Comparison of rule- and ordinary differential equation-based dynamic model of DARPP-32 signalling network

Preprint, compiled March 26, 2022

Emilia M. Wysocka[1], Matthew Page[2], James Snowden[2], and T. Ian Simpson [1]

[1]School of Informatics, University of Edinburgh, Edinburgh
[2]UCB Celltech, 208 Bath Road, Slough, Berkshire SL1 3WE

Abstract

Dynamic modelling has considerably improved our understanding of complex molecular mechanisms. Ordinary differential equations (ODEs) are the most detailed and popular approach to modelling the dynamics of molecular systems. However, their application in signalling networks, characterised by multi-state molecular complexes, can be prohibitive. Contemporary modelling methods, such as rule-based (RB) modelling, have addressed these issues. Although the advantages of RB modelling over ODEs have been presented and discussed in numerous reviews, no direct comparison of the time courses of a molecular system encoded in the two frameworks has been made before. To make such a comparison, a set of reactions that underlie an ODE model by Fernandez et al. [1] was manually encoded in the Kappa language, one of the RB frameworks. A comparison of the models was performed at the level of model specification and results were acquired through model simulations. We found that the Kappa model recapitulated the general dynamics of its ODE counterpart with minor differences. These differences occur whenever molecules have multiple sites binding the same interactor. The notation of such rules requires a complete listing of all possible binding configurations. Furthermore, activation of these molecules in the RB model is slower than in the ODE one but can be corrected by revision of the rate constants used in the relevant rules. We conclude that the RB representation offers a more expressive and flexible syntax that eases access to fine-grain details of the model, facilitating model reuse. In parallel with these analyses, this manuscript reports a refactored model of a DARPP-32 interaction network that can serve as a canvas for the development of a more complex interaction network to study this particular molecular system.

Keywords modeling molecular dynamics · rule-based modelling · ordinary differential equations · DARPP-32 · dopamine-dependent synaptic plasticity · Kappa

1 Introduction

Computational dynamic modelling frameworks probe mechanistic and quantitative aspects of molecular interactions, which can grant the development of mechanism-based therapies with more predictive power on outcomes of therapeutic interventions [2]. Such interventions often target molecular signalling [3–5], characterised as a complex system of coupled interacting components, often proteins, forming networks that activity leads to non-additive effects [6]. Defining molecular reactions as a set of coupled ODEs has traditionally enabled dynamic modelling of molecular pathways [7]. ODE-based modelling is a powerful and acclaimed formalism with a long tradition. It has extensive established standards, and various software tools that support and facilitate the formulation and analysis of ODE models [7–9]. However, the requirement to explicitly enumerate the molecular species in ODE-based modelling precludes modelling of molecules that assemble into multivalent protein complexes; typically these have multiple functionally divergent states, a common characteristic of molecules involved in cell signalling [10]. The current development of modelling methods derived from computer science has mainly addressed expressivity and the growing complexity of the systems to be modelled. An example of such a method is RB modelling, designed to model interacting proteins. The potential of the method has been extensively discussed [11–14] and demonstrated with examples [11] or with models answering new biological questions [13, 15–17]. These models are often founded on existing ones, which in the great majority are categorised as ODEs. However, to the authors' knowledge, any direct and systematic comparison of the same biological system defined in the light of two formalisms has not been presented before. Such a comparison could reveal the impact of the differences between the two types of model and provide a better understanding of RB modelling. For instance, ODE-based modelling represents a molecular system as concentrations of molecular species and focuses on their reaction kinetics. In contrast, RB modelling is an agent-centred method in which the unfolding of molecular compositions can be studied alongside their abundances [11]. As it is still a relatively new paradigm, we need to further develop RB modelling research, tools, and applications within the modelling community to better understand the strengths and weaknesses of the two different approaches.

