# Single Molecule Simulation of Diffusion and Enzyme Kinetics

Gael Pérez-Rodríguez,[†] Denise Gameiro,[‡] Martín Pérez-Pérez,[†] Anália Lourenço,[*,†,§] and Nuno F. Azevedo[‡]

[†]ESEI: Escuela Superior de Ingeniería Informática, University of Vigo, Edificio Politécnico, Campus Universitario As Lagoas s/n, 32004 Ourense, Spain

[‡]LEPABE — Department of Chemical Engineering, Faculty of Engineering, University of Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal

[§]CEB - Centre of Biological Engineering, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

**S** *Supporting Information*

**ABSTRACT:** This work presents a molecular-scale agent-based model for the simulation of enzymatic reactions at experimentally measured concentrations. The model incorporates stochasticity and spatial dependence, using diffusing and reacting particles with physical dimensions. We developed strategies to adjust and validate the enzymatic rates and diffusion coefficients to the information required by the computational agents, i.e., collision efficiency, interaction logic between agents, the time scale associated with interactions (e.g., kinetics), and agent velocity. Also, we tested the impact of molecular location (a source of biological noise) in the speed at which the reactions take place. Simulations were conducted for experimental data on the 2-hydroxymuconate tautomerase (EC 5.3.2.6, UniProt ID Q01468) and the Steroid Delta-isomerase (EC 5.3.3.1, UniProt ID P07445). Obtained results demonstrate that our approach is in accordance to existing experimental data and long-term biophysical and biochemical assumptions.
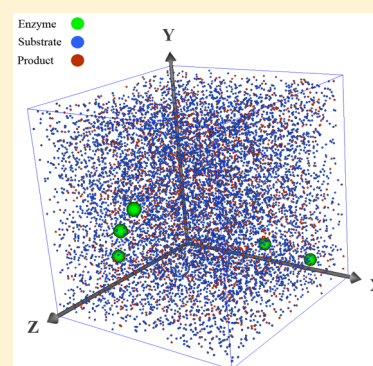
## INTRODUCTION

Due to our lack of sufficient understanding about cellular mechanisms, results of *in vitro* experiments often differ considerably from those observed *in vivo*.[1] Possible explanations include mechanism-specific uncertainty, unknown and non-specific interactions, and molecular crowding. Cells and biomolecular systems are also subject to noise, which is often masked *in vitro* in ensemble experiments and difficult to incorporate in traditional models of biological phenomena.

Cellular noise may be described as the (small) variation of the physiological state of individual cells belonging to an isogenic population and exposed to a similar microenvironment.[2] Ultimately, noise may have a profound effect in the system as these variations may impact the survival of individual cells when exposed to detrimental microenvironmental conditions such as, for instance, high concentrations of antimicrobial agents.[3]

Sources of noise are varied and include stochastic gene expression and the quantity and spatiotemporal variation of molecules (most notably proteins) that are present within the cell.[4,5] However, due to their complexity and interrelationship at the nanoscale, the relative contributions of different sources of noise to the overall cellular behavior have been hard to evaluate in the laboratory. Let us, for instance, consider a simple case where we want to assess the impact of the quantity and spatiotemporal variation of a protein or substrate in a biochemical reaction and, subsequently, in the cellular behavior. In such an experiment, we would have to follow the position

and determine the identity of multiple molecules at the nanotime scale and at a minute volume. While single particle tracking and other advanced experimental methods are already delivering data at single molecule precision, these assessments are very time-consuming.[6] Furthermore, at present, no technique allows the simultaneous observation of different molecule types.[7]

*In silico* simulation offers a valid alternative of analysis since it can be implemented and adjusted more easily. The noise of a biochemical reaction associated with spatiotemporal variation can be assessed by a three-dimensional modeling strategy that accounts for the quantity, diffusion and spatial location of the intervening molecules together with the rules of collision between them. As such, the assembly of biomolecular models has the obvious potential to elucidate structure and auto-organization between molecules as well as complex molecular interplay that are difficult to observe *in vivo* or *in vitro*. The construction of molecular scale models is challenged by the integration of spatial and temporal scales of different orders of magnitude, and the significant computational costs associated with higher levels of modeling detail.

Previous platforms, consisting mainly of agent-based modeling approaches, such as ReaDDy,[8] Smoldyn,[9] and Cellular Dynamic Simulator,[10] are feature-rich tools geared
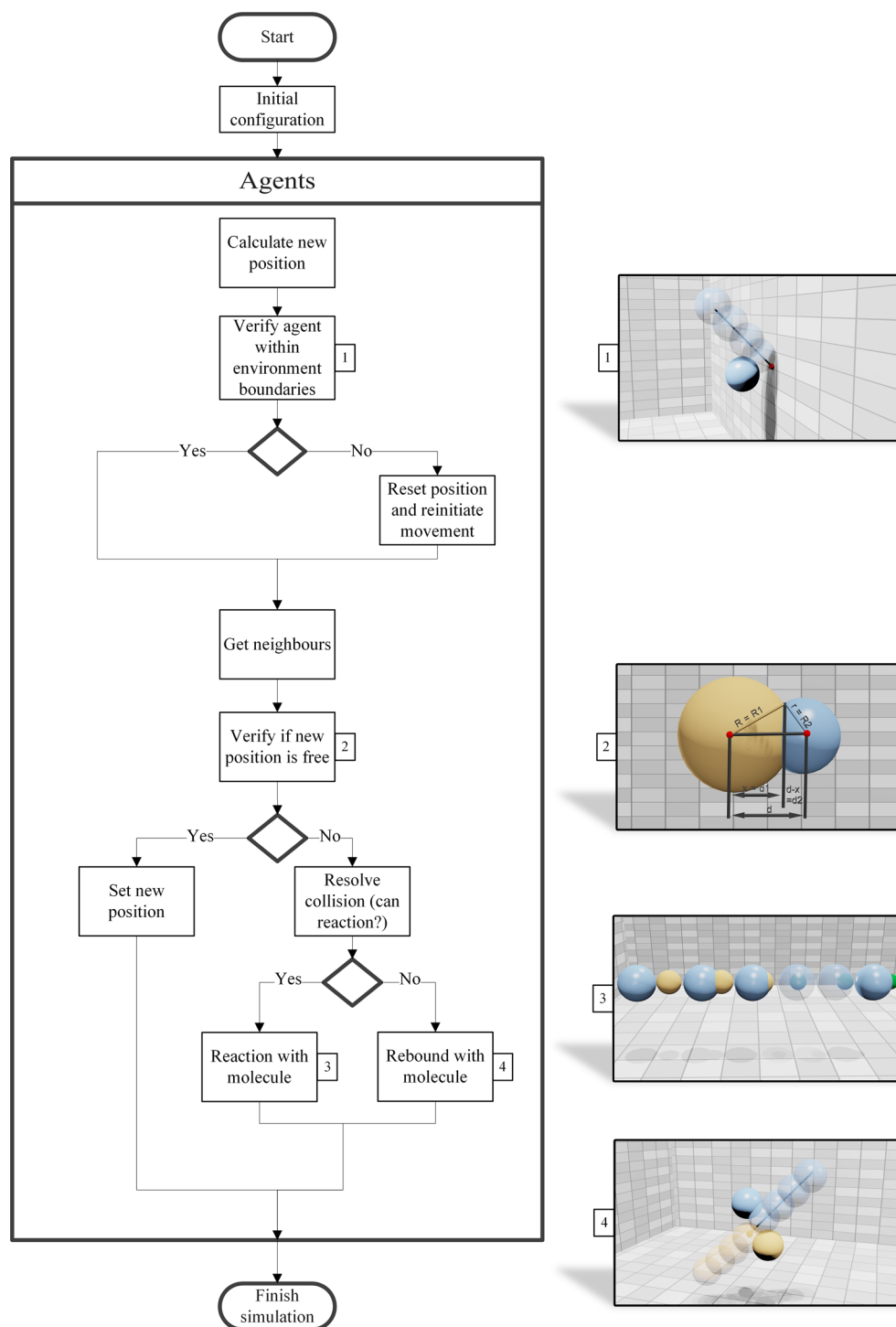
**Figure 1.** Agent-based approach to reaction-diffusion simulation. The basic steps include the description of the simulation environment, the indication of molecular species and corresponding molecular weight, the calculation of the radii and diffusion coefficients of the agents, the definition of the behavioral rules associated with each agent species, and the indication of the number of agents at simulation start.

toward cellular and biomolecular simulation. However, there is often a gap between the data required as inputs in these models and the data that is widely available, either at public databases or routinely determined in published experiments. A paradigmatic case is the kinetics of enzymatic reactions. While individual rates are commonly used in biomolecular modeling, most of the published literature uses Michaelis−Menten parameters to quantify catalytic properties.[11] In fact, databases maintaining information on kinetic parameters, such

as BRENDA[12] and BioCyc,[13] collect kinetic data in terms of the Michaelis−Menten parameters. Therefore, interest is set on addressing the realistic representation and calibration of simulation parameters according to these experimental data.

