Agent-based Models for Detecting the Driving Forces of Biomolecular Interactions

Stefano Maestri
University of Camerino
Emanuela Merelli (emanuela.merelli@unicam.it)
University of Camerino
Marco Pettini
Aix-Marseille University

Research Article

Keywords: biological systems, biomolecules, electromagnetic potentials, glycolytic oscillations

Posted Date: September 22nd, 2021

DOI: https://doi.org/10.21203/rs.3.rs-911183/v1

Version of Record: A version of this preprint was published at Scientific Reports on February 3rd, 2022.
See the published version at https://doi.org/10.1038/s41598-021-04205-8.

Agent-based models for detecting the driving forces of biomolecular interactions

Stefano Maestri[1,2], **Emanuela Merelli**[1,*], **and Marco Pettini**[2]

[1]School of Science and Technology, University of Camerino, Camerino, 62032, Italy
[2]Aix-Marseille Univ, Université de Toulon, CNRS, Centre de Physique Théorique,163 Avenue de Luminy, 13288 Marseille Cedex 9, France
[*]emanuela.merelli@unicam.it

ABSTRACT

Agent-based modelling and simulation have been effectively applied to the study of complex biological systems, especially when composed by a large number of interacting entities. Representing biomolecules as autonomous agents allows this approach to bring out the global behaviour of biochemical processes as resulting from local molecular interactions. In this paper, we leverage the capabilities of the agent paradigm to construct an *in silico* replica of the glycolytic pathway of baker's yeasts; the aim is to detect the role that long-range electrodynamic forces might have on the rate of glucose oxidation. Experimental evidences have shown that random encounters and short-range potentials might not be sufficient to explain the high efficiency of biochemical reactions in living cells. However, while the latest *in vitro* studies are limited by the present-day technology, agent-based simulations provide an *in silico* support to the outcomes hitherto obtained and shed light on behaviours not yet well understood. Our results reveal to be able to grasp properties hard to uncover through other computational methods, such as the effect of electromagnetic potentials on glycolytic oscillations.

Introduction

Long-distance electrodynamic interactions between two small molecules have been largely studied within the framework of quantum electrodynamics, since long-range forces can be detected among excited atoms with similar transition frequencies[1,2]. However, interactions beyond the Debye screening length ($\simeq 10$Å in biological systems[3]), carried out by the molecular cognate-partners of a biochemical reaction, are not well investigated. Nonetheless, experimental evidences for the existence of collective excitations of biological macromolecules are available in the Raman and Far-infrared spectroscopic domains[4,5]. If long-distance electrostatic interactions have been considered unlikely, electrodynamic interactions, occurring between oscillating electric dipoles, might have a long-range nature; deterministic selective forces can thus be activated at a distance, when the molecules undergo coherent collective oscillations[6]. The existence of forces of this kind might justify the efficiency of biochemical reactions more than the solely effect of stochastic short-range interactions, which rely just on Brownian diffusion and chemical affinity. Numerical studies proved that the overall interaction potential $\mathsf{U}(\mathsf{r})$ between cognate-partners (with r being the intermolecular distance) is generally composed of a short-range term ($1/\mathsf{r}^6$) and a resonant long-range term ($1/\mathsf{r}^3$), meaning that, when the dipole moments of two molecules oscillate at the same frequency, an attractive resonant potential $\mathsf{U}(\mathsf{r}) \sim \mathsf{r}^{-3}$ should be added to the random Brownian force[7].

These phenomena have been lately analysed, theoretically and experimentally, in the interactions among lysozyme molecules and oppositely charged dyes[8]. However, detecting long-range molecular recruitments in biosystems is still held back by the current technology; even these recent results, gained through Fluorescence Correlation Spectroscopy, are limited to systems where the long-range interactions are built-in (by setting up a solution in which the electrostatic interactions are non-screened).

Computational approaches might be able to overcome some of these hurdles, allowing to test *in silico* the existing theoretical models. Indeed, numerical simulations, such those performed though Molecular Dynamics, have been already successfully carried out[8], taking into account an *a priori* knowledge of

numerous physical parameters characterising the molecular interactions under study. From the one hand, a model that considers the largest amount of empirical information allows generating a faithful representation of the biological system and provides a reliable *in silico* support for theoretical and experimental analyses; from the other hand, the lack of empirical data may limit the complexity of the system simulated.

In this regard, Systems Biology represents a complementary approach. When oriented to provide computational models of metabolic pathways, it focuses mostly on the kinetic properties of the enzymatic reactions; it can thus be used to simulate a complex network of interacting macromolecules basing on a relatively small amount of experimental data[9]. Nonetheless, heavily relying on kinetic information might also represent its major flaw, especially since the *in vitro* procedures adopted to sample kinetic parameters might not faithfully capture the actual behaviour that enzymes carry out in living cells, where they dynamically interact with their environment[10]. Furthermore, the standard kinetic modelling is based on the Michaelis-Menten formalism, which assumes diffusion-limited enzymatic reactions and a homogeneous environment, overlooking the effects of molecular crowding on diffusion processes[11]. Other modelling approaches in Systems Biology (such as those based on Brownian dynamics or spatial partial differential equations) also struggled to deal with this problem, due to the complexity of representing molecules' 3D shape and diffusion at different scales[12].

In this article, we aim to address most of these issues, by exploiting an alternative way to define a computational model of molecular interactions in a metabolic pathway. Specifically, we construct an agent-based model (ABM) of a well-studied process, the glycolysis of yeasts, in order to simulate the effect of the long-distance electrodynamic interactions among the biomolecules involved in the pathway. Agent-based simulations make use of autonomous software pieces (agents) able to interact with one another in a concurrent and asynchronous fashion; they can thus fairly faithfully replicate *in silico* the behaviour of the entities interacting in a real biological system. Differently from other computational methods, however, ABMs require to instruct the agents representing the simulated molecules with minimal empirical information, letting the global behaviour of the process to result from local interactions, which are generated dynamically at each step of the simulation. The system evolves due to the ability of every agent to perceive and respond to the states of its environment, which is unpredictable and populated by other agents; the agent's perception results in performing an appropriate action (if any) able to modify the environment[13]. The agent-based approach allows both the environment and the molecules to be three-dimensional (as shown in Fig. 1); molecular shapes can thus affect the diffusion processes.

ABMs have been already successfully applied in the analysis of several biological systems and used to develop tools for *in silico* supporting experimental studies[14–16]. With the present work, we leverage the flexibility of the agent-based modelling to construct *in silico* biochemical systems; this approach is intended to simulate the glycolytic process by taking into account different types of forces driving molecular interactions. We aim to abstract the core features of biochemical systems characterised by purely random molecular encounters and compare them to those where the cognate-partners' interactions are mainly driven by deterministic long-range forces. ABMs allow us to reproduce these phenomena in a network of mutually conditioning reactions without knowing *a priori* all the parameters needed in a numerical simulation, which might be missing or difficult to assay experimentally.

By analysing the concentration changes of the molecular species during each step of the agent-based simulation, we are able to hypothesise how long-distance interactions may quantitatively and qualitatively affect the glycolysis process. This way, we can also hint what might be the physical phenomena underlying the kinetic parameters if they would be assayed *in vivo* and highlight possible discrepancies with the values obtained *in vitro*. These results would provide the basis for setting up further experimental studies.

