}{\pi} e^{-(\frac{\pi^2}{4}+\beta)E} \\
\mathrm{Res}(\alpha g^{-1}(\alpha,\beta)e^{\alpha E}, \alpha_1 = -\frac{-9\pi^2}{4} - 4\beta) &= -\frac{9\pi^2+4\beta}{5\pi} e^{-(\frac{9\pi^2}{4}+\beta)E} \\
&\vdots
\end{aligned} \tag{S30}$$

The first term above cancels with the last term in Eq. S27. Combining the remaining terms, we are left with the overall enzyme exposure distribution

$$P_{\mathrm{u}}(E) = \sum_{n=0}^{\infty} (-1)^n \left(\frac{(2n+1)^2\pi^2 + 4\beta}{(2n+1)\pi}\right) e^{-(\pi^2(n+\frac{1}{2})^2+\beta)E}. \tag{S31}$$

IV. THREE DIMENSIONS

In three dimensions we impose rotational symmetry and reduce the reaction diffusion equation to depend only on the radial coordinate,

$$r^{-2}\partial_r(r^2\partial_r\rho(r)) - \alpha e(r)\rho(r) = 0. \tag{S32}$$

We apply the following boundary conditions, $(4\pi r^2 \partial_r \rho(r))|_{r=0} = -1$ and the outer sphere is absorbing $\rho(r)|_{r=1} = 0$.

A. Clustered enzyme profile with absorbing boundary

We again begin by considering a clustered distribution of enzymes, $e_{\mathrm{c}}(r) = \frac{\delta(r-r_0)}{3r_0^2}$, that has been normalized such that $\int_0^1 4\pi r^2 e_{\mathrm{c}}(r)\mathrm{d}r = 4\pi/3$. As in section II A 1 we divide the system into two parts, part I for $r < r_0$ and part II for $r > r_0$. In each part, due to the absence of E_2 enzymes, the solution of Eq. S32 is

$$\rho_i(r) = \frac{A_i}{r} + B_i \tag{S33}$$

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

$$A_I = \frac{1}{4\pi} \qquad \text{and} \qquad A_{II} = -B_{II}. \tag{S34}$$

The remaining two conditions come again from the matching of the concentration at $r = r_0$, $\rho_I(r_0) = \rho_{II}(r_0)$, and the discontinuity of the derivative of the concentration $\rho(r)$ at $r = r_0$. Hence we get

$$\frac{1}{4\pi r_0} + B_I = A_{II}(\frac{1}{r_0} - 1) \tag{S35}$$

and

$$-A_{II} + \frac{1}{4\pi} - A_{II}\left[\frac{\alpha}{3}\left(\frac{1}{r_0} - 1\right)\right] = 0. \tag{S36}$$

With these four conditions Eqs. S34-S36 we obtain after some straightforward algebra the expressions for the four constant $\{A_I, A_{II}, B_I, B_{II}\}$. Since we do not consider a competing pathway in this model the efficiency of the pathway can also be calculated as

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{c}} = 1 - 4\pi(r^2\partial_r\rho(r))|_{r=1} = 4\pi A_{II} = \frac{\frac{\alpha}{3}(1-r_0)}{r_0 + \frac{\alpha}{3}(1-r_0)}. \tag{S37}$$

B. Uniform enzyme profile with absorbing boundary

For the uniform enzyme profile $e_{\mathrm{u}}(r) = 1$ the reaction-diffusion equation Eq. S32 reads

$$r^{-2}\partial_r(r^2\partial_r\rho(r)) - \alpha\rho(r) = 0. \tag{S38}$$

This is solved by

$$\rho(r) = \frac{1}{r}\left(Ae^{\sqrt{\alpha}r} + Be^{-\sqrt{\alpha}r}\right). \tag{S39}$$

With the boundary conditions we arrive at the following two conditions: at the origin,

$$A + B = \frac{1}{4\pi}, \tag{S40}$$

and at the absorbing outer boundary,

$$Ae^{\sqrt{\alpha}} + Be^{-\sqrt{\alpha}} = 0. \tag{S41}$$

In the same way as the case of clustered enzymes the efficiency is given by

$$\left(\frac{J_2}{J_1}\right)_{\mathrm{u}} = 1 - 4\pi(r^2\partial_r\rho(r))|_{r=1} = 1 - \sqrt{\alpha}\mathrm{csch}(\sqrt{\alpha}). \tag{S42}$$