With the new modeling approach reported here, we analyze the noise resulting from arbitrary initial location and Brownian dynamics, in particular how molecular diffusion may affect enzymatic kinetics. We start by adjusting the information from classical theories, namely the Michaelis constant, $K_m$, the

turnover number, $k_{cat}$, and the diffusion coefficient, $D$, to the information required by the computational agents, i.e., the collision efficiency, the interaction logic, and the velocity of each type of agent.

Overall, simulation results demonstrate that our approach is able to describe the molecular interplay behind enzymatic reactions and account for stochasticity and spatial dependence successfully. For the first time, a biomolecular model has been built on top of database records and experimental data, considering the accurate modeling of biological features as well as the necessary trade-offs between biological details and computational costs.

## ■ MATERIALS AND METHODS

**Simulation Environment.** The present work describes and validates the application of agent-based reaction-diffusion algorithms to the modeling of diffusional motion of individual biomolecules and intermolecular interactions in continuous and fixed time steps three-dimensional cellular environments. Enzymes and metabolites are represented by agents with physical dimensions, which are based on spherical approximations of the excluded volume of the biomolecules. 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.[14] The excluded volume, approximated by the hydrodynamic or van der Waals radius, is a good measure of the actual space occupied by molecules in the reaction volume of biomolecular systems (such as enzymatic assays or reactors and cells).[15] This volume is superior to the purely structural dimensions of biomolecules, due to interactions with the solvent media.

An agent-based simulation system consisting of enzymes and compounds was constructed in a 0.00088 $\mu m^3$ cubic environment. Agents capable of diffusion and reaction take on physical dimensions based on the data reported by scientific literature and the BRENDA database.[16] The hydrodynamic radii and diffusion coefficients of the agents are calculated as described in ref 15. The radius of each particle is $r = 0.0515 M_w^{0.392}$ (nm) for proteins and

$$r = \sqrt[3]{\left(\frac{3}{4\pi}\right)\left[\left(\sum_i^{atom} V_{vdWi}\right) - 5.92N_B - 14.7R_A - 3.8R_{NA}\right]}$$

(nm) for compounds

where $M_w$ corresponds to the molecular weight of the enzyme (including quaternary structure and number of subunits in the multimers, when applicable), $V_{vdWi}$ accounts for the van der Waals volume of each of the atoms of the metabolite, $N_B$ refers to the number of bonds in the chemical structure of the metabolite, and $R_A$ and $R_{NA}$ denote the number of aromatic and nonaromatic rings in the chemical structure. Agents are represented by an equivalent sphere with their Stokes radii.

The simulator implements reaction-diffusion algorithms for modeling the diffusion of individual molecules and the reactions between them in continuous and fixed time steps three-dimensional cellular environments (Figure 1).

Every agent is randomly initialized with a given three-dimensional orientation and tracked continuously throughout the simulation. In each time step, the model checks the current position of the agents, determines the occurrence of agent collisions, and, if so, determines which rule(s) should be triggered. The behavior of the agent is determined by the

corresponding set of behavioral rules and its local surroundings. To facilitate visual inspection, enzymes and metabolites are represented as spheres, differentiating the species by color and size.

Model construction and simulation was conducted in the multi-agent simulator of neighborhoods (MASON), a Java-based and open-source agent-based modeling framework.[17]

**Theoretical Diffusion and Brownian Displacement.** The diffusion coefficients were calculated using the Stokes–Einstein equation for diffusion of spherical particles in a liquid:[15]

$$D_c = \frac{k_B T}{6\pi\eta r}$$

where $k_B$ is Boltzmann's constant, $T$ is the absolute temperature, $\eta$ is the viscosity, and $r$ is the hydrodynamic radius of the spherical particle.

Then, the square-root law of Brownian motion is used to verify that time equivalence and theoretically calculated diffusion rates comply with the expected random walk of the molecules, namely:

$$\langle R^2 \rangle = 6D_c t$$

where $\langle R^2 \rangle$ is the average squared displacement over time, $D_c$ is the diffusion constant, and $t$ is the time interval of simulation.

**Collision Detection.** Given the spherical shape of the agents, the detection of a collision between agents is based on the Pythagorean theorem for triangles:[18,19]

$$D = \sum_{i=1}^{3} a_{1i} a_{2i}$$

where $a_{1i}$ and $a_{2i}$ are the coordinates of two agents and $D$ is the square distance between the two distances (the calculation of square roots is avoided to optimize computational cost). So, a collision is deemed to occur whenever $D \leq radius_{a1} + radius_{a2}$, i.e., collision is detected by knowing that if the distance between the centers of the agents is less than their combined radius the agents collide.

Collision events are 2-fold: agent against environment boundary, and agent against agent. The rule-based behavioral engine decides on whether agents are simply reoriented or additional actions should be taken (Figure 2). In the present scenario, only enzyme–substrate collisions are actionable and represent the occurrence of an enzymatic reaction.

**Enzymatic Reactions.** Enzymatic reactions are defined by the computational parameters $simk_{cat}$, $simK_m$, and reaction radius. The $simk_{cat}$ is defined as the number of time steps between the formation of enzyme–substrate reaction complex and the product release (and return of the enzyme to its free state), mirroring the true $k_{cat}$ parameter as a measure of catalytic efficiency of the enzyme under substrate saturation. The $K_m$, which is related to the affinity of the enzyme toward the substrate and represents the concentration of substrate at which the velocity of a reaction is half of the maximum velocity, was translated in the $simK_m$, which in here it was considered to quantify the probability of a successful collision between an interacting enzyme and substrate. More specifically, a $simK_m$ of 50% means that when the enzyme and substrate collide, half of the times an enzyme–substrate complex is formed and ultimately converted to a product agent, and the remaining times the agents rebound. When a $simK_m$ of 100% was not sufficient to portray a high affinity toward substrate, the

$$1/2m_1\vec{V}_{A1}^2 + 1/2m_2\vec{V}_{B1}^2 = 1/2m_1\vec{V}_{A2}^2 + 1/2m_2\vec{V}_{B2}^2$$
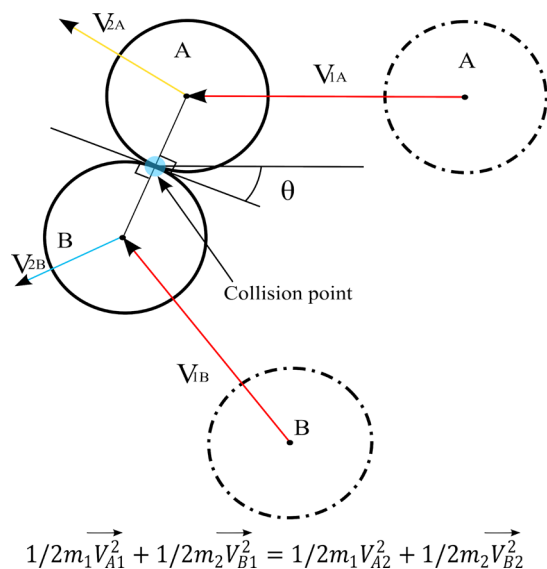
**Figure 2.** Detection and resolution of collisions.

reaction radius was extended so that the enzyme agent could detect substrate agents off the immediate vicinity and test more possible interactions.