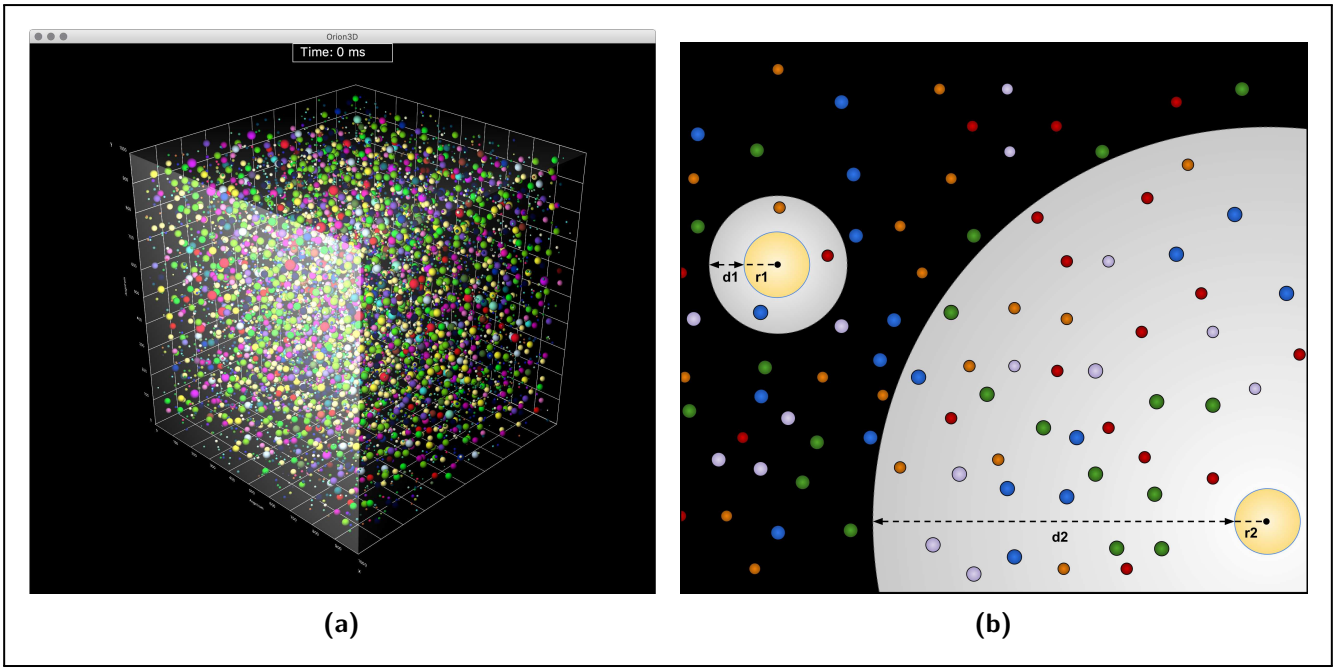

Figure 1. **(a)** Three-dimensional representation of the agents environment. Specifically, it is a one-attolitre cubic portion of cell cytoplasm, populated by enzymes and metabolites, each modelled as an autonomous agent and represented as a sphere. The figure is obtained from the 3D interface of the simulator we developed over our glycolysis ABM. It shows the position of every molecule instant by instant. The software makes also possible to highlight the metabolites perceived by each enzyme at each time step of the simulation. **(b)** Graphical representation of the agent's perception, by which every modelled enzyme detects the cognate metabolites in its surrounding environment. Each enzyme, depicted as a sphere of radius `r`, is able to perceive its neighbouring metabolites at different distances `d`. Such a process is fundamental to reproduce *in silico* the effects of the long-range forces on biochemical reactions, as it will be discussed throughout this article.

Methods

From a Kinetic to an Agent-based Model

To construct an ABM able to represent the molecular interactions of a metabolic pathway, some information is required on the pathway itself and the environment where it takes place. In particular, we need to know the sequence of reactions to simulate, or a subset of those that are relevant for our analysis, and some quantitative data, such as the initial concentrations of the species involved (as it will be better explained later in this section). In this perspective, a kinetic model can serve as a source of such data and also represent a reference to which compare our results.

We remark that we cannot completely base our study on a kinetic model, since it uses experimental parameters, often assayed *in vitro*, to directly describe the global properties of the system through a set of differential equations. However, we aim to understand if kinetic data actually underlie processes related to the ability of molecules to perceive each other, even from a long distance. An ABM of molecular interactions permits to not push *a priori* in the system most of these parameters and thus provides a better baseline over which carrying out our *in silico* studies. ABMs describe the molecular interactions at a local level, but they also posses *compositionality*, that is the capability of recursively applying the rules characterising agents interactions to progressively define higher abstraction levels. In this way, we are able to hide the unnecessary details of a specific level and, at the same time, to observe its global behaviour[17,18].

Considering the case of a metabolic pathway, a kinetic model treats enzymatic reactions as mathematical functions that relate the concentrations of reactants to those of products, assuming that they incorporate the role carried out by each molecular interaction. Conversely, in our ABM, each enzyme is represented by a dedicated agent able to perceive the environment and its cognate partners; the interactions among the molecules are explicit in the definition of the model. The compositionality of ABMs makes also possible to conduct the study at an abstraction level that can be represented with a small amount of empirical data, without loosing in accuracy when reproducing macromolecular behaviours.

Nonetheless, not all the kinetic parameters can be overlooked: in order for a modelled saturated enzyme to generate the products of the reaction, faithfully to its biological counterpart, it must wait for a time interval corresponding to the reciprocal of its k_{cat} value (or turnover number). The k_{cat} represents the number of molecules converted by an enzyme in the time unit, therefore, its reciprocal provides the interval after which the reaction products are released in the environment and the enzyme returns free.

Several kinetic models have already been constructed over metabolic pathways, mainly because the properties of metabolism at steady state simplify the model definition[9]. However, by considering the enzymatic reactions as just mathematical functions from reactants to products, they mostly focus on changes in metabolite concentrations and do not provide the actual number of enzyme molecules in the simulated environment. Conversely, for the reasons explained above, this information is fundamental for constructing our ABM. Basing on this requirement, we identified in the "Smallbone2013 - Iteration 18"[19] a model particularly suitable to serve as a source for the ABM, since it contains a complete set of experimental data on the isoenzymes involved in a well-studied metabolic process, the glycolysis of Saccharomyces cerevisiae. The Smallbone2013 model provides a detailed description of the chain of reactions that generates energy from glucose by breaking it into two molecules of pyruvate. In addition to the main branch of glycolysis, it includes the glycerol, glycogen and trehalose branches and also considers the alcoholic fermentation steps, which lead to the formation of ethanol (see Figure 2). It therefore defines a system of interacting molecules sufficiently complex to allow bringing out, through an ABM, the global effect of the long-range forces on the dynamics of the pathway.

Designing the Model of Glycolysis

The ABM is the basis for a simulator whose input is in the form of a SBML (Systems Biology Markup Language) file filled with experimental data[20]. It contains information about the molecules involved in the metabolic pathway and their initial concentrations; the data related to the reactions carried out are also taken from this source. The Smallbone2013 model is represented in the SBML format.

A dedicated interface of the simulator converts the SBML model to an XML (Extensible Markup Language) file specifically formatted to be interpreted by the simulator itself but also to be human-readable[21]. Therefore, its main function is to translate the kinetic representation of the metabolic reactions into our agent-based model. To do this, for every reaction in the SBML model, it gets the reactants and products, and generates a XML code for each of its interactions. It also associates, to each modelled molecule and reaction, the physical and kinetic parameters needed to setup the simulation; for the study proposed in this paper, they are limited to the molecular weights (automatically retrieved from online databases, such as UniProt[22] and ChEBI[23]) and the turnover numbers. A generalisation of the XML input and its construction, along with the actual file generated for running the simulations, is provided in the Supplementary Information.

