FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Spatial modelling of biomolecular systems

Denise Neves Gameiro

Master in Bioengineering - Biological Engineering

Supervisor: Prof. Nuno Azevedo

Co-supervisor: Prof. Anália Lourenço

July 31, 2015

Abstract

Simulations of cells at the biomolecular scale yield accurate representations of the metabolic behaviour, cellular phenomena and culture inter-variance. Using the interactions between enzymes and metabolites as the functional unit of biomolecular systems, the agent-based modelling (ABM) paradigm has the potential to develop such model.

This dissertation approaches how to translate conceptual and mathematical models from Physics, Chemistry and Biology to an agent-based computational model. Specifically, the present work focuses on the curation of biological data available at either public databases or scientific literature, and in adapting the modelling strategy so that this data can be used as modelling inputs.

The developed modelling approach was used to simulate volumes of enzyme and substrate solutions in different biological scenarios, such as diffusion-controlled reactions, substrate saturation, Brownian motion, and reactions limited by substrate concentration. An adaptation of the Michaelis-Menten kinetic model for different isomerases was developed. The behaviour and kinetic parameters were successfully recreated for two isomerases, a 2-hydroxymuconate tautomerase and a steroid δ-isomerase. With iterative refinements to the model inputs, the parameters of 2-hydroxymuconate tautomerase were recreated with deviations of 6% for the K_m and 19% for the k_{cat}. In a first estimate of the modelling inputs for steroid δ-isomerase, the resulting K_m was within the same order of magnitude as the published value. Additionally, it was assessed the potential of the present model to quantify intrinsic metabolic noise due to differences in spatial location of molecules.

Additionally, to evaluate the availability of data at the metabolic scale, a proof-of-concept curation of the necessary modelling inputs was conducted for the glycolysis pathway of *E. coli*. The retrieved data regards mainly the size of biomolecules and kinetic parameter values for the enzymes in the pathway. The procedure of curation and criteria to choose among differing available data are discussed, and data to simulate the glycolysis pathway and pyruvate metabolism of *E. coli* is presented.

The underlying assumptions of the physical and biological behaviour of agents was validated and the present modelling approach can be used as a foundation for the agent-based modelling of more complex systems, such as metabolic pathways, in more complex scenarios, such as molecular crowding and physiological conditions.

Resumo

Simulações da célula à escala biomolecular conseguem produzir uma representação exacta do comportamento metabólico, fenómenos celulares e variância entre culturas celulares. Usando interações entre enzimas e metabolitos como unidade funcional de sistemas biomoleculares, o paradigma de modelação por agentes (ABM) tem potencial para desenvolver tal modelo.

Esta dissertação aborda como transpôr modelos conceptuais e matemáticos das áreas da Física, Química e Biologia para um modelo computational baseado em agentes. Especificamente, o trabalho desenvolvido foca-se na curação de informação e dados biológicos disponíveis, em bases de dados públicas ou na literatura científica, e em adaptar a estratégia de modelação para que estes dados possam ser introduzidos como variáveis no modelo.

A modelação desenvolvida foi utilizada para simular volumes de soluções de enzimas e metabolitos em cenários biológicos diferentes, tais como reacções controladas pela difusão e em saturação de substrato, movimento Browniano e reacções limitadas por concentração de substrato. Uma adaptação computacional do modelo cinético de Michaelis-Menten para diferentes isomerases foi desenvolvida. O comportamento e parâmetros cinéticos de Michaelis-Menten foram recriados com sucesso para duas isomerases, uma *2-hydroxymuconate tautomerase* e uma *steroid δ-isomerase*. Com refinamentos iterativos aos *inputs* do modelo, os parâmetros para a *2-hydroxymuconate tautomerase* foram recriados com um desvio de 6% para o K_m e 19% para o k_{cat}. Numa primeira estimativa dos *inputs* de modelação da enzima *steroid δ-isomerase*, o K_m resultante é da mesma ordem de magnitude do publicado. Adicionalmente, foi avaliado o potencial do modelo em questão para quantificar ruído intrínseco metabólico devido a diferenças na localização espacial de moléculas.

Adicionalmente, para avaliar a disponibilidade de informação relevante para este método de simulação à escala metabólica, foram recolhidos dados sobre a via metabólica da glicólise em *E. coli*, como teste de conceito. Os dados necessários consistem principalmente em dados sobre tamanho de biomoléculas e dados cinéticos das enzimas da via, e recolha de várias fontes da sua grande maioria é apresentada neste trabalho. São ainda discutidos os procedimentos de curação de dados e critérios de escolha entre diferentes dados disponíveis.

As hipóteses subjacentes ao comportamento físico e biológico dos agentes foram validadas e a presente abordagem de modelação pode ser usada como uma base para a modelação baseada em agentes em sistemas mais complexos, como vias metabólicas, em cenários complexos como *crowding* molecular e condições fisiológicas.

Acknowledgements

I would like to thank my supervisors, Professors Nuno Azevedo and Anália Lourenço, for all their guidance, insight, availability, interest in the project and optimism. I would also like to thank Gael and Martin, who handled the informatics and without whose work this Dissertation would not be possible.

My thanks to the people in Bioengineering and Biological Engineering that put up with me, even during my "exile" to the library; to Tiago, for all his proof-reading, support, help, patience and saying "it'll all be OK" so many times; and to my family, who remained supportive throughout these five years, down to the last moody deadline.

Denise

"When you eliminate the impossible, whatever remains,
however improbable, must be the truth."

Sir Arthur Conan Doyle

Contents

List of Figures

List of Tables

Nomenclature

Abbreviations

ABM	Agent-based modelling
ES	Enzyme substrate complex
E_t	Total enzyme
MW	Molecular weight
N/A	Not available

Variables and constants

Symbol	Description	Common units	Dimensions
D_c	Diffusion coefficient	$\mu m^2.s^{-1}$	$[L^2.T^{-1}]$
dP/dt	Linear regression of product agent number *vs.* time step	$P.ts^{-1}$	-
k_{cat}	Turnover number	s^{-1}	$[T^{-1}]$
K_{eq}	Collision rate constant of a real enzyme-substrate solution	$M^{-1}.s^{-1}$	$[N^{-1}.L^3.T^{-1}]$
K_m	Michaelis constant	mM	$[N.L^{-3}]$
$simk_{cat}$	Number of time steps from ES formation to product formation	ts	-
$simK_m$	Probability of a collision between enzyme and substrate forming ES	-	-
R_h	Hydrodynamic radius	nm	[L]
t	time	s	[T]
T	Temperature	K	$[\theta]$

Chapter 1

Work Outline

1.1 Project background

The objective of the this project is to apply agent-based modelling to biomolecular systems. The proposed model focuses on rendering an accurate representation of the interactions between biomolecules, namely enzyme and metabolites. Such approach has the potential to correctly model the emergence of biological phenotypes and capture its complexity, as well as account for the spatial scale of cells. Ultimately, it can become a powerful tool to simulate metabolic pathways and cellular systems.

This project is the result of a current collaboration between LEPABE/FEUP and U. Vigo. The work of this thesis is centred in translating conceptual models of biological phenomena into a computational model, and the retrieval the curation of pertinent biological data.

A previous instantiation of the proposed biomolecular ABM model was made in 2D [1], in which the enzymatic behaviour was correctly simulated, reproducing the behaviour observed in real enzymes during kinetic assays. Notably, two simulation variables were calibrated in order to match the kinetic parameters by the mathematical modelling of Michaelis-Menten [2].

This is also the core of the 3D version of the proposed biomolecular model, with space considerations updated for the additional dimension. It is not within the scope of this thesis to focus on the purely computational aspects of creating such model, but to devise strategies to adapt current conceptual and mathematical models of physical, chemical and biological phenomena to a computational representation. The model structure is designed with the intent of simulating scenarios of interest to real world applications. Given this, the necessary inputs for the model should be widely available data. This depends essentially of selecting data that can be obtained through standard laboratory experiments and accessible in biological databases. Attention is paid to computational aspects mainly to establish a working compromise between biological accuracy and processing constraints, besides assuring a realistic behaviour of the agents as physical biomolecules.

1.2 Main objectives

The contribution to the overall project of this dissertation followed two distinct angles to approach the ultimate goal, for the main project, of creating an ABM at the metabolic scale.

The first objective was to design simulations based on biological knowledge that could validate the computational representation of a biomolecular environment. The scope of this part of the work includes the definition of the simulation's time step as a fraction of real time in a biological scenario, implementation of diffusional behaviour, and the modelling of enzymes according to their reported kinetic parameters.

The second objective was to assess the availability of biological data, regarding the input requirements of the designed modelling approach at the metabolic scale, and thus assess the feasibility of constructing a simulation based on these data. The curation of data for the glycolysis pathway of *E. coli* was used as a proof-of-concept. Searches on databases and scientific literature were conducted to gather all the relevant data required to enable the modelling of the glycolysis pathway. This part of the work aims to discuss the details of the available biological data regarding the physical dimensions, physical behaviour and intermolecular interactions. The gathered insights can guide both the implementations of the project's modelling approach, in order to reflect the reality of data availability, and the future collection of data for models at the metabolic or cellular scale.

1.3 Organization

Adding to the present Chapter, that introduces the scope and objectives of the thesis and presents an overall view of the organization of the thesis, the 2nd chapter covers the state of the art in biological simulation and the basis of ABM, and how it was previously used in modelling biological phenomena.

The 3rd chapter covers the methodology followed in iteratively validating model assumptions and reviewing literature, from basic physical rules to the simulation of an enzymatic assay.

The 4th chapter explores the results obtained in the experiments previously described and their significance.

Chapter 5 provides the main conclusions of the dissertation work, and frames its relevance within the main project.

Appendix A provides a description of the supplementary material, available in the CD accompanying this Dissertation.

Chapter 2

Introduction

2.1 Modelling Biological Systems

Modelling biological systems has the purpose of both aiding scientific understanding and serving as practical tool to enable an *in silico* laboratory approach. By building a model of biological phenomena according to the current understanding of the phenomena, one can predict the behaviour of the modelled system and see to what degree it depicts reality. Once validated, the model becomes a valuable tool to simulate different scenarios. This can reduce the number of experimental possibilities to be tested in a wet lab, or traditional lab, to the ones with most promise, saving time and financial resources. In particular, modelling is instrumental in Systems Biology, a discipline that aims to integrate biological data and enable a system-level understanding of organisms and their biochemistry [3, 4].

Modelling efforts started by systematizing biological data that was obtained by traditional reductionist methods in molecular biology. Their scope and detail increased with the popularization of high-throughput methods and "omics" data sets [5]. Models exploiting a systematic perspective of the cell are useful to survey useful genetic modifications to organisms relevant to industrial biotechnology [6, 7].

Biological modelling can be performed at different scales, like at the level of multicellular organisms, at the level of the cell, for instance to simulate microbial colonies or animal tissues, or at the biomolecular level, representing components inside the cell with varying degrees of abstraction. Within these scales, modelling can focus on different biochemical and cellular processes. The main processes occurring in biological systems are based on the underlying signalling, regulatory and metabolic networks [8]. The present work focuses on the biomolecular scale and on metabolic networks.

The designation of biomolecular modelling can apply to different modelling approaches, with diverse points of focus. This focus can range from fully atomistic models, which depict fully the atomic constitution and interactions and have a resolution below the nanoscale, to coarse-grained models of solutes, which are able to render larger simulation volumes and longer periods of time [9].

Coarse-grained models represent the system by a reduced and essential number of degrees of freedom and interactions. The use of coarse-graining models is the most viable strategy for developing physically accurate models that, at the same time, cover both time and length scales of biological processes [10, 11]. The key issue is to determine what approximations can be tolerated without comprising the overall level of realism and, inherently, the predictive ability.

Biologically relevant time scales range from nanosecond to microsecond time scales for the internal dynamics of individual molecules to time scales of seconds to hours for entire biological processes [9].

However, these two approaches focus largely on the physical aspects of biomolecules, such as conformation and diffusion [9]. The aim of the present model would be to ultimately bridge the spherical coarse-grained model of solutes in the cell with reaction models, retaining the capability of the former to accurately represent diffusional behaviour in the crowded cellular environment, with kinetic information of enzyme-catalysed reactions. To implement such model, agent-based modelling (ABM) was considered among the different types of modelling methods.

2.1.1 Types of modelling

Different classes of modelling can be separated with basis on different criteria, such as whether it considers dynamic or steady-state responses, or is deterministic or stochastic. In terms of methods, they can be in mathematical or direct computational models [12, 13].

Mathematical models encompass a large range of models, including some of the most traditional and popular methods in biological and metabolic modelling. Kinetic metabolic models are examples of the application of differential equations, while constraint-based methods, like flux-balance analysis, are based on linear algebra [14]. Kinetic models can take into account mass action laws or stochastic processes. They are dependent on a detailed characterization of enzymes and their mechanisms, which are commonly modelled with ordinary differential equations (ODEs) or partial differential equations (PDEs). Mogilevskaya et. al approaches the building of kinetic models for the metabolism of *E. coli* [15], and Link et al. reviews the recent advances and challenges in integrating kinetic information in kinetic models [16]. Despite being able to represent the dynamics of biological phenomena in a quantitive manner, they typically imply handling complex mathematical formulas that are difficult to implement. This happens especially when the number of enzymes and reactions increases, which is necessary to create relevant biological models of pathways.

Constraint-based stoichiometric methods have the significant advantage of scalability and their use is widespread in metabolic modelling [17, 18]. Genome-scale information of metabolic pathways can be incorporated in stoichiometric models and analysed with linear algebra. However, this approach overlooks many important kinetics of biomolecular interaction, such as regulation and inhibition phenomena. Furthermore, it is only valid under assumption of steady-state and the inherent linearity of stoichiometric models does not correctly characterize the non-linear behaviour of biological systems [14].

Mathematical models differ from direct computational models, which represent a constituent part of a system as an "agent". In terms of application, a key difference is that mathematical models treat populations as homogeneous [19]. Due to this feature, it is hard to incorporate or study the biological phenomena related to individual variation or complex interactions.

2.1.2 Agent-based modelling (ABM)

Agent-based modelling (ABM) is an example of direct computational modelling method. It resembles cellular automata (CA) models, but represents agents in a continuous environment and is capable of asynchronous interactions . It is also known as individual or particle based model. It is related to multi-agent systems (MAS). ABM has been applied to distinct fields like social sciences, economics, operational research and management, ecology, environmental studies and biology [20].

The main potentialities of ABM when applied to the biological scenario are its intuitive nature, suitability to depict emergence of biocomplexity with the establishment of a set of relatively simple rules [21], and implicit modelling of the spatial scale [1].

Intuitiveness is an important feature of the model, because it increases the usefulness of the model by easing its interpretation [14]. This is particularly important, as models are created and used by multidisciplinary teams [8]. The intuitive nature of direct computational models is due to the fact that the modelling efforts occur mainly at the level of the attributes of individual agents, which matches a reductionist perception of biomolecular systems. In fact, at the biomolecular level a large part of the current understanding and available data arises from reductionist research, focusing on individual biomolecules, their properties and their most immediate interactions. Due to the ABM capacity to represent complex behaviours emerging from simple rules and individual interactions, it becomes an ideal methodology to utilize available reductionist data and long-standing conceptual models that focus on individual molecules. ABM can be used concurrently with recent experimental technologies, such as single-molecule enzymology and single-molecule tracking [22, 23], which explore the stochastic nature of enzymatic catalysis [24] and the motion of molecules across cellular membranes and in the cytoplasm [25, 26].

ABM was applied in many of the different levels of biological modelling. For example, at the multicellular organism level, there are ABMs of the behaviour of ant colonies [27, 28]. ABM was applied to the cell-cell interaction in the epithelial tissue [29, 30], the inflammation response in diabetic ulcers [31] and to model the bacterial colonizations of plant's roots [32]. Hellweger et. al published a review on 46 ABMs applied to microbe populations, including bacteria in wastewater treatment, biofilms and food [33]. Previous works using ABM at the biomolecular scale included the modelling of the oxygen response in *E. coli* [34], the emergence of the competency phenotype in *Bacillus subtilis* [35] and an assessment of the effect of intracellular crowding in the eukariotic metabolism [36]. General tools to perform ABM in biomolecular scenarios include software programs like AgentCell [37], M-Cell [38], Smoldyn [39], the Cellular Dynamic Simulator [40] and ReaDDY [41]. Despite the advances in the area of agent-based biomodelling, there are some recurring limitations. These include a gap between the required inputs by the models and the

biological data that is actually available, the lack of validation or application of the proposed modelling approach to real biological scenarios, or the development of models of excessively complex biological scenarios, reducing the generality of the modelling approach and hampering the application of the developed model to other biological scenarios.

2.2 Structure of a biomolecular ABM

ABM have simple structures, comprised of three main elements: agents, rules and behaviour [42]. The application of this structure will vary with the biological level and process modelled. In a biomolecular metabolic ABM, the agents are the biomolecules involved in reactions and pathways, with special focus on enzymes and metabolites, and the model is focused on their behaviour and interactions in the cellular environment (Table 2.1). Alternatively, the biomolecular system can be an *in vitro* assay, an environment in which the catalytic and kinetic properties of enzymes are commonly determined [43, 44].

Table 2.1: ABM structure applied to the biomolecular scenario

GENERIC ABM	BIOMOLECULAR ABM
Agents	Molecules (enzymes, metabolites, co-factors)
Behaviour rules	Molecular diffusion
Interaction rules	Enzyme kinetics
Shared environment	Cytosol, membranes

The rules aim to portray computationally conceptual models of biochemistry and biophysics (Figure 2.1). The motion of agents should be equivalent to the diffusion of the modelled biomolecules. Interaction between agents varies in nature, with its core being defined by enzyme's physiological reactions, arranged in metabolic networks. Enzyme kinetics characterize the dynamics of part of the interactions, namely the binding between enzyme and metabolites. The binding of co-factors to the enzyme influences the kinetics of binding between substrate. The definition of specific rules require data of chemical and biological nature, specific to the enzyme, pathway or organism which is desired. These information and data should be easily and routinely determined in the laboratory, publicly available and, preferably, organised in programmatically accessible databases.

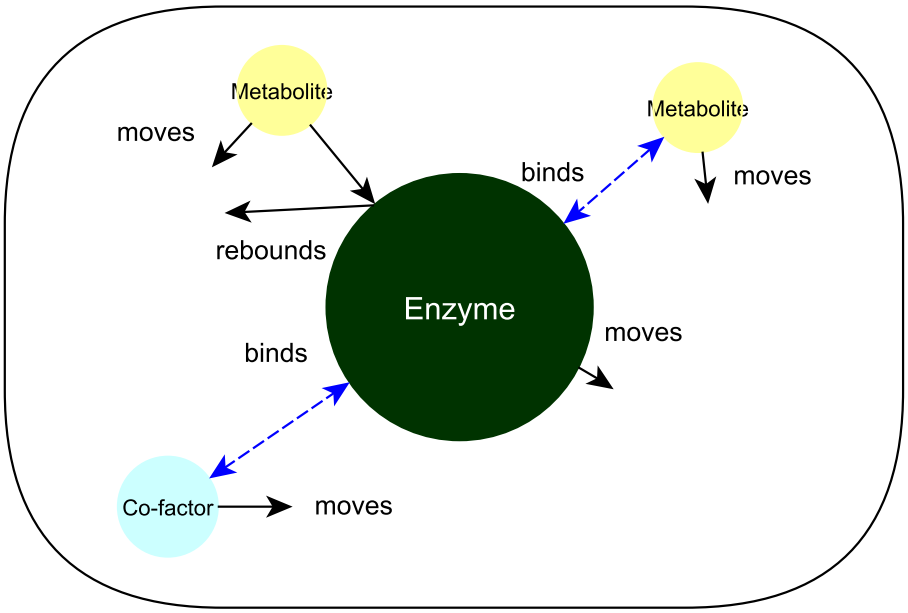

Figure 2.1: Interaction logic of a biomolecular ABM, in which all agents move, co-factors and substrates can bind to the enzyme, and agents rebound of each other.