This study compares an existing ODE model to a new RB model, both based on the same reaction network. We specifically ask: (1) whether the dynamics of an ODE-based model can be replicated with an RB model? (2)

if differences between the two are observed, what is the underlying cause? (3) if the dynamics of the system defined by ODEs are replicated by the RB method, what advantages are there to using an RB model? We present the results of the models' comparison at the level of notation and using model dynamics under different conditions. The advantages and disadvantages of the two model representations are discussed, alongside suggestions for future research.

2 Background

2.1 *Objectives of dynamic modelling*

Dynamic computational modelling frameworks consist of model specification schema and simulation methods. The model specification is a set of equations or instructions written in a machine-interpretable language. These languages are based on mathematical formalisms that define relationships between variables whose quantities vary over time. Models are run as numerical simulations using algorithms that interpret the model's mathematical formulation and calculate changes in the quantities of model variables. By adopting a suitable level of abstraction and with the use of a sufficiently expressive language, systems can be modelled such that experimentally-derived evidence can be incorporated to improve the model's quality. A formal approach to the generation of models is desirable to (i) encode facts in an unambiguous and explicit manner, (ii) facilitate the understanding of models, (iii) allow easy modification of models to accommodate more than one hypothesis, (iv) aid interpretation of the underlying biological phenomena, and (v) provide a standard approach to the integration of novel data sources [6]. These features are especially important because model generation often requires knowledge spanning multiple disciplines; the existence of formal modelling frameworks enforces a common understanding of the explicit meaning of model components.

2.2 *Modelling signalling systems with ordinary differential equations*

Ordinary differential equations (ODEs) are defined according to a set of chemical reactions that describe how reactants turn into products over time [19] (Figure 1). ODE-based models are routinely solved with numerical procedures, e.g. the Runge-Kutta method [19]. Time courses obtained by solving ODE models are continuous and deterministic, characterised by smooth and gradual change of species concentrations over time [20], and identical simulation outcomes if initial molecular states and input parameters are unchanged. This setup does not reflect the actual characteristics of subcellular events driven by random collisions between discrete molecules [21]. Nevertheless, the deterministic and continuous approach is correct as long as abundances of reactants are large enough to render random fluctuations as negligible [22]. Molecular processes in synapses are characterised by high numbers of divergent types of molecular species but low numbers of instances per type. In these conditions, magnitudes of fluctuations

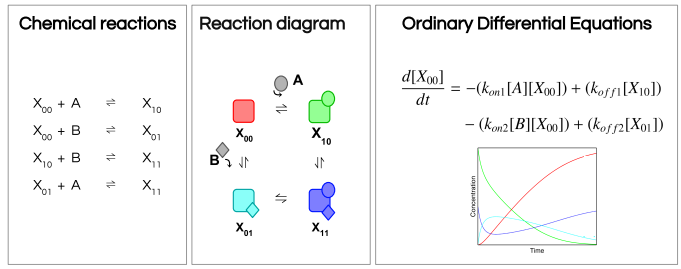

Figure 1: The first step in building a kinetic model of molecular interactions is the definition of *chemical reactions* between molecules. These reactions describe chemical transitions between states of one reactant catalysed by the other. An arrow denotes the direction of the transformation and a doubled arrow reaction reversibility. A set of chemical reactions is denoted with a *reaction diagram*. Quantitative evaluation of model behaviour is achieved by converting coupled chemical reactions (simultaneously solved) to a set of ODEs. Each rate equation expresses the change of concentration of a single molecular species over time, formulated with reaction rates that directly take part in the creation and elimination of this species [7]. On the right-hand side of the figure, a rate equation for X_{00} is defined with 4 rate laws of mass action. The law of mass action states that the speed of reaction is proportional to the product of the concentrations of the reactants [18, p.141]. Each reaction rate is weighted by a *rate constant*, specific to each reaction. Positive and negative signs of reaction rates denote the direction of arrows pointing respectively towards and away of X_{00} [7]. Example adapted from Hlavacek et al. [17].