By importing the reactions of the SMBL file as the input of our agent-based simulations, we excluded all those for which the Smallbone2013 model does not provide the enzymatic concentrations. Our simulator can actually handle this kind of reactions, since we can model them in terms of their bulk effects; however, for the aim of observing the global behaviour of glycolysis as resulting from the the local molecular interactions, introducing any bulk reaction would perturb the environment and hide the absence of actual interactions among the molecules, making the ABM close to a standard kinetic model. Basing on this idea, we do not consider the *adenylate kinase* reaction, the *ATPase* reactions, the *UDP to UTP* reaction and the

glucose transport (between the extracellular environment and the cytosol). The most significant of these reactions is the adenylate kinase, since it controls the ratio of `ATP`, `ADP` and `AMP` (also called energy charge), which in turn affects the allosteric regulation of important enzymes, such as `phosphofructokinase` and `hexokinase`[24]. However, the length of the simulated process (1 second, as better discussed in the Results) makes the allosteric regulation and the whole energy charge effects negligible[25,26]. Suppressing glucose transport and enzyme regulation also prevents, de facto, the achievement of a steady state, helping us to emphasise the effects of the various types of interactions on the concentration changes in the simulation interval.

The initial concentrations of the molecular species are gained from the SBML file as millimoles per litre (mmol/l). The simulator's interface converts these values into the initial particle numbers, needed to instantiate the agents at the beginning of the simulation. In this regard, we point out that, although agent-based simulations have a fairly light computational load, reproducing a metabolic pathway involves thousands of molecules, and therefore as many agents running concurrently. The resulting resources demand conditioned the molecular concentrations we were able to simulate. More precisely, we scaled the concentrations provided by the Smallbone2013 model to values less than 1 mmol/l. In Supplementary Table S1, we report the initial concentrations of all the simulated species.

Our agent-based model is intended to study the glycolytic pathway from the general perspective of the oxidation of one molecule of `glucose` to two molecules of `pyruvate`; for this reason, we consider the `pyruvate` as the end product of the process and excluded the fermentation-related reactions, catalysed by the `pyruvate decarboxylase isoenzymes` (`PDC1`, `PDC5`, `PDC6`) and by the two `alcohol dehydrogenase isoenzymes` (`ADH1` and `ADH5`). Therefore, the branches acting on `pyruvate`, that is the `succinate` and `acetate` branches of glycolysis, are not taken into account in our model.

To complete the list of changes we made to the original kinetic model, we report that, according to most of the literature, we modelled the reactions catalysed by `hexokinase` (and `glucokinase`), `phosphofructokinase` and `pyruvate kinase` as irreversible[27–29], since they function as control points of the whole glycolysis process, despite the Smallbone2013 model considers irreversible only the reaction performed by the `phosphofructokinase`.

The subset of reactions characterising the model at the basis of our simulations, as resulting from the above described adaptations, can be found in Fig. 2 of this manuscript and in Supplementary Table S2.

Orion - Agent-based Simulator for Metabolic Pathways

The study proposed in this manuscript has been carried out with the aid of Orion, a spatial simulator for metabolic pathways. It has been developed in Java starting from a prototype project[30,31]. In what follows, we introduce the basic principles over which the simulator has been designed; this information can help the reader to better understand the results of our work. Further details on modelling and implementation choices are provided in Section 2 of the Supplementary Information.

Orion is a agent-based simulator, this means that the molecules involved in the pathway are represented by agents, autonomous software peaces able to perceive changes in the environment and react to them. Formally, a *reactive agent* is defined by a 6-tuple $\langle E, Per, Ac, see, action, do \rangle$ where:

- E is the set of all states for the environment
- Per is a partition of E
- Ac is a set of actions
- see: $E \rightarrow Per$
- $action$: $Per \rightarrow Ac$
- do: $Ac \times E \rightarrow E$

Per represents the *perception* of the environment from the agent's point of view. An agent observes the environment (see), selects the appropriate action ($action$), and acts (do) on the environment itself[13].

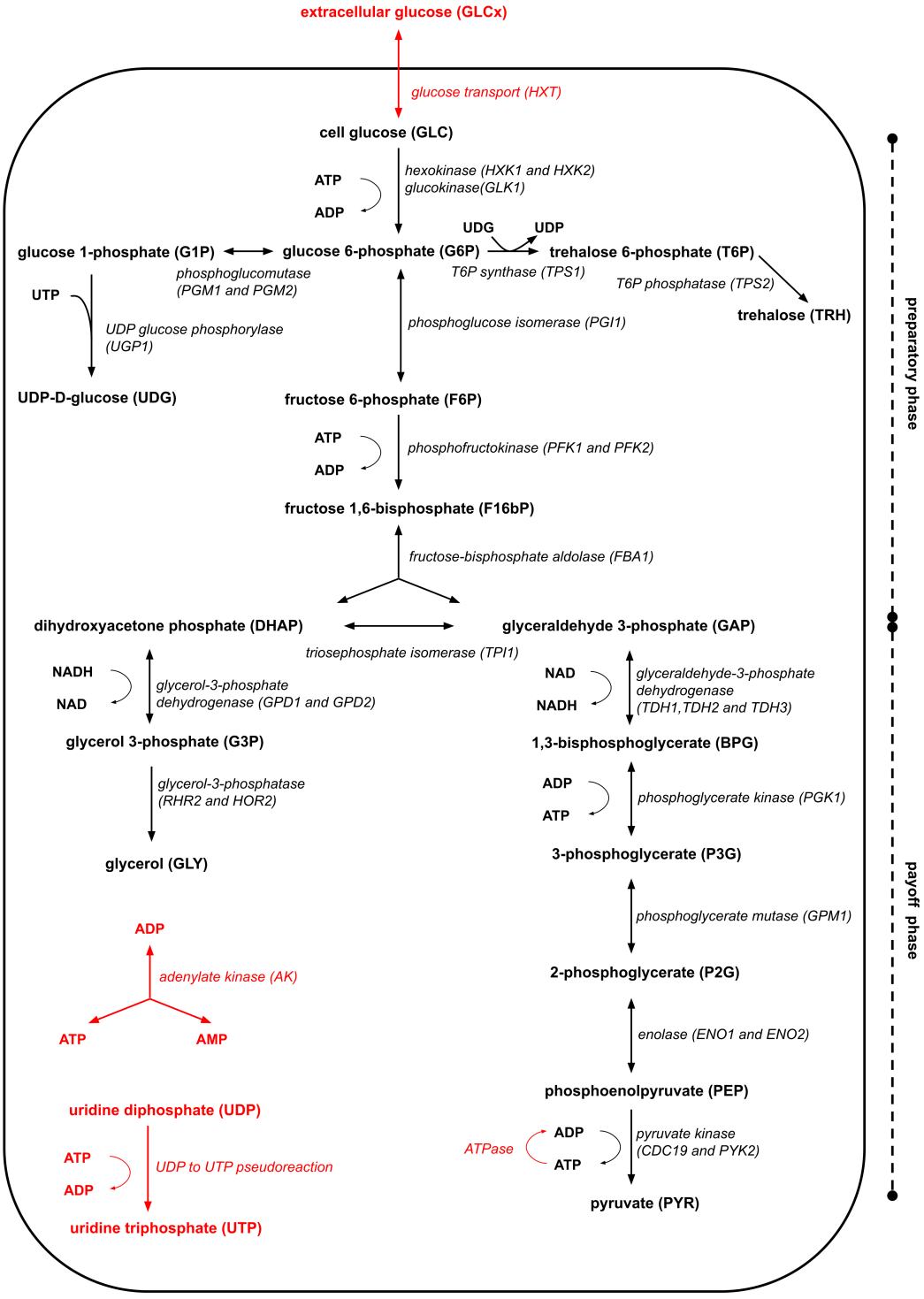

Figure 2. Schematic representation of the glycolysis steps and branches taken into account in our ABMs. They are extracted and adapted, through a dedicated interface of Orion, from the SBML of the Smallbone2013 kinetic model[19]; the reactions in red are those excluded during the conversion (see the Methods section for details). For each metabolite involved, we report both the name and the acronym (in bold), while, for every reaction, we indicate the abbreviation of each isoenzyme carrying it out (in italics). On the right side of the image, we highlight the two main phases of the process; since the ethanol fermentation has not been simulated, we preferred to not show this phase, in order to preserve the readability of the figure.