V. ENZYME EXPOSURE PROBABILITY DISTRIBUTION IN 3D

Similarly to the one dimensional case the enzyme exposure probability distribution is obtained by the inverse Laplace transformation of the loss current through the boundary. The loss current of the clustered enzyme profile reads

$$\frac{J_{\mathrm{loss}}}{J_1} = \frac{r_0}{r_0 + \frac{\alpha}{3}(1-r_0)} \tag{S43}$$

which has the general form discussed in Section III. Hence the enzyme exposure distribution is

$$P_{\mathrm{c}}(E) = \frac{3r_0}{1-r_0}e^{-\frac{3r_0}{1-r_0}E}. \tag{S44}$$

For the uniform enzyme profile, the calculation again proceeds in the same way as described above. We firstly calculate the singularities of the loss current, the roots of $\sinh(\sqrt{\alpha})$, which are given by

$$\alpha_n = -(n\pi)^2 \qquad \text{with} \quad n \in \mathbb{N}. \tag{S45}$$

This leads then to the following residues

$$\begin{aligned}
\mathrm{Res}(\frac{\sqrt{\alpha}}{\sinh\sqrt{\alpha}}e^{\alpha E}, \alpha_0 = 0) &= 0 \\
\mathrm{Res}(\frac{\sqrt{\alpha}}{\sinh\sqrt{\alpha}}e^{\alpha E}, \alpha_1 = -\pi^2) &= 2\pi^2 e^{-\pi^2 E} \\
\mathrm{Res}(\frac{\sqrt{\alpha}}{\sinh\sqrt{\alpha}}e^{\alpha E}, \alpha_2 = -4\pi^2) &= -8\pi^2 e^{-4\pi^2 E} \\
&\vdots
\end{aligned} \tag{S46}$$

We assemble all the individual terms and get for the enzyme exposure probability distribution

$$P_{\mathrm{u}}(E) = 2\sum_{n=1}^{\infty}(-1)^{n+1}(\pi n)^2 e^{-(n\pi)^2 E}. \tag{S47}$$

Bibliography

[1] Hans M. Warrick and James A. Spudich. Myosin structure and function in cell motility. *Ann. Rev. Cell Biol.*, 3:379–421, 1987.

[2] I. Rayment, H.M. Holden, M. Whittaker, C.B. Yohn, M. Lorenz, K.C. Holmes and R.A. Milligan. Structure of the actin-myosin complex and its implications for muscle contraction. *Science*, 261:58–65, 1993.

[3] Lodi P.J., Ernst J.A., Kuszewski J., Hickman A.B., Engelman A., Craigie R., Clore G.M. and Gronenborn A.M. Solution structure of the DNA binding domain of HIV-1 integrase. *Biochemistry.*, 31:9826–9833, 1995.

[4] Allan Svendsen. Lipase protein engineering. *Biochimica et Biophysica Acta*, 1543:223238, 2000.

[5] Lizbeth Hedstrom. Serine protease mechanism and specificity. *Chem. Rev.*, 102:4501?4523, 2002.

[6] Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts and Peter Walter. *Molecular Biology of the Cell*. Garland Science, Taylor and Francis Group, New York: Fifth Edition, 2008.

[7] Leif Dehmelt and Philippe I.H. Bastiaens. Spatial organization of intracellular communication: insights from imaging. *Curr Opin Cell Biol*, 11:440–452, 2010.

[8] Robert A. Copeland. *Enzymes: A Practical Introduction to Structur*. Wiley-VCH Inc: Second Edition, 2000.

[9] Chen L.H., Kenyon G.L., Curtin F., Harayama S., Bembenek M.E., Hajipour G. and Whitman C.P. 4-oxalocrotonate tautomerase, an enzyme composed of 62 amino acid residues per monomer. *J Biol Chem.*, 25:17716–17721, 1992.

[10] Smith S. The animal fatty acid synthase: one gene, one polypeptide, seven enzyme. *FASEB J.*, 15:1248–1259, 1994.

[11] Jennifer A. Doudna and Jon R. Lorsch. Ribozyme catalysis: not different, just worse. *Nature structural and Molecular biology*, 12:395–402, 2005.

[12] Arthur J. Zaug and Thomas R. Cech. The intervening sequence RNA of tetrahymena is an enzyme. *Science*, 231:470–475, 1986.

[13] Mark G. Snider, Brenda S. Temple and Richard Wolfenden. The path to the transition state in enzyme reactions: a survey of catalytic efficiencies. *J. Phys. Org. Chem*, 17:586–592, 2004.

[14] Thomas Bulter Edgardo T. Farinas and Frances H. Arnold. Directed enzyme evolution. *Current Opinion in Chemical Biology*, 12:545–551, 2001.

[15] Haiyan Tao and Virginia W. Cornish. Milestones in directed enzyme evolution. *Current Opinion in Chemical Biology*, 6:858864, 2002.