Our simulations were 2-fold: a virtual enzyme with a very small $K_m$ and extreme substrate affinity (a percentage of reactive collisions of 100%) and values of sim$k_{cat}$ ranging from 1 to 30 and two enzymes whose kinetic parameters were available in the literature.[20,21] For simplicity, we performed this analysis for enzymes that use only one substrate.

The simulation of enzymatic reactions accounts for the number of substrate, enzyme–substrate complex, and product agents. Simulations were made in triplicate for scenarios with low concentrations of substrate (1.41 mM and below, for both xylH and ksi isomerases simulations) and duplicate for the rest. The velocity of reaction is obtained by linear regression of the number of product agents vs the number of time steps and later converted to biological units of mM s$^{-1}$. Simulations where the substrate was replenished as it was consumed and where the substrate varied with time were performed. For the determination of Michaelis–Menten parameters, it is followed an equivalent procedure to the wet lab experiment, i.e., the measurement of initial velocity of enzymatic assays with constant enzyme concentration and variable substrate concentration.[22]

Noise assessment, in the form of the relative deviation of product formation, was determined by the relation between the standard deviation of total product formed at a given time step and the average of total product. Final averages were obtained from the interval of time steps from the end of the first turnover event to the linear limit of catalysis of the simulation scenario with lower concentration of substrate. For the enzymes 2-hydroxymuconate tautomerase (EC 5.3.2.6) and Steroid Delta-isomerase, the tracked time of simulations were of $5.30 \times 10^{-7}$ to $2.93 \times 10^{-6}$ s (4914 time steps) and $2.68 \times 10^{-5}$ to $3.42 \times 10^{-5}$ s (125 000 time steps), respectively.

## ■ RESULTS AND DISCUSSION

**General Simulation of a Virtual Enzymatic Reaction.** The first aim of this work was to define the duration of one simulation time step in seconds. The time step should represent a sufficiently small fraction of the second in order to be able to simulate superefficient enzymes, i.e., diffusion-controlled enzymes with second-order rate constants of $10^8$ to $10^{10}$ M$^{-1}$ s$^{-1}$.[23] 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, we started by 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. 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. The simulation environment takes on 0.00088 $\mu$m$^3$. The system was populated with 5 enzyme agents (a concentration of $9.38 \times 10^{-3}$ mM) and 20 000 substrate agents (a concentration of $3.75 \times 10^1$ mM), which are realistic relative concentrations compared to literature values.[24−26] While the concentration difference between molecule types is maintained, the concentrations of enzyme and substrate are lower than it is usual in laboratory assays. It is important to notice, however, that the periods of time studied in the simulation are also shorter (below 1 s) than the usual time spans of enzymatic assays. In these shorter periods of time, it is reasonable to consider that phenomena like substrate limitation do not affect the enzymatic reaction in the simulation. The computational parameters sim$k_{cat}$ and sim$K_m$ define the molecular behavior derived for an enzymatic reaction. For a virtual superefficient enzyme, the sim$k_{cat}$, considered to be the number of time steps needed for a reaction to occur once the enzyme–substrate complex is formed, took values from 1 to 30 time steps. The sim$K_m$, defined as the probability of one enzyme and one substrate located within the reaction radius form an enzyme–substrate complex, was set to 100%. For superefficient enzymes, the reversibility of the formation of the enzyme–substrate couples was not considered.

**Diffusion-Controlled Enzymes: sim$k_{cat}$ Cutoff Point.** Initially, we performed a simulation where the transformation of substrate into product would take only one time step to occur, which is equivalent in our system to the minimum value of $k_{cat}$. This implies that, under conditions of substrate saturation, the simulated reaction should be limited by diffusion. We used different criteria to establish the limit between a superefficient, diffusion-limited virtual enzyme, which represents the maximum catalytic efficiency achievable in the simulation environment, and enzymes whose reactions are limited by the catalytic step.

In a first approach, as sim$k_{cat}$ increases, the rate of formation of product agents during the time steps becomes increasingly linear (Figure 3A). This is to be expected of reactions limited by the catalytic step.

The relative occupancy of enzymes, defined as the percentage of the total number of enzymes that are bound in the enzyme–substrate complex (ES/$E_t$ %), was also used to distinguish simulations where the enzymatic reaction was limited by diffusion, as opposed to being limited by the product release step. The relative occupancy of enzymes is calculated based on the average number of ES complexes in the simulation for an increasing number of time steps. This behavior can be observed in Figure 3B, where systems with higher sim$k_{cat}$ are constant at near 100% occupancy of enzymes. Simulations with sim$k_{cat}$ higher than 6 show very similar behaviors to those observed for sim$k_{cat}$ of 5 and 6 (see Supporting Information S1 for full simulation data).
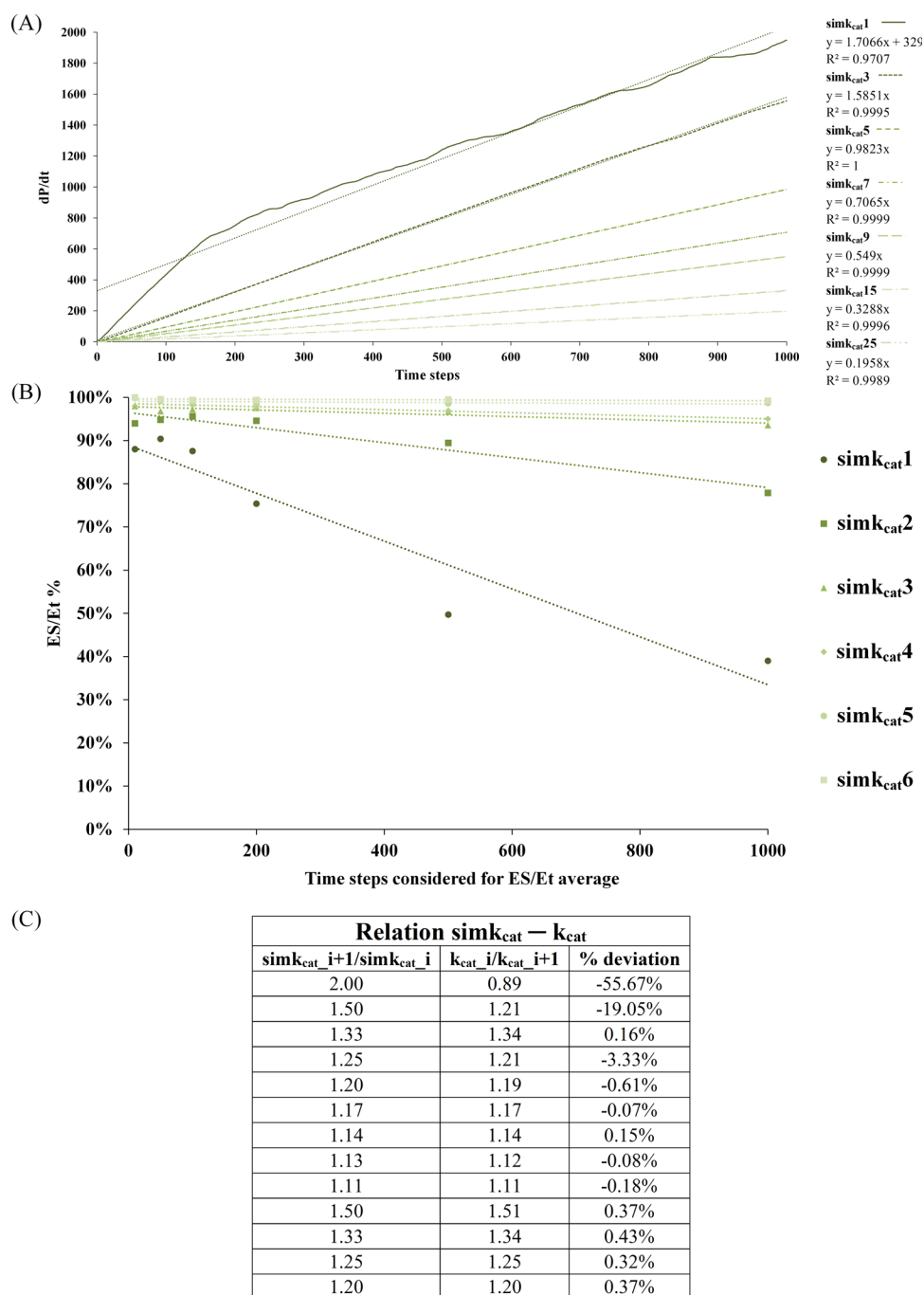
**Figure 3.** Simulation of a virtual enzymatic assay of an isomerase. (A) Velocity of the enzymatic catalysis of the virtual isomerase considering different $simk_{cat}$. Velocity is calculated as $dP/d$ (time step), i.e., the linear regression coefficient of the curve of product formation as a function of the number of time steps. (B) Evolution of the relative occupancy of enzymes ($ES/E_t$) in simulations with increasing $simk_{cat}$. (C) Deviations from a linear relationship between an increase of $simk_{cat}$ ($simk_{cati+1}/simk_{cati}$) and a decrease in the $k_{cat}$ of equivalent simulations ($k_{cati}/k_{cati+1}$).