The simulations are performed in the three-dimensional space, which represents a portion of the cytoplasm, that is the environment perceived by the agents. Each molecule is modelled as a sphere, whose radius is estimated from its molecular weight and the average value of the molar specific volume of a protein in solution[32–34]. These modelling choices produce a fairly realistic molecular crowding in the simulated portion of the cytoplasm. Moreover, by making every molecule spherical, we can correlate its shape to its diffusion coefficient through the Stokes-Einstein equation for the Brownian motion of a spherical particle:

$$D = \frac{k_B T}{6\pi\eta r} \tag{1}$$

where, k_B is the Boltzmann constant, T is the temperature, η the viscosity of the environment and r is the radius of the molecule. For our simulations, we set $T = 298.15$ kelvin and $\eta = 0.0011$ pascal-second.

Each molecule can freely move, inside the simulation volume, according to a vector applied to the centre of its sphere: its direction is generated randomly basing on polar coordinates, while its module is calculated from the ambient diffusion coefficient D, obtained via Equation 1, as the average value of the square of the molecule's displacement x in a time t:

$$< x^2 > = 2Dt \tag{2}$$

A dedicated agent monitors the position of all the molecules to ensure that every movement ends in an empty space of the environment, avoiding collisions and overlaps.

The simulator enables to set the space unit and the time scale as per requirement; in this study we considered the angstrom (Å, equivalent to 10^{-10} m) for space and 10^{-4} second for time (corresponding to one tick of the simulation clock). A cube of 1 attolitre (10^{-18} L, having a side of 1000 Å) represents the best option for the aim of our analysis and meets the computational demand of the simulations (Figure 1a shows the 3D space visualised through the interface of the simulator).

The model at the basis of the simulator classifies molecules in three types: *enzymes*, *complexes* and *metabolites*. The property that distinguish enzymes and complexes from metabolites is that the latter are only able to move while the first two classes of molecules can act on the environment.

Although molecular movements are modelled basing on Brownian diffusion, this study pushes forward the capabilities of the agent paradigm by not limiting molecular interactions to just those allowed by random encounters. Indeed, enzymatic reactions are modelled basing the ability of agents to perceive and interact with one another: each enzyme identifies the cognate metabolites in its proximity thanks to a perception-sphere that it projects on the environment. Such an approach is the simulator key-feature that allows studying the effects of the long-distance interactions among biomolecules; for this reason, it will be deepened in the Results section.

Every molecular interaction may lead to the formation of a complex, modelled in the ABM as a new agent. If such a complex represents a saturated enzyme, it waits an amount of time corresponding to the reciprocal of its k_{cat} value and then release the final product (or products) of the reaction; otherwise, it moves and acts on the environment to bind the metabolite needed to reach saturation. This modelling approach is based on the construction of a *reaction automaton*, introduced in a previous work[35] and formalised in Section 2.1 of the Supplementary Information.

Results

Modelling Short and Long-range Forces among Biomolecules

To simulate the effects of the molecular interactions operating at different distances, we endowed agents with perception capabilities specifically designed. Their core property lies on the definition of a *perception sphere* that surrounds each active molecule (enzymes and complexes, as explained in the Methods section). By setting the *perception radius*, that is the radius of the perception sphere, we can model various lengths

at which enzymes and complexes are able to interact with their cognate partners. Therefore, *the capability of agents to perceive and interact with one another allows us to abstract the effects of the electrostatic and electrodynamic potentials among the molecules of the simulated environment*; this can be achieved without taking into account all the physical parameters (such as the actual potential values or the forces generated by molecular collisions) usually required in molecular dynamics simulations[8].

Each perception radius is obtained by summing the radius of the enzyme to the *perception distance* at which we want the enzyme to be able to find a cognate metabolite; the perception distance extends beyond the surface of the sphere representing the enzyme. As the distance of the metabolite from the enzyme increases, the intensity of the forces acting on a metabolite diminishes; for this reason, each perception sphere is characterised by different interaction probabilities, depending on its size (more details on the perception spheres' implementation are provided in Section 2.4 of the Supplementary Information).

We simulated three different systems, in which the interactions characterising the glycolytic process are driven by the specific kinds of forces whose effects on the pathway we aim to compare. Going into details, the agent-based modelling approach makes us able to define:

- A system in which molecular encounters are driven only by the Brownian motion and dynamic complementarities (e.g., lock-and-key or induced-fit phenomena). We modelled this system allowing enzymes and complexes to identify a cognate metabolite within a *perception distance of 5 Å*; this sets the space on which electrostatic forces, such those resulting from a van der Waals-like potential, operate.
- A system where a *10 Å perception distance* models the effects of electromagnetic potentials limited by the Debye screening[3]; it restricts the interactions to just those allowed by stochastic short-range forces.
- A system characterised by *perception distances of 300 Å*, chosen as the average length to simulate the existence of long-range electrodynamic forces among biomolecules (considering that the size of the simulation volume of our study is 1000 cubic angstroms). As mentioned in the Introduction, these are deterministic attractive forces activated by a long-range potential between two dipolar molecules A and B, if they vibrate at frequencies $\omega_A \simeq \omega_B$ (that is, if they are at resonance). In real cells, this phenomenon might be observed because a macromolecule oscillating at a high frequency (in the range of $10^{10} - 10^{11}$ Hz) does not suffer Debye screening effect by the ions of the medium[6,7].

In Fig. 1b, we provide a graphical representation of how the perception radii project on the environment. Each of them is given as the sum of the enzyme radius (`r`) and the perception distance (`d`) at which the molecule is able to detect its cognate metabolites. The perception radius of the `r1+d1` type schematises the constraint that limits the enzyme interactions to those allowed by short-range forces (both *5 and 10 Å perception distances*), while a `r2+d2` type radius models the effects of the long-distance electrodynamic interactions. We point out that, the figure arranges side by side two different types of radii just for comparative purposes; in the ABMs defined for glycolysis, *only one type of radius is allowed per modelled system.*

Outcome of the Simulations

By setting the local rules that determine movements and interactions of the molecules involved in our model of yeast glycolysis (as detailed in the Methods section), the global behaviour of the pathway can be observed, during the simulation, in the form of molecular concentration changes (mmol/l) over time (s).

To balance the computational demand of dealing with thousands of molecules and the need of producing worthwhile outputs, we run each type of simulation for an interval of 1 second (about ten days of actual running time); it turned out to be sufficiently long for us to observe and compare the specific features of each of the three modelled systems.

In Fig. 3, we report some of the concentration changes characterising each type of system. For generating these plots, we selected the metabolites whose amount variations during the simulation has

been accounted as the most meaningful for our analysis (a complete set of plots, covering all the species considered in our models, are provided in Section 3 of the Supplementary Information).

The simulation performed by setting *perception distances of 300 Å*, which represents a system where we hypothesise the existence of selective long-range molecular recruitments, has the highest reactivity and efficiency (Figure 3a); already after 0.9 s, all the `glucose` in the environment has been consumed and the `pyruvate` (one of the main products of the pathway) increases from an initial concentration of 0.2 mmol/l to about 1 mmol/l.

In the system where we limited the electromagnetic forces to those below the Debye screening (*perception distances of 10 Å* – Fig. 3b), we do not observe utterly different concentration changes in comparison to the previous one; however, they clearly show a lower efficiency in the production of `pyruvate` from glucose, which a system of that type is unable to completely deplete in the chosen simulation interval.

These two types of simulations can also be compared in terms of variations of the other main products of glycolysis and related branches. Indeed, they both report a clear, yet similar, increase of the `glycerol` amount. Conversely, if we take into account the effects of long-distance interactions, the `trehalose` branch shows a change of its end-product from 0.015 to 0.76 mmol/l, a concentration 56% higher compared the 0.48 mmol/l resulting from a system limited by a 10 Å perception distance. Regarding `ATP` and `NADH`, their concentrations reach almost immediately a value close to zero and then oscillate around it during the entire simulation. This behaviour is observable due to the short interval of glycolysis we are analysing: at this stage of the process, the reactions that use `ATP` as energy donor, as well as the redox conversion of `dihydroxyacetone phosphate` to `glycerol 3-phosphate`, which is coupled with the oxidation of `NADH`, still have abundance of substrate to consume; as a consequence, the related enzymes continuously bind the `ATP` and `NADH` in the environment to perform their catalytic activity.