[16] Lubert Stryer. *Biochemie*. Heidelberg; Berlin; New York: Spekturm Akad. Verlag: Vierte Auflage, 1991.

[17] R.D. Bauer u.a. *Biophysik*. Springer-Verlag Heidelberg: Zweite Auflage, 1982.

[18] Jiali Gao, Shuhua Ma, Dan T. Major, Kwangho Nam, Jingzhi Pu, and Donald G. Truhlar. Mechanisms and free energies of enzymatic reactions. *Chem. Rev.*, 106:3188–3209, 2006.

[19] Bruice T.C. and Benkovic S.J. Chemical basis for enzyme catalysis. *Biochemisty*, 39:6267–6274, 2000.

[20] Gordon G. Hammes. Multiple conformational changes in enzyme catalysis. *Biochemisty*, 41:8221–8228, 2002.

[21] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, 3rd edition, 2004.

[22] Richard Wolfenden and Mark J. Snider. The depth of chemical time and the power of enzymes as catalysts. *Acc. Chem. Res.*, 34:938–945, 2001.

[23] Gregory A. Petsko and Dagmar Ringe. *Protein Structure and Functions*. Qxford University Press, 2009.

[24] Sven Lindskog. Structure and mechanism of carbonic anhydrase. *Pharmacol. Ther.*, 74:1–20, 1997.

[25] M.E. Stroppolo, M. Falconi, A.M. Caccuri and A. Desideri. Superefficient enzymes. *Cell. Mol. Life Sci.*, 58:1451–1460, 2001.

[26] Alex Gutteridge and Janet M. Thornton. Understanding natues's catalytic toolkit. *TRENDS in Biochemical Science*, 30:622629, 2005.

[27] Nicolas C. Price. What is meant by 'competitive inhibition'? *TRENDS in Biochemical Science*, 4:1112, 1979.

[28] R. H. Behal, D. B. Buxton, J. G. Robertson, and M. S. Olson. Regulation of the pyruvate dehydrogenase multienzyme complex. *Annu. Rev. Nutr.*, 13:497–520, 1993.

[29] Fischer E. Einfluss der Configuration auf die Wirkung der Enzyme. *Chem. Ges.*, 27:29852993, 1894.

[30] Cooper G.M. *The Cell. A Molecular Approach.* Sunderland MC: Sinauer Assiciates, 2nd edition, 2000.

[31] Jeremy R. Knowles. Enzyme catalysis: not different, just better. *Nature*, 350:121–124, 1991.

[32] H. Peter Lu, Luying Xun and X. Sunney Xie. Single-moleclue enzymatic dynamics. *Science*, 282:1877–1882, 1998.

[33] U.F. Müller. Re-creating an RNA world. *Cell. Mol. Life Sci.*, 63:12781293, 2006.

[34] Walter Gilbert. The RNA world. *Nature*, 319:618–619, 1986.

[35] Johnston W., Unrau P., Lawrence M., Glasner M. and Bartel D. . RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science*, 292:1319–1325, 2001.

[36] Lester J. Reed. Multienzyme complexes. *Acc. Chem. Res*, 7:40–46, 1974.

[37] Alejandro Martin Aart de Kok, Annechien F. Hengeveld and Adrie H. Westphal. The pyruvate dehydrogenase multi-enzyme complex from gram-negative bacteria.

[38] Z. Hong Zhou, Diane B. McCarthy, Catherine M. OConnor, Lester J. Reed and James K. Stoops. The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes. *PNAS*, 98:14802–14807, 2001.

[39] Roche T.E., Baker J.C., Yan X., Hiromasa Y., Gong X, Peng T., Dong J., Turkan A. and Kasten S.A. Distinct regulatory properties of pyruvate dehydrogenase kinase and phosphatase isoforms. *Prog Nucleic Acid Res Mol Biol.*, 70:33–75, 2001.

[40] Z. Hong Zhou , Wangcai Liao, R. Holland Cheng, J. E. Lawson, D. B. McCarthy, Lester J. Reed and James K. Stoops. Direct evidence for the size and conformational variability of the pyruvate dehydrogenase complex revealed by three-dimensional electron microscopy. *The Journal of biological chemistry*, 276:2170421713, 2001.

[41] Axel Berg, Adrie H. Westphal, Hans J. Bosma and Aart de Kok. Kinetics and specificity of reductive acylation of wild-type and mutated lipoyl domains of 2-oxo-acid dehydrogenase complex from azotobacter vinelandii. *Eur. J. Biochem.*, 252:45–50, 1998.