The matching ratios of $simk_{cat}$ and $k_{cat}$ are also useful to evaluate diffusion control and the coherency of simulation behavior (Figure 3C). It is expected that the $k_{cat}$ will decrease in a linear inverse proportion to the $simk_{cat}$. This rational was applied to the $simk_{cat}$ and the corresponding calculated $k_{cat}$. The deviation between the ratio of consecutive $simk_{cat}$ values ($simk_{cat}$ and $simk_{cati+1}$) and the inverted ratio of the equivalent $k_{cat}$ values corroborate the observation of relative occupancy: the first scenario whose reaction velocity is controlled by the product release time ($k_{cat}$) corresponds to the enzyme with a $simk_{cat}$ of 5 (see details in the Supporting Information, S1).

**Relation between Real Time and Simulation Time Steps.** As stated above, a very important aspect of this work was to find equivalence between each time step and real time (in seconds). For diffusion-limited reactions, $K_{eq}$ constants are known to have a value between $10^8$ and $10^{10}$ $M^{-1}$ $s^{-1}$.[23] Therefore, for the conditions described in our simulation, $5.94 \times 10^3$ M $s^{-1}$ of product should be formed (see the Supporting Information, S1 for full data).

Considering that the cutoff point for the relative occupancy of enzymes was previously defined for a $simk_{cat}$ of 5, the simulation scenarios with lower $simk_{cat}$ should have a

**Table 1. Configuration Data of the Virtual and Real-World Enzyme Simulations**[a]

| molecule | molecular weight ($M_w$ in Da) | spherical radius (nm) | diffusion rate ($\mu m^2$/s) | conc (mM) |
|---|---|---|---|---|
| virtual enzyme | 50 000 | 3.58 | 91.7 | $9.38 \times 10^{-3}$ |
| virtual substrate/product | 70 | 0.59 | 556 | 1.41 |
| xylH enzyme | 22 500 | 2.62 | 125 | $9.38 \times 10^{-3}$ |
| 2-hydroxy-muconate/2-oxo-3-hexenodiate (the substrate and product of the 2-hydroxymuconate tautomerase, respectively) | 158.11 | 0.93 | 353 | $4.69 \times 10^{-2}$ to $18.8 \times 10^{1}$ |
| ksi enzyme | 450 000 | 8.47 | 38.8 | $9.38 \times 10^{-3}$ |
| 3-oxo-delta(5)-steroid/3-oxo-delta(4)-steroid (the substrate and product of the Steroid Delta-isomerase, respectively) | 286.41 | 2.37 | 138.5 | $1.88 \times 10^{-2}$ to 3.75 |

[a]Parameters include the molecular weight, the spherical radius, the diffusion rate, and the concentration of the agents, per molecular species.
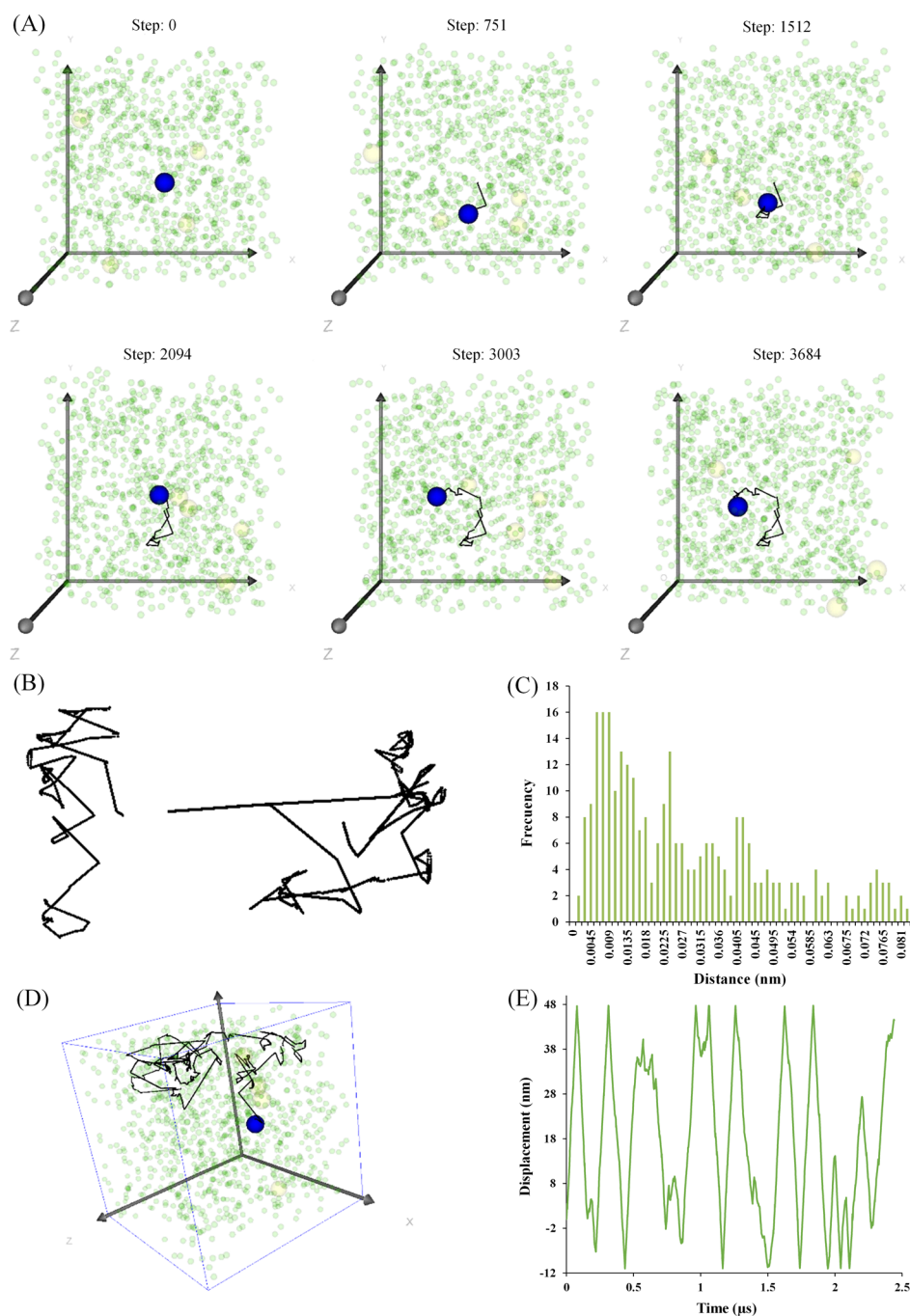