in copy numbers of molecules become important [22]. Among others, the inherent presence of noise in signalling systems is attributed to transient, low-affinity and promiscuous protein interactions [23]. Variations of this type of protein count can only be captured with stochastic simulation [22]. The current development in standard formats for encoding biological models, in particular Systems Biology Markup Language (SBML), allows to obtain the trajectories of the same model with different simulation methods, both deterministic solvers and stochastic simulators, e.g. Gillespie's Stochastic Simulation Algorithm (SSA) [9, 24]. A more critical shortcoming of ODE-based models applied to signalling networks lies in the requirement for explicit enumeration of all molecular species [11]. This drawback precludes mechanistic modelling of systems that include molecules that can adopt combinatorially complex states for example those with multiple binding partners, e.g. in Epidermal Growth Factor Receptor (EGFR) signalling [25–28]. In such cases only a small fraction of all possible EGFR molecular species can be represented with ODEs [13].

2.3 *Rule-based modelling - formal computational method for simulating molecular interactions*

The development of formal methods in computer science has facilitated their application to the modelling of molecular processes. These approaches have expanded the number of observed properties of biological systems that can be dynamically and quantitatively modelled [29] (reviewed in [29–33]). As this study examines an alter-

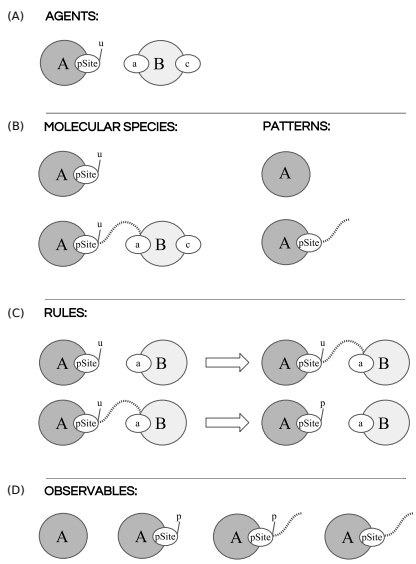

FIGURE 2: Main concepts and elements of the Kappa language (description in the main text).

native to ODEs, the focus lies on non-spatial single-scale mechanistic methods. Among methods that fit these characteristics are *Petri nets* [34], *process algebras* [35–38]), and *rule-based (RB) modelling*. We chose rule-based (RB) modelling as the most suitable framework for representing molecular systems containing components that take part in combinatorial interactions with other molecules and/or have multiple molecular states.

RB modelling, which in some implementations is partially derived from process algebras [39], is based on a formal concept of *graph rewriting*, that represent a molecule-centred view, where molecules are structured graph objects, i.e. nodes with sites. Reactions occurring between molecules are represented as graph transformations, where bonds are formed between the node sites. Graph transformations are encoded as rules that are instructions for local and conditioned transformations. Therefore, a rule represents either a set of reactions or an exact reaction instance, dependent on details of said conditions. By this, a rule can express an infinite number of reactions with a small and finite number of generalised rules [12]. A protein abstracted as a node with binding sites (e.g. protein domains) that have internal states (e.g. phosphorylated) gives a sufficiently expressive system to capture the principal mechanisms of signalling processes (e.g. dissociation, synthesis, degradation, binding, state change, and complex formation [40]), providing insight into site-specific details of molecular interactions (e.g. affinities, dynamics of post-translational modifications, domain availability, competitive binding, causality) and the intrinsic structure of interaction networks. As an example of the RB modelling approach [41], the Kappa framework [39] was chosen as a well-designed modelling environment, where the specification language (Kappa) is separated from the simulator (KaSim). The language is composed of 4 fundamental concepts: agents, patterns, rules and observables (FIGURE 2)[42]. An *agent* represents a molecular entity, e.g. a protein (FIGURE 2A), as a named node (*"A"*, *"B"*) with *interface*, that has a set of *sites* (*"pSite"*, *"a"*, *"c"*). A site has a unique *label* that can carry 2 types of information: *internal state* (*"u"*) and *binding state* [42]. Two or more bound agents form a complex. All agents and complexes at a given point of simulation form *mixture* of existing *molecular species* in a system. Molecular species are agents and complexes that contain a full description of their states and site occupancy (FIGURE 2B). Agents can also have a partially specified interface, i.e. expressed as a *pattern* [43], thereby matching larger groups of molecular species. *Rules* define state transformations or actions of agents, i.e. reactions. Dependent on the generality of agent description, a single rule can represent a set of reactions (FIGURE 2C). On the other hand, a rule specified with complete information about the agent's interface is in one-to-one correspondence with a reaction. *Observables* are desired simulation outputs, as time-courses, that can be represented as patterns or molecular species (FIGURE 2D). An observable that is declared as an incomplete pattern denotes multiple molecular species and these are represented in the time courses as a single summed trajectory. For instance, the second node/observable in FIGURE 2D represents a sum of abundances of the agent "A" that are phosphorylated and in an unknown binding state. The Kappa language also provides means to induce perturbations during the simulation to test experimental hypotheses. Perturbations can be applied once or repeatedly according to the indicated conditions, to update rate constants, or add molecules. *Snapshots* of a molecular mixture, taken during the simulation, provide information on molecular species and their quantities at specified time-points.