[42] L. J. Reed and M. L. Hackert. Structure-function relationships in dihydrolipoamide acyltransferases. *The Journal of Biological*, 265:8971–8974, 1990.

[43] Markus Fries, Kathrine M. Stott, Stephen Reynolds and Richard N. Perham. Distinct modes of recognition of the lipoyl domain as substrate by the E1 and E3 components of the pyruvate dehdrogenase multienzyme complex. *J. Mol. Biol.*, 366:132–139, 2007.

[44] Pei X.Y., Titman C.M., Frank R.A., Leeper F.J. and Luisi B.F. Snapshots of catalysis in the E1 subunit of the pyruvate dehydrogenase multienzyme complex. *Structure*, 16:1860–1872, 2008.

[45] Richard N. Perham. Swinging arms and swinging domains inmultifunctional enzymes: Catalytic machines formultistep reactions. *Annu. Rev. Biochem.*, 69:9611004, 2000.

[46] Santiago Schnella, Michael J. Chappellc, Neil D. Evansc and Marc R. Rousseld. The mechanism distinguishability problem in biochemical kinetics: The single-enzyme, single-substrate reaction as a case study. *Comptes Rendus Biologies*, 329:5161, 2006.

[47] Boris Rotman. Measurement of activity of single molecules of beta-d-galactosidase. *PNAS*, 47:1981–1991, 1961.

[48] Takashi Funatsu, Yoshie Harada, Makio Tokunaga, Kiwamu Saito and Toshio Yanagida. Imaging of single fluorescent molecules and individual atp turnovers by single myosin molecules in aqueous solution. *PNAS*, 374:555–559, 1995.

[49] Kerstin Blank, Cert De Cremer and Johan Hofkens. Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol. J.*, 4:465–479, 2005.

[50] Rahul Roy, Sungchul Hohng and Taekjip Ha. A practical guide to single-molecule FRET . *Nature Methods*, 5:507 – 516, 2008.

[51] Taekjip Ha, Alice Y. Ting, Joy Liang, W. Brett Caldwell, Ashok A. Deniz, Daniel S. Chemla, Peter G. Schultz and Shimon Weiss. Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism. *PNAS*, 96:893–898, 1999.

[52] X. Sunney Xie and Jay K. Trautman. Fluorescence-based analysis of enzymes at the single-molecule level. *Biotechnol. J.*, 4:441–480, 1998.

[53] Carlos Bustamante. In singulo biochemistry: When less is more. *Annu. Rev. Biochem.*, 77:4550, 2008.

[54] Victor I. Claessen,1 Hans Engelkamp, Peter C.M. Christianen, Jan C. Maan, Roeland J.M. Nolte, Kerstin Blank and Alan E. Rowan. Single-biomolecule kinetics: The art of studying a single enzyme. *Annu. Rev. Anal. Chem.*, 3:319440, 2010.

[55] Sunney Xie. Single-molecule approach to enzymology. *Single Mol.*, 2:229–236, 2001.

[56] Brian P. English, Wei Min, Antoine M van Oijen, Kang Taek Lee, Guobin Luo, Hongye Sun, Binny J. Cherayil, S. C. Kou and X. Sunney Xie. Ever-fluctuating single enzyme molecules: Michaelis-menten equation revisited. *Nature Chemical Biology*, 2:87–94, 2006.

[57] Binning, G., Quate, C.F. and Gerber, C.H. Atomic force microscopy. *Phys. Rev. Lett.*, 56:930–933, 1986.

[58] Drake B, Prater CB, Weisenhorn AL, Gould SA, Albrecht TR, Quate CF, Cannell DS, Hansma HG and Hansma PK. Imaging crystals, polymers, and processes in water with the atomic force microscope. *Science*, 243:1586–1589, 1989.

[59] A. Ashkin, J. M. Dziedzic, J. E. Bjorkholm and Steven Chu. Observation of a single-beam gradient force optical trap for dielectric particles. *Optical Letters*, 11:288–290, 1986.

[60] Zoldk G and Rief M. Force as a single molecule probe of multidimensional protein energy landscapes. *Curr Opin Struct Biol.*, 23:48–57, 2013.

[61] Dill K.A., Ozkan S.B., Shell M.S. and Weikl T.R. The protein folding problem. *Annu. Rev. Biophys.*, 37:289–316, 2008.

[62] Ken A. Dill and Justin L. MacCallum. The protein-folding problem, 50 years on. *Science*, 338:1042–1046, 2012.

[63] David D. Boehr, H. Jane Dyson and Peter E. Wright. An nmr perspective on enzyme dynamics. *Chem. Rev.*, 106:30553079, 2006.