The agents are represented as spheres, based on their hydrodynamic volume. Hydrodynamic volume is a good measure of the actual space occupied by the molecule in its reaction media, be it water in an enzymatic assay or the intracellular space. The spherical approximation is an effective way to create middle-out, coarse-grained models of the cell, in a compromise between realistic dimensions and computer tractability [9]. The volume of the spheric agents was determined by the hydrodynamic radii (R_h), or van der Waals radii (R_vdW). This radius was approximated by correlations with molecular weight (MW, in Dalton units), in the case of enzymes (eq. 2.1) [45] , and with the formula and rings in the structure, in the case of small organic metabolites (eq. 2.2) [46].

$$R_h(nm) = 0.0515MW^{0.392} \tag{2.1}$$

$$R_h(nm) = \sqrt[3]{\frac{3}{4.\pi}.\left[(\sum_i^{atom} V_{vdW_i}) - 5.92N_B - 14.7R_A - 3.8R_{NA}\right]} \tag{2.2}$$

where V_{vdW_i} is the van der Walls volume (*i.e.*, hydrodynamic volume) of each atom of the metabolite, the N_B the number of bonds, R_A the number of aromatic rings and R_{NA} the number of non-aromatic rings. The van der Walls radii of each atom was calculated by Bondi [47, 46]

2.2.1 Motion of biomolecular agents

In the computational model, the motion of the biomolecule agents is defined by velocity (number of simulation unit distance moved per time step). However, in biological terms, the motion of biomolecules is quantified in terms of diffusion. This relates to the concept of random-walk, or Brownian motion [48]. Due to collisions with the molecules of the solvent (water or air), molecules have a random trajectory and a null average displacement. The squared average displacement ($<R^2>$) is a measure of the absolute distance the molecule is likely to travel in any given direction. Eq. 2.3 shows its relation with the diffusion coefficient (D_c) and the time (t) during which the displacement occurred.

$$< R^2 >= 6D_c.t \tag{2.3}$$

where t is the time of motion and D_c the diffusion coefficient.

The diffusion coefficient can be calculated using the Einstein-Stokes equation (eq. 2.4), where k_B is the Boltzmann constant, T the absolute temperature and η the viscosity of the solvent media.

$$D_c = \frac{k_B.T}{6\pi.\eta.R_h} \tag{2.4}$$

2.2.2 Kinetics of biomolecular agents

Enzyme kinetics are related to a key issue in biomolecular ABMs, being one of the fields where the gap between computational inputs and common biological data is wider. Most models and tools for creating biomolecular ABMs require the intrinsic (or individual) rates of each catalytic mechanism step [39, 40], which are not commonly determined in the lab [49]. Therefore, there is a shortage of information. There is also no current efforts to build a systematized repository of individual rates values.

Most kinetic parameters of enzymes are determined in terms of Michaelis-Menten parameters, taken from the historical derivation of enzyme's catalytic behaviour. This mechanism features a simple isomerization reaction, in which a substrate molecule (*S*) binds to an enzyme (*E*), giving origin to an enzyme-substrate complex (*ES*). Under the assumption that the intermediate enzyme-product complex is negligible, the *ES* complex dissociates, forming one product molecule (*P*) and freeing the enzyme (eq. 2.5). k_1, k_{-1} and k_2 are the individual rate constants of each step of the catalysis mechanism.

$$E + S \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} E + P \tag{2.5}$$

Given this reality, it would be particularly interesting to be able to implement a biomolecular ABM in such a way that Michaelis-Menten kinetic parameters can be used as inputs to model the interactions between enzymes and substrates. There are two Michaelis-Menten parameters, the

Michaelis constant, K_m (eq. 2.6), and the maximum velocity, V_{max}. V_{max} is related to the turnover number, k_{cat} , and the concentration of enzymes, [Et] (eq. 2.7).

$$K_m = \frac{k_{-1} + k_2}{k_1} \tag{2.6}$$

$$V_{max} = k_{cat}.[E_t] = k_2.[E_t] \tag{2.7}$$

The velocity is limited by diffusion when the conversion of substrate to product is fast (high k_{cat}) and substrate is present at saturating concentration, making K_m influence on reaction velocity negligible. Diffusion-limited enzymes are close to kinetic perfection, with the parameter of k_{cat}/K_m being equivalent to the encounter rate constant, K_{eq}. The values of K_{eq} for enzyme-substrate systems is reported as between 10^9 and 10^{10} $M^{-1}.s^{-1}$ [50].

2.3 Data availability and assessment: proof of concept

The previous sections outlined how agents can be characterised and the key conceptual models of basic biological phenomena to be portrayed in a biomolecular ABM, such as diffusion and reactions. A second part of the work consists of a review to provide a systematic and comprehensive overview of available resources for the modelling of metabolic events at the molecular scale. The objective is to evaluate the feasibility of the application of the proposed model to a complete metabolic pathway, from a point of view of data availability. In order to be useful, the required data by the model in its present implementation should be available for all or most biomolecules involved in a given pathway. Gathering data on a specific pathway also provides insight of the prevalence of more complex biological phenomena, whose implementation is not covered in the scope of the present work, and thus provides a direction for future work.

The glycolysis and pyruvate metabolism pathways of *E. coli* serves as case study. *E. coli* is one of the most well-studied microorganisms, and is widely used in both laboratorial setting and in industrial applications. Similarly, glycolysis and pyruvate metabolism are central to the carbon metabolism, and are well-characterized pathways. The required data by the present implementation of the model is surveyed for the enzymes and metabolites of the two pathways. The procedure of collecting such data, the resources available, the effort involved in curation and the availability and uniformity of the data are discussed.

Chapter 3

Methods

The Methods chapter cover the model building, the simulation scenarios and the guidelines followed to retrieve cellular and biomolecular data in a systematic manner. The rationale behind the order of simulations performed was:

1. Defining a relation between computational and biological time, using a virtual enzyme with maximum catalytic efficiency;
2. Defining a relation between the kinetic parameter k_{cat} (s^{-1}) and its computational equivalent, a simk_{cat} (ts);
3. Defining a procedure to assign a velocity to each agent, so that the average displacement of the agents is equivalent to the theoretically predicted by the Stokes-Einstein equation;
4. Assessing which computational variables could be equivalent to the aggregated kinetic parameter K_m;
5. Recreating the Michaelis-Mente behaviour and kinetic parameters for two enzymes, the 2-hydroxymuconate tautomerase [51] and steroid δ-isomerase [52].

3.1 Model building

The proposed model was developed in the MASON (Multiagent Simulation of Neighborhoods) [53]. A previous implementation for the biomolecular scenario, considering a two-dimensional environment, is available [1].

The physical scale of the model was scaled relative to the hydrodynamic radius of a water molecule, which is of 0.16 nm [46]. Therefore, one simulator distance unit is equivalent to 1.6×10^{-10} m [54]. For each simulation, the shape and size of the simulation environment must be defined. To simulate a reactor carrying an enzymatic assay, the simulation environment was cubic. The edges of the reactor took the values of 550 and 600 simulator distance units for different simulations. Thus, the volume of the simulation environment was 6.81×10^5 nm^3, for the first set of simulation with a virtual enzyme, and 8.85×10^5 nm^3 for the remaining simulations. Enzymes and

substrates are represented in the model as spherical agents. The size of agents is defined based on the correlations for the hydrodynamic radii of enzymes (eq. 2.1) and for substrate hydrodynamic radius (eq. 2.2).

The hydrodynamic radius of the molecules defines the physical radius of the agent. An additional reaction radius can be defined, as a multiple of the physical radius of the agent. Within the volume defined by this reaction radius, a enzyme agent can "capture" substrate to form an *ES* agent.

Collisions between the spherical agents are detected based on the Pythagorean Theorem for triangles, which calculates the squared distance (d^2) taking into account the coordinates in the ith dimension of a first agent, a_1, and a second agent, a_2 (eq. 3.1).

$$d^2 = \sum_{i=1}^{3} (a_{1i} - a_{2i})^2 \tag{3.1}$$

For all the simulations, a simple enzymatic isomerization reaction is defined. A collision between enzyme agent and substrate agent can form another agent, the enzyme-substrate complex (*ES*) - a "successful" collision. To simplify the computational implementation, the *ES* agent was considered to have the same size as the enzyme agent alone. The fraction of collisions between enzyme and substrate that form *ES* agents is subject to the ABM rules, and can vary from 0 to 100% of all collisions. The *ES* can either dissociate back into enzyme and substrate, or enzyme and product. The ratio of these outcomes and time steps that it takes to dissociate are also subject to ABM rules.

The product and substrate have the same size and velocity, as the reactions represent merely structural changes in the substrate molecule, without loss or gain of mass. Structural changes that could alter the size of the molecule, such as change in the ring structures, were not considered.

Non-reacting collisions can occur between two enzyme agents, two substrate agents, enzyme and product agents, substrate and product agents, non-reacting enzymes and substrates (a fraction of reacting collision of 0%), or randomly between enzymes and substrates agents with a ratio of reacting collisions above 0% and below 100%. Non-reacting collisions result in the rebound of the agents, they are reoriented in a new randomly chosen direction.

3.2 Simulations

3.2.1 Relation between computational and biological time

The simulation environment was populated with 5 virtual enzymes (equivalent to 1.22×10^{-2} mM) and 20 000 molecules of a generic substrate (equivalent to 4.87×10^{-2} mM). This relation of 4 orders of magnitude between enzyme and substrate concentrations is similar to the ones verified in enzymatic assays where the concentrations are specified [55, 56, 57]. The relatively few number of enzymes in the simulation allowed for the maintenance of an appropriate ratio, while keeping within computational constraints in the possible number of agents. The initial number of product

agents is zero. Upon a collision between a substrate and enzyme, an enzyme-substrate complex enzyme is formed

The radius of the enzyme agent was defined as 3.579 nm, based on a molecular weight (MW) of 50 kDa, which is a reasonable size for an enzyme (for a comparison with enzymes of common metabolic pathways, refer to Table 4.11 or Table 4.12 in Chapter 4) The substrate agent has a radius of 0.589 nm, which would be roughly equivalent to a chemical compound with 70 Da and similar to pyruvate or ethanol (Table 4.13). The proper definition of velocities of agents is discussed in the next section, in the simulation of a real enzyme and substrate. For this first group of simulations, a velocity was estimated from the numeric value of the of the diffusion coefficient an equivalent molecule would have. This velocity assured that the distance covered by the agent in each discrete time step was inferior to the diameter of the smallest agent. The practical reason behind this was so that no collision could be missed, in a given time step where the discrete "leap" of the agent could not go over entirely another agent, without testing a possible reaction or altering its direction.

The reaction radius for the virtual enzyme and substrate was set to 1.15 (a 15% increase relative to the hydrodynamic radius). The ratio of formation of *ES* agent was set to 100%. Thus, all the collisions between enzyme and substrate formed an *ES* agent. The focus of this simulation was the computational variable of the time between the formation of *ES* agent and dissociation to enzyme and product (simk$_{cat}$). Small values of simk_{cat} were tested, to simulate a very fast isomerase: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25 and 30 ts. In a total of 14 individual simulations were run for each of the simk$_{cat}$. For each simulation, it was tracked the number of substrate, enzyme, *ES* and product agents at every time step, during a total of 1 000 time steps. The reaction's velocity for each simulation was determined as the slope of a linear regression of the product formed as a function of time steps. The linear regression was forced through the origin (*0,0* coordinates), except in the simulations with simk$_{cat}$ of 1 and 2 ts.

Additionally, the relative occupancy of enzymes was calculated for given points of each simulation (10, 50, 100, 200, 500 and 1 000 time steps). The relative occupancy of enzymes, ES/E_t, is defined as the fraction of enzymes engaged in enzyme-substrate complexes, within the total population of enzymes.

The real velocity of a diffusion-limited reaction was calculated as in eq. 3.2.

$$\begin{aligned} V &= K_{eq}.[E].[S] = \\ &= 10^{10}M^{-1}.s^{-1} \times 1.22 \times 10^{-5}M \times 4.87 \times 10^{-2}M = \\ &= 5.94 \times 10^{-3}M.s^{-1} \end{aligned} \tag{3.2}$$

in which K$_{eq}$ is the collision constant of a enzyme-substrate system [50], and [E] and [S] the concentration of enzyme and substrate in the simulation environment.

The simulations' velocities (dP/dt) were converted from the units of agent/ts to M/ts by diving by the reactor volume and the Avogadro constant (6.022×10^{23}).

The relation between biological and computational time was calculated, as in eq. 3.3.

$$\frac{Real\ time\ velocity}{Simulation\ velocity} = \frac{5.94 \times 10^{3} M.s^{-1}}{2.89 \times 10^{-6} M.ts^{-1}} = \\ = 2.05 \times 10^{9} ts.s^{-1} \tag{3.3}$$

This relation between the time step and second was used to calculate the k_{cat} of the simulations from their velocities, by multiplying the velocity value by 2.05×10^{9} ts/s and the number of enzymes in the simulation (5 agents).

The relative occupancy of enzymes (ES/E_t) was calculated by the average of ES agents in a simulation, up to several point designed by time steps (10, 50, 100, 200, 500 and 1 000), and divided by the total enzyme agents in the simulation environment, either free or engaged in the enzyme-substrate complex.

To assess the relation between $simk_{cat}$ and k_{cat}, the relation between two consecutive (denoted by the i index) $simk_{cat}$ and the inverse of consecutive k_{cat} were calculated. Without limitation by diffusion, the two relations should be equivalent (eq. 3.4). The deviation between the two resulting values was calculated.

$$\frac{simk_{cat\,i+1}}{simk_{cat\,i}} = \frac{k_{cat\,i}}{k_{cat\,i+1}} \tag{3.4}$$

3.2.2 Velocity of agents, diffusion and Brownian motion

The assignment of agent velocities in simulation scenarios with different number of substrate agents was carried out in a cubic simulation environment, with a volume of 8.85×10^{5} nm^3.

All reactions between enzyme and substrate agents were disabled, by setting the ratio of reacting collisions to 0%.

The enzyme and substrate agents physical dimensions were defined according to a real enzyme and its substrate, a 2-hydroxymuconate tautomerase (EC 5.3.2.6) from *Pseudomonas putida*, coded by the gene 2-hydroxymuconate, that catalyses the conversion of 2-hydroxy-muconate to 2-oxo-3-hexenodiate. The enzyme's molecular weight is 225 kDa, and the radius 2.62 nm. The product agent has a radius of 0.93 nm.

Simulations with 750, 1 000, 1 250, 2 500, 5 000, 10 000 and 20 000 substrate agents were run. For each simulation, it was tracked for each type of agent the initial and final positions in each of the 3 cartesian coordinates (x, y and z), during 1 000 time steps, and whether it had hit the reactor boundary or not.

The displacement (R) of each agent was calculated by the eq. 3.5.

$$R = \sqrt[2]{(x_f - x_i)^2 + (y_f - y_i)^2 + (z_f - z_i)^2} \tag{3.5}$$

where x, y and z are the cartesian coordinates, and the subscript i denotes the initial coordinate and f the final coordinate.

Averages of the squared displacement (in μm^2) were calculated for enzymes and substrates, enzymes and substrates that didn't hit the reactor boundary, and enzyme and substrates that had hit the reactor boundary.

The simulation diffusion coefficient was calculated using the squared displacement of enzymes and substrates that didn't hit the reactor boundary, by the eq. 2.3, in which t is the time during which the agents were moving (1000 time steps, or 4.88×10^{-7} s).

This simulation diffusion coefficient was compared to the theoretical diffusion coefficient, calculated by the Stokes-Einstein equation (eq. 2.4, taking into account the radii of the agents, a temperature of 37 °C (310.15 K), and the viscosity of water at 37 °C.

Differences between the theoretical and simulation diffusion coefficients were expressed as a deviation from theoretical diffusion and by division.

If the absolute deviation of the substrate agents was superior to 15%, a new iteration of the simulation was run. The new values of velocity for the agents were calculated by multiplying the current velocities by the square root of the division between theoretical and simulation diffusion coefficients. The number of iterations for each of the initial substrate scenarios varied from 3 to 5.

3.2.3 Calibration of K_m

The calibration of K_m was performed in groups of simulations with different initial substrate agents, resembling wet lab kinetic assays.

The simulation environment was cubic, with a volume of 8.85×10^5 nm^3. The size of the agents was modelled after 2-hydroxymuconate and its substrate, as discussed in the previous subsection.

Using the inverse linear relation established between $simk_{cat}$ and k_{cat}, in the simulations with the virtual enzyme, the k_{cat} of 2-hydroxymuconate of 1.39×10^6 s^{-1} [51] was converted to a $simk_{cat}$ of 1086 ts.

Three implementations were tested to assign a modelling variable to the $simK_m$, a computational equivalent to the kinetic parameter K_m.

In the first and the second implementations, $simK_m$ was defined as the probability of a collision between an enzyme and substrate agent originating a *ES* agent. The difference between them is that, in the first implementation, each possible collision was tested in each time step. This implementation was tested for $simK_m$ values of 100%, 75%, 50% and 25%, with initial substrates of 750, 1000, 1250, 2500, 5000 and 10 000. In the second implementation of $simK_m$, only one collision was resolved in each time step. This implementation was testes for $simK_m$ values of 100%, 75%, 50% and 25%, 10%, 1% and 0.1% with initial substrates of 750, 1000, 1250, 2500 and 5000 agents.

A third implementation was tested, in which $simK_m$ was the probability of a *ES* agent dissociate into enzyme and product agents, as opposed to enzyme and substrate. Both dissociations

happened after the computational time defined by simk_{cat}. In this case, the dissociation happened at the 1 086th time step after the formation of the *ES*.

For each individual simulation, with a different combination of simK_m and initial substrate agents, the reaction velocity, dP/dt, was calculated and converted to real units ($\text{mM}_{Prod}.\text{s}^{-1}$) as discussed previously.

The velocity of reaction as a function of initial substrate was fit to the Michaelis-Menten equation, by non-linear least squares regression [58], and by linear transformations of Lineweaver-Burke, Eadie-Hofstee and Hanes-Woolf [59]. Values determined by the non-linear method were used for comparison between simulations with different simK_m.

3.2.4 Reaction radii and non-reactive agents

2-hydroxymuconate was simulated again in a cubic environment, with 8.85 nm^3, and 10 000 substrate-sized agents. Different scenarios of initial substrate concentrations were achieved by replacing substrate agents with non-reactive agents (simK_m of the enzyme 0% towards these agents), in the necessary ratio. Scenarios with 25, 50, 100, 250, 500, 750, 1 000, 1 250, 2 500, 5 000 and 10 000 substrate agents were tested. This group of scenarios were tested for different reaction radius: 2, 4 and 10.