KaSim is the Kappa language simulator, based on Gillespie's SSA [21, 24] that numerically simulates individual stochastic trajectories of molecules defined with reaction networks [20, 24]. To model multi-state proteins, the KaSim authors [11, 44] proposed a generalised variant of Gillespie's algorithm that does not enumerate all possible molecular species; therefore, its efficiency is independent of the network size [45].

2.4 *Representative ODE model*

To directly compare models using different formalisms, we need to be able to run model simulations quantifying their molecular abundances over time under comparable conditions. To achieve this, the ODE model must be encoded in a machine-readable format that can be numerically simulated in non-obsolete software, thus satisfying the reproducibility criterion. A model of the immediate interactors of dopamine- and cAMP-regulated neuronal phosphoprotein with molecular weight 32 kDa (DARPP-32) network by Fernandez et al. [1] satisfies this requirement, as it was created with this consideration in mind. The model exists in 2 formats: SBML (attached to the paper and deposited in the BioModels database: BIOMD0000000153 [46]), and E-Cell (version 3). Furthermore, from the point of view of scientific quality the model is considered a valuable study that can serve as a solid core for the construction of larger and more complex models in the community interested

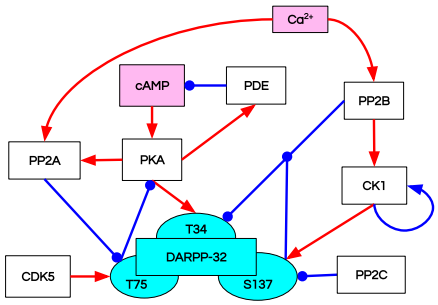

FIGURE 3: Reaction diagrams representing different aspects of the DARPP-32 network included in the ODE model by Fernandez et al. [1] Nodes: DARPP-32 *(cyan)*, second messengers *(magenta)*, kinases/phosphatases *(white)*. Edges: inhibition reactions *(blue)*, activation reactions *(red)*.

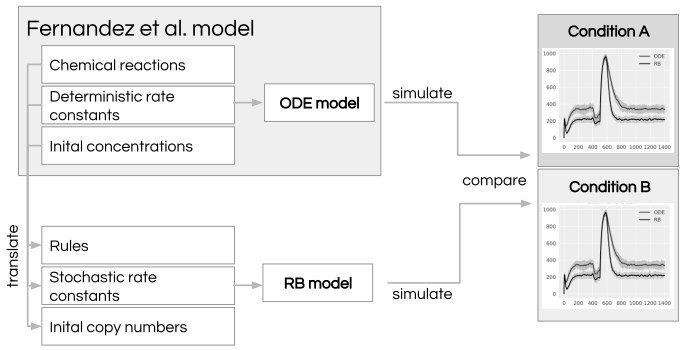

FIGURE 4: Approach to comparison of ODE and RB modelling frameworks.

in the modelling of dopamine-dependent synaptic plasticity [47]. Finally, it is a study widely cited not only by modellers [48–50] but also by experimentalists [51–53].