We obtained remarkably opposed results in the case of simulations based on *perception distances of 5 Å*, which model a system affected only by short-range van der Waals-like potentials (Figure 3c). Despite the certainty that a metabolite will be bound by its cognate enzyme when it enters a so small perception sphere (as detailed in Section 2.4 of the Supplementary Information), at the end of the simulation we can observe negligible increases in the concentration of the the pathway end-products as well as in the consumption of `glucose`. In particular, the curve representing this last metabolite reaches a plateau after a small depletion in its concentration, a behaviour we would observe at steady state; however, mostly because we did not implement enzyme regulation and glucose transport, such a condition is unlikely in our simulated systems. We can also observe similar concentration changes for `glycerol` (in this case, it increases before reaching a plateau). In both the situations, these anomalous behaviours are explainable if we observe the curves of `ATP` and `NADH`: the amount of `ATP` never decreases, because, during the preparation phase of glycolysis, neither the `hexokinases` nor the `phosphofructokinases` are able to bind this molecule and complete the catalysis of their respective reactions. Indeed, `glucose` molecules are bound at the beginning of the simulation, but then the environment maintains the same concentration of `hexokinase-glucose` and `glucokinase-glucose` complexes for the entire simulated interval (plots reporting the concentrations changes of complexes are provided in Section 3.2 of the Supplementary Information). This phenomenon also justifies why `fructose 1,6-bisphosphate` (product of the `phosphofructokinase`) can only decrease, consumed by the `fructose-bisphosphate aldolase`. `NADH`, instead, remains stable at its initial concentration of 0.086 mmol/l, because the `glycerol-3-phosphate dehydrogenase` is not able to bind it; when all the `glycerol 3-phosphate` in the environment has already been converted in `glycerol`, the latter can no longer be produced (causing the observed plateau of its curve).

We compared the results described so far with the output obtained through a numerical time-course simulation; this has been performed via the Copasi software[36] over the Smallbone2013 model[19]. We modified the original SBML with the same changes applied to our ABM, relating to the reactions accounted and initial molar concentrations (see Methods). However, we left unchanged the functions associated to enzymes' regulation, because a system of differential equation resulted to be less flexible than an ABM and removing this feature would have compromise its consistency, making the numerical simulation impossible.

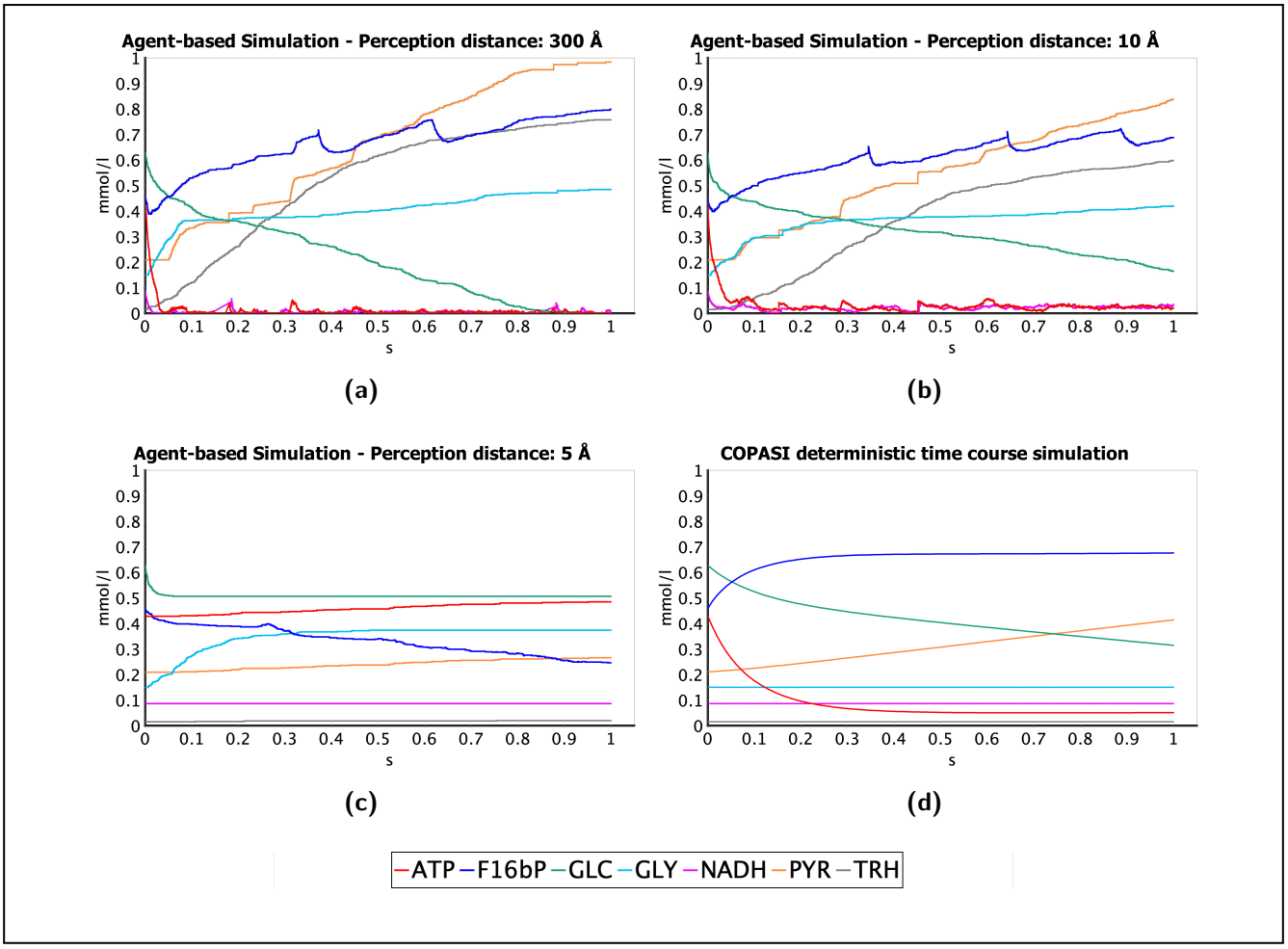

Figure 3. Concentration changes over time, in simulations of 1 second, of a selection of metabolites particularly relevant for our study (for the complete set of plots, representing all the metabolites simulated, see Section 3.1 of the Supplementary Information). Through this figure, we provide a comparison of the plots generated by three agent-based simulations – with perception distances set to 300 Å **(a)**, 10 Å **(b)** and 5 Å **(c)** respectively – and by a deterministic time course simulation based on the Smallbone2013 kinetic model **(d)**. The selected metabolite species are: `glucose` (`GLC`), the source of the glycolytic pathway; `pyruvate` (`PYR`), `NADH` and `ATP`, the end products of glycolysis; `trehalose` (`TRH`) and `glycerol` (`GLY`), respectively, the products of the two main glycolysis branches; `fructose 1,6-bisphosphate` (`F16bP`), the product of the most important control-point reaction of the glycolytic pathway, that is the one catalysed by the `phosphofructokinase`. In plot **(a)**, it is possible to notice how the simulation that takes into account long-range electrodynamic forces (300 Å perception distance) also shows a higher reactivity and an evident increase in the amounts of the pathway end-products. In comparison, the simulation that limits the electromagnetic forces to those affected by the Debye screening (10 Å perception distance), shown in **(b)**, is not able to consume the whole `glucose` in the environment and generates significantly smaller amounts of `pyruvate` and `thralose`. Simulating a system driven by van der Waals-like potentials (5 Å perception distance), whose plot is represented in **(c)**, causes negligible changes in metabolite concentrations and the `glucose` consumption reaches a plateau; the agent-based approach allows us to attribute this behaviour to the inability of the reactions that use `ATP` or `NADH` as energy donor to bound these types of metabolites (see the Results section for further details). The plot **(d)** is generated through the deterministic time-course simulation of the Smallbone2013 model using the software Copasi[36].