[64] Katherine A. Henzler-Wildman, Ming Lei, Vu Thai, S. Jordan Kerns, Martin Karplus and Dorothee Kern. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*, 450:913918, 2007.

[65] Franz Schwabl. *Statistische Mechanik*. Springer Verlag Berlin Heidelberg New York, Zweite Auflage, 2004.

[66] Frauenfelder H., Sligar S.G. and Wolynes P.G. The energy landscapes and motions of proteins. *Science.*, 254:1598–1603, 1991.

[67] Ernst-Ludwig Florin, Vincent T. Moy and Hermann E. Gaub. Adhesion forces between individual ligand-receptorpairs. *Science*, 264:415–417, 1994.

[68] Merkel R., Nassoy P., Leung A., Ritchie K. and Evans E. Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature*, 397:50–53, 1999.

[69] Falk Schwesinger, Robert Ros, Torsten Strunz, Dario Anselmetti, Hans-Joachim Güntherodt, Annemarie Honegger, Lutz Jermutus, Louis Tiefenauer, and Andreas Plückthun.

[70] Vincent T. Moy, Ernst-Ludwig Florin and Hermann E. Gaub. Adhesion forces between individual ligand-receptorpairs. *Science*, 264:257–259, 1994.

[71] G.I. Bell. Models for the specific adhesion of cells to cells. *Science*, 200:618–627, 1978.

[72] Gerhard Hummer and Attila Szabo. Kinetics from nonequilibrium single-molecule pulling experiments. *Biophysical Journal*, 85:5–15, 2003.

[73] Olga K. Dudko, Gerhard Hummer and Attila Szabo . Intrinsic rates and activation free energies from single-molecule pulling experiments. *Phys. Rev. Lett.*, 96:108101, 2006.

[74] Olga K. Dudko, Gerhard Hummer and Attila Szabo. Theory, analysis, and interpretation of single-molecule force spectroscopy experiments. *PNAS*, 105:15755–15760, 2008.

[75] Yohichi Suzuki and Olga K. Dudko. Single-molecule rupture dynamics on multidimensional landscape. *Phys. Rev. Let.*, 104:048101–1–048101–4, 2010.

[76] Olga K. Dudko, Thomas G. W. Graham and Robert B. Best. Locating the barrier for folding of single molecules under an external force. *Phys. Rev. Let.*, 107:208301–1–208301–4, 2011.

[77] Xiaowei Zhuang, Laura E. Bartley, Hazen P. Babcock, Rick Russell, Taekjip Ha, Daniel Herschlag and Steven Chu. A single-molecule study of RNA catalysis and folding. *Science*, 288:2048–2051, 2000.

[78] Jan Liphardt, Bibiana Onoa, Steven B. Smith, Ignacio Tinoco Jr. and Carlos Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, 2001.

[79] José Nelson Onuchic and Peter G. Wolynes. Theory of protein folding. *Current Opinion in Structural Biology*, 14:70–75, 2004.

[80] Michael Schlierf and Matthias Rief. Single-molecule unfolding force distribution reveal a funnel-shaped energy landscape. *Biophysical Journal*, 90:33–35, 2005.

[81] H. Gumpp, E. M. Puchner, J. L. Zimmermann, U. Gerland, H. Gaub and K. Blank. Triggering enzymatic activity with force. *Nano Lett.*, 9(9):3290–3295, 2009.

[82] Akihiko Ishijima, Hiroaki Kojima, Takashi Funatsu, Makio Tokunaga, Hideo Higuchi, Hiroto Tanaka, and Toshio Yanagida. Simultaneous observation of individual atpase and mechanical events by a single myosin molecule during interaction with actin. *Cell*, 92:161171, 1998.

[83] Thorsten Hugel, Nolan B. Holland, Anna Cattani, Luis Moroder, Markus Seitz, and Hermann E. Gaub. Single-molecule optomechanical cycle. *Science*, 296:1103–1106, 2002.

[84] Matthew J. Lang, Polly M. Fordyce, Anita M. Engh, Keir C. Neuman and Steven M. Block. Simultaneous, coincident optical trapping and single-molecule fluorescence. *Nature Methods*, 1:133–139, 2004.

[85] Takashi Kodama, Hiroyuki Ohtani, Hideo Arakawa and Atsushi Ikai. Mechanical perturbation-induced fluorescence change of green fluorescent protein. *Appl. Phys. Lett.*, 86:043901, 2005.

[86] Yufan He, Maolin Lu, Jin Cao and H. Peter Lu. Manipulating protein conformations by single-molecule AFM-FRET nanoscopy. *ACS nano*, 6:1221–1229, 2012.