It ought to be noted that the translation of the model of Fernandez et al. [1] into the Kappa-based framework does not aim to demonstrate that a stochastic model is better to study this particular reaction network. In this case, the ODE-based model is a benchmark against which we can measure other model's performance. In this ODE model the copy number of molecules are sufficiently high that we avoid the previously stated drawback of ODE models when copy numbers are low. Therefore, stochastic modelling in this instance would not be expected to lead to any substantial difference in performance to that of an ODE-model.

DARPP-32 and its interaction network DARPP-32, officially named as Protein Phosphatase 1 Regulatory Inhibitor subunit 1B isoform 1 (PPP1R1B) [54], is an important multistate and intrinsically disordered protein [55] regulating dopamine-dependent synaptic plasticity in medium spiny projection neurons (MSPN) of the striatum. Around 95% of human striatal cells are MSPNs, in which signalling cascades activated simultaneously by glutamatergic and dopaminergic stimuli are a necessary condition for the long-term potentiation (LTP) that underlies context and reward-related learning [56]. Furthermore, DARPP-32 malfunction and abundance relate to multiple central nervous systems (CNS) disorders. Among these are Alzheimer's disease [57], addiction [58], affective disorders [59], and schizophrenia [59, 60]. Glutamatergic and dopaminergic signals are integrated by DARPP-32 which is involved in a complex network of interactions regulating its multiple phosphorylation sites, of which 4 are known to have a regulatory impact on DARPP-32 itself [61]: Threonine 34 (Thr34), Threonine 75 (Thr75), Serine 137 (Ser137), and Serine 102 (Ser102) (positioned on the *Rattus norvegicus* protein sequence). The multiplicity of phosphorylation states leads to numerous interacting partners for DARPP-32 that affect these states. The Threonine sites (Thr34, Thr75) have major regulatory roles in signal processing. The Serine sites (Ser137, Ser102) have a supporting role in Thr34 signal enhancement.

Fernandez et al. [1] studied the integrative effect and sensitivities of DA- and Glu-mediated signals on the DARPP-32 network (FIGURE 3). Their model examined the particular effect of cyclic adenosine monophosphate (cAMP)-pulse followed by calcium ions (Ca^{2+}) spike trains whilst varying the distance between the stimuli. The study showed that DARPP-32 is a robust integrator, indifferent to its initial concentration and delay between the stimuli, far more complex than a bistable switch between DA and Glu signals. To reproduce the system's behaviour, the authors included two main pathways that mediate these signals, cAMP–PKA–DARPP-32 phosphorylated at Threonine 34 (D34) and Ca^{2+}–PP2B–DARPP-32 phosphorylated at Threonine 75 (D75). Contrary to the majority of previous models of Glu and DA signal integration [48, 62–64], DARPP-32 included three phosphorylation sites: Thr34, Thr75 and Ser137. The authors performed two *in silico* mutagenesis experiments modifying the role of Ser137. The first mutation inhibits site phosphorylation by changing Serine to Alanine at the 137 position (Ser137Ala). The second mutation leads to permanent phosphorylation of the Ser137 site (constSer137).

3 MATERIALS AND METHODS

The first step of comparison of RB and ODE frameworks involved encoding reactions underlying the ODE model of Fernandez et al. [1] ("model B") into the Kappa language (3.5 version [65]). Then, models were simulated in different variants to obtain time courses of equivalent observables to compare (FIGURE 4).