As shown in Fig. 3d, the kinetic model thus generates results closer to a steady state condition, a property that, at a first glance, may mislead the observer to find analogies with the simulations accounting 5 Å perception distances. However, excluding the fluctuations of metabolites' concentrations, which are better captured by the ABM and more evident in the related plots, most of the shown concentration changes are loosely similar to those identified when we simulated 10 Å and 300 Å perception distances. This can be verified at least for the consumption of `glucose` and `ATP`, and for the increase of `pyruvate` and `fructose 1,6-bisphosphate`; nonetheless, they show significantly smaller variations from their initial molar concentrations. Although we consider noteworthy identifying such properties in some of the most relevant species of the pathway, we also point out that the last observations do not apply to all the simulated metabolites (as explained in Section 3.1 of the Supplementary Information).

Discussion

The outcomes of the agent-based simulations detailed above suggest that the two systems reproducing an off-resonance situation, where molecular interactions rely only on van der Waals-like potentials or, at least, on electromagnetic forces shorter than the Debye length, are not able to oxidise glucose at a high rate. This property is particularly true when we limit the *perception distance* to 5 Å, resulting in negligible changes in metabolite concentrations. By analysing the complexes formed by specific enzymes, such as `hexokinases` and `phosphofructokinases`, we attributed this behaviour to the inability of the electrostatic forces to guarantee the interaction of these enzymes with the needed energy donors. In this regard, the agent-based approach shows one of its major capabilities: it reproduces the dynamics of the local interactions among the molecules (modelled as autonomous agents) and "captures" the formation of complexes, even when they are made by partly-saturated enzymes.

Such a possibility allowed us to observe, in the system limited by short-range electrostatic interactions, a condition that might be detrimental for the cell anaerobic metabolism, which commits the production of energy (in the form of `ATP`) only to a fast-paced glycolytic process. At the current stage of our work, this consideration represents just a hypothesis: in real cells, glycolysis' processes take place in times ranging from few seconds to hours[25,26,37,38], making our one-second interval of simulation just a testbed to validate the capability of ABMs to support the study of the above-described forces in biological systems. However, if confirmed by further analyses, this result might suggest the non-feasibility of the *lock-and-key* model for enzymes in metabolic processes.

Interestingly, despite the fact that enzymes' regulation has not been modelled, the systems driven by electromagnetic forces (included those below the Debye screening length) produce oscillatory-like fluctuations in the concentrations of `fructose 1,6-bisphosphtate`, the main product of `phosphofructokinase`. Moreover, as shown in Fig. 4, these fluctuations are synchronised with the concentration changes of `DHAP` and `GAP`, the products of the subsequent reaction in the glycolytic pathway, especially due to its reversibility. Conversely, such a behaviour is almost unnoticeable in the output of the simulation that allows only short-range van der Waals-like potentials (5 Å perception distance). `Phosphofructokinase` has a central role in the regulation of glycolysis and, pivoting around this enzyme, an oscillatory behaviour has been experimentally observed during the oxidation of glucose (even if at much lower frequencies)[38,39]. Considering the high level of abstraction of the current glycolysis ABM, this result might be considered another clue that, by not limiting molecular interactions to just shape complementarities and chemical affinities, we generated processes more faithful to those occurring in cellular glycolysis.

We could not have reached such a conclusion if we based our analysis on a standard kinetic model, which derives the changes over time of the concentrations (often of just the metabolites) through rate and balance equations: lacking the ability to represent the granularity of a molecular system, this approach hardly grasps the fluctuations in the species' amounts, generating several discrepancies with the results we gained through our agent-based approach. Although differential equations best suit modelling the continuum or macroscale level[14], such divergences might also be attributed to the possible inaccuracy

through which kinetic parameters are essayed *in vitro*. Indeed, already in the early 2000, Teusink et al. questioned that the *in vitro* kinetics could be able to faithfully describe an *in vivo* behaviour[10].

Molecular Dynamics simulations, which may not be affected by the limitations of the standard kinetic approach, require a high number of physical parameters to be performed. Numerical simulations of this kind have been carried out to detect long-range interactions among biomolecules through the molecular diffusion behaviour[8]. In this case, simulating just one type of protein (the white egg Lysozyme) and one oppositely charged dye (the Alexa Fluor 488), required an *a priori* knowledge of several data; applying the same approach to a complex pathway of many reactions would be significantly more difficult than performing *in silico* studies through agent-based simulations.

We think that the results provided in this article support the reliability of ABMs in capturing the essential features of a complex biological process and faithfully reproduce different aspects of its behaviour, even on the basis of few empirical data. This approach identified in the long-range electrodynamic forces some of the fundamental "ingredients" necessary for glycolysis to function in an efficient way.

With this manuscript, however, we just laid the groundwork for further *in silico* and experimental studies that would explore those aspects of metabolism's dynamics overlooked at the current stage of our analysis. An optimised implementation of Orion would allow longer simulations that, complemented by an experimental validation of the present results, might highlight if some of our outcomes would be biased by the abstraction level of the agent-based models. Once reinforced the robustness of this approach, it might pave the way for a better comprehension of those phenomena associated to cellular metabolism that are still not well understood. As an example, it can be applied in the study of the Warburg effect, which describes the preference of cancer cells for the anaerobic (and energetically inefficient) consumption of glucose through glycolysis, even in presence of a high oxygen concentration[40]. Recent studies have linked such a process to the effect of glycolytic oscillations[41] and to the rate of glycolysis, increased to provide a selective advantage over metabolic competition in the tumour environment[42]. In this paper, we have shown how the long-range electrodynamic forces may affect rate and efficiency of glucose oxidation, as well as the oscillations in glycolysis intermediates; therefore, additional studies might enlighten their potential involvement in such an anomalous behaviour of tumour cells.

Similar results might also be reached by empowering the capabilities of the agent-based approach with methods from other disciplines. Among them, the topological data analysis, already used to better understand enzymatic reactions through ABMs[35], may provide the current model with a many-body perspective. Shape calculus and hierarchical structures can be applied to represent molecular conformations and increase the accuracy of the interactions between cognate-partners; these approaches would allow modelling the geometry of molecules' shapes and collisions with a higher precision[43,44]. Moreover, we can better capture the collective synchronisation properties of a population of molecules behaving as coupled oscillators by using BOSL, the biological oscillators synchronisation logic[45]. Putting efforts in these directions might provide a new standpoint in our comprehension of molecular interactions and disclose aspects of biological systems that are still unexplored.