Simulation reaction velocity (dP/dt) was calculated as previously, in the linear velocity portion of the scenario of 25 substrate agents. Reaction velocity was calculated for each of the three replicate simulations. The linear velocity portion was deemed to occur from the beginning of the simulation up to 6000 time steps.

The velocity of reaction (average and standard deviation of the triplicates) as a function of initial substrate was fit to the Michaelis-Menten equation, by non-linear least squares regression [58], and by linear transformations of Lineweaver-Burke, Eadie-Hofstee and Hanes-Woolf [59]. Values determined by the non-linear method were used for comparison between simulations with different simK_m.

Noise assessment consisted on following the rate of product formation of each of the replicates, in the total of the followed time steps and in the first 6 turnover events (6 000 time steps).

The standard deviation of samples from product existence at a given time step was calculated, and the relative deviation was the result of diving the standard deviation by the average of product. This was measured from the first turnover event (at time step of 1 086) to the end of the linear portion of velocities (at time step of 6 000). The average of relative deviation was calculated within this portion.

3.2.5 Simulation of steroid δ-isomerase enzyme

Steroid δ-isomerase was simulated again in a cubic environment, with 8.85 nm^3, and 2 000 substrate-sized agents. 5 enzyme agents were used, as in the previous simulations. Different scenarios of initial substrate concentrations were achieved by replacing substrate agents with non-reactive agents (simK_m of the enzyme 0% towards these agents), in the necessary ratio. Scenarios

with 10, 15, 20, 25, 50, 100, 250, 500, 1 000 and 2 000 substrate agents were tested. Reaction radius was set to 4. simk_{cat} was set to 54122, in accordance to the published value of k_{cat} [52].

Simulation reaction velocity (dP/dt) was calculated as previously, in the linear velocity portion of the scenario of 10 substrate agents. Reaction velocity was calculated for each of the three replicate simulations. The linear velocity portion was deemed to occur from the first turnover event (time step of 55 000) of the simulation up to 180 000 time steps.

The velocity of reaction (average and standard deviation of the triplicates) as a function of initial substrate was fit to the Michaelis-Menten equation, by non-linear least squares regression [58], and by linear transformations of Lineweaver-Burke, Eadie-Hofstee and Hanes-Woolf [59]. Values determined by the non-linear method were used for comparison between simulations with different simK_m.

Noise assessment consisted on following the rate of product formation of each of the replicates, in the total of the followed time steps and in the first 6 turnover events (180 000 time steps).

The standard deviation of samples from product existence at a given time step was calculated, and the relative deviation was the result of diving the standard deviation by the average of product. This was measured from the first turnover event (at time step of 55 000) to the end of the linear portion of velocities (at time step of 180 000). The average of relative deviation was calculated within this portion.

3.3 Data retrieval and curation

The data of the simulated enzymes, 2-hydroxymuconate and steroid δ-isomerase, were initially extracted from the BRENDA database. Additional information was collected from the respective source publications identified in BRENDA [51, 52].

In order to assess the availability of data to perform ABM simulations of a full pathway, the enzymes and metabolites involved in glycolysis and pyruvate metabolism of *E. coli* were collected, according to map00010 (Glycolysis/Gluconeogenesis) of KEGG [60], "Pathway: glycolysis I (from glucose-6-phosphate)" and "Pathway: mixed acid fermentation" from EcoCyc [61]. Enzymes are identified by the Enzyme Commission (EC) number and with the corresponding gene identifier. Metabolites were identified by a non-standard readable name and the PubChem CID [62]

After collecting the participating biomolecules, the following data was collected on the enzymes: substrates and products of in vivo reaction, the kinetic parameters of the in vivo reaction, the molecular weight of the enzymes and the number of subunits.

Substrates and products were specified in the pathway maps or the source publication. The kinetic parameters were initially collected from the BRENDA [63] and EcoCyc [61] databases, and checked individually in the respective source. The molecular weight was extracted from EcoCyc [61]. Molecular weight values were available either as direct experimental values of the multimer (when the quaternary structure of the protein is composed of more than one of the individually coded polypeptide), experimental values of the polypeptide or values inferred from the

aminoacid sequence. When molecular weights were available only for the individual polypeptide, a final molecular weight was calculated by multiplying the molecular weight of the polypeptide by the number of subunits. Molecular weight of the enzymes was used to calculate the hydrodynamic radius, according the correlation described by The number of subunits is available in EcoCyc and UniProt [64]. UniProt data was preferred, as it keeps references for the stored value. There was one practical case of conflict in the curated data, the ackA enzyme.

From the substrate and product list, a final list of metabolites was collected. Size data was collected for each of the biomolecules. The method to calculate the V_{vdW} required the chemical formula and the number of aromatic and non-aromatic rings in the structure [46].

3.3.1 Kinetic data

Kinetic data was available in terms of Michaelis-Menten parameters, k_{cat} (s^{-1}) and K_m (mM). BRENDA and EcoCyc have kinetic data values. EcoCyc was searched by the coding gene ID and BRENDA by the EC number. To distinguish isoenzymes (enzymes catalysing the same reaction, but with different aminoacid sequence, size and kinetic properties), the UniProt identifier in the BRENDA register of literature sources was checked. When the field is empty, the information was searched for in the original source.

Kinetic data is very heterogeneous. In this work, the original works cited as sources for the data were checked individually, to assess the existing variables in the parameter determination. Table 3.1 organizes the possible criteria in the individual experiments of parameter determination, and the desired values for an hypothetical simulation of the pathway. These criteria were followed to choose parameters' values from all the available sources (for full data, see supplementary material A.7).

Table 3.1: Criteria for selecting kinetic data from different published literatures and experimental variables

Order	**Criteria**	1st preference	Other acceptable values (ordered)
1	Substrate of enzyme assay	Same as in vivo reaction	-
2	Origin of corresponding gene	*E. coli* K-12 MG1655	*E. coli* K-12 (any substrain), *E. coli* (any strain), proteins mentioned in literature to be very similar, mutated enzymes from *E. coli* genes, gene from the phylogenetically closest organism to *E. coli*
3	Available information on a given paper	Pair of K_m and k_{cat} for the enzyme in direct and reverse direction, for the same conditions.	K_m and k_{cat} pair, direct and reverse direction pairs for one of the parameters, isolated parameters
4	pH of enzymatic assay	7.4 (physiological)	Mesophilic range and higher than 7.4, mesophilic range and below 7.4, physiological range
5	Temperature of enzymatic assay	37 °C	Closest to 37 °C, mesophilic range
6	Detection method	Direct assays	NADPH-detecting coupled-enzyme spectrophotometry assay, others
7	Parameter determination method	Non-linear fit to Michaelis-Menten constant	Linear plots
8	Co-factor presence	Physiological concentrations	Closest available

Chapter 4

Results and Discussion

This chapter covers the results obtained from simulations and the literature review. The individual results are discussed, as well as their implications in the methods of the following subsections.

4.1 Simulations

The simulations tested a simple reaction, with a single substrate and product, emulating the behaviour of a generic isomerase in an *in vitro* enzymatic assay, carried in a microscale biorreactor and aqueous media. This simple reaction was used to validate the physical, chemical and biological assumptions of the model without unnecessary complexity.

The model aims to replicate an enzymatic assay, mimicking the conditions under which the kinetic parameters are determined in laboratory settings, and as such molecular crowding was not accounted for.

Whether the assumptions of the Michaelis-Menten catalytic mechanism, such as a relative higher concentration of free substrate, and the kinetic parameters determined *in vitro* can capture the kinetics of enzymes inside the cell is debatable [65, 66, 67]. However, the determination of kinetic constants in *in vitro* assays that resemble the cytoplasmic environment has been shown to improve significantly the kinetic modelling of yeast [67]. If published kinetic parameters of *in vitro* experiments allow for the determination of intrinsic enzyme characteristics, the simulation of the same enzymes in an environment closer to the cytoplasmic composition could offer insight into the kinetic behaviour of enzymes *in vivo*, and by extension the kinetics of metabolic pathways and cellular systems.

4.1.1 Relation between computational and biological time

The first aim of this work was to define the duration of one simulation time step in seconds. This relation is important because the temporal scale of reactions, determined by enzyme kinetics, should be consistent with the movement velocity of the agents.

The strategy to get such result was to simulate a diffusion-controlled enzymatic reaction. The time step should represent a sufficiently small fraction of the second in order to be able to simulate

super-efficient enzymes, *i.e.* diffusion-controlled enzymes with second-order rate constants of 10^8 to 10^{10} $M^{-1}.s^{-1}$ [50].

However, if the time step represents an excessively small fraction of biological time, it will create an unnecessary computational burden in the simulation. As such, the first step was testing a virtual enzymatic reaction which consisted of an isomerization, with a single substrate and single product, to which the Michaelis-Menten equation can be applied directly, and occurring in the limit of the catalytic efficiency.

From the value of K_{eq}, the maximal velocity of reaction in a real biological system of enzyme and substrate can be estimated. The real reaction velocity can be then compared to the simulation reaction velocity of an equivalent system, with a fast enzyme whose reaction rate is limited by diffusion.

In order to simulate enzymes with high catalytic efficiencies, a virtual enzyme with different $simk_{cat}$ ranging from 1 to 30 ts.

The rate of product formation, dP/dt, corresponds to the number of product agents produced per time step - the simulation's reaction velocity. It will decrease as the $simk_{cat}$ increases (Figure 4.1). The parameters of the linear regression, namely the slope and regression coefficient, are presented in Table 4.1.

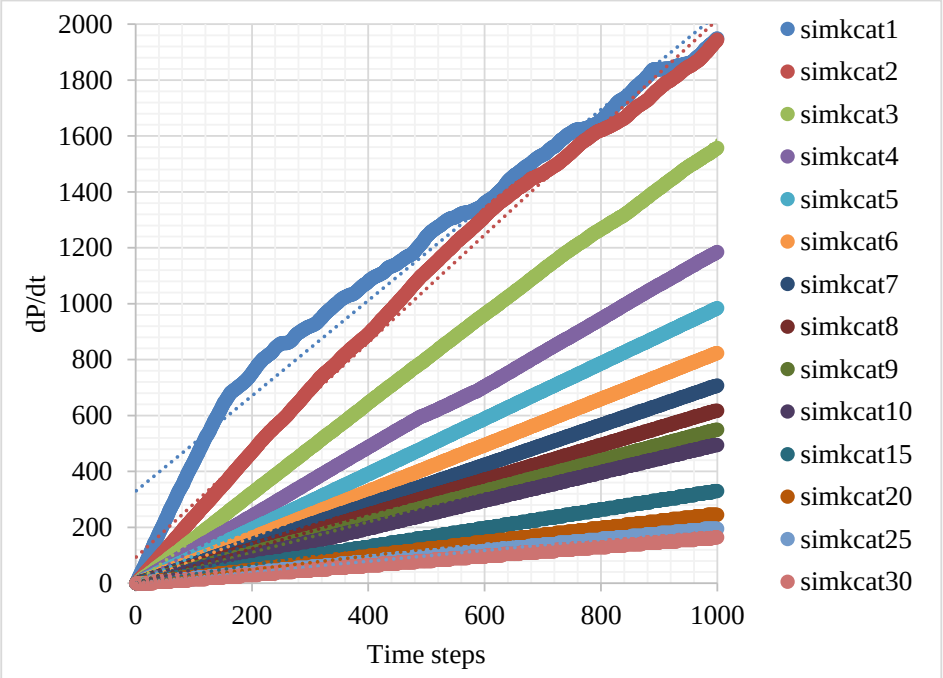

Figure 4.1: Plot of product agents as a function of time steps, in simulations with the virtual enzyme and several $simk_{cat}$ (1 to 30 ts).

Table 4.1: Linear regression parameters of *dP/dt* of simulations for different simk$_{cat}$

simk$_{cat}$ (ts)	*dP/dt* (P.ts^{-1})	R^2
1	1.707	97.07%
2	1.925	99.22%
3	1.585	99.99%
4	1.187	99.99%
5	0.982	100%
6	0.824	100%
7	0.706	100%
8	0.617	100%
9	0.549	100%
10	0.495	100%
15	0.329	99.99%
20	0.246	99.98%
25	0.196	99.97%
30	0.163	99.96%

The boundary of diffusion control was established by three different analysis: the linearity of the *dP/dt* regression, as measured by the regression coefficient, the relative occupancy of enzymes and the influence on velocity of simk_{cat}. Relative occupancies reaching and maintaining 100% are indicative that turnover number is the limiting step, and thus indicate the boundary of diffusion control. From the simk$_{cat}$ of 5 ts on, the relative occupancies are maintained at 100% throughout the simulation (Figure 4.2).

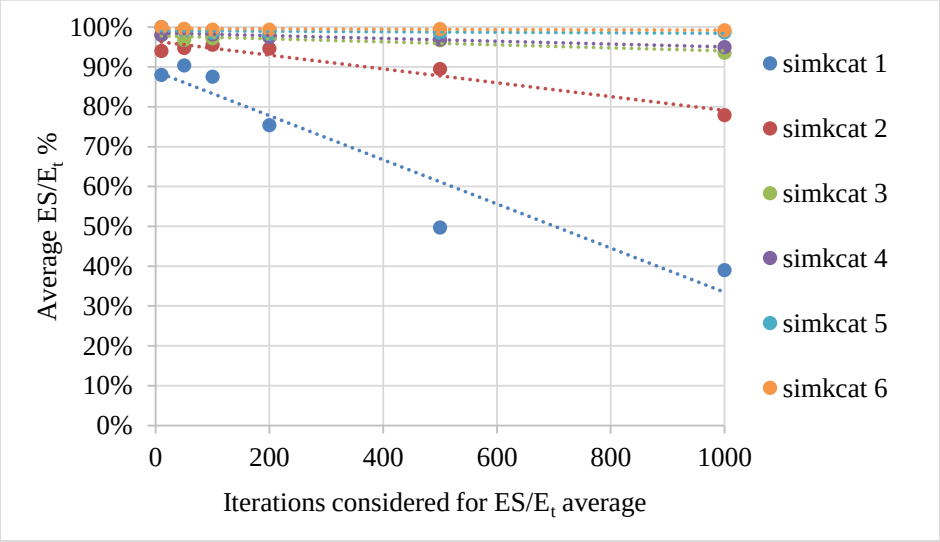

Figure 4.2: Plot of relative occupancy of enzymes from simulation of virtual enzymes with different simk_{cat}, at different points in the simulation.

The influence of simk_{cat} on velocity was assessed on basis of the following reasoning: if turnover number is the limiting step, the velocity of the reaction on the simulation scenario varies in a direct inverse proportion of the simk_{cat}. The deviation between the relation of values of calculated k_{cat} and its inverse of simk_{cat} approaches 0% as the reaction ceases to be controlled by diffusion. The results (Table 4.2) corroborate the conclusions extracted from the relative occupancies.

Notably, the existence of such relation between two simk$_{cat}$ and their equivalent k$_{cat}$ allows to estimate the resulting k$_{cat}$ from using a certain simk$_{cat}$ as input, and vice-versa, with reasonable accuracy. The practical significance of this is that the k$_{cat}$ can be used directly as an input in the simulations. A pair of simk_{cat} and k_{cat} values from simulated enzymes, above the diffusion control limit, and a simple proportion is an efficient way to calculate the simk_{cat} of a new enzyme to be simulated.

Table 4.2: Relation between simk$_{cat}$ and equivalent k$_{cat}$

simk$_{cat}$ (ts)	k$_{cat}$ (P.E^{-1}.s^{-}1)	simk$_{cat\,i+1}$/ simk$_{cat\,i}$	k$_{cat\,i}$/ k$_{cat\,i+1}$	Deviation (%)
1	7.01×10^{8}	2.00	0.89	-55.7%
2	7.90×10^{8}	1.50	1.21	-19.0%
3	6.51×10^{8}	1.33	1.34	0.16%
4	4.87×10^{8}	1.25	1.21	-3.33%
5	4.03×10^{8}	1.20	1.19	-0.61%
6	3.38×10^{8}	1.17	1.17	-0.07%
7	2.90×10^{8}	1.14	1.14	0.15%
8	2.53×10^{8}	1.13	1.12	-0.08%
9	2.25×10^{8}	1.11	1.11	-0.18%
10	2.03×10^{8}	1.50	1.51	0.37%
15	1.35×10^{8}	1.33	1.34	0.43%
20	1.01×10^{8}	1.25	1.25	0.32%
25	8.04×10^{7}	1.20	1.20	0.37%
30	6.67×10^{7}			

The value of k$_{cat}$ for the enzymes with simk$_{cat}$ below 5 should be comparable to the k$_{cat}$ of real enzymes that are very close to kinetic perfection, such as catalase. The highest reported value of k$_{cat}$ for a catalase in the BRENDA database is around $3\times10^{6}s^{-1}$ (EC number 1.11.1.6), *i.e.* there is a difference of two orders of magnitude between the perfect simulation k$_{cat}$ and the maximum k$_{cat}$ occurring in reality.

Full data is available as supplementary material S1 (A.1).

4.1.2 Velocity of agents, diffusion and Brownian motion

The motion of biomolecules in aqueous or intracellular media follows a random-walk behaviour (Brownian motion). In the simulation space, the solvent molecules of water are not explicit. This is due to the computational cost of creating agents, and representation all solvent molecules as agents would render the simulations intractable.

In this simulation environment, Brownian motion is created by collision between the agents, and the velocity of each agent species is iterated upon until the resulting diffusive behaviour matches the one expected for that particular molecule sizes and environment constants, such as temperature and viscosity of the simulated reaction medium. Achieving a realistic diffusive behaviour is relevant to simulate a real enzyme. In this section of the work, the diffusive behaviour was approximated to that of a real isomerase, 2-hydroxymuconate tautomerase (EC 5.3.2.6), taking into account the real dimensions of the enzyme and its substrate, the chemical compound 2-hydroxymuconate.

Given this, each simulation scenario with a different number of agents needs to have the agent's velocity calibrated to, within that scenario, yield an average displacement of each type of agent that is equivalent to the one predicted by the diffusion coefficient, as calculated by the Stokes-Einstein equation.

The random trajectories of the agents arise from collisions between the agents, as can be seen in the time-lapse tracking of an enzyme agent trajectory (Figure 4.3A). The trajectories of agents in the simulation are consistent with the expected random movement (Figure 4.3B,C), and are comparable to the path reconstructed from single-enzyme tracking (Figure 4.3D) [68].