Figure 4. Trajectory of a particle in the three-dimensional continuous and fixed time steps simulation environment. (A) 4D perspective, (B) 2D perspective, (C) distribution of the distance traveled by enzymes, (D) 3D perspective, and (E) 1D perspective.

simulation velocity that can be related to the real velocity expected for an equivalent real system, as calculated by the $K_{eq}$ constant. This relation is a way of comparing real time, in seconds, with simulation time, expressed in time steps (ts):

$$\frac{\text{real time velocity}}{\text{simulation velocity}} = \frac{5.94 \times 10^3 \text{ M s}^{-1}}{2.89 \times 10^{-8} \text{ M ts}^{-1}}$$

$$= 2.05 \times 10^9 \text{ ts s}^{-1}$$

So, the value of $k_{cat}$ can be calculated from simulation velocity, as exemplified in

$$\frac{\text{product formation rate}}{\text{enzymes}} = \frac{1.187 \frac{\text{P molecule}}{\text{ts}}}{5 \text{E molecule}} \times 2.05 \times 10^9$$

$$\text{ts s}^{-1} = 3.65 \times 10^8 \text{ P E}^{-1} \text{ s}^{-1}$$

The value of $k_{cat}$ for the enzymes with $simk_{cat}$ equal or below 5 should be higher than 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 $3 \times 10^6 \text{ s}^{-1}$ (EC number 1.11.1.6), i.e., there is a difference of 2 orders of magnitude between the perfect $simk_{cat}$ and the maximum $k_{cat}$ occurring in reality. This allows the simulation framework to accommodate the fastest enzymes already described and even enzymes that may be identified in the future with fastest kinetics. Hence, the value of $2.05 \times 10^9$ ts $s^{-1}$ was used for the rest of this work, including for the calculation of $k_{cat}$ values (the product produced per enzyme per second) from simulation velocity.

It is important to bear in mind that the simulation framework is sufficiently flexible to modify the correlation between time steps and seconds in several ways, which allows for the correspondence between diffusion and reaction to be adaptable. In fact, we can either define that the minimum $k_{cat}$ takes more than one step (hence adjusting the maximum speed at which a reaction may occur), alter the speed of the molecules in the simulation (hence adjusting the maximum speed at which each molecule can move), or even change the enzyme radii at which a reaction is considered to be able to occur.

**Simulation of Real Isomerases with Defined Kinetic Parameters.** After studying a virtual enzymatic reaction, we focused on simulating the behavior of real isomerases that have been previously described in the literature (Table 1). The first real isomerase to be studied was the 2-hydroxymuconate tautomerase from *Pseudomonas putida* (EC 5.3.2.6, UniProt ID Q01468). This enzyme is reported to have a molecular weight of 22.5 kDa and converts the chemical compound 2-hydroxymuconate to 2-oxo-3-hexenodiate, with a value of $k_{cat}$ of $1.39 \times 10^6 \text{ s}^{-1}$ and a value of $K_m$ of 0.1449 mM.[20] Based on the previous linear relation established between $k_{cat}$ and $simk_{cat}$, the $simk_{cat}$ input for the enzyme was of 1086 time steps ($5.30 \times 10^{-07} \text{ s}^1$).

Additional simulations were conducted for the 2-hydroxymuconate tautomerase. Specifically, these simulations accounted for a nonreacting agent strategy to characterize diffusion, substrate concentrations ranging from $4.69 \times 10^{-2}$ to $18.8 \times 10^1$ mM and the $simk_m$ of 100%, and tested reaction radii 2, 4, and 10 times greater than the hydrodynamic radius of 2-hydroxymuconate tautomerase enzyme (see details in the Supporting Information, S3).

**Brownian Dynamics Simulations.** After defining the equivalence between the time step of the simulation and time,

we were able to calculate the diffusion of the enzyme, product, and substrate agents.

The simulation of Brownian motion, which makes the molecules undergo random-walk motion, follows a square-root law involving the average displacement over time and the diffusion coefficient. For the sake of computational tractability, the simulation does not portray the reaction medium (such as water molecules or other molecules that might eventually be part of the laboratorial experiment) as explicit agents. Random motion is created by collisions between the agents, and the velocity of each agent species is iterated upon until the resulting diffusive behavior matches the one expected for that particular molecule sizes and environment constants, such as temperature and viscosity of the simulated reaction medium. Our results are consistent with the expected random movement (see graphical illustration of simulated molecular motion in Figure 4).

Besides visual inspection of the trajectory of the agents, the convergence of displacements of individual agents was evaluated. This was achieved by following the evolution of the average value for an increasing number of displacement observations, from a single observation to the average of displacements of all the identical agents. The minimum number of agents tested was 750 substrate agents and 5 enzyme agents, which in a simulation volume of 0.00088 $\mu m^3$ correspond to concentrations of 1.41 mM of substrate and $9.38 \times 10^{-3}$ mM of enzyme. For a lower number of agents, the visual inspection of the trajectories of each individual agent indicated that the Brownian motion was no longer applicable, a perception that was strengthened by the fact that agent velocities had to be sharply decreased in these scenarios for a suitable diffusion value to be obtained. As single molecules trajectories can be analyzed, a distribution of the distances traveled by the molecules can also be obtained (Figure 4C).

To determine the Michaelis−Menten parameters, reactions velocities are calculated for different initial substrate concentrations and under scenarios of substrate saturation and substrate limitation. In order to simulate the $K_m$ of the 2-hydroxymuconate tautomerase, simulations with initial substrate concentration below 1.41 mM were executed to emulate reaction velocities in scenarios under substrate limitation, closer to the $K_m$ value of 0.144 mM. In order to obtain a more robust simulation framework, a second strategy was devised, in which part of the substrate-like agents were converted in noninteracting agents, i.e., these agents became obstacles. The velocity of the agents was determined for the scenario with highest number of substrate agents and remained constant for the rest of the simulation runs.

**Calibration of $K_m$.** The calibration of $simK_m$ to $K_m$ was based on the reproduction of experimental assays of kinetic parameters, which measure the velocity of the reaction for different concentrations of substrate, below the substrate saturation level.

Different implementations of the model were tested to refine the meaning of $simK_m$. The first scenarios ran with the same exact implementation that was simulated for the previous experiments of $k_{cat}$ and time to time step relation, i.e., considering a $simK_m$ of 100%. So, in each time step, the enzyme agent looked into its local surroundings for possible interactions with a suitable substrate agent, and if such interaction was possible, the binding was determined by a probability, defined as $simK_m$.