3.1 *Model translation*

As a single reaction can be written as a single rule, the translation could have been accomplished in a one-to-one manner. However, to fully benefit from rule patterns, multiple reactions can be condensed into fewer rules by removing irrelevant context, i.e. *decontextualised*. The context of a reaction in the RB model is defined as the information about the agent's binding sites, partners and internal

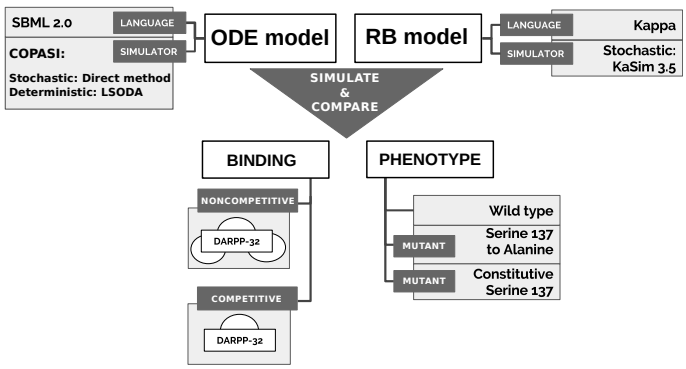

FIGURE 5: Approach to comparisons of time-series.

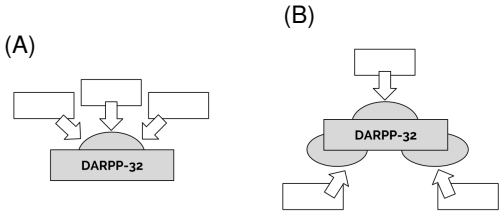

FIGURE 6: Two scenarios of binding patterns illustrate the implications of changing the number of binding sites. In the non-competitive case (A), the phosphorylation sites can bind to their respective kinases and phosphatases. In the other scenario (B), only one site can connect. This leads to competition between the interactors.

states. Based on this definition of reaction context, the following criteria guided decisions about condensing reactions into rules. Given a set of reactions of the same type (forward, backward, or catalytic) between the same reactants (agents), if the difference between reactions lies in agent states (internal or binding) that do not change after the transition from reactants to products, and reaction constants (rates) in all these reactions have the same values, then information about agent states does not define reaction conditions; hence, it can be removed from the reaction notation, i.e. a set of reactions become a rule pattern.

Among the least intuitive cases in encoding reactions into rules are complex substrate activations of PKA and PP2B. PKA is activated by the binding of 4 cAMP molecules, whereas PP2B activation requires 4 Ca^{2+} ions. In other words, multiple molecules of the same type bind substrates on different sites that have to be uniquely named, which requires explicit encoding of all possible binding combinations on four different sides (approach after Danos et al. [66]), called hereafter as *combinatorial binding*.

We also need to translate molecular concentrations, rate constants and initial molecular abundances to copy-numbers [65, 67]. Lastly, the cAMP pulse and the Ca^{2+} spiking are reproduced by the addition of molecular copy numbers and modification of rate constants during the simulation, respectively.

3.2 *Approach to comparison of models*

The visual comparison of the results of two dynamic models required (FIGURE 5) the simulation of both models in a stochastic scheme and the alignment of the trajectory of the corresponding observables under varying conditions to allow more comprehensive comparisons between the frameworks.

Selecting and pairing observables The plots in the original publication show aggregated variables that are summed trajectories of multiple molecular species. For instance, "D34" denotes DARPP-32 phosphorylated at Thr34, regardless of its state of binding or other phosphorylation sites. The concept of aggregated variables corresponds to *observables* in RB modelling, and therefore, we use the term *observable* hereafter to denote aggregated variables. Observables of the ODE model were aggregated here based on their names matching partial strings representing the observables of interest. To verify this approach, obtained observables of the ODE model with this method were visually compared with the six observables plotted in the original publication. The choice of other observables follows these principles: (1) if an agent has internal states, the activated state is set as its observable form, e.g. "CK1u"; (2) if an agent is created and degraded over the simulation, its observable is set to its least specific form, e.g. "PKA"; (3) if an agent is not created and degraded during the simulation, i.e. its level remains constant throughout the simulation and has no internal states, its observable is set to its bound form, e.g. "_CDK5" (see TABLE S1 for the complete list of RB and ODE observables with definitions).