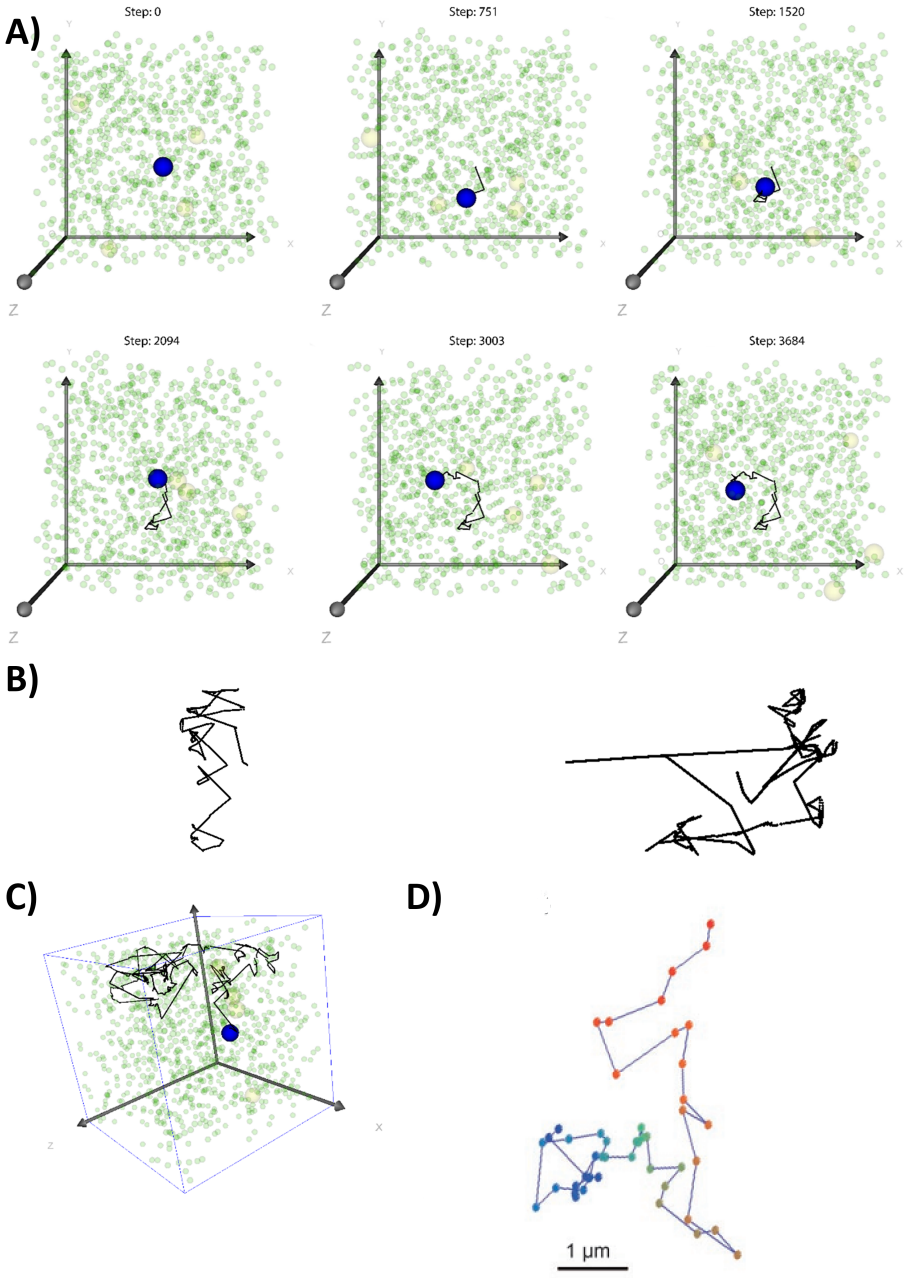

Figure 4.3: Trajectories of an agent in the three dimensional simulation environment in a) 4D perspective, b) 3D perspective, c) 2D perspective. For comparison, d) trajectory reconstruction of a fluorescent-labeled enzyme from experimental tracking (Source: Peneva et al, 2008).

Besides visual inspection of the trajectory of the agents, the convergence of the average value or an increasing number of displacement observations, from a single observation to the average of displacements of all the identical agents.

The convergence of the accumulated average can be seen in Figure 4.4 for each of the simulations with different substrate number. From the graph, it is possible to observe that convergence to an average displacement happens before 750 substrate agents are considered to the accumulated average. This number was the minimum of substrate agents in a simulation scenario.

Figure 4.4 shows the same data, with more detail of the initial convergence around 300 substrate agents considered.

The iterative process of assigning agents with velocity was applied to the following groups of simulation, which related reaction velocity with substrate concentration.

However, to determine Michaelis-Menten parameters, simulations with substrates below 1.38 mM were needed, as the K_m values for the simulated enzymes were 0.145 mM and 0.050 mM. Simulations with initial substrate concentrations near the K_m value are needed to obtain velocities limited by substrate availability. So, a complementary strategy was used, in which part of the substrate-like agents were converted in non-reactive agents, *i.e.* these agents became obstacles. The velocity of the agents was determined iteratively for the scenario with highest number of substrate agents, and remained so for the rest of simulation runs.

Table 4.3: Velocities defined for enzyme and substrate agents, in 2-hydroxymuconate tautomerase simulation scenarios with different numbers of initial substrate, and deviation of the resulting diffusion coefficient from theoretical values

Initial substrate agents	Velocity enzyme (μm.s^{-1})	Deviation (enzyme)	Velocity substrate (μm.s^{-1})	Deviation (substrate)
20 000	8444	22%	16271	-5%
10 000	6170	5%	10663	9%
5000	4930	10%	7939	0%
2500	4205	9%	6770	5%
1250	3386	7%	6248	5%
1000	3641	10%	5942	2%
750	3414	14%	5750	0%

Full data can be found in the supplementary material S2 (A.2).

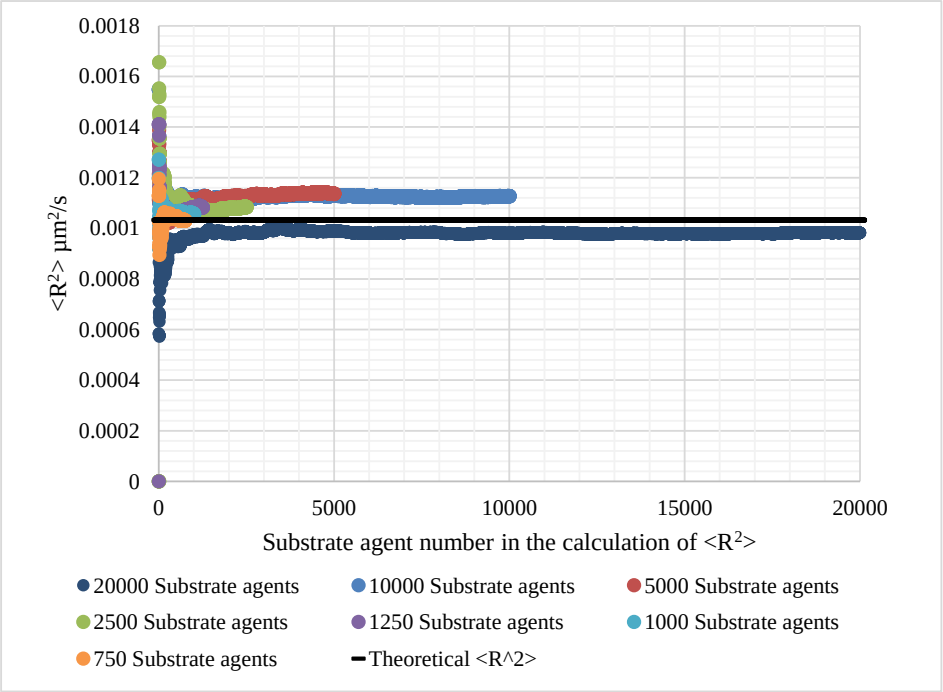

Figure 4.4: Accumulated average of displacements of substrate agents, plotted against the number of substrate agents included in the average calculation, for simulation scenarios with different substrate agents.

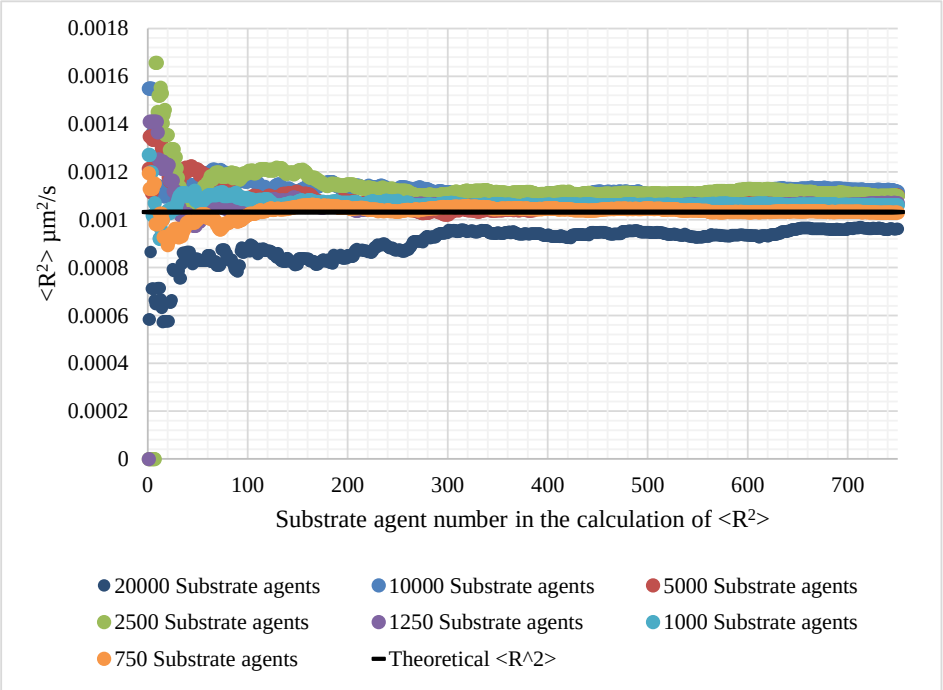

Figure 4.5: Accumulated average of displacements of substrate agents, plotted against the number of substrate agents included in the average calculation (up to 750 substrate agents), for simulation scenarios with different substrate agents.

4.1.3 Calibration of K_m

The calibration of simK_m to K_m was based on the reproduction of experimental assays to determine Michaelis-Menten kinetic parameters, which measure the velocity of the reaction for different initial concentrations of substrate, at the substrate saturation level and lower.

The strategy to find a relation between the simK_m and K_m uses the biological meaning of the parameter, that represents the substrate concentration at which the velocity of the reaction is half of the maximal velocity. Maximal velocity is given by k_{cat}, which quantifies the conversion velocity of the enzyme's active site, multiplied by the concentration of enzymes.

Three implementations were tested to assign a modelling variable to the simK_m, a computational equivalent to the kinetic parameter K_m.

In the first and the second implementations, simK_m was defined as the probability of a collision between an enzyme and substrate agent originating a ES agent. The difference between them is that, in the first implementation, each possible collision was tested in each time step. The distinction between the two implementations is irrelevant at probabilities of 100%, as it was implemented in the previous simulations. In the first implementation, the resulting Michaelis-Menten kinetic parameters were not sufficiently different across different simK_m, as can be seen in the Table 4.4.

Table 4.4: Michaelis-Menten kinetic coefficients for different values of simK_m, implemented as the probability of an enzyme and substrate forming a ES agent, with all possible collisions tested per each time step

simK_m	K_m (mM)	k_{cat} (s^{-1})
100%	2.39	2.09×10^6
75%	1.08	1.99×10^6
50%	1.10	1.99×10^6
25%	1.21	1.99×10^6

A third implementation was tested, in which simK_m was the probability of a ES agent dissociate into enzyme and product agents, as opposed to enzyme and substrate. Both dissociations happened after the computational time defined by simk_{cat}.

However, this implementation of simK_m altered the resulting k_{cat} significantly, instead of limiting its influence to the affinity of the enzyme agent towards the substrate (Table 4.5).

Table 4.5: Michaelis-Menten kinetic parameters resulting from simulations with different values of simK_m, implemented as the probability of an ES agent dissociating into enzyme and product agents

simK_m	K_m (mM)	k_{cat} (s^{-1})
100%	1.48	2.05×10^6
75%	0.897	1.46×10^6
50%	0.485	9.72×10^5

Table 4.5: Michaelis-Menten kinetic coefficients for different values of simK_m, implemented as the probability of an ES agent dissociating into enzyme and product agents

simK_m	K_m (mM)	k_{cat} (s^{-1})
25%	0.278	4.66×10^5

The most successful implementation of simK_m was its definition as the probability of an ES-forming collision, with one collision resolved per time step. The results suggest that a logarithmic scale would be more appropriate in the manipulation of this computational parameter, as in the results for simK_m from 75% to 25% there is no clear difference or trend between the resulting K_m in the simulation. Figures 4.6 and 4.7 show the corresponding Michaelis-Menten (Figure 4.6) and Lineweaver-Burke graphs (Figure 4.7) show the increase of substrate concentration influence over the velocity of reaction, as the simK_m decreases. The numeric values of the Michaelis-Menten parameters are presented in the Table 4.6.

Full data is available in supplementary material S4 (A.4), namely the original simulation data and the subsequent calculation of velocities and Michaelis-Menten parameters.

Table 4.6: Kinetic parameters resulting from simulations with different simK_m, defined as the probability of a collision of an enzyme and substrate agent reacting and forming an ES agent, with one collision resolved per time step

simK_m	k_{cat} (s^{-1})	K_m (mM)
100%	2.09×10^6	1.00
75%	2.01×10^6	0.69
50%	1.99×10^6	0.69
25%	2.06×10^6	0.94
10%	2.03×10^6	0.92
1%	1.93×10^6	1.18
0.1%	2.93×10^6	6.16

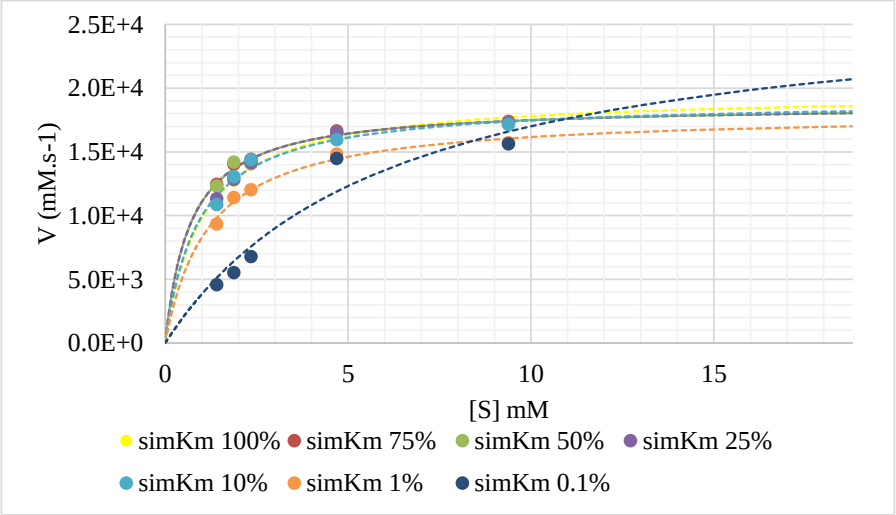

Figure 4.6: Plot of velocity *vs.* substrate concentration, with non-linear fit to Michaelis-Menten for simulations with simK_m implemented as the probability of a collision between enzyme and substrate forming a ES complex.

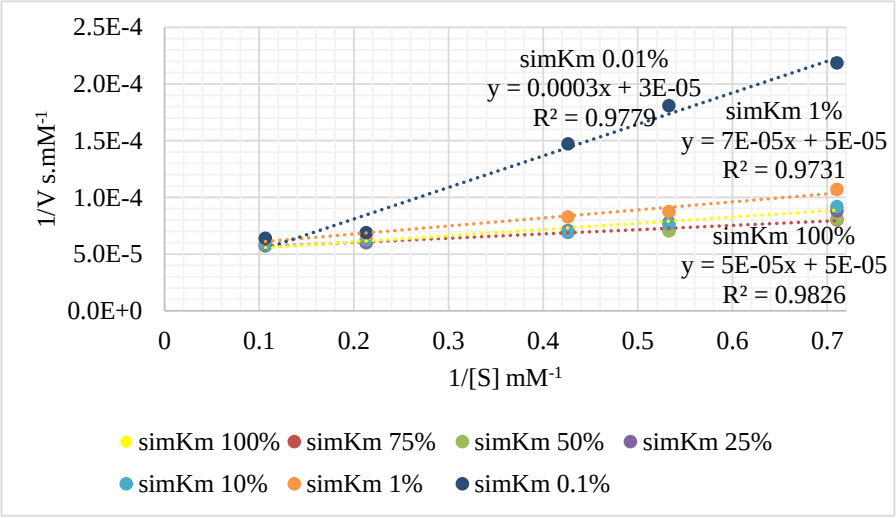

Figure 4.7: Lineweaver-Burke plot for simulations with simK_m implemented as the probability of a collision between enzyme and substrate forming a ES complex.

4.1.4 Reaction radii and non-reactive agents

It was observed that while there was a variation of the resulting K_m for different values of simK_m, as intended, the set of simulation runs with simK_m of 100% resulted in a K_m superior to the published value of 0.145 mM. To increase the affinity of the enzyme agent, the concept of reaction radius was applied.

The reaction radius defines a volume external to the enzyme and within which this agent anticipates potential reactive collisions. The input of this parameter was inserted in the model as a

multiple of the hydrodynamic radius of the enzyme. The concept is similar to the encounter radius of Smoldyn, a point-like particle-based modelling tool [39].

As previously discussed, to recreate low K_m values there is a need to simulate scenarios with low initial substrate concentrations. Without compromising the random motion of the agents, low substrate concentrations can be accomplished by taking a high substrate simulation scenario and replacing a fraction of substrate agents for non-reactive agents.

To reproduce the Michaelis-Menten kinetic parameters for 2-hydroxymuconate, the reaction radius of 1.15 was increased to 2, 4 and 10 times that of the hydrodynamic radius of the enzyme. The resulting Michaelis-Menten plots are shown in the Figure 4.8, and the values of the kinetic parameters in the Table 4.7. The increase of reaction radii lowered the resulting K_m, without altering k_{cat} values in a significant way.

The simulation with a reaction radius of 2 comes close to reproducing the kinetic behaviour of the 2-hydroxymuconate enzyme.

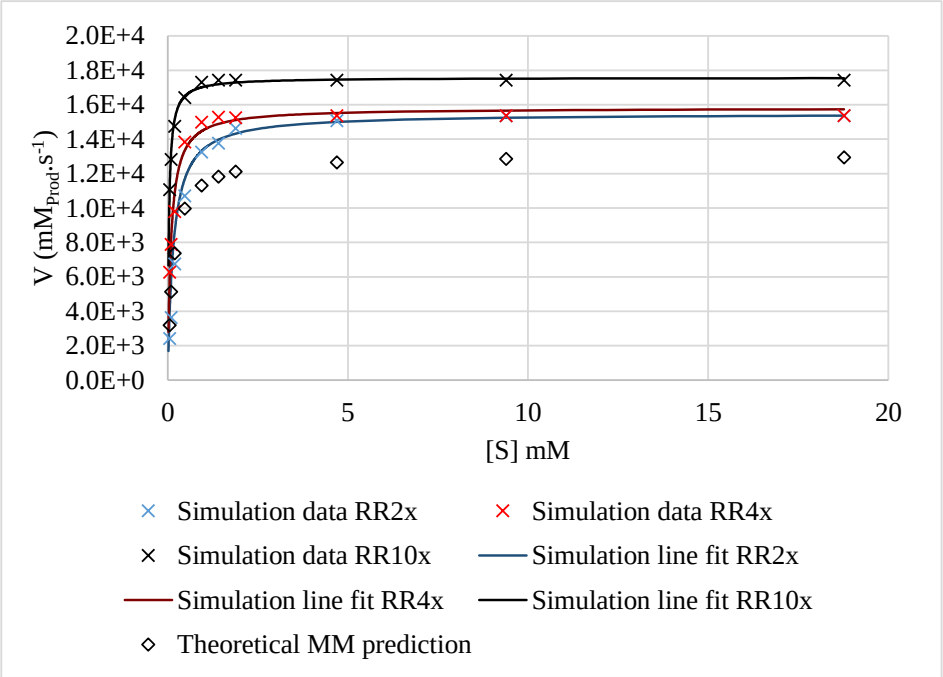

Figure 4.8: Plot of velocity *vs.* substrate concentration for the 2-hydroxymuconate tautomerase, with non-linear fit to Michaelis-Menten for simulations with different reaction radii (RR).