The second implementation changed the enzyme agent behavior so that only one possible enzyme−substrate binding
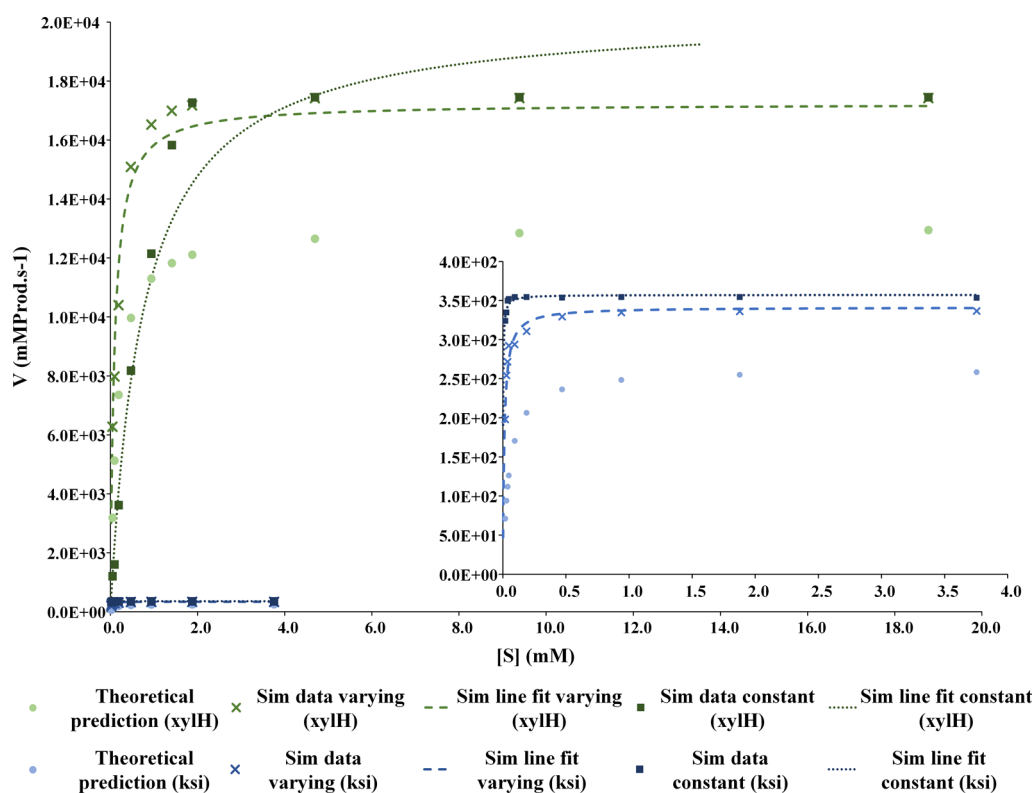
| Enzyme | Theoretical values | | Simulation under varying substrate | | Simulation under constant substrate | |
|---|---|---|---|---|---|---|
| | $K_m$ (mM) | $k_{cat}$ (s$^{-1}$) | $K_m$ (mM) | $k_{cat}$ (s$^{-1}$) | $K_m$ (mM) | $k_{cat}$ (s$^{-1}$) |
| xylH | $1.45 \times 10^{-1}$ | $1.39 \times 10^{6}$ | $8.86 \times 10^{-2}$ | $1.84 \times 10^{6}$ | $7.480 \times 10^{-1}$ | $2.16 \times 10^{6}$ |
| ksi | $5.03 \times 10^{-2}$ | $2.79 \times 10^{4}$ | $1.17 \times 10^{-2}$ | $3.64 \times 10^{4}$ | $1.63 \times 10^{-3}$ | $3.81 \times 10^{4}$ |

**Figure 5.** Final calibration of values for the 2-hydroxymuconate tautomerase (EC 5.3.2.6) and the Steroid Delta-isomerase (EC 5.3.3.1). Plots show the simulation data points for constant and varying substrate, the fitting of the Michaelis−Menten rate equation to the simulation data, and the predicted data points by the published Michaelis−Menten parameters.

was tested in each time step, affecting the probability of successful binding (sim$K_m$) as before. This implementation was tested for values of sim$K_m$ of 100%, 75%, 25%, 10%, 1%, and 0.1%.

In the third implementation, the meaning of sim$K_m$ was changed from a probability of successful enzyme−substrate binding to a probability of release of product from the enzyme−substrate complex (considering a value of sim$K_m$ of 100%, as in the previous implementations). This implementation was tested for values of sim$K_m$ of 100%, 75%, 50%, and 25%.

The final strategy was selected on the basis of the sensibility of the resulting values for Michaelis−Menten parameters and a lack of effect in the $k_{cat}$ parameter (see details in the Supporting Information, S2). Hence, the sim$K_m$ for this enzyme was

defined as the probability of a collision between substrate and enzyme generating an enzyme−substrate complex with only one collision tested per time step. It is noteworthy that, for slower enzymes, $K_m$ is also affected by other characteristics of the system, such as the reversibility of the enzyme−substrate complex, a characteristic that will not be dealt with here. Hence, this implementation was tested for different concentrations of substrate (1.41, 1.88, 2.35, 4.69, 9.38, and 18.8 mM) and different sim$K_m$ probabilities (100%, 75%, 50%, 25%, 10%, 1%, and 0.1%). For the resulting velocities, the $K_m$ was determined using the least-squares nonlinear regression,[27−29] and the linear transformation of Lineweaver−Burke, Hanes−Woolf, and Eadie−Hofstee,[30] yielding similar values of kinetic parameters.

It was observed that while there was a variation of the resulting $K_m$ for different values of sim$K_m$, as intended, the set
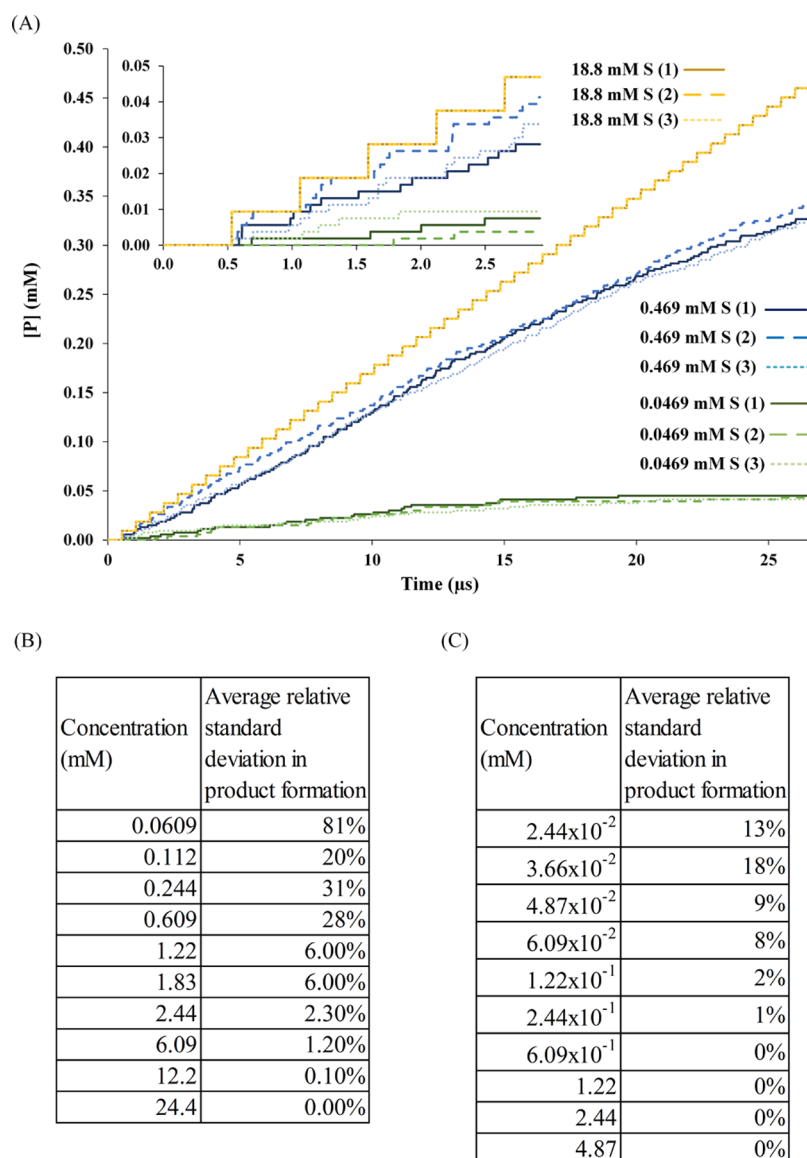
**Figure 6.** Observation of intrinsic stochastic noise at the level of stochastic substrate fluctuations. (A) Evolution of product formation for replicates of the final simulation of 2-hydroxymuconate tautomerase (reaction radius of 2×) for low, medium, and high concentrations of initial substrate. (B) Average relative deviation in product formation for all the initial substrate concentrations tested for 2-hydroxymuconate tautomerase. (C) Average relative deviation in product formation for all the initial substrate concentrations tested for Steroid Delta-isomerase.

of simulation runs with sim$K_m$ of 100% resulted in a higher $K_m$ than the published value of 0.1449 mM. To increase the affinity of the enzyme agent, the concept of reaction radius was applied. Reaction radius is defined as a volume external to the enzyme and within which this agent looks for (anticipates) potential reactive collisions. This radius is defined 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 modeling tool.[9]