Model simulation We simulate both models in the stochastic scheme but within different simulation environments. The RB model was simulated with KaSim. The SBML format of the ODE model was run with COPASI (version 4.20), a common simulation environment for SBML-formated models [9], using the deterministic solver (LSODA) and stochastic simulator (implementation of *direct method* by Gillespie [24]).

Model perturbations The first type of perturbation is based on a modification of rate constants. We used this modification to induce site-directed mutations: Ser137Ala and constSer137, replicating the original study [1]. In both cases, the alteration of the model involved the inactivation of 4 reactions by zeroing their rate constants. In the RB model, they are represented by 1 rule, i.e. a change of a single constant induced each mutagenesis.

We additionally tested the RB model with two different binding schemes, applicable only to the RB model, and further called noncompetitive (FIGURE 6B) and competitive binding (FIGURE 6A). In the noncompetitive binding, all interactors of DARPP-32 can bind simultaneously to three different sites. The competitive binding assumes one interaction with DARPP-32 at a time, which reflects the ODE model assumption. The Fernandez et al. [1] study does not discuss if DARPP-32 binding partners bind to the same or different active sites, though DARPP-32 is an intrinsically disordered protein with an unknown

binding interface [68–71]. The ODE model specification demonstrates that DARPP-32 forms at most heterodimers. This type of modification in the ODE model would require enumeration of additional molecular species, the addition of new equations and updating the existing ones. Contrary to the ODE model specification, a definition of such a binding scenario in the RB notation requires the same number of rules, provided that concurrently bound interactors do not influence each other.

4 Results

Comparison of models was performed on two levels, model notation and simulation results. The model notation was analysed by dividing the model into components and comparing their sizes. We expect the set of reactions underlying the ODE model to be represented with fewer rules since a single one can constitute a pattern representing several reactions. The comparison of simulation results involved the alignment of equivalent time courses obtained by model simulations. We performed the comparison of time courses between three variants of each model: (1) base-line condition (wild-type) and two site-directed mutations: (2) Ser137Ala and (3) constSer137. Finally, we compared two RB model variants, representing: (1) DARPP-32 with a single binding site; and (2) DARPP-32 with three independent binding sites.

4.1 Rule patterns reduce reaction number in a certain type of model components

Table 1 juxtaposes the total counts of model components in each model. The RB model represents 152 reactions with 132 rules, each parameterised by one of 62 unique rate constants. This number is lower than the total number of rate constants used to parameterise the ODE model (152). The final rule set is more than twice as large as the unique number of rate constants, meaning that more than one rule is parameterised by the same rate constant. The number of molecular species in the RB model, obtained with snapshots capturing the state of the molecular mixture over simulation time (every 10000th event), is 91 for the competitive RB model, and 137 for the non-competitive one. In both cases, the sum of molecular species is higher than in the ODE model (75). As expected, the number of rules corresponding to reactions is lower, and the number of molecular species is much higher, confirming that expression patterns reduce the number of rules needed to represent a reaction system. Nonetheless, the number of rules is only slightly lower than the number of reactions (152 to 131). If we closely compare models by parts representing more general molecular mechanisms, rule representation reduces the reaction number in some components but extends it in others (Table 2). The reduction occurred only in "DARPP-32 phosphorylation" and "PP2A activation by Ca^{2+}" components, where combinations of states of DARPP-32 phosphorylation sites do not have to be explicitly named. In contrast, the increase in the number of reactions compared to the reactions occurred in the components "PKA activation" and "PP2B activation". They both have four sites that bind the same molecules, Ca^{2+} and cAMP, respectively, which requires the expression of combinatorial binding in the rule notation.

4.2 RB model recapitulates dynamics of ODE model with minor discrepancies

The RB model recapitulates the principal dynamics of the ODE model, albeit there are some observable differences (compare Figures 7B and Figure 7C). For instance, there is a comparable variability of trajectories of Thr34 phosphorylated isoform ("D34" and "D34*") that can be observed in the relaxation phase (after the 600th time point). "D34" in the ODE model needs 100 more time steps to reach the second peak, and it is weaker than its RB counterpart ("D34*"). Worth noting is that the standard deviation in the stochastically simulated ODE model reveals a distinctive variation in abundance of the "D34" observable during the relaxation phase. We further use the stochastic trajectories of the ODE model for comparison with the RB model.