Table 4.7: Values of kinetic parameters for the simulation of 2-hydroxymuconate tautomerase with different reaction radius

Reaction radius	K_m (mM)	Deviation from published value (K_m)	k_{cat} (s^{-1})	Deviation from published value (k_{cat})
2	1.50×10^{-1}	6%	1.65×10^{6}	19%
4	1.10×10^{-1}	22%	1.68×10^{6}	21%
10	3.00×10^{-2}	79%	1.87×10^{6}	35%

4.1.5 Simulation of steroid δ-isomerase enzyme

Like the previous enzyme, steroid δ-isomerase is an isomerase enzyme with a high k_{cat}, which expedited the validation of the modelling hypothesis at the present stage of the work.

The process developed for reproducing the kinetic parameter K_m of the 2-hydroxymuconate enzyme was applied to another simple isomerase, the steroid δ-isomerase. The kinetic parameters for the conversion of 5-androstene-3,17-dione to 4-androstene-3,17-dione by steroid δ-isomerase were reported to be 0.050 mM for K_m and 2.79×10^{4} for k_{cat} [52]. The k_{cat} was converted to a $simk_{cat}$ with the previously established relation. Unlike k_{cat}, the K_m can not be simply converted to a single computational variable. However, an iterative approach can be used.

Published Michaelis-Menten parameters were used to predict the how the velocity of reaction would vary with various concentrations, and from that data it was possible to select the number of initial substrate agents for different simulation scenarios. Based on the previous results for 2-hydroxymuconate, the simulations of steroid δ-isomerase enzyme had their $simK_m$ defined to 100% and a reaction radius of 4. The Michaelis-Menten plot of these simulations is shown in Figure 4.9. The K_m obtained in the simulation was of 9.32×10^{-3} mM, and the k_{cat} was of 3.60×10^{4} s^{-1}.

4.1.6 Assessment of spatial noise

A spatial agent-based model at the biomolecular scale is suitable for assessing intrinsic stochastic noise. ABM would be a suitable method to portray the fluctuations in conformational structure of enzymes, that lead to stochastic catalytic activities[69]. However, along with that source of variability, an ABM allows for the assessment of local substrate concentrations fluctuations an its role in metabolic intrinsic noise [70, 71].

The average relative deviation of the number of product agents was calculated to assess the impact that the initial substrate concentration has on product formation (Table 4.8). Relative deviations at each time step vary with the catalytic cycle, with small increases at higher concentrations, and higher and delayed perturbations at lower concentrations (Figure 4.10).

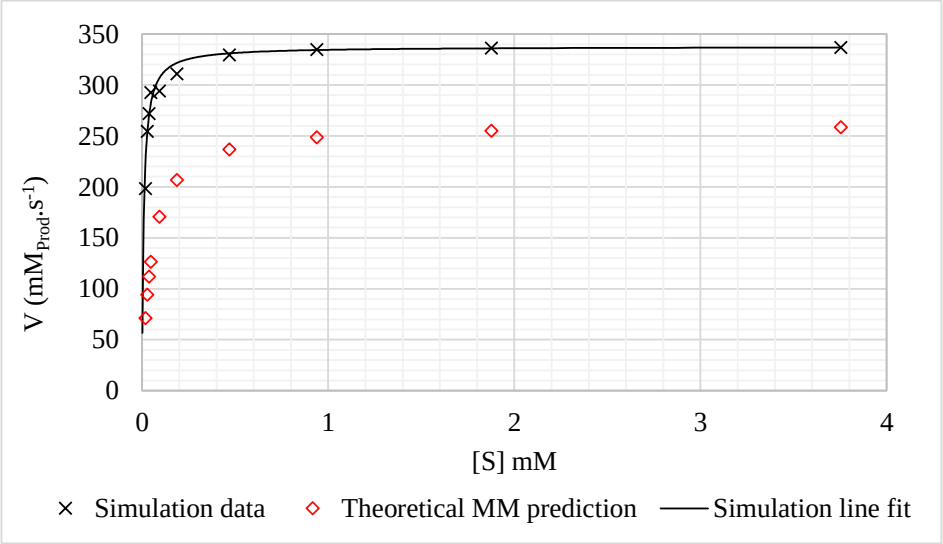

Figure 4.9: Plot of velocity *vs.* substrate concentration for the steroid δ-isomerase enzyme, with non-linear fit to Michaelis-Menten

Table 4.8: Average of relative deviation in product formation for simulations with different initial concentrations of substrate, for the 2-hydroxymuconate tautomerase simulation (with reaction radius of 2)

Concentration substrate (mM)	Average of relative deviation in product formation
0.0469	81%
0.0938	20%
0.188	31%
0.469	28%
0.938	6.0%
1.41	6.0%
1.88	2.3%
4.69	1.2%
9.38	0.1%
18.8	0.0%

Figure 4.11, from the simulation 2-hydroxymuconate enzyme, shows product formation for triplicate simulations at low, intermediate and high initial substrate concentrations. While at high concentrations product formation is highly homogeneous at the high concentration scenario, it becomes more random as the initial substrate concentration lowers to an intermediate concentration of 0.469 mM, while the simulations of lowest concentration stabilize early in the simulation, due to substrate limitations. The effect of variability is more noticeable in the detail of 6 initial turnovers (Figure 4.12). The same behaviour can be observed in the simulations of steroid δ-isomerase (Figure 4.13).

At high concentrations, the turnover events of replicate simulations are highly synchronized, with 5 additional product agents appearing near the time step of possible turnover (multiples of simk$_{cat}$), due to the simultaneous capture of substrate by the 5 enzyme agents. Turnover events are spread across time steps at lower concentrations, with the predictable cyclic nature of turnovers practically disappearing at the lowest concentration.

In the 2-hydroxymuconate tautomerase simulation, for a small concentration of substrate, 6.09×10^{-2} mM, the deviation in product formed is above 50%, and the deviation decreases as more substrate is available in the simulation. Given that the enzyme agent will form an enzyme-substrate complex if it detects any substrate agent, as specified by a simK$_m$ of 100%, the deviations in product formation are due to the unavailability of substrate agents within the reaction radius. While the mathematical modelling of enzymatic reactions accounts for the lowering of product formation rate resulting from lower availability of substrate, it does not account for the higher variability. The results show that there are considerable fluctuations in the availability of substrate, despite the fact that the three simulations had identical starting conditions, including same initial substrate concentration and homogeneous spatial distribution. Low substrate concentrations are not negligible in *in vivo* reactions, where as intracellular metabolite concentrations in the order of 10^{-2} mM or below are common, including metabolically important metabolites such as glycerol-3- phosphate and NADP$^+$ [72]. It is worth to note that the present assessment of variability takes place in the simulation of an *in vitro* assay, with water as its reaction medium. It is expectable that the intracellular environment is a more crowded medium, with higher viscosity, making the heterogeneity of local substrate concentrations more relevant [73].

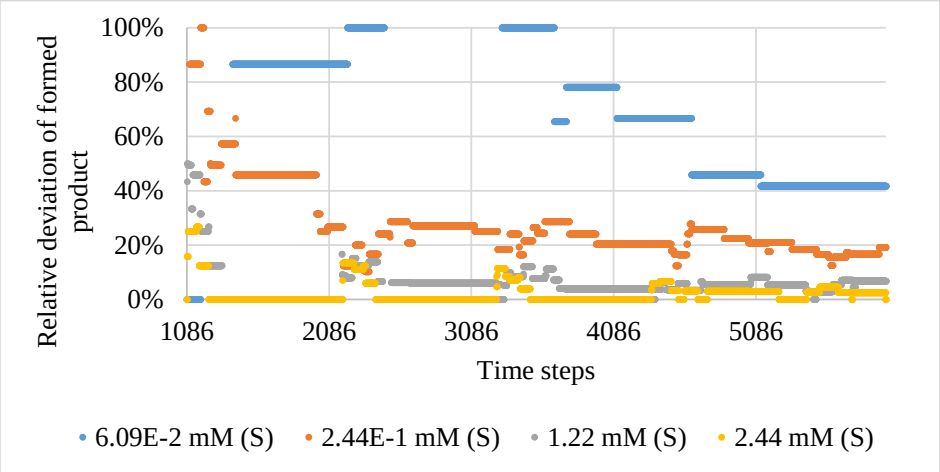

Figure 4.10: Relative deviation in product formation as a function of time steps, in the linear portion of velocity in the simulation of 2-hydroxymuconate

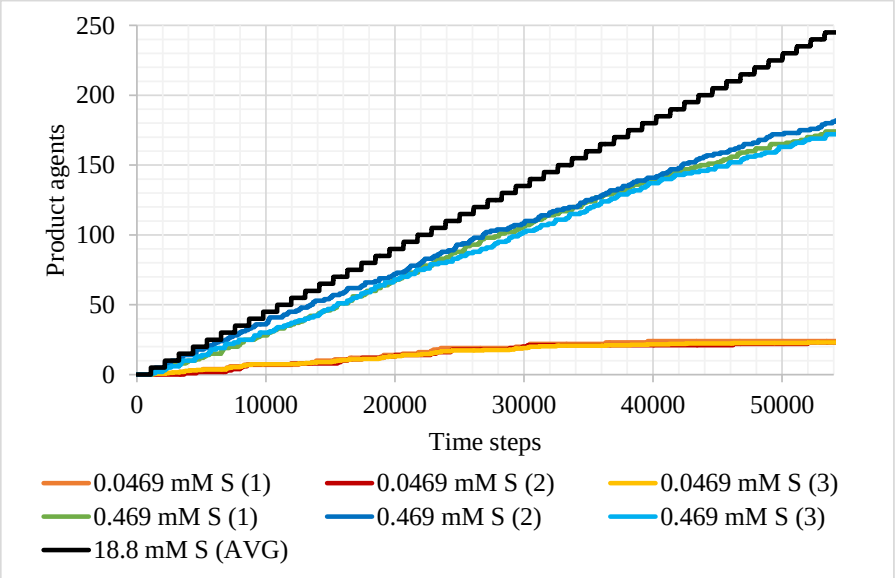

Figure 4.11: Product agents as a function of time steps, for high concentrations of substrate (18.8 mM), and replicate simulations of intermediate (0.469 mM) and low (0.0469 MM) concentrations, for the 2-hydroxymuconate tautomerase (reaction radius of 2).

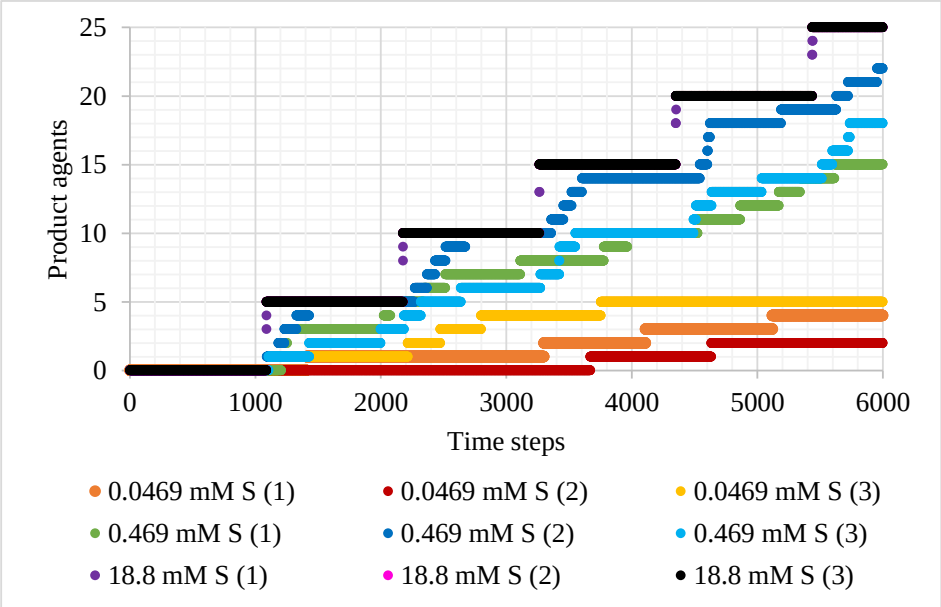

Figure 4.12: Product agents as a function of time steps in the first 6 turnover events, for high concentrations of substrate (18.8 mM), and replicate simulations of intermediate (0.469 mM) and low (0.0469 MM) concentrations, for the 2-hydroxymuconate enzyme.

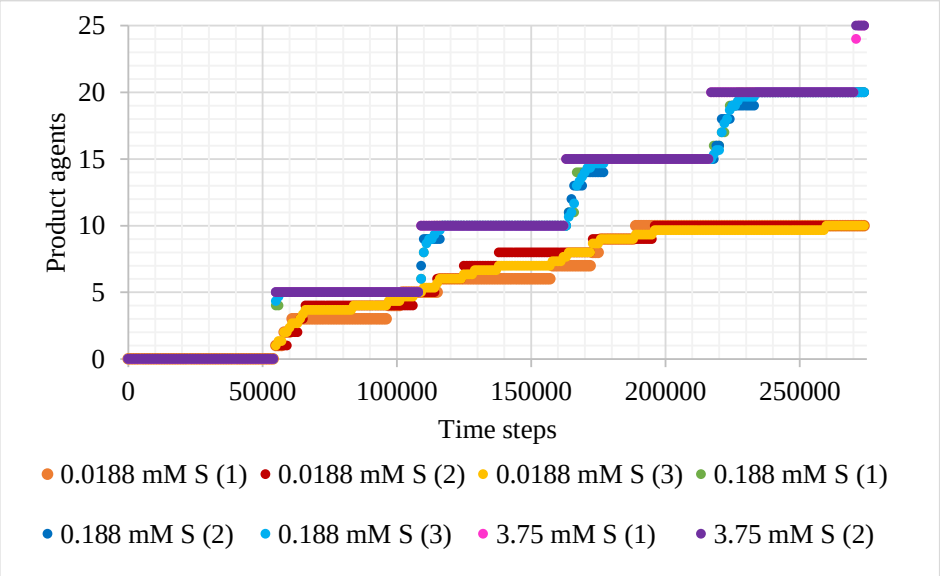

Figure 4.13: Product agents as a function of time steps in the first 6 turnover events, for high concentrations of substrate (3.75 mM), and replicate simulations of intermediate (0.188 mM) and low (0.0188 MM) concentrations, for the steroid δ-isomerase.

4.2 Data retrieval and curation

In order to determine the feasibility of the implemented model from a standpoint of the available biological data on the enzymes and metabolites was collected or estimated. Presented in the following tables, the collected values review are the necessary the primary biological data needed for an agent-based model of the glycolysis pathway and pyruvate metabolism for *E. coli*. It also contains the transformations into usable inputs, as is the case of size-related data. Of all the data collected, values on most parameters are fairly straightforward to find and that can be retrieved programatically. Manual curation was required for the kinetic data for enzymes and the determination of the metabolites' radii.

To define the logic of interactions, it is necessary to collect data on the pathway's enzymes and their *in vivo reactions* (Table 4.9 and Table 4.10). The computational parameters regarding reactions would be set to zero for all possible interactions that are not specified by the pathway's reactions. These parameters include reaction radius, simK$_m$ and simk$_{cat}$.

Table 4.9: Reaction data for glycolysis pathway in *E. coli*

EC	**Enzyme's gene**	**Substrates**	**Products**
2.7.1.2	glk	β-D-glucose + ATP	β-D-glucose-6-P + ADP
5.3.1.9	pgi	β-D-glucose-6-P	β-D-fructose-6-P
2.7.1.11	pfkB	β-D-fructose-6-P + ATP	β-D-fructose-1,6-2P + ADP
3.1.3.11	glpX	β-D-fructose 1,6-2P + H_2O	β-D-fructose-6-P + P_i
3.1.3.11	fbp	β-D-fructose-1,6-2P + H_2O	β-D-fructose-6-P + P_i
4.1.2.13	fbaB	β-D-fructose-1,6-2P	glycerone-P + D-glyceraldehyde-3-P
4.1.2.13	fbaA	β-D-fructose-1,6-2P	glycerone-P + D-glyceraldehyde-3-P
5.3.1.1	tpiA	glycerone-P	D-glyceraldehyde-3-P
1.2.1.12	gapA	D-glyceraldehyde-3-P + NAD^+ + P_i	1,3-2P-D-glycerate + NADH + H^+
2.7.2.3	pgk	ADP + 1,3-2P-D-glycerate	ATP + 3-P-D-glycerate
5.4.2.11	gpmA	3-P-D-glycerate	2-P-D-glycerate
5.4.2.12	gpmM	3-P-D-glycerate	2-P-D-glycerate
4.2.1.11	eno	2-P-D-glycerate	P-enolpyruvate + H_2O

Table 4.9: Reaction data for glycolysis pathway in *E. coli*

EC	**Enzyme's gene**	**Substrates**	**Products**
2.7.1.40	pykF	P-enolpyruvate + ADP + H^+	pyruvate + ATP
2.7.1.40	pykA	P-enolpyruvate + ADP + H^+	pyruvate + ATP
2.7.9.2	ppsA	P-enolpyruvate + AMP + P_i	pyruvate + ATP + H_2O

Table 4.10: Reaction for pyruvate metabolism in *E. coli*

EC	**Enzyme's gene**	**Substrates**	**Products**
1.1.1.28	ldhA	pyruvate + NADH + H^+	(R)-lactate + NAD^+
2.3.1.54	tdcE	pyruvate + CoA	acetyl-CoA + formate
2.3.1.54	pflB	pyruvate + CoA	acetyl-CoA + formate
2.3.1.8	pta	acetyl-CoA + P_i	CoA + acetyl-P
2.3.1.8	eutD	acetyl-CoA + P_i	CoA + acetyl-P
2.7.2.1	purT	ADP + acetyl-P	ATP + acetate
2.7.2.1	ackA	ADP + acetyl-P	ATP + acetate
1.2.1.10	mhpF	acetyl-CoA + NADH + H^+	acetaldehyde + CoA + NAD^+
1.1.1.1	adhP	acetaldehyde + NADH	ethanol + NAD^+
1.1.1.1	adhE	acetaldehyde + NADH	ethanol + NAD^+

The physical dimensions of the agents, as well as their motion, is defined by their shape. Tables 4.11, 4.12 and 4.13 present relevant size-related information for the participating molecules.

Precise measurements and descriptions of size of molecules are not systematized. Furthermore, molecule shape is not standard [74]. Depicting the molecule as a spherical agent is a compromise between detail and computational tractability [9]. In order to estimate a radii that would render an agent with a reasonable approximation of the hydrodynamic radius of the molecule, two correlations were used for enzymes and small organic chemical compounds, as detailed in the Introduction and Methods chapters. From the perspective of data curation, it was important that the correlated variables were consistently determined for the majority of molecular species, and it was simple to automatize.