The strategy developed for the 2-hydroxymuconate tautomerase was applied to a different isomerase, the Steroid Delta-isomerase, also from *Pseudomonas putida* (EC 5.3.3.1, UniProt ID P07445). The kinetic parameters for the conversion of 5-androstene-3,17-dione to 4-androstene-3,17-dione are reported to be $5.03 \times 10^{-2}$ mM for the $K_m$ and $2.79 \times 10^4$ s$^{-1}$ for the $k_{cat}$.[21] Model inputs were set accordingly: a sim$k_{cat}$ of 54122 time steps ($2.64 \times 10^{-5}$ s), a sim$K_m$ of 100%, and a reaction radius 4 times greater than the hydrodynamic

radius of the Steroid Delta-isomerase. The sim$k_{cat}$ was determined with the proportion previously described in the virtual enzyme section, and the computational parameters related to the sim$K_m$ were selected according to the aimed $K_m$ and the previous results obtained for the 2-hydroxymuconate tautomerase. These values rendered simulation results that are in the same order of magnitude of the values of Michaelis–Menten parameters (Figure 5). A smaller deviation could be achieved by fine-tuning the reaction radius parameter, in an iterative fashion. While we are working with only 5 enzymes, it has been previously demonstrated that the steady-state of reaction rates of single enzymes still obey the Michaelis–Menten equation.[31] It is therefore no surprise that, in spite of individual on–off events, the overall equation holds valid. In addition, we have also tested this strategy with varying concentrations of substrate (i.e., allowing the substrate to be consumed without adding new substrate agents to the

simulation). As expected, this modification had a higher impact in the $K_m$ than in the $K_{cat}$ simulated for both enzymes.

**Assessing Intrinsic Noise.** A spatial agent-based model at the biomolecular scale is suitable for assessing intrinsic stochastic noise, especially at the level of stochastic substrate fluctuations. The low concentrations of some of the key enzymes and metabolites inside the cell can make local substrate fluctuations important sources of cellular variability. While at high concentrations replicate simulations have practically the same exact behavior, with very few local discrepancies of the moment of catalytic turnover, at lower concentrations the turnover events are more dispersed across time steps (Figure 6A).

Additionally, 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, both for the 2-hydroxymuconate tautomerase enzyme (Figure 6B) and steroid-delta isomerase enzyme (Figure 6C).

In the 2-hydroxymuconate tautomerase simulation, for a small concentration of substrate, $4.69 \times 10^{-2}$ mM, the deviation in product formed is above 50%, and the deviation decreases as more substrate is available in the simulation. For a low concentration of substrate, the time difference between equivalent turnover events (a first conversion of substrate into product, for example) can be up to $1.21 \times 10^{-6}$ s, a period of time superior to the rate at which the enzyme can convert a bound substrate to a product molecule. 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 modeling 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 in every respect, including initial substrate concentration and homogeneous spatial distribution. Conditions of low substrate concentration are not negligible in *in vivo* reactions, where 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+.[32]

Relative deviations are smaller for equivalent substrate concentrations in the simulation of the Steroid Delta-isomerase. This could be due to differences in Michaelis−Menten behavior, since the steroid delta-isomerase has a lower $K_m$, i.e., has more affinity toward the substrate, and has a lower $k_{cat}$, i.e., slower catalytic conversion of substrate.

Previous work in intrinsic metabolic noise attributed variation to two main sources, namely, the fluctuations on the three-dimensional structure of proteins and the substrate concentrations in the immediate surroundings of the enzyme. The first type of fluctuations give origin to several interconvertible conformational states with different catalytic activities.[33] The second type of noise, that is explored in the present work, arises from the uneven spatial distribution of substrate, which is particularly relevant in scenarios with low concentrations of biomolecules or crowding, such as the cell.[34,35]

**Final Remarks.** Ultimately, the goal of molecular-scale computational models of cellular environments is to grow our understanding about the differences between biomolecular behavior observed *in vitro* and *in vivo*. The simulations described here represent the first attempt to build such a model using diffusing and reacting particles with realistic physical dimensions, and incorporating stochasticity and spatial dependence in a three-dimensional environment.

The biophysical assumptions of the model were validated, namely the biological temporal scales portrayed by simulation runs and diffusive behavior. We modeled the computational parameters of the enzyme agents in such a way that replicated their kinetic behavior in an equivalent to the reaction environment of the enzymatic assay, namely, with water as the solvent of the reaction. If published kinetic parameters of *in vitro* experiments allow us to determine intrinsic enzyme characteristics, the simulation of the same enzymes in an environment closer to the cytoplasmic composition could offer insight into the kinetic behavior of enzymes *in vivo*, and by extension, into the kinetics of metabolic pathways and cellular systems. A molecular-based model allows the intuitive modeling of different sources of cellular and biomolecular noise. Notably, the spatial scale implicitly incorporates the noise related to the spatial distribution of biomolecules. Uneven spatial distribution, at low concentrations, can originate diverging results in identical biological scenarios and, at the cellular scale, diverging phenotypes.

In this model, spherical molecules of different sizes diffuse through the three-dimensional space and are capable of reaction upon collision. The realism of the model in terms of its diffusive and catalytic properties was evaluated according to known empirical behavior and available experimental data. For the most part, results show that the devised agent-based approach is consistent with experimentally validated results and, thus, may be used for *in silico* metabolic simulation.

In order to validate core assumptions without unnecessary complexity, the simulated enzymes were both isomerases, following the simplest enzymatic mechanism of irreversible substrate binding and single substrate. Also, we selected two enzymes with $k_{cat}$ values above the average to keep computation time manageable. However, the present approach should be able to reproduce the behavior of any isomerase enzyme, granted that the computational parameters are tuned accordingly. The model may also be expanded to portray more complex kinetics, such as reversibility, multisubstrate reactions, enzymatic activation, and inhibition by other metabolites.

This experiment exposes the computational requirements imposed by a realistic scenario and raises discussion about future lines of research and development for agent-based biomodelling. The coarse-grain modeling approach proposed here can also be further adapted to the simulation of other known enzymatic behaviors, such as in the case of fluctuating enzymes,[36,37] or expanded to simulate full metabolic pathways.

## ■ ASSOCIATED CONTENT

**Ⓢ Supporting Information**

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcb.5b12544.

Data on the determination of $simk_{cat}$ cutoff point for diffusion-controlled enzymes and equivalence between real time and simulation time steps (XLSX)

Data on the calibration of $K_m$ with the computational parameter of probability of successful reactions (XLSX)

Data on the calibration of $K_m$ with the computational parameter of reaction radius (XLSX)

■ **AUTHOR INFORMATION**

**Corresponding Author**
*Phone: +34 988 387013. Fax: +34 988 387001. E-mail: analia@uvigo.es.

**Notes**
The authors declare no competing financial interest.

■ **REFERENCES**

(1) Pfaffelhuber, P.; Popovic, L. How Spatial Heterogeneity Shapes Multiscale Biochemical Reaction Network Dynamics. *J. R. Soc., Interface* **2015**, *12*, 20141106−20141106.

(2) Raser, J. M.; O'Shea, E. K. Noise in Gene Expression: Origins, Consequences, and Control. *Science* **2005**, *309*, 2010−2013.

(3) Eldar, A.; Elowitz, M. B. Functional Roles for Noise in Genetic Circuits. *Nature* **2010**, *467*, 167−173.

(4) Snijder, B.; Pelkmans, L. Origins of Regulated Cell-to-Cell Variability. *Nat. Rev. Mol. Cell Biol.* **2011**, *12*, 119−125.

(5) Swain, P. S.; Elowitz, M. B.; Siggia, E. D. Intrinsic and Extrinsic Contributions to Stochasticity in Gene Expression. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12795−12800.

(6) Ladbury, J. E.; Arold, S. T. Noise in Cellular Signaling Pathways: Causes and Effects. *Trends Biochem. Sci.* **2012**, *37*, 173−178.

(7) McGuffee, S. R.; Elcock, A. H. Diffusion, Crowding & Protein Stability in a Dynamic Molecular Model of the Bacterial Cytoplasm. *PLoS Comput. Biol.* **2010**, *6*, e1000694.