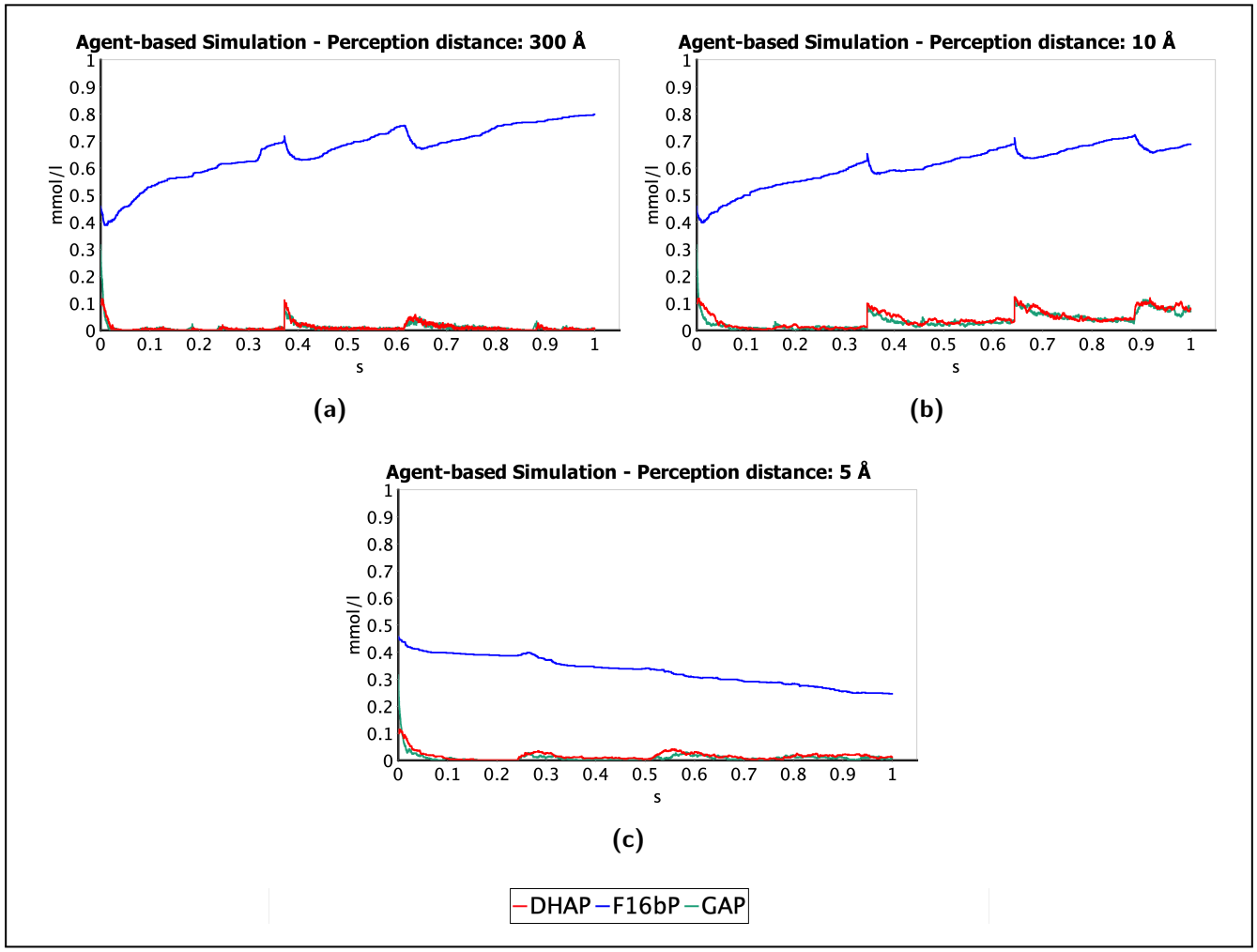

Figure 4. Synchronised oscillation-like fluctuations observed in `fructose 1,6-bisphosphate` (`F16bP`), `dihydroxyacetone phosphate` (`DHAP`) and `glyceraldehyde 3-phosphate` (`GAP`). The first metabolite is the product of the phosphorylation of `fructose-6-phosphate`, catalysed by `phosphofructokinase`, while the other two are generated by the subsequent reaction in the glycolytic pathway, carried out by `fructose-bisphosphate aldolase`. `DHAP` and `GAP` are also interconverted by the `triosephosphate isomerase`. In **(a)** and **(b)**, that is, the plots of the simulations that take into account the electromagnetic forces (limited or not by the Debye screening), we can observe a marked oscillation-like trend with a frequency of about 2.8 s^{-1}, synchronised in all the three curves. Conversely, in the simulation that considers just short-range electrostatic interactions, shown in the plot **(c)**, these oscillations are almost unnoticeable. The higher frequency measured experimentally in yeast's glycolysis[38] is of 30 s^{-1}; therefore, at the time scale of our simulations, these can be considered more as micro-oscillations, which give us a clue of the higher faithfulness to the real glycolytic process of the models whose interactions are not limited to just random encounters and chemical affinities.

Data Availability

All data generated or analysed during this study are included in this published article (and its Supplementary Information files). The SBML of the Smallbone2013 - Iteration 18 model is accessible at http://identifiers.org/biomodels.db/MODEL1303260018; the modified version of this SBML, generated for running the time-course simulations with Copasi, is available at https://bit.ly/orion-simulator.

Code availability

The simulations described in this manuscript have been performed with the agent-based simulator Orion 2.0.0. The simulator has been developed at the University of Camerino as a non-open and commercially-oriented software. For this reason, the source code is not openly released; however, we agree to make it available, to editors and reviewers, only upon request and exclusively for the purpose of reviewing this manuscript. To guarantee the methods reproducibility, a demo version of the simulator is accessible at https://bit.ly/orion-simulator.

References

1. McLachlan, A. A variational solution of the time-dependent Schrodinger equation. *Mol. Phys.* **8**, 39–44, DOI: 10.1080/00268976400100041 (1964).
2. Stephen, M. J. First-Order Dispersion Forces. *The J. Chem. Phys.* **40**, 669–673, DOI: 10.1063/1.1725188 (1964).
3. Cherstvy, A. G., Kolomeisky, A. B. & Kornyshev, A. A. Protein-DNA Interactions: Reaching and Recognizing the Targets. *J. Phys. Chem. B* **112**, 4741–4750, DOI: 10.1021/jp076432e (2008).
4. Painter, P. C., Mosher, L. & Rhoads, C. Low-frequency modes in the raman spectrum of DNA. *Biopolymers* **20**, 243–247, DOI: 10.1002/bip.1981.360200119 (1981).
5. Fischer, B. M., Walther, M. & Jepsen, P. U. Far-infrared vibrational modes of DNA components studied by terahertz time-domain spectroscopy. *Phys. Med. Biol.* **47**, 3807–3814, DOI: 10.1088/0031-9155/47/21/319 (2002).
6. Preto, J., Floriani, E., Nardecchia, I., Ferrier, P. & Pettini, M. Experimental assessment of the contribution of electrodynamic interactions to long-distance recruitment of biomolecular partners: Theoretical basis. *Phys. Rev. E* **85**, 041904, DOI: 10.1103/PhysRevE.85.041904 (2012).
7. Preto, J., Pettini, M. & Tuszynski, J. A. Possible role of electrodynamic interactions in long-distance biomolecular recognition. *Phys. Rev. E* **91**, 052710, DOI: 10.1103/PhysRevE.91.052710 (2015).
8. Nardecchia, I. *et al.* Detection of long-range electrostatic interactions between charged molecules by means of fluorescence correlation spectroscopy. *Phys. Rev. E* **96**, 022403, DOI: 10.1103/PhysRevE.96.022403 (2017).
9. Westerhoff, H. V. *et al.* From Silicon Cellsilicon cell to Silicon Humansilicon human. In Booß-Bavnbek, B., Klösgen, B., Larsen, J., Pociot, F. & Renström, E. (eds.) *BetaSys*, 437–458, DOI: 10.1007/978-1-4419-6956-9_19 (Springer New York, New York, NY, 2011).
10. Teusink, B. *et al.* Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? Testing biochemistry: Do we understand yeast glycolysis? *Eur. J. Biochem.* **267**, 5313–5329, DOI: 10.1046/j.1432-1327.2000.01527.x (2000).
11. Berry, H. Monte Carlo Simulations of Enzyme Reactions in Two Dimensions: Fractal Kinetics and Spatial Segregation. *Biophys. J.* **83**, 1891–1901, DOI: 10.1016/S0006-3495(02)73953-2 (2002).
12. Takahashi, K., Arjunan, S. N. V. & Tomita, M. Space in systems biology of signaling pathways - towards intracellular molecular crowding in silico. *FEBS Lett.* **579**, 1783–1788, DOI: 10.1016/j.febslet.2005.01.072 (2005).