The molecular weight of enzymes is an ideal candidate. Its determination is common to identify and characterize proteins, and relies on laboratory methods such as SDS-Page, mass chro-

matography or size-exclusion spectrophotometry [75]. In case of multimeric proteins (composed of more than one of the polypeptide that was individually translated from the genetic sequence, resulting in a quaternary structure), the experimental determination of molecular weight can be either of the multimer or of its subunits. In addition to the experimental methods of determination, a reasonable estimate of molecular weight can be inferred from the aminoacid sequence. This is featured, for example, in the EcoCyc [61] and UniProt [76] databases.

To compare estimates of molecular weight from amino acid sequence with experimental determinations, the reader can refer to the full curated data for inputs in the glycolysis pathway, in the supplementary material S7 (A.7).

The selected choice to determine the radius of the sphere-approximation of metabolites is the method described by Zhao et. al [46]. The equation requires the number of aromatic and non aromatic rings in the chemical structure, which in this work was determined by visualization of the chemical structure as available in databases of chemical compounds, such as PubChem [62] and ChEBI [77]. InChl and SMILES are structure identifiers, from which may be possible to deduce the structure and the number and type of intramolecular cyclizations, which can be aromatic or non-aromatic. Further manual curation might be necessary to select close chemical compounds, specially in matters of cyclized structures and conjugate acid and bases. For example, D-glucose exists both with an open-chain and cyclized structure. The cyclized structure is prevalent in aqueous and intracellular media [78]. Additionally, the exact formula of organic acids and compounds with phosphate is dependent on the pH of the solvent media and the compound's pKa. ChEBI Onthol ogy systematizes information on conjugated acid base pairs, but none of the surveyed databases collects data on pK_a. As a rule of thumb, organic acids were modelled after the conjugated bases, such as glycerate, pyruvate, formate and acetate. Cases like this raise problems in the accurate au tomatic collection of metabolite data, though the error should be negligible when compared to the spheric approximation of the metabolites and other computational simplifications. For example, the difference between open-chain and cyclized glucose is of 0.077 nm, which would translate in an additional 0.365 nm^3 in volume if the open-chain structure was used to model the metabolite.

Table 4.11: Size and motion data for enzymes of the glycolysis pathway in *E. coli*

EC	**Enzyme's gene**	**MW (kDa)**	**R_h (nm)**	**D ($m^2.s^{-1}$)** ***water, 37 °C***
2.7.1.2	glk	35 †	1.80	1.82×10^{-19}
5.3.1.9	pgi	125 †	6.44	5.10×10^{-20}
2.7.1.11	pfkB	71 †	3.66	8.98×10^{-20}
3.1.3.11	glpX	80 †	4.12	7.97×10^{-20}
3.1.3.11	fbp	150 †	7.73	4.25×10^{-20}
4.1.2.13	fbaB	305 §	15.70	2.09×10^{-20}
4.1.2.13	fbaA	78 §	4.03	8.14×10^{-20}
5.3.1.1	tpiA	54 §	2.78	1.18×10^{-19}
1.2.1.12	gapA	142 §	7.32	4.48×10^{-20}
2.7.2.3	pgk	41 §	2.12	1.55×10^{-19}
5.4.2.11	gpmA	49 †	2.52	1.3×10^{-19}
5.4.2.12	gpmM	61 †	3.14	1.04×10^{-19}
4.2.1.11	eno	91 §	4.70	6.98×10^{-20}
2.7.1.40	pykF	203 §	10.45	3.14×10^{-20}
2.7.1.40	pykA	190 †	9.79	3.35×10^{-20}

Table 4.12: Size data for enzymes of the pyruvate metabolism in *E. coli*

EC	**Enzyme's gene**	**MW (kDa)**	**R_h (nm)**	**D ($m^2.s^-1$)** ***water, 37 °C***
2.7.9.2	ppsA	168 ‡	4.33	7.59×10^{-20}
1.1.1.28	ldhA	37 §	1.88	1.74×10^{-19}
2.3.1.54	tdcE	85 †	4.38	7.50×10^{-20}
2.3.1.54	pflB	170 ‡	4.38	7.50×10^{-20}
2.3.1.8	pta	484 †	24.93	1.32×10^{-20}
2.3.1.8	eutD	36 §	1.86	1.77×10^{-19}
2.7.2.1	purT	84 ‡	4.32	7.61×10^{-20}
2.7.2.1	ackA	87 §	4.46	7.36×10^{-20}
1.2.1.10	mhpF	54 †	2.76	1.19×10^{-19}
1.1.1.1	adhP	142 ‡	7.29	4.50×10^{-20}
1.1.1.1	adhE	3845 ‡	198.02	1.66×10^{-21}

‡ from experimental values of the polypeptide monomer and multiplied by subunit number
§ from values inferred from aminoacid sequence for the polypeptide and multiplied by number of subunits
† from experimental values

Table 4.13: Size data for the metabolites involved in the glycolysis and pyruvate metabolism pathways in *E. coli*

Metabolites	**PubChem CID**	**R$_v$*dW* (nm)**	**D (m^2.s^{-1})**
H^+	1038	0.058	5.70×10^{-18}
P_i	1061	0.194	1.69×10^{-18}
H_2O	962	0.138	2.38×10^{-18}
CO_2	280	0.304	1.08×10^{-18}
ATP	5957	2.799	1.17×10^{-19}
ADP	6022	2.431	1.35×10^{-19}
AMP	6083	2.064	1.59×10^{-19}
NADH	439153	4.045	8.12×10^{-20}
NAD^+	925	4.024	8.16×10^{-20}
β-D-glucose	5793	1.215	2.70×10^{-19}
β-D-glucose-6-P	439427	1.583	2.07×10^{-19}
β-D-fructose-6-P	440641	1.583	2.07×10^{-19}
β-D-fructose-1,6-2P	10267	1.951	1.68×10^{-19}
glycerone-P	4643300	1.038	3.16×10^{-19}
D-glyceraldehyde-3-P	439168	1.038	3.16×10^{-19}
1,3-2P-D-glycerate	439191	1.475	2.23×10^{-19}
3-P-D-glycerate	439183	1.108	2.96×10^{-19}
2-P-D-glycerate	439278	1.108	2.96×10^{-19}
P-enolpyruvate	1005	1.017	3.23×10^{-19}
pyruvate	1060	0.649	5.06×10^{-19}
formate	283	0.314	1.04×10^{-18}
acetyl-CoA	444493	5.118	6.41×10^{-20}
CoA	6816	4.647	7.07×10^{-20}
acetyl-P	186	0.830	3.96×10^{-19}
acetate	175	0.382	8.60×10^{-19}
acetaldehyde	177	0.392	8.37×10^{-19}
ethanol	702	0.483	6.79×10^{-19}

4.2.1 Kinetic data

Kinetic data is highly heterogeneous, as the determination of Michaelis-Menten kinetic parameters has many variables in their experimental procedures.

The procedure used to collect the data in Table 4.14 and Table 4.15 is summarized in the Methods chapter (see Table 3.1). It was defined along the collection of kinetic data and the survey of the original sources in published literature.

Table 4.14: Kinetic data for enzymes of the glycolysis pathway in *E. coli*

EC	**Enzyme's gene**	**K_m (mM) forward**	**K_m (mM) reverse**	**k_{cat} (s^{-1}) forward**	**k_{cat} (s^{-1}) reverse**	**Source**
2.7.1.2	glk	(0.78, ATP), (3.76, β-D-glucose)	N/A	92.17	N/A	Meyer1997 [79]
5.3.1.9	pgi	1.018	0.078	684	N/A	Ogawa2007 [80]
2.7.1.11	pfkB	(0.018, β-fructose-6-P), (0.012, ATP)	(0.14, β-fructose-1,6-2P)	62	N/A	Rivas-Pardo2011 [81], Babul1978 [82]
3.1.3.11	glpX	(0.1 β-fructose-1,6-2P)	N/A	14.6	N/A	Kelley-Loughane2002 [83]
3.1.3.11	fbp	(0.0154, β-fructose-1,6-2P)	N/A	24	N/A	Hines2007 [84]
4.1.2.13	fbaB	(0.19, β-fructose-1,6-2P)	N/A	8.17	N/A	Platter1999 [85]
4.1.2.13	fbaA	(0.02 β-fructose-1,6-2P)	N/A	0.35	N/A	Platter1999 [85]
5.3.1.1	tpiA	1.03	N/A	5.4×10^4	N/A	Alvarez1998 [55]
1.2.1.12	gapA	(0.045, NAD), (0.89, 3-PGA), (0.53 P_i)	N/A	268	N/A	Eyschen1999 [86]
2.7.2.3	pgk	N/A	(0.24, MgATP)	N/A	328.94	Fifis1978 [87]
5.4.2.11	gpmA	(200, 3-PGA)	(190, 2-PGA)	330	220	Fraser1999 [88]
5.4.2.12	gpmM	(210, 3-PGA)	(97, 2-PGA)	22	10	Fraser1999 [88]
4.2.1.11	eno	0.1	N/A	197.8	N/A	Spring1971 [89]
2.7.1.40	pykF	(0.13, PEP)	N/A	N/A	N/A	Malcovati1969 [90]
2.7.1.40	pykA	(0.82, PEP)	N/A	N/A	3480	Somani1977 [91]
2.7.9.2	ppsA	N/A	(0.83, pyruvate), (0.028, ATP)	N/A	7.84	Berman1970 [92]

Table 4.15: Kinetic data for enzymes of the pyruvate metabolism in *E. coli*

EC	**Enzyme's gene**	**K_m (mM) forward**	**K_m (mM) reverse**	**k_{cat} (s^{-1}) forward**	**k_{cat} (s^{-1}) reverse**	**Source**
1.1.1.28	ldhA	(2.6, pyruvate)	N/A	410	N/A	Furukawa2014 [93]
2.3.1.54	tdcE	N/A	N/A	N/A	N/A	
2.3.1.54	pflB	(2.05, pyruvate), (0.0068, CoA)	(24.5, formate), (0.051, acetyl-CoA)	1100	280	Campos-Bermudez2010 [94]
2.3.1.8	pta	(2.1 P_i, 0.0449 acetyl-CoA)	(0.9, acetyl-P), (0.0672 CoA)	29.6	227.6	Campos-Bermudez2010 [94]
2.3.1.8	eutD	(0.0095, acetyl-CoA)	(0.3117, acetyl-P), (0.0327, CoA)	119	415.5	Bologna2010 [95]
2.7.2.1	purT	(0.16, acetyl-P), (0.5, ADP)	(7, acetate), (0.07, ATP)	3033	2333	Fox1986 [96]
2.7.2.1	ackA	N/A	N/A	N/A	N/A	
1.2.1.10	mhpF	(38, acetaldehyde), (90, CoA), (0.25, NAD^+)	N/A	15.70	N/A	Fischer2013 [97]
1.1.1.1	adhP	(0.03, acetaldehyde)	(0.7, ethanol)	163.33	67.5	Shaqfat1999 [98]
1.1.1.1	adhE	(5.4, acetaldehyde)	(240, ethanol)	256.34	640.85	Membrillo-Hernández2000 [99]

When choosing the right kinetic parameters, the most basic requirement is that the enzymatic assay was performed with the same substrate that is consumed in the physiological reaction that takes place in the cells. Despite enzyme's high specificity, the consumption of other substrates may be studied for industrial interests.

In this specific proof-of-concept curation, the second most important criteria to obey is that the enzyme was codified by a gene from *E. coli* K-12 MG1655. This strain of *E. coli* is one of the most widespread strains for laboratory studies, and as such it is likely that a greater number of enzymatic studies use enzymes of this specific strain, as opposed to other strains of E. coli.

However, due to the general similarity of the amino acid sequence in proteins of strains of the same organism, genes originating from other *E. coli* K-12 strains are a subsequent good option,

followed by any *E. coli* strain. In isolated occasions, literature covering the enzyme will refer that the enzyme structure of another species, such as *Saccharomyces cerevisae*, is very closely related to that of *E. coli* [89]. In those cases, and in the absence of more precise information, kinetic parameters of the mentioned organism can be used. Despite the fact that enzyme mutants are usually produced or studied under the premise that their kinetic parameters are superior to wild-type enzyme, most mutants differ by a small number of amino acids. Given this, mutated enzymes were considered better sources of information than genes belonging to other species when no information is known about similarity. As a last resort, enzymatic kinetics are taken from the phylogenetically closest organism for which information is available.

Origin organism is not be confused with host organism, also referred to in many experiments that use recombinant DNA methods to obtain significant amounts of the enzyme to be studied. A common host organism is the competent strain *E. coli* BL21(DE3).

The second criterion is obtaining paired information on K_m and k_{cat} on a single paper. It places importance in defining one given enzyme with data obtained under the same conditions and experimental procedure. Besides maintaining a more accurate relation between kinetic parameters, papers that satisfy this requisite will usually have this analysis at its core. This entails that they will have a more detailed description of all the experimental variables that potentially affect the values of kinetic parameters, and more careful experiment design and data analysis in what regards the kinetic parameter determination.

pH and temperature are experimental conditions that directly affect the catalytic efficiency of enzymes, and are referred in almost all determination experiments. The preferential values aim to mimic the *in vivo* conditions of an *E. coli* culture and to match the conditions of experimental quantification of intracellular metabolites, such as a temperature of 37 °C at which *E. coli* was grown [72, 100].

The pH of *E. coli* cytoplasm is reported to be between 7.2 and 7.8 [101]. When values in this range were not found, slightly superior pH was preferred to acidic conditions.

The methods of detection of the reaction rate can be widely variable, and thus hard to compare. It is considered here that direct assays of substrates or products are less affected by experimental error and variation than more complex methods, such as coupled-enzyme spectrophotometric assays. Direct methods can refer to the detection of either a substrate or product by spectrophotometry, mass spectrometry, radioactive labelling or any other analytical method that detects and quantifies the substance of interest from the assay mixture.

The next criterion evaluates the quality of data treatment, originating from the measurement of reaction velocity as a function of substrate concentration. Non-linear fitting to the original Michaelis-Menten equation yields the best results, though linear methods are more popular. Gao et al. performed both determinations, obtaining parameters from a non-linear regression and including Lineweaver-Burke plots for visualization [102].

Co-factor concentrations has a significant influence in catalytic efficiency. This is, however, hard to compare between different enzymes, as experimental design varies in determining the

influence of co-factors. Ideally, the kinetic parameters would be determined in an assay in which the co-factor concentration was close to the intracellular concentration [103].

Chapter 5

Final remarks

5.1 Conclusions

The current implementation of the model can successfully reproduce the kinetic behaviour of isomerases, using their Michaelis-Menten kinetic parameters as inputs. Specifically, two real isomerases, 2-hydroxymuconate and steroid δ-isomerase, were successfully simulated.

This was accomplished by establishing a relation between computational and real biological time. This ensured a match the time scales in diffusion and kinetics.

A simple procedure was developed to use k_{cat} as an input in the simulation, along with an iterative procedure to convert molecule's diffusion coefficient.

Different ways of implementing a computational equivalent to the K_m were tested. The probability of a collision between enzyme and substrate forming an enzyme-substrate complex (simK_m) and the reaction radius, as an artificial abstraction of the enzyme's affinity to substrate, were considered. Initial values of these computational variables can be estimated from the K_m, and the results tuned further in an iterative manner.

The model's required inputs for the glycolysis pathway were successfully collected for most of the parameters, including kinetic ones. For some enzymes, no kinetic studies were found. All of them have isoenzymes catalysing the same in vivo reaction. The remaining missing parameters are K_m for different substrates for a few number of multisubstrate reactions.

5.2 Future work

Future work will bridge the gap between the carried simulations and the curated data.

The simulated isomerases, 2-hydroxymuconate and steroid δ-isomerase, were used to develop and test a transferable strategy to use Michaelis-Menten kinetics as input. Other isomerases, with different k_{cat} and K_m, should be tested to further tune the strategy and validate it for a representative range of Michaelis-Menten parameters [104].

To achieve the metabolic scale, other type of enzymes have to be simulated besides isomerases. The existence of reaction with multiple substrates also requires a more advanced implementation

of the K_m, as each substrate is characterized by its affinity to the enzyme. This also raises the question of mechanism of reaction. Multiple substrates can bind to enzyme in a random order, compulsory order or in a double displacement mechanism (also known as ping-pong) [105]. There are some attempts to systematize information regarding enzymatic mechanisms (MaCIE [106] and EzCatDB [107] databases), but most of the glycolytic enzymes are not yet covered. The articles reviewed for kinetic parameter curation are a potential good source for this information.

Cooperative kinetics should also be implemented in the future. These are modelled after Hill equations, and cannot be explicited in the form of of Michaelis-Menten parameters.

Regulatory effects, such as activation and inhibition, are also present in some of the enzymes.

Finally, to move from a biorreactor, enzymatic assay scenario and simulate the *in vivo* behaviour of enzymes, the simulation volume has to contain agents that resemble the cytoplasmic concentrations and types of biomolecules.

Bibliography

[1] G. Pérez-Rodríguez, M. Pérez-Pérez, D. Glez-Peña, F. Fdez-Riverola, N. F. Azevedo, and A. Lourenço, "Agent-Based Spatiotemporal Simulation of Biomolecular Systems within the Open Source MASON Framework," *BioMed Research International*, vol. 2015, pp. 1–12, 2015.

[2] K. A. Johnson and R. S. Goody, "The original Michaelis constant: Translation of the 1913 Michaelis-Menten Paper," *Biochemistry*, vol. 50, no. 39, pp. 8264–8269, 2011.

[3] H. Kitano, "Systems biology: a brief overview." *Science (New York, N.Y.)*, vol. 295, no. 5560, pp. 1662–1664, 2002.

[4] ——, "Computational systems biology." *Nature*, vol. 420, no. 6912, pp. 206–210, 2002.

[5] H. V. Westerhoff and B. O. Palsson, "The evolution of molecular biology into systems biology." *Nature biotechnology*, vol. 22, no. 10, pp. 1249–1252, 2004.

[6] G. Stephanopoulos, H. Alper, and J. Moxley, "Exploiting biological complexity for strain improvement through systems biology." *Nature biotechnology*, vol. 22, no. 10, pp. 1261–1267, 2004.

[7] S. Y. Lee, D.-y. Lee, and T. Y. Kim, "Systems biotechnology for strain improvement," *Trends in Biotechnology*, vol. 23, no. 7, 2005.

[8] D. Machado, R. S. Costa, M. Rocha, E. C. Ferreira, B. Tidor, and I. Rocha, "Modeling formalisms in Systems Biology," p. 45, 2011.

[9] M. Feig and Y. Sugita, "Reaching new levels of realism in modeling biological macromolecules in cellular environments," *Journal of Molecular Graphics and Modelling*, vol. 45, pp. 144–156, Sep. 2013.

[10] J. Schöneberg, A. Ullrich, and F. Noé, "Simulation tools for particle-based reaction-diffusion dynamics in continuous space," *BMC Biophysics*, vol. 7, no. 1, p. 11, Dec. 2014.

[11] S. Riniker, J. R. Allison, and W. F. van Gunsteren, "On developing coarse-grained models for biomolecular simulation: a review," *Physical Chemistry Chemical Physics*, vol. 14, no. 36, p. 12423, 2012.

[12] J. Fisher and T. A. Henzinger, "Executable cell biology." *Nature biotechnology*, vol. 25, no. 11, pp. 1239–1249, 2007.

[13] C. A. Hunt, G. E. P. Ropella, S. Park, and J. Engelberg, "Dichotomies between computational and mathematical models." pp. 737–738; author reply 738–739, 2008.