[87] Pak-Wing Fok and Tom Chou. Reconstruction of bond energy profiles from multiple first passage time distributions. *Proc. Royal Soc. A*, 466:13479–3499, 2010.

[88] Christophe H. Schilling, Stefan Schuster, Bernhard O. Palsson and Reinhart Heinrich. Overview of regulatory strategies and molecular elements in metabolic engineering of bacteria. *Biotechnol. Prog.*, 199:296–303, 1999.

[89] Judit Ovádi and Valdur Saks. On the origin of intracellular compartmentation and organized metabolic systems. *Molecular and Cellular Biochemistry*, 256/257:5–12, 2004.

[90] Tianwen Wang, Xingyuan Ma, Guocheng Du and Jian Chen. Overview of regulatory strategies and molecular elements in metabolic engineering of bacteria. *Mol. Biotechno.*, 52:300–308, 2012.

[91] Paul A. Srere and Judit Ovadi. Enzyme-enzyme interactions and their metabolic role. *FEBS Letters*, 268:360–364, 1990.

[92] http://commons.wikimedia.org/wiki/File:Biological cell.svg.

[93] Salih J. Wakil, James K. Stoops, and Vasudev C. Joshi. Fatty acid synthesis and its regulation. *Ann. Rev. Biochem.*, 52:537–579, 1983.

[94] Edward A. Bayer, Henri Chanzy, Taphael Lamed and Yuval Shoham. Cellulose, cellulases and cellulosomes. *Current Opinion in Structural Biology*, 8:548–557, 1998.

[95] I. A. Rose and J. V. B. Warms. Mitochondrial hexokinase. *J. Biol. Chem.*, 242:1635–1645, 1967.

[96] John E. Wilson. Isozymes of mammalian hexokinase: structure, subcellular localization and metabolic function. *Journal of Experimental Biology*, 104:435–443, 2003.

[97] Jurczak M.J., Danos A.M., Rehrmann V.R. and Brady M.J. The role of protein translocation in the regulation of glycogen metabolism. *Journal of Cellular Biochemistry*, 104:435–443, 2008.

[98] Ribalet B. John S, Weiss J.N. Subcellular localization of hexokinase I and II directs the metabolic fate of glucose. *PLOS one*, 6:1–13, 2011.

[99] Reinhart Heinrich, Stefan Schuster and Hermann-Georg Holzhütter. Mathematical analysis of enzymic reaction systems using optimization principles. *Eur. J. Biochem.*, 201:1–21, 1991.

[100] Judit Ovádi. Physiological significance of metabolic channelling. *J. theor. Biol.*, 152:1–22, 1991.

[101] Athel Cornish-Bowden. Failure of channelling to maintain low concentrations of metabolic intermediates. *Eur. J. Biochem.*, 195:103–108, 1991.

[102] Pedro Mendes, Douglas B. Kell and Hans V. Westerhoff. Channelling can decrease pool size. *Eur. J. Biochem.*, 204:257–266, 1992.

[103] Athel Cornish-Bowden and Maria Luz Cardenas. Channelling can affect concentrations of metabolic intermediates at constant net flux: artefact or reality? *Eur. J. Biochem.*, 213:87–92, 1993.

[104] Pedro Mendes, Douglas B. Kell and Hans V. Westerhoff. Why and when channelling can decrease pool size at constant net flux in a simple dynamic channel. *Biochmica et Biophysica Acta*, 289:175–186, 1996.

[105] Edith Wilson Miles, Sangkee Rhee and David R. Davies. The molecular basis of substrate channeling. *The Journal of the Biological Chemistry*, 274:12193–12196, 1999.

[106] Xinyi Huang, Hazel M. Holden, and Frank M. Raushel. Channeling of substrates and intermediates in enzyme-catalyzed reactions. *Ann. Rev. Biochem.*, 70:149180, 2001.

[107] Samanta Raboni, Stefano Bettati and Andrea Mozzarelli. Tryptophan synthase: a mine for enzymologists. *Cellular and Molecular Life Sciences*, 66:23912403, 2009.

[108] Dueber J.E., Wu G.C., Malmirchegini G.R., Moon T.S., Petzold C.J., Ullal A.V., Prather K.L. and Keasling J.D. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol.*, 27:753–759, 2009.

[109] Krishnan Radhakrishnan1, Ádám Halász, Dion Vlachos and Jeremy S, Edwards. Quantitative understanding of cell signaling: the importance of membrane organization. *Current Opinion in Biotechnology*, 21:677682, 2010.

[110] Boris N. Kholodenko. Spatially distributed cell signaling. *FEBS Letters*, 583:40064012, 2009.