(8) Schöneberg, J.; Noé, F. ReaDDy–a Software for Particle-Based Reaction-Diffusion Dynamics in Crowded Cellular Environments. *PLoS One* **2013**, *8*, e74261.

(9) Andrews, S. S.; Addy, N. J.; Brent, R.; Arkin, A. P. Detailed Simulations of Cell Biology with Smoldyn 2.1. *PLoS Comput. Biol.* **2010**, *6*, e1000705.

(10) Byrne, M. J.; Waxham, M. N.; Kubota, Y. Cellular Dynamic Simulator: An Event Driven Molecular Simulation Environment for Cellular Physiology. *Neuroinformatics* **2010**, *8*, 63−82.

(11) Gameiro, D.; Pérez-Pérez, M.; Pérez-Rodríguez, G.; Monteiro, G.; Azevedo, N. F.; Lourenço, A. Computational Resources and Strategies to Construct Single-Molecule Metabolic Models of Microbial Cells. *Briefings Bioinf.* **2015**, 1−14.

(12) Chang, A.; Schomburg, I.; Placzek, S.; Jeske, L.; Ulbrich, M.; Xiao, M.; Sensen, C. W.; Schomburg, D. BRENDA in 2015: Exciting Developments in Its 25th Year of Existence. *Nucleic Acids Res.* **2015**, *43*, D439−D446.

(13) Caspi, R.; Altman, T.; Dale, J. M.; Dreher, K.; Fulcher, C. A.; Gilham, F.; Kaipa, P.; Karthikeyan, A. S.; Kothari, A.; Krummenacker, M.; et al. The MetaCyc Database of Metabolic Pathways and Enzymes and the BioCyc Collection of Pathway/genome Databases. *Nucleic Acids Res.* **2010**, *38*, D473−D479.

(14) Feig, M.; Sugita, Y. Reaching New Levels of Realism in Modeling Biological Macromolecules in Cellular Environments. *J. Mol. Graphics Modell.* **2013**, *45*, 144−156.

(15) Kalwarczyk, T.; Tabaka, M.; Holyst, R. Biologistics–Diffusion Coefficients for Complete Proteome of Escherichia Coli. *Bioinformatics* **2012**, *28*, 2971−2978.

(16) Scheer, M.; Grote, A.; Chang, A.; Schomburg, I.; Munaretto, C.; Rother, M.; Söhngen, C.; Stelzer, M.; Thiele, J.; Schomburg, D. BRENDA, the Enzyme Information System in 2011. *Nucleic Acids Res.* **2011**, *39*, D670−D676.

(17) Luke, S.; Cioffi-Revilla, C.; Panait, L.; Sullivan, K.; Balan, G. MASON: A Multiagent Simulation Environment. *Simul. Trans. Soc. Model. Simul. Int.* **2005**, *82*, 517−527.

(18) Millington, I. *Game Physics Engine Development: How to Build a Robust Commercial-Grade Physics Engine for Your Game*; CRC Press: Boca Raton, FL, 2010.

(19) Palmer, G. *Physics for Game Programmers*; Apress: New York, 2005.

(20) Whitman, C. P.; Aird, B. A.; Gillespie, W. R.; Stolowich, N. J. Chemical and Enzymic Ketonization of 2-Hydroxymuconate, a Conjugated Enol. *J. Am. Chem. Soc.* **1991**, *113*, 3154−3162.

(21) Choi, G.; Ha, N. C.; Kim, M. S.; Hong, B. H.; Oh, B. H.; Choi, K. Y. 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* **2001**, *40*, 6828−6835.

(22) Brooks, H. B.; Geeganage, S.; Kahl, S. D.; Montrose, C.; Sittampalam, S.; Smith, M. C.; Weidner, J. R. *Basics of Enzymatic Assays for HTS*; Sittampalam, G., Coussens, N., Nelson, H., Arkin, M., Auld, D., Austin, C., Bejcek, B., Glicksman, M., Inglese, J., Iversen, P., et al., Eds.; NCBI Bookshelf: Bethesda, MD, 2004.

(23) Alberty, R. A.; Hammes, G. G. Application of the Theory of Diffusion-Controlled Reactions to Enzyme Kinetics. *J. Phys. Chem.* **1958**, *62*, 154−159.

(24) Donahue, J. L.; Bownas, J. L.; Niehaus, W. G.; Larson, T. J. Purification and Characterization of glpX-Encoded Fructose 1,6-Bisphosphatase, a New Enzyme of the Glycerol 3-Phosphate Regulon of Escherichia Coli. *J. Bacteriol.* **2000**, *182*, 5624−5627.

(25) Alvarez, M.; Zeelen, J. P.; Mainfroid, V.; Rentier-Delrue, F.; Martial, J. A.; Wyns, L.; Wierenga, R. K.; Maes, D. Triose-Phosphate Isomerase (TIM) of the Psychrophilic Bacterium Vibrio Marinus: Kinetic and Structural Properties. *J. Biol. Chem.* **1998**, *273*, 2199−2206.

(26) Poyner, R. R.; Laughlin, L. T.; Sowa, G. a.; Reed, G. H. Toward Identification of Acid/base Catalysts in the Active Site of Enolase: Comparison of the Properties of K345A, E168Q, and E211Q Variants. *Biochemistry* **1996**, *35*, 1692−1699.

(27) Shone, C. C.; Fromm, H. J. Steady-State and Pre-Steady-State Kinetics of Coenzyme A Linked Aldehyde Dehydrogenase from Escherichia Coli. *Biochemistry* **1981**, *20*, 7494−7501.

(28) Yang, D.-F.; Wei, Y.-T.; Huang, R.-B. Computer-Aided Design of the Stability of Pyruvate Formate-Lyase from Escherichia Coli by Site-Directed Mutagenesis. *Biosci., Biotechnol., Biochem.* **2007**, *71*, 746−753.

(29) Berry, A.; Marshall, K. E. Identification of Zinc-Binding Ligands in the Class II Fructose-1,6-Bisphosphate Aldolase of Escherichia Coli. *FEBS Lett.* **1993**, *318*, 11−16.

(30) Dowd, J. E.; Riggs, D. S. A Comparison of Estimates of Michaelis-Menten Kinetic Constants from Various Linear Transformations. *J. Biol. Chem.* **1965**, *240*, 863−869.

(31) Cao, J. Michaelis-Menten Equation and Detailed Balance in Enzymatic Networks. *J. Phys. Chem. B* **2011**, *115*, 5493−5498.

(32) Bennett, B. B. D.; Kimball, E. E. H.; Gao, M.; Osterhout, R.; Van Dien, S. J.; Rabinowitz, J. D. Absolute Metabolite Concentrations and

Implied Enzyme Active Site Occupancy in Escherichia Coli. *Nat. Chem. Biol.* **2009**, *5*, 593−599.

(33) English, B. P.; Min, W.; van Oijen, A. M.; Lee, K. T.; Luo, G.; Sun, H.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Ever-Fluctuating Single Enzyme Molecules: Michaelis-Menten Equation Revisited. *Nat. Chem. Biol.* **2006**, *2*, 87−94.

(34) Yi, M.; Liu, Q. Michaelis−Menten Mechanism for Single-Enzyme and Multi-Enzyme System under Stochastic Noise and Spatial Diffusion. *Phys. A* **2010**, *389*, 3791−3803.

(35) Qian, H.; Elson, E. L. Single-Molecule Enzymology: Stochastic Michaelis−Menten Kinetics. *Biophys. Chem.* **2002**, *101−102*, 565−576.

(36) Yang, S.; Cao, J. Direct Measurements of Memory Effects in Single-Molecule Kinetics. *J. Chem. Phys.* **2002**, *117*, 10996−11009.

(37) Witkoskie, J. B.; Cao, J. Single Molecule Kinetics. I. Theoretical Analysis of Indicators. *J. Chem. Phys.* **2004**, *121*, 6361.