[14] N. V. Torres and E. O. Voit, *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press, 2002.

[15] E. A. Mogilevskaya, K. V. Peskov, E. A. Metelkin, G. V. Lebedeva, T. Y. Plyusnina, I. I. Goryanin, and O. V. Demin, "Kinetic Modeling of *E. coli* Enzymes: Integration of in vitro Experimental Data," in *Systems Biology and Biotechonology for Escherichia coli*, 1st ed., S. Y. Lee, Ed. Springer, 2009, ch. 10, pp. 177–209.

[16] H. Link, D. Christodoulou, and U. Sauer, "Advancing metabolic models with kinetic information," pp. 8–14, 2014.

[17] Z. A. King, C. J. Lloyd, A. M. Feist, and B. O. Palsson, "Next-generation genome-scale models for metabolic engineering," *Current Opinion in Biotechnology*, 2015.

[18] M. R. Long, W. K. Ong, and J. L. Reed, "Computational methods in metabolic engineering for strain design," *Current Opinion in Biotechnology*, vol. 34, pp. 135–141, Aug. 2015.

[19] M. Holcombe, S. Adra, M. Bicak, S. Chin, S. Coakley, A. I. Graham, J. Green, C. Greenough, D. Jackson, M. Kiran, S. MacNeil, A. Maleki-Dizaji, P. McMinn, M. Pogson, R. Poole, E. Qwarnstrom, F. Ratnieks, M. D. Rolfe, R. Smallwood, T. Sun, and D. Worth, "Modelling complex biological systems using an agent-based approach," *Integrative Biology*, vol. 4, no. 1, p. 53, 2012.

[20] M. Niazi and A. Hussain, "Agent-based computing from multi-agent systems to agent-based models: A visual survey," *Scientometrics*, vol. 89, no. 2, pp. 479–499, 2011.

[21] H. Kaul and Y. Ventikos, "Investigating biocomplexity through the agent-based paradigm." *Briefings in bioinformatics*, vol. 44, no. 1, 2013.

[22] E. Meijering, O. Dzyubachyk, and I. Smal, *Methods for cell and particle tracking*, 1st ed. Elsevier Inc., 2012, vol. 504.

[23] N. Chenouard, I. Smal, F. de Chaumont, M. Maška, I. F. Sbalzarini, Y. Gong, J. Cardinale, C. Carthel, S. Coraluppi, M. Winter, A. R. Cohen, W. J. Godinez, K. Rohr, Y. Kalaidzidis, L. Liang, J. Duncan, H. Shen, Y. Xu, K. E. G. Magnusson, J. Jaldén, H. M. Blau, P. Paul-Gilloteaux, P. Roudot, C. Kervrann, F. Waharte, J.-Y. Tinevez, S. L. Shorte, J. Willemse, K. Celler, G. P. van Wezel, H.-W. Dan, Y.-S. Tsai, C. Ortiz de Solórzano, J.-C. Olivo-Marin, and E. Meijering, "Objective comparison of particle tracking methods." *Nature Methods*, vol. 11, no. 3, pp. 281–289, 2014.

[24] A. E. Knight, "Single Enzyme Studies: A Historical Perspective," in *Single Molecule Enzymology*, G. Mashanov and C. Batters, Eds. Humana Press, 2011, ch. 1.

[25] N. Ruthardt, D. C. Lamb, and C. Bräuchle, "Single-particle tracking as a quantitative microscopy-based approach to unravel cell entry mechanisms of viruses and pharmaceutical nanoparticles." *Molecular therapy : the journal of the American Society of Gene Therapy*, vol. 19, no. 7, pp. 1199–1211, 2011.

[26] Y. Lill, W. a. Kaserer, S. M. Newton, M. Lill, P. E. Klebba, and K. Ritchie, "Single-molecule study of molecular mobility in the cytoplasm of *Escherichia coli*," *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 86, pp. 1–7, 2012.

[27] S. C. Pratt, D. J. T. Sumpter, E. B. Mallon, and N. R. Franks, "An agent-based model of collective nest choice by the ant Temnothorax albipennis," *Animal Behaviour*, vol. 70, no. 5, pp. 1023–1036, 2005.

[28] E. J. H. Robinson, F. L. W. Ratnieks, and M. Holcombe, "An agent-based model to investigate the roles of attractive and repellent pheromones in ant decision making during foraging," *Journal of Theoretical Biology*, vol. 255, no. 2, pp. 250–258, 2008.

[29] D. C. Walker, J. Southgate, G. Hill, M. Holcombe, D. R. Hose, S. M. Wood, S. Mac Neil, and R. H. Smallwood, "The epitheliome: Agent-based modelling of the social behaviour of cells," in *BioSystems*, vol. 76, no. 1-3, 2004, pp. 89–100.

[30] R. Mukhopadhyay, S. V. Costes, A. V. Bazarov, W. C. Hines, M. Barcellos-Hoff, and P. Yaswen, "Promotion of variant human mammary epithelial cell outgrowth by ionizing radiation: an agent-based model supported by in vitro studies," *Breast Cancer Research*, vol. 12, no. 1, p. R11, 2010.

[31] Q. Mi, B. Rivière, G. Clermont, D. L. Steed, and Y. Vodovotz, "Agent-based model of inflammation and wound healing: Insights into diabetic foot ulcer pathology and the role of transforming growth factor-β1," *Wound Repair and Regeneration*, vol. 15, no. 5, pp. 671–682, 2007.

[32] A. L. Muci, M. A. Jorquera, A. I. Ávila, Z. Rengel, D. E. Crowley, M. De, and L. Mora, "A combination of cellular automata and agent-based models for simulating the root surface colonization by bacteria," *Ecological Modelling*, vol. 247, pp. 1–10, 2012.

[33] F. L. Hellweger and V. Bucci, "A bunch of tiny individuals-Individual-based modeling for microbes," *Ecological Modelling*, vol. 220, no. 1, pp. 8–22, 2009.

[34] K. Bettenbrock, H. Bai, M. Ederer, J. Green, K. J. Hellingwerf, M. Holcombe, S. Kunz, M. D. Rolfe, G. Sanguinetti, O. Sawodny, P. Sharma, S. Steinsiek, and R. K. Poole, *Towards a systems level understanding of the oxygen response of Escherichia coli.*, 1st ed. Elsevier Ltd., Jan. 2014, vol. 64.

[35] S. M. Stiegelmeyer and M. C. Giddings, "Agent-based modeling of competence phenotype switching in *Bacillus subtilis*." *Theoretical biology & medical modelling*, vol. 10, no. 1, p. 23, Jan. 2013.

[36] M. T. Klann, A. Lapin, and M. Reuss, "Agent-based simulation of reactions in the crowded and structured intracellular environment: Influence of mobility and location of the reactants." *BMC systems biology*, vol. 5, p. 71, 2011.

[37] T. Emonet, C. M. Macal, M. J. North, C. E. Wickersham, and P. Cluzel, "AgentCell: A digital single-cell assay for bacterial chemotaxis," *Bioinformatics*, vol. 21, no. 11, pp. 2714–2721, 2005.

[38] R. A. Kerr, T. M. Bartol, B. Kaminsky, M. Dittrich, J.-C. J. Chang, S. B. Baden, T. J. Sejnowski, and J. R. Stiles, "Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces," pp. 3126–3149, 2008.

[39] S. S. Andrews, N. J. Addy, R. Brent, and A. P. Arkin, "Detailed Simulations of Cell Biology with Smoldyn 2.1," *PLoS computational biology*, vol. 6, no. 3, 2010.

[40] M. J. Byrne, M. N. Waxham, and Y. Kubota, "Cellular Dynamic Simulator: An Event Driven Molecular Simulation Environment for Cellular Physiology," *Neuroinformatics*, vol. 8, no. 2, pp. 63–82, 2010.

[41] J. Schöneberg and F. Noé, "ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments." *PloS one*, vol. 8, no. 9, p. e74261, 2013.

[42] E. Bonabeau, "Agent-based modeling: Methods and techniques for simulating human systems," *Proceedings of the National Academy of Sciences*, vol. 99, no. Supplement 3, pp. 7280–7287, 2002.

[43] A. Illanes, Ed., *Enzyme Biocatalysis - Principles and Applications*. Springer, 2008.

[44] H. B. Brooks, S. Geeganage, S. D. Kahl, C. Montrose, S. Sittampalam, M. C. Smith, and J. R. Weidner, "Basics of Enzymatic Assays for HTS," in *Assay Guidance Manual*, S. Sittampalam, N. P. Coussens, and H. Nelson, Eds. Eli Lilly & Company and the National Center for Advancing Translational Sciences, 2012.

[45] T. Kalwarczyk, M. Tabaka, and R. Holyst, "Biologistics-Diffusion coefficients for complete proteome of *Escherichia coli*," *Bioinformatics*, vol. 28, no. 22, pp. 2971–2978, 2012.

[46] Y. H. Zhao, M. H. Abraham, and A. M. Zissimos, "Fast calculation of van der Waals volume as a sum of atomic and bond contributions and its application to drug compounds." *The Journal of organic chemistry*, vol. 68, no. 19, pp. 7368–73, Sep. 2003.

[47] A. Bondi, "van der Waals Volumes and Radii," *The Journal of Physical Chemistry*, vol. 68, no. 3, pp. 441–451, 1964.

[48] J. T. Mika and B. Poolman, "Macromolecule diffusion and confinement in prokaryotic cells," *Current Opinion in Biotechnology*, vol. 22, no. 1, pp. 117–126, 2011.

[49] Y. M. Ayala and E. Di Cera, "A simple method for the determination of individual rate constants for substrate hydrolysis by serine proteases." *Protein science : a publication of the Protein Society*, vol. 9, no. 8, pp. 1589–1593, 2000.

[50] R. A. Alberty and G. G. Hammes, "Application of the theory of diffusion-controlled reactions to enzyme kinetics," *The Journal of Physical Chemistry*, vol. 62, pp. 154–159, 1958.

[51] C. P. Whitman, B. a. Aird, W. R. Gillespie, and N. J. Stolowich, "Chemical and enzymic ketonization of 2-hydroxymuconate, a conjugated enol," *Journal of the American Chemical Society*, vol. 113, no. 8, pp. 3154–3162, 1991.

[52] G. Choi, N. C. Ha, M. S. Kim, B. H. Hong, B. H. Oh, and K. Y. Choi, "Pseudoreversion of the catalytic activity of Y14F by the additional substitution(s) of tyrosine with phenylalanine in the hydrogen bond network of Δ5-3-ketosteroid isomerase from *Pseudomonas putida* biotype B," *Biochemistry*, vol. 40, no. 23, pp. 6828–6835, 2001.

[53] S. Luke, "MASON: A Multiagent Simulation Environment," *SIMULATION*, vol. 81, no. 7, pp. 517–527, Jul. 2005.

[54] J. L. Finney, "The water molecule and its interactions: the interaction between theory, modelling, and experiment," *Journal of Molecular Liquids*, vol. 90, no. 1-3, pp. 303–312, 2001.

[55] M. Alvarez, J. P. Zeelen, V. Mainfroid, F. Rentier-Delrue, J. A. Martial, L. Wyns, R. K. Wierenga, and D. Maes, "Triose-phosphate Isomerase (TIM) of the Psychrophilic Bacterium *Vibrio marinus*: Kinetic and Structural Properties," *Journal of Biological Chemistry*, vol. 273, no. 4, pp. 2199–2206, Jan. 1998.

[56] J. L. Donahue, J. L. Bownas, W. G. Niehaus, and T. J. Larson, "Purification and characterization of glpX-encoded fructose 1,6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*," *Journal of Bacteriology*, vol. 182, no. 19, pp. 5624–5627, 2000.

[57] R. R. Poyner, L. T. Laughlin, G. a. Sowa, and G. H. Reed, "Toward identification of acid/base catalysts in the active site of enolase: Comparison of the properties of K345A, E168Q, and E211Q variants," *Biochemistry*, vol. 35, no. 1991, pp. 1692–1699, 1996.

[58] J. C. Pezullo and A. Herráez, "Michaelis-Menten Nonlinear Regression."

[59] J. Dowd and D. Riggs, "Constants of Estimates from Various of Michaelis-Menten Kinetic Linear Transformations," *Journal of biological chemistry*, vol. 240, no. 2, 1965.

[60] M. Kanehisa, S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe, "Data, information, knowledge and principle: Back to metabolism in KEGG," *Nucleic Acids Research*, vol. 42, no. D1, 2014.

[61] I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martínez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muñiz Rascado, Q. Ong, S. Paley, I. Schröder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen, and P. D. Karp, "EcoCyc: Fusing model organism databases with systems biology," *Nucleic Acids Research*, vol. 41, no. D1, 2013.

[62] E. E. Bolton, Y. Wang, P. A. Thiessen, and S. H. Bryant, "PubChem: Integrated Platform of Small Molecules and Biological Activities," *Annual Reports in Computational Chemistry*, vol. 4, pp. 217–241, 2008.

[63] A. Chang, I. Schomburg, S. Placzek, L. Jeske, M. Ulbrich, M. Xiao, C. W. Sensen, and D. Schomburg, "BRENDA in 2015: exciting developments in its 25th year of existence," *Nucleic Acids Research*, vol. 43, no. D1, pp. D439–D446, 2014.

[64] The UniProt Consortium, "UniProt: a hub for protein information," *Nucleic Acids Research*, vol. 43, no. D1, pp. D204–D212, 2014.

[65] K. R. Albe, M. H. Butler, and B. E. Wright, "Cellular concentrations of enzymes and their substrates." *Journal of theoretical biology*, vol. 143, no. 2, pp. 163–195, 1990.

[66] B. Teusink, J. Passarge, C. a. Reijenga, E. Esgalhado, C. C. Van Der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. Van Dam, H. V. Westerhoff, and J. L. Snoep, "Can yeast glycolysis be understood terms of vitro kinetics of the constituent enzymes? Testing biochemistry," *European Journal of Biochemistry*, vol. 267, no. February, pp. 5313–5329, 2000.

[67] K. van Eunen, J. a. L. Kiewiet, H. V. Westerhoff, and B. M. Bakker, "Testing biochemistry revisited: How in vivo metabolism can be understood from in vitro enzyme kinetics," *PLoS Computational Biology*, vol. 8, no. 4, 2012.

[68] K. Peneva, G. Mihov, F. Nolde, S. Rocha, J. I. Hotta, K. Braeckmans, J. Hofkens, H. Uji-i, A. Herrmann, and K. Müllen, "Water-soluble monofunctional perylene and terrylene dyes: Powerful labels for single-enzyme tracking," *Angewandte Chemie - International Edition*, vol. 47, pp. 3372–3375, 2008.

[69] B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie, "Ever-fluctuating single enzyme molecules: Michaelis-Menten equation revisited," *Nature Chemical Biology*, vol. 2, no. 2, pp. 87–94, Feb. 2006.

[70] H. Qian and E. L. Elson, "Single-molecule enzymology: Stochastic Michaelis-Menten kinetics," *Biophysical Chemistry*, vol. 101-102, pp. 565–576, 2002.

[71] M. Yi and Q. Liu, "Michaelis-Menten mechanism for single-enzyme and multi-enzyme system under stochastic noise and spatial diffusion," *Physica A: Statistical Mechanics and its Applications*, vol. 389, no. 18, pp. 3791–3803, 2010.

[72] B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz, "Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*." *Nature chemical biology*, vol. 5, no. 8, pp. 593–9, Aug. 2009.

[73] R. J. Ellis, "Macromolecular crowding: An important but neglected aspect of the intracellular environment," *Current Opinion in Structural Biology*, vol. 11, no. 00, pp. 114–119, 2001.

[74] H. P. Erickson, "Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy," *Biological Procedures Online*, vol. 11, no. 1, pp. 32–51, 2009.

[75] A. S. Bommarius and B. R. Riebel, *Biocatalysis*. Wiley-VCH, 2004.

[76] UniProt Consortium, "Activities at the Universal Protein Resource (UniProt)." *Nucleic acids research*, vol. 42, no. Database issue, pp. D191–8, Jan. 2014.

[77] J. Hastings, P. De Matos, A. Dekker, M. Ennis, B. Harsha, N. Kale, V. Muthukrishnan, G. Owen, S. Turner, M. Williams, and C. Steinbeck, "The ChEBI reference database and ontology for biologically relevant chemistry: Enhancements for 2013," *Nucleic Acids Research*, vol. 41, no. D1, 2013.

[78] D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry*, 4th ed. W. H. Freeman, 2005.

[79] D. Meyer, C. Schneider-Fresenius, R. Horlacher, R. Peist, and W. Boos, "Molecular characterization of glucokinase from *Escherichia coli* K-12." *Journal of bacteriology*, vol. 179, no. 4, pp. 1298–1306, Mar. 1997.

[80] T. Ogawa, H. Mori, M. Tomita, and M. Yoshino, "Inhibitory effect of phosphoenolpyruvate on glycolytic enzymes in *Escherichia coli*," *Research in Microbiology*, vol. 158, no. 2, pp. 159–163, 2007.

[81] J. A. Rivas-Pardo, A. Caniuguir, C. a. M. Wilson, J. Babul, and V. Guixé, "Divalent metal cation requirements of phosphofructokinase-2 from *E. coli*. Evidence for a high affinity binding site for Mn2+," *Archives of Biochemistry and Biophysics*, vol. 505, no. 1, pp. 60–66, 2011.

[82] J. Babul, "Phosphofructokinases from *Escherichia coli*: Purification and Characterization of the Nonallosteric Isozyme," *Journal of Biological Chemistry*, vol. 253, no. 12, pp. 4350–4355, 1978.

[83] N. Kelley-Loughnane, S. a. Biolsi, K. M. Gibson, G. Lu, M. J. Hehir, P. Phelan, and E. R. Kantrowitz, "Purification, kinetic studies, and homology model of *Escherichia coli* fructose-1,6-bisphosphatase," *Biochimica et Biophysica Acta - Protein Structure and Molecular Enzymology*, vol. 1594, no. 1, pp. 6–16, 2002.

[84] J. K. Hines, H. J. Fromm, and R. B. Honzatko, "Structures of activated fructose-1,6-bisphosphatase from *Escherichia coli*: Coordinate regulation of bacterial metabolism and the conservation of the R-state," *Journal of Biological Chemistry*, vol. 282, no. 16, pp. 11 696–11 704, 2007.

[85] A. R. Plater, S. M. Zgiby, G. J. Thomson, S. Qamar, C. W. Wharton, and A. Berry, "Conserved residues in the mechanism of the *E. coli* Class II FBP-aldolase." *Journal of molecular biology*, vol. 285, no. 2, pp. 843–855, 1999.

[86] J. Eyschen, B. Vitoux, M. Marraud, M. T. Cung, and G. Branlant, "Engineered glycolytic glyceraldehyde-3-phosphate dehydrogenase binds the anti conformation of NAD+ nicotinamide but does not experience A-specific hydride transfer." *Archives of biochemistry and biophysics*, vol. 364, no. 2, pp. 219–227, 1999.

[87] T. Fifis and R. K. Scopes, "Purification of 3-phosphoglycerate kinase from diverse sources by affinity elution chromatography." *The Biochemical journal*, vol. 175, no. 1, pp. 311–319, 1978.

[88] H. I. Fraser, M. Kvaratskhelia, and M. F. White, "The two analogous phosphoglycerate mutases of *Escherichia coli*," *FEBS letters*, vol. 455, pp. 344–348, 1999.