[111] W. Richard Burack and Andrey S. Shaw. Signal transduction: hanging on a scaffold. *Curr Opin Cell Biol*, 12:211–216, 2000.

[112] Jason W. Locasale, Andrey S. Shaw and Arup K. Chakraborty. Scaffold proteins confer diverse regulatory properties to protein kinase cascades. *PNAS*, 104:13307–13312, 2007.

[113] Hideji Murakoshi, Ryota Iino, Takeshi Kobayashi, Takahiro Fujiwara, Chika Ohshima, Akihiko Yoshimura and Akihiro Kusumi. Single-molecule imaging analysis of RAS activation in living cells. *PNAS*, 101:7317–7322, 2004.

[114] Sarah J. Plowman, Cornelia Muncke, Robert G. Parton and John F. Hancock. H-ras, K-ras, and inner plasma membrane raft proteins operate in nanoclusters with differential dependence on the actin cytoskeleton. *PNAS*, 102:15500–15505, 2005.

[115] Siebe B. van Albada and Peter Rein ten Wolde. Enzyme localization can drastically affect signal amplification in signal transduction pathway. *PLOS Computational Biology*, 10:1925–1934, 2007.

[116] Andrew Mugler, Aimee Gotway Baily, Koichi Takahashi and Pieter Rein ten Wolde. Membrane clustering and the role of rebinding in biochemical signaling. *Biophysical Journal*, 10:1925–1934, 2012.

[117] John S. Easterby. A generalized theory of the transition time for sequential enzyme reactions. *Biochem. J.*, 199:155–161, 1981.

[118] J. Ovádi, P. Tompa, B. Vértessy, F. Orosz, T. Keleti, G. R. Welch. Transient-time analysis of substrate-channelling in interacting enzyme systems. *Biochem. J.*, 257:187–190, 1989.

[119] Judit Ovádi and Pauel A. Srere. Channel your energy. *Trends in Biochemical Sciences*, 17:445–447, 1992.

[120] G. Rickey Welch and John S. Easterby. Metabolic channeling versus free diffusion: transition-time analysis. *Trends in Biochemical Sciences*, 19:193–197, 1994.

[121] David A. Fell. Metabolic control analysis-a survey of its theoretical and experimental development. *Biochem. J.*, 286:313–330, 1992.

[122] Athel Cornish-Bowden. Metabolic control analysis in theory and practice. *Advances in Molecular and Cell Biology*, 11:21–64, 1995.

[123] J. H. S. Hofmeyr. Metabolic control analysis in a nutshell. *In Proceedings of the 2nd International Conference on Systems Biology*, pages 291–300, 2001.

[124] Paul A. Srere, Bo Mattiasson and Klaus Mosbach. An immobilized three-enzyme system: A model for microenvironmental compartmentation in mitochondria. *PNAS*, 70:2534–2538, 1973.

[125] M. O. Mnsson, N. Siegbahn and K. Mosbach. Site-to-site directed immobilization if enzymes with bis-NAD analogs. *PNAS*, 80:1487–1491, 1983.

[126] Robert J. Conrado, Jeffry D. Varner and Matthew P. DeLisa. Engineering the spatial organization of metabolic enzymes: mimicking natures's synergy. *Current Opinion in Biotechnology*, 19:492–499, 2008.

[127] Nadrian C. Seeman. Nucleic acid junctions and lattices. *J. theor. Biol.*, 99:237–347, 1982.

[128] Watson J.D. and Crick F.H.C. A structure for deoxyribose nucleic acid. *Nature*, 171:737–738, 1953.

[129] Paul W. K. Rothemund. Folding dna to create nanoscale shapes and patterns. *Nature*, 440:297–302, 2006.

[130] Shawn M. Douglas, Hendrik Dietz, Tim Liedl, Bjrn Hgberg, Franziska Graf and William M. Shih. Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature*, 459:414–418, 2009.

[131] Qian Lulu, Wang Ying, Zhang Zhao, ZhaoJian, Pan Dun, Zhang Yi, Liu Qiang, Fan Chunhai, Hu Jun and He Lin. Analogic China map constructed by DNA. *Chin. Sci. Bull.*, 51:2973–2976, 2006.

[132] S.K. Kufer, E.M. Puchner,H.Gumpp and H.E.Gaub. Single-molecule cut-and-paste surface assembly. *Science*, 319:594–596, 2008.

[133] Niels V. Voigt, Thomas Tørring, Alexandru Rotaru, Mikkel F. Jacobsen, Jens B. Ravnsbæk, Ramesh Subramani, Wael Mamdouh, Jørgen Kjems, Andriy Mokhir, Flemming Besenbacher and Kurt Vesterager Gothel. Single-molecule chemical reactions on DNA origami. *Nature Nanotechnology*, 5:200–203, 2010.