13. Genesereth, M. R. & Nilsson, N. J. *Logical Foundations of Artificial Intelligence* (Morgan Kaufmann, Los Altos, Calif, 1987).

14. Cannata, N., Corradini, F., Merelli, E. & Tesei, L. Agent-Based Models of Cellular Systems. In Reisfeld, B. & Mayeno, A. N. (eds.) *Computational Toxicology*, vol. 930 of *Methods in Molecular Biology*, 399–426, DOI: 10.1007/978-1-62703-059-5_18 (Humana Press, Totowa, NJ, 2013).

15. Belenchia, M. *et al.* Agent-Based Learning Model for the Obesity Paradox in RCC. *Front. Bioeng. Biotechnol.* **9**, 642760, DOI: 10.3389/fbioe.2021.642760 (2021).

16. Norton, K.-A., Gong, C., Jamalian, S. & Popel, A. Multiscale Agent-Based and Hybrid Modeling of the Tumor Immune Microenvironment. *Processes* **7**, 37, DOI: 10.3390/pr7010037 (2019).

17. Brazier, F., Jonker, C. & Treur, J. Compositional design and reuse of a generic agent model. *Appl Artif. Intell.* **14**, 491–538, DOI: 10.1080/088395100403397 (2000).

18. Dastani, M., Arbab, F. & de Boer, F. Coordination and composition in multi-agent systems. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems - AAMAS '05*, 439, DOI: 10.1145/1082473.1082540 (ACM Press, The Netherlands, 2005).

19. Smallbone, K. *et al.* A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes. *FEBS Lett.* **587**, 2832–2841, DOI: 10.1016/j.febslet.2013.06.043 (2013).

20. Hucka, M. *et al.* The systems biology markup language (SBML): A medium for representation and exchange of biochemical network models. *Bioinformatics* **19**, 524–531, DOI: 10.1093/bioinformatics/btg015 (2003).

21. World Wide Web Consortium (W3C). Extensible Markup Language (XML) 1.0 (Fifth Edition) (2013).

22. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515, DOI: 10.1093/nar/gky1049 (2019).

23. Hastings, J. *et al.* ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res* **44**, D1214–D1219, DOI: 10.1093/nar/gkv1031 (2016).

24. Guimarães, P. M. R. & Londesborough, J. The adenylate energy charge and specific fermentation rate of brewer's yeasts fermenting high- and very high-gravity worts. *Yeast* **25**, 47–58, DOI: 10.1002/yea.1556 (2008).

25. Laurent, M., Seydoux, F. J. & Dessen, P. Allosteric regulation of yeast phosphofructokinase. Correlation between equilibrium binding, spectroscopic and kinetic data. *J. Biol. Chem.* **254**, 7515–7520 (1979).

26. van den Brink, J. *et al.* Dynamics of Glycolytic Regulation during Adaptation of Saccharomyces cerevisiae to Fermentative Metabolism†. *AEM* **74**, 5710–5723, DOI: 10.1128/AEM.01121-08 (2008).

27. De la Fuente, I. M. & Cortes, J. M. Quantitative Analysis of the Effective Functional Structure in Yeast Glycolysis. *PLoS ONE* **7**, e30162, DOI: 10.1371/journal.pone.0030162 (2012).

28. Berg, J., Tymoczko, J. & Stryer, L. The Glycolytic Pathway Is Tightly Controlled. In *Biochemistry* (New York: W H Freeman, 2002), 5th edition edn.

29. Jurica, M. S. *et al.* The allosteric regulation of pyruvate kinase by Fructose-1,6-Bisphosphate. *Structure* **6**, 195–210, DOI: 10.1016/S0969-2126(98)00021-5 (1998).

30. Angeletti, M. *et al.* Spatial behavioral modeling and simulation of metabolic pathways with Orion. In *IV Bioinformatics ITalian Society Meeting (BITS 2007)*, 70 (Napoli, Italy, 2006).

31. Cannata, N., Corradini, F. & Merelli, E. Multiagent modelling and simulation of carbohydrate oxidation in cell. *Int J Model. Identif. Control.* **3**, DOI: 10.1504/IJMIC.2008.018191 (2008).

32. Richards, F. M. Areas, volumes, packing and protein structure. *Annu. Rev. Biophys. Bioeng.* **6**, 151–176, DOI: 10.1146/annurev.bb.06.060177.001055 (1977).

33. Zamyatnin, A. Protein volume in solution. *Prog. Biophys. Mol. Biol.* **24**, 107–123, DOI: 10.1016/0079-6107(72)90005-3 (1972).

34. Harpaz, Y., Gerstein, M. & Chothia, C. Volume changes on protein folding. *Structure* **2**, 641–649, DOI: 10.1016/S0969-2126(00)00065-4 (1994).

35. Piangerelli, M., Maestri, S. & Merelli, E. Visualising 2-Simplex formation in metabolic reactions. *J. Mol. Graph. Model.* **97**, 107576, DOI: 10.1016/j.jmgm.2020.107576 (2020).

36. Hoops, S. *et al.* COPASI–a COmplex PAthway SImulator. *Bioinformatics* **22**, 3067–3074, DOI: 10.1093/bioinformatics/btl485 (2006).

37. Nielsen, K., Sørensen, P., Hynne, F. & Busse, H.-G. Sustained oscillations in glycolysis: An experimental and theoretical study of chaotic and complex periodic behavior and of quenching of simple oscillations. *Biophys. Chem.* **72**, 49–62, DOI: 10.1016/S0301-4622(98)00122-7 (1998).

38. Richard, P. The rhythm of yeast. *FEMS Microbiol Rev* **27**, 547–557, DOI: 10.1016/S0168-6445(03)00065-2 (2003).

39. Wolf, J. *et al.* Transduction of Intracellular and Intercellular Dynamics in Yeast Glycolytic Oscillations. *Biophys. J.* **78**, 1145–1153, DOI: 10.1016/S0006-3495(00)76672-0 (2000).

40. Warburg, O. The Metabolism of Carcinoma Cells. *The J. Cancer Res.* **9**, 148–163, DOI: 10.1158/jcr.1925.148 (1925).

41. Rietman, E. A. *et al.* An integrated multidisciplinary model describing initiation of cancer and the Warburg hypothesis. *Theor Biol Med Model.* **10**, 39, DOI: 10.1186/1742-4682-10-39 (2013).

42. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* **41**, 211–218, DOI: 10.1016/j.tibs.2015.12.001 (2016).

43. Buti, F., Cacciagrano, D., Corradini, F., Merelli, E. & Tesei, L. BioShape: A spatial shape-based scale-independent simulation environment for biological systems. *Procedia Comput. Sci.* **1**, 827–835, DOI: 10.1016/j.procs.2010.04.090 (2010).

44. Quadrini, M., Daberdaku, S. & Ferrari, C. Hierarchical Representation and Graph Convolutional Networks for the Prediction of Protein–Protein Interaction Sites. In Nicosia, G. *et al.* (eds.) *Machine Learning, Optimization, and Data Science*, vol. 12566, 409–420, DOI: 10.1007/978-3-030-64580-9_34 (Springer International Publishing, Cham, 2020).

45. Bartocci, E., Corradini, F., Merelli, E. & Tesei, L. Detecting synchronisation of biological oscillators by model checking. *Theor. Comput. Sci.* **411**, 1999–2018, DOI: 10.1016/j.tcs.2009.12.019 (2010).

Acknowledgements

This article is the result of the research project funded by the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme (FP7) for Research of the European Commission, under the FET-Proactive grant agreement TOPDRIM (`www.topdrim.eu`), number FP7-ICT- 318121.

Author contributions statement

EM and MP designed and supervised the research. SM implemented the method, performed the research and wrote the paper. All authors revised the paper and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SciRepSupplementaryInformation.pdf