[89] T. G. Spring and F. Wold, "The purification and characterization of *Escherichia coli* enolase." *The Journal of biological chemistry*, vol. 246, no. 22, pp. 6797–6802, 1971.

[90] M. Malcovati and H. L. Kornberg, "Two types of pyruvate kinase in *Escherichia coli* K12." *Biochimica et biophysica acta*, vol. 178, no. 2, pp. 420–423, 1969.

[91] B. L. Somani, G. Valentini, and M. Malcovati, "Purification and molecular properties of the AMP-activated pyruvate kinase from *Escherichia coli*." *Biochimica et biophysica acta*, vol. 482, no. 1, pp. 52–63, 1977.

[92] K. M. Berman and M. Cohn, "Phosphoenolpyruvate synthetase of *Escherichia coli*. Purification, some properties, and the role of divalent metal ions." *The Journal of biological chemistry*, vol. 245, no. 20, pp. 5309–5318, 1970.

[93] N. Furukawa, A. Miyanaga, M. Togawa, M. Nakajima, and H. Taguchi, "Diverse allosteric and catalytic functions of tetrameric d-lactate dehydrogenases from three Gram-negative bacteria," *AMB Express*, vol. 4, no. 1, p. 76, 2014.

[94] V. A. Campos-Bermudez, F. P. Bologna, C. S. Andreo, and M. F. Drincovich, "Functional dissection of *Escherichia coli* phosphotransacetylase structural domains and analysis of key compounds involved in activity regulation." *The FEBS journal*, vol. 277, no. 8, pp. 1957–66, May 2010.

[95] F. P. Bologna, V. a. Campos-Bermudez, D. D. Saavedra, C. S. Andreo, and M. F. Drincovich, "Characterization of *Escherichia coli* EutD: A phosphotransacetylase of the ethanolamine operon," *Journal of Microbiology*, vol. 48, no. 5, pp. 629–636, 2010.

[96] D. K. Fox and S. Roseman, "Isolation and characterization of homogeneous acetate kinase from Salmonella typhimurium and *Escherichia coli*," *Journal of Biological Chemistry*, vol. 261, no. 29, pp. 13 487–13 497, 1986.

[97] B. Fischer, S. Boutserin, H. Mazon, S. Collin, G. Branlant, A. Gruez, and F. Talfournier, "Catalytic properties of a bacterial acylating acetaldehyde dehydrogenase: Evidence for several active oligomeric states and coenzyme A activation upon binding," *Chemico-Biological Interactions*, vol. 202, no. 1-3, pp. 70–77, 2013.

[98] J. Shafqat, J. O. Höög, L. Hjelmqvist, U. C. T. Oppermann, C. Ibáñez, and H. Jörnvall, "An ethanol-inducible MDR ethanol dehydrogenase/acetaldehyde reductase in *Escherichia coli*: Structural and enzymatic relationships to the eukaryotic protein forms," *European Journal of Biochemistry*, vol. 263, no. 2, pp. 305–311, 1999.

[99] J. Membrillo-Hernández, P. Echave, E. Cabiscol, J. Tamarit, J. Ros, and E. C. C. Lin, "Evolution of the adhE gene product of *Escherichia coli* from a functional reductase to a dehydrogenase: Genetic and biochemical studies of the mutant proteins," *Journal of Biological Chemistry*, vol. 275, no. 43, pp. 33 869–33 875, 2000.

[100] B. D. Bennett, J. Yuan, E. H. Kimball, and J. D. Rabinowitz, "Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach." *Nature protocols*, vol. 3, no. 8, pp. 1299–1311, 2008.

[101] J. C. Wilks and J. L. Slonczewski, "pH of the cytoplasm and periplasm of *Escherichia coli*: Rapid measurement by green fluorescent protein fluorimetry," *Journal of Bacteriology*, vol. 189, no. 15, pp. 5601–5607, 2007.

[102] H. Gao, Y. Chen, and J. a. Leary, "Kinetic measurements of phosphoglucose isomerase and phosphomannose isomerase by direct analysis of phosphorylated aldose-ketose isomers using tandem mass spectrometry," *International Journal of Mass Spectrometry*, vol. 240, no. 3 SPEC. ISS., pp. 291–299, 2005.

[103] S. Sundararaj, A. Guo, B. Habibi-Nazhad, M. Rouani, P. Stothard, M. Ellison, and D. S. Wishart, "The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate in silico modeling of *Escherichia coli*." *Nucleic acids research*, vol. 32, no. Database issue, pp. D293–D295, Jan. 2004.

[104] A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo, "The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters," *Biochemistry*, vol. 50, no. 21, pp. 4402–4410, 2011.

[105] H. Sauro, "Basic Enzyme Kinetics," in *Enzyme Kinetics for Systems Biology*. Ambrosius Publishing, 2012, vol. 2, ch. 3, pp. 55–84.

[106] G. L. Holliday, C. Andreini, J. D. Fischer, S. A. Rahman, D. E. Almonacid, S. T. Williams, and W. R. Pearson, "MACiE: Exploring the diversity of biochemical reactions," *Nucleic Acids Research*, vol. 40, no. D1, 2012.

[107] N. Nagano, N. Nakayama, K. Ikeda, M. Fukuie, K. Yokota, T. Doi, T. Kato, and K. Tomii, "EzCatDB: the enzyme reaction database, 2015 update," *Nucleic Acids Research*, vol. 43, no. D1, pp. D453–D458, 2014.

Appendix A

Index of supplementary material

The supplementary material is available in the adjoining CD. Main spreadsheets contain an initial sheet of contents and notes relevant to that content. All spreadsheets contain a "Simulation specifications" page, detailing the size of agents (enzymes, substrates, products and "dead" agents), rules (simk$_{cat}$, simK$_m$, reaction radius), and shape and size of the environment.

A.1 S1 Time relation and kcat calibration

1. S1 Time relation and kcat calibration.xlsx
   (a) Contents
   (b) Simulation specifications
   (c) Results: Product formation for several simkcat values (from 1 to 30 ts). Plot and calculation of simulation values.
   (d) ES Et %: Table of relative occupancy of enzymes for different simkcat values at different points in the simulation (time step). Graph of relative occupancy of enzymes for different simkcat values and points in the simulation (time steps).
   (e) simkcat #: Tracked data in the simulation: substrate, free enzyme, ES, total product. Averages of ES along the simulation (for calculation of relative occupancy of enzymes).

A.2 S2 Diffusion calibration for xylH

1. S2 Diffusion calibration for xylH - 750S-20000S.xlsx
   (a) Contents
   (b) Simulation specifications
   (c) Diffusion calibration results: velocity of agents (enzyme and substrate) for the different scenarios with different number of substrate, and the deviation of the resulting diffusion coefficient and the theoretical coefficient.

2. S2.*x*.*y* Diffusion calibration for xylH - *x*S (iteration *y*).xlsx

   (a) Contents

   (b) Simulation specifications

   (c) Simulation data: Data tracked during simulations, for each of the agents during 1000 time steps: initial and final positions (in X,Y and Z cartesian coordinates), and whether they hit the reaction volume boundaries at any time during simulation (HasCrash variable, true or false). Calculation of the displacement (difference between initial and final position) of each agent in X, Y and Z cartesian coordinates.

   (d) Displacement calculus: Calculation of the displacement (difference between initial and final position) of each agent in X, Y and Z cartesian coordinates. Calculation of average displacements for enzymes and substrates: all agents, agents that did not hit the reactor boundary (HasCrash==false), agents that his the reactor boundary (HasCrash==true).

   (e) Diffusion comparison: Calculation of diffusion coefficient of the simulation agents, in simulation and real units, for: all agents, agents that did not hit the reactor boundary (HasCrash==false), agents that his the reactor boundary (HasCrash==true). Calculation of the diffusion coefficient of the equivalent biomolecules (real enzyme and real substrate) in a reaction medium (water at 37 ºC) with Stokes-Einstein formula. Calculation of deviation between the simulation diffusion coefficient and the theoretical diffusion coefficient.

   (f) Next V_{sim} (iteration): Values of Vsim for enzyme and substrate agents (current and for a new iteration of the simulation), in real and simulation units

A.3 S3 Diffusion calibration for ksi

1. S3.*y* Diffusion calibration for ksi (iteration *y*)

   (a) Contents

   (b) Simulation specifications

   (c) Simulation data: Data tracked during simulations, for each of the agents during 1000 time steps: initial and final positions (in X,Y and Z cartesian coordinates), and whether they hit the reaction volume boundaries at any time during simulation (HasCrash variable, true or false). Calculation of the displacement (difference between initial and final position) of each agent in X, Y and Z cartesian coordinates.

   (d) Displacement calculus: Calculation of the displacement (difference between initial and final position) of each agent in X, Y and Z cartesian coordinates. Calculation of average displacements for enzymes and substrates: all agents, agents that did not hit the reactor boundary (HasCrash==false), agents that his the reactor boundary (HasCrash==true).

(e) Diffusion comparison: Calculation of diffusion coefficient of the simulation agents, in simulation and real units, for: all agents, agents that did not hit the reactor boundary (HasCrash==false), agents that his the reactor boundary (HasCrash==true). Calculation of the diffusion coefficient of the equivalent biomolecules (real enzyme and real substrate) in a reaction medium (water at 37 ºC) with Stokes-Einstein formula. Calculation of deviation between the simulation diffusion coefficient and the theoretical diffusion coefficient.

(f) Next V_{sim} (iteration): Values of Vsim for enzyme and substrate agents (current and for a new iteration of the simulation), in real and simulation units

A.4 S4 simKm implementation - probability of E+S to ES

1. S4 simKm implementation - probability of E+S to ES

   (a) Contents

   (b) Simulation specifications

   (c) Individual Michaelis-Menten: velocities of product agent formation (dP/dt) in each scenario with different substrate, for all the tested simKm (100%, 75%, 50%, 25%, 10%, 1%, 0.1%). Results of least squares non-linear regression and linear transformations (parameter values and standard error associated with regression) of Lineweaver-Burke, Eadie-Hofstee and Hanes-Woolf. Michaelis-Menten and Lineweaver-Burke plots.

   (d) Aggregated Michaelis-Menten: table of resulting values for Michaelis-Menten parameters (least squares non-linear regression). Michaelis-Menten and Lineweaver-Burke plots for all the values. *Auxiliary values to plot the lines of non-linear fit.*

2. S4.*x.y* simKm implementation - probability of E+S to ES - #% #S

   (a) Simulation specifications

   (b) Simulation data: Tracked data (number of substrate, enzyme, ES and product agents per time step). Calculation of dP/dt linear regression parameters (slope and regression coefficient) in the linear portion of dP/dt of the simulation with 750 substrate agents (0 to 100000 ts). Plot of dP/dt.

A.5 S5 Velocity *vs.* substrate for xylH

1. S5.1 Velocity vs. substrate for xylH - reaction radius 4x - Michaelis-Menten and noise.xlsx

   (a) Contents

   (b) Simulation specifications

(c) MM plots: Compilation of average velocity (dP/dt) and standard deviation, for individual simulations with different numbers of substrate agent. Conversion to real units. Plot of Velocity vs. Substrate in real units. Reported Michaelis-Menten kinetic parameter values (Whitman1999) and fit to the Michaelis-Menten equation through 1) non-linear least squares regression, 2) Lineweaver-Burke linear transformation, 3) Hanes-Woolf linear transformation and 4) Eadie-Hofstee linear transformation.

(d) Noise - Product formation: Compilation of number of product agents number at each time step, for individual simulations of 25, 250 and 10000 initial substrate agents (triplicates). Graph of evolution of product agents number with time steps for each of the individual simulations.

(e) Noise - Deviation: Relative standard deviation (standard deviation over the average of substrates) of product number for each of the individual simulations (25 substrate agents to 10000 substrate agents) at each time step, during the linear velocity part of simulation for the simulation with 25 substrate agents (0 to 6000 time steps). Calculation of the average relative standard deviation for simulation scenarios with the different substrate agents number.

2. S5.1.*y* xylH simulation RR4x #S.xlsx

(a) Simulation specifications.

(b) Simulation data: Tracked data (substrate, enzyme, ES, product agents per time step) for the simulation replicates. Calculation of the average product agents formed, standard deviation and relative standard deviation (standard deviation divided by average).

(c) Results: Linear regression dP/dt parameters (slope and regression coefficient) for each of the replicates, calculation of the average dP/dt and standard deviation. Plot of dP/dt.

3. S5.2 Velocity vs. substrate for xylH - reaction radius 2x - Michaelis-Menten and noise.xlsx

(a) Contents

(b) Simulation specifications

(c) MM plots: Compilation of average velocity (dP/dt) and standard deviation, for individual simulations with different numbers of substrate agent. Conversion to real units. Plot of Velocity vs. Substrate in real units. Reported Michaelis-Menten kinetic parameter values (Whitman1999) and fit to the Michaelis-Menten equation through 1) non-linear least squares regression, 2) Lineweaver-Burke linear transformation, 3) Hanes-Woolf linear transformation and 4) Eadie-Hofstee linear transformation.

(d) Noise - Product formation: Compilation of number of product agents number at each time step, for individual simulations of 25, 250 and 10000 initial substrate agents (triplicates). Graph of evolution of product agents number with time steps for each of the individual simulations.

(e) Noise - Deviation: Relative standard deviation (standard deviation over the average of substrates) of product number for each of the individual simulations (25 substrate agents to 10000 substrate agents) at each time step, during the linear velocity part of simulation for the simulation with 25 substrate agents (0 to 6000 time steps). Calculation of the average relative standard deviation for simulation scenarios with the different substrate agents number.

4. S5.2.*y* xylH simulation RR2x #S

   (a) Simulation specifications.

   (b) Simulation data: Tracked data (substrate, enzyme, ES, product agents per time step) for the simulation replicates. Calculation of the average product agents formed, standard deviation and relative standard deviation (standard deviation divided by average).

   (c) Results: Linear regression dP/dt parameters (slope and regression coefficient) for each of the replicates, calculation of the average dP/dt and standard deviation. Plot of dP/dt.

5. S5.3 Velocity vs. substrate for xylH - reaction radius 2x - Michaelis-Menten

   - Contents
   - Simulation specifications
   - MM plots: Compilation of average velocity (dP/dt) and standard deviation, for individual simulations with different numbers of substrate agent. Conversion to real units. Plot of Velocity vs. Substrate in real units. Reported Michaelis-Menten kinetic parameter values (Whitman1999) and fit to the Michaelis-Menten equation through 1) non-linear least squares regression, 2) Lineweaver-Burke linear transformation, 3) Hanes-Woolf linear transformation and 4) Eadie-Hofstee linear transformation.
   - Aggregated MM: Kinetic parameter values for reaction radius of 4x and 2x. Plot containing the simulation data of S5.1 MM Plots (reaction radius of $4\times R_h$), S5.2 MM Plots (reaction radius of $2\times R_h$) and S5.3 MM Plots (reaction radius $10\times R_h$).
   - Noise - Product formation: Compilation of number of product agents number at each time step, for individual simulations of 25, 250 and 10000 initial substrate agents (triplicates). Graph of evolution of product agents number with time steps for each of the individual simulations.
   - Noise - Deviation: Relative standard deviation (standard deviation over the average of substrates) of product number for each of the individual simulations (25 substrate agents to 10000 substrate agents) at each time step, during the linear velocity part of simulation for the simulation with 25 substrate agents (0 to 6000 time steps). Calculation of the average relative standard deviation for simulation scenarios with the different substrate agents number.

A.6 S6 velocity *vs.* substrate for ksi

1. S6 Velocity *vs.* substrate for ksi - Michaelis-Menten and noise
   (a) Contents
   (b) Simulation specifications
   (c) MM plots: Compilation of average velocity (dP/dt) and standard deviation, for individual simulations with different numbers of substrate agent. Conversion to real units. Plot of Velocity vs. Substrate in real units. Reported Michaelis-Menten kinetic parameter values (Whitman1999) and fit to the Michaelis-Menten equation through 1) non-linear least squares regression, 2) Lineweaver-Burke linear transformation, 3) Hanes-Woolf linear transformation and 4) Eadie-Hofstee linear transformation.
   (d) Noise - Product formation: Compilation of number of product agents number at each time step, for individual simulations of 10 substrate agents (triplicates), 100 substrate agents (triplicates) and 2000 substrate agents (duplicates).
   (e) Noise - Deviation: Graph of evolution of product agents number with time steps for each of the individual simulations.
2. S6.*y* ksi simulation #S
   (a) Simulation specifications
   (b) Simulation data: Tracked data (substrate, enzyme, ES, product agents per time step) for the simulation replicates. Calculation of the average product agents formed, standard deviation and relative standard deviation (standard deviation divided by average).
   (c) Results: Linear regression dP/dt parameters (slope and regression coefficient) for each of the replicates, for the linear portion (55000 to 180000 time steps). Calculation of the average dP/dt and standard deviation. Plot of dP/dt.

A.7 S7 Data curation

1. S7.1 Data curated for input_high kcat isomerases: Identification, enzyme size, reaction, kinetics, simulation inputs, expected behaviour for four different isomerases (2-hydroxymuconate tautomerase, steroid δ-isomerase, isopentenyl-biphosphate δ-isomerase and triose-phosphate isomerase).
2. S7.2 Data curated for input_glycolysis E. coli
   (a) Contents
   (b) Reactions: EC, enzyme's gene ID, pathway, substrates, products, reaction image (link to BRENDA database).

(c) Kinetics: EC, enzyme's gene ID, Km and kcat in the forward and reverse directions, Km and kcat reference (author-date), paired Km and kcat references.

(d) Size: EC, enzyme's gene ID, monomers, molecular weight (polypeptide, inferred from sequence), molecular weight (polypetide, experimental), molecular weight (multimer, experimental), source of experimental values (Pubmed ID), aminoacid number (polypeptide), calculation of hydrodynamic radius (R_h) according to Kalwarczyk's correlation [45], calculation of diffusion coefficient in water at 37 °C by the Stokes-Equation.

(e) Substrate size: metabolite name, PubChem CID, molecular weight (MW), chemical formula, aromatic rings (number), non-aromatic rings (number), $R_v dW$ (nm) calculated by Zhao's correlation [46], calculation of the diffusion coefficient in water at 37 °C by the Stokes-Einstein equation.

3. S7.3 Sources of kinetic data_glycolysis E. coli

(a) Main: all information, organized by Source ID (author-date) and enzyme's gene ID.

(b) Data: source ID, enzyme's gene ID, data field, data value, data value's standard deviation, data value's units

(c) Enzyme production and purification: source ID, enzyme's gene ID, origin organism, , recombinant protein expression (T/F), protein expression organism, plasmid expression vector, culture medium for growth of the expression organism, culture duration for growth of expression organism, culture temperature for growth of expression organism, separation process, final purity of the purified enzyme.

(d) Kinetic parameter assay: source ID, enzyme's gene ID, assay volume, enzyme concentration, substrate concentration range, temperature at which the kinetic assay was carried, pH of the kinetic assay medium, co-factor, co-factor presence (T/F), co-factor concentration, chemical medium of the kinetic assay, data regression, Hill behaviou (T/F), Hill constant, inhibitor, inhibitor presence (T/F), inhibitor concentration.

(e) Source ID and information: source ID, enzyme's gene ID, doi, PMID, Open access (T/F), scope of the literature source, year of publication