[134] Sherri Rinker, Yonggang Ke, Yan Liu, Rahul Chhabra and Hao Yan. Self-assembled DNA nanostructures for distance-dependent multivalent ligandprotein binding. *Nature Nanotechnology*, 3:418–422, 2008.

[135] Friedrich C. Simmel. DNA-based assembly lines and nanofactories. *Current Opinion in Biotechnology*, 23:516–521, 2012.

[136] Jinglin Fu, Minghui Liu, Yan Liu, Neal W. Woodbury and Hao Yan. Interenzyme substrate diffusion for an enzyme cascade organized on spatially addressable DNA nanostructures. *Journal of the American Chemical Society*, 134:5516–5519, 2012.

[137] Joerg Koehler Christof M. Niemeyer and Chris Wuerdemann. Dna-directed assembly of bienzymic complexes from in bivo biotinylated nad(p)h:fmn oxidoreductase and luciferase.

[138] Ofer I. Wilner, Yossi Weizmann, Ron Gill, Oleg Lioubashevski, Ronit Freeman and Itamar Willner. Enzyme cascades activated on topologically programmed DNA scaffolds. *Nature Nanotechnology*, 4:249–254, 2009.

[139] Jason Gorman and Eric C. Greene. Visualizing one-dimensional diffusion of proteins along DNA. *Nature Structural and molecular Biology*, 15:768–774, 2008.

[140] Carsten Teller and Itamar Willner. Organizing proteinDNA hybrids as nanostructures with programmed functionalities. *Trends in Biochemical Sciences*, 28:619–628, 2010.

[141] Jinglin Fu, Minghui Liu, Yan Liu and Hao Yan. Spatially-interactive biomolecular networks organized by nucleic acid nanostructures. *Accounts of chemical research*, 45:1215–1226, 2012.

[142] Camille J. Delebecque, Ariel B. Lindner, Pamela A. Silver and Faisal A. Aldaye. Organization of intracellular reactions with rationally designed RNA assemblies. *Science*, 333:470–474, 2008.

[143] H. Risken. *The Fokker-Planck Equation.* Springer Verlag: Berlin, Heidelberg, New York, Second Edition, 1996.

[144] C.W. Gardiner. *Handbook of Stochastic Processess for physics, chemistry and natural science, 4th ed.* Springer Verlag, 2009.

Danksagung

Zu allererst möchte ich meinem Prof. Ulrich Gerland für die intensive Betreuung meiner Doktorarbeit danken. Seine Begeisterungsfähigkeit für biologische Fragestellungen in der Physik waren für mich immer eine Quelle der Motivation und Inspiration. Vor allem aber danke ich ihm für die konstruktiven und überaus offenen wissenschaftlichen Diskussionen ohne die diese Arbeit nicht möglich gewesen wäre.

Als nächstes möchte ich mich beim kompletten Lehrstuhl Frey bedanken. Die außuniversitären Zusammenkommen, wie das Lehrstuhlgrillen oder das gelegentliche Feierabend Bier habe ich sehr genossen. Zudem möchte ich mich für die spontanen wissenschaftlichen Diskussion im Kaffeeraum bedanken, die von Zeit zu Zeit durchaus länglichen Charakter hatten. Im Speziellen danke ich Louis und Patrick für das Korrekturlesen meiner Arbeit. Ein großer Dank geht an meinen ehemaligen Zimmerkollegen Ben Klünder für die unzähligen Diskussionen. Seine Ratschläge und seine verlässliche Hilfe bei wissenschaftlichen Problemen waren ungemein förderlich für das Vorankommen der Arbeit.

Daneben gilt mein Dank Dr. Filipe Tostevin von seinem Wissen und seiner Erfahrung im Wissenschaftsbetrieb konnte ich sehr profitieren. Ausserdem danke ich Fabienna Arends und Florian Hinzpeter die Projekte und die Zusammenarbeit mit Euch haben mir sehr viel Freude bereitet.

Mein ganz besonderer Dank gilt natürlich meiner Familie die mich während des gesamten Studiums und der Promotionszeit unterstützt haben, insbesondere meinen Eltern und meinem Bruder die immer für mich da waren. Ebenfalls danke ich meinen beiden Omas, die mich während der kreativen Trockenzeit meiner Arbeit mit kulinarischen Köstlichkeiten ausgezeichnet versorgten, meinem Opa der mir mit so manchen edlem Tropfen den Kopf befreite. Abschliessend möchte ich mich noch ganz besonders bei Michaela für ihren täglichen Beistand bedanken.