For a closer examination, traces of 15 observables (defined in Table S1) obtained from ODE and RB simulations were paired and superimposed (Figure 8). Next to the clear matches (e.g. Figure 8: B, E, H, N), there are discrepancies between paired curves. Five of these 15 observables (Figure 8: C, F, I, J, O) are examples of the largest divergence between models following a similar pattern of behaviour. They are directly connected in a chain of activation reactions that begins with Ca^{2+} (Figure 9). Higher abundance of all Ca^{2+} ions present in the system of the ODE model (Figure 8C) could explain differences between the remaining 4 observables. However, the trajectory of "all_Ca" remains at the 0 level during steady states rising only in the spiking interval, which resembles the abundance of free Ca^{2+}. (Figure 8B). Ca^{2+} activates PP2B represented by the trajectory "PP2Bactive". The higher level of "PP2Bactive" is consistent with the other three observables, suggesting that this is a factor generating the divergences between the models. Based on the curves of the ODE model, we can reason that a stronger activation of PP2B results in proportionally more copies of the unphosphorylated CK1 and phosphorylated D137. This, in turn, increases substrate availability for PP2C; therefore, more copy-numbers of its bound form. This effect is inverted in the trajectories derived from the RB model.

4.3 RB language allows for detailed dissection of observed molecular species facilitating reuse and reanalysis of dynamic models

It remains unclear why the "all_Ca" observable trajectory produced by the ODE model is much lower than in the RB model at the steady state. Moreover, "PP2Bactive" appears to dictate the higher effect on the other three observables ("D137", "CK1u", "_PP2C"). Therefore, "all_Ca" and "PP2Bactive" are closer analysed in further steps.

According to the reaction system underlying both models, the activated PP2B is a complex of 4 Ca^{2+} ions and PP2B. This detail is not explicitly stated in the variable name of the ODE model. Therefore, to obtain the trajectory of all Ca^{2+} ions, the sum of the copy numbers of the

ODE model			RB model
Model Component	Total counts	Total counts	Model Component
Reaction instances	152	132	Reaction rules
Concentration-based rate constants	152	62	Stochastic rate constants
Initial concentrations	75	8	Initial copy numbers
Molecular species	75	91/137	Molecular species
Stimuli events	21	21	Stimuli events

TABLE 1: The specification of a model can be divided into components. The total number of elements in each component is shown for both the ODE model and the RB model.

	Model component	Reactions	Rules	Unique rate constants
1.	DARPP-32 phosphorylation	84	27	27
2.	CK1 phosphorylation	4	4	4
3.	PDE phosphorylation	4	4	4
4.	PP2A phosphorylation	4	4	4
5.	PP2B activation	4	24	4
6.	PKA activation	12	56	7
7.	cAMP & Ca^{2+} degradation	8	8	8
8.	PP2A activation by Ca^{2+}	32	4	4

TABLE 2: The list of reactions in the Fernandez et al. [1] publication was divided into components based on more general molecular processes represented by subsets of reactions, such as phosphorylation or activation. We can closely examine reaction-rule relation by comparing models by components. The table shows the number of reaction rules versus reaction instances and a unique number of rates per model component. It is noticeable that the reduction in the number of reaction instances due to the translation of reactions into Kappa language occurred in only two model components (1. & 8.), while in two others, it resulted in an expansion of the rule number (5. & 6.).

molecular species whose variable names contain "Ca^{2+}", would have to be replaced by a more thorough analysis of the relevant reaction context of the ODE model. This is not the case in the RB model, where an observable of interest is obtained with an automated procedure that sums the trajectories of molecular species containing the specified expression pattern. Since the trajectory of all Ca^{2+} in the RB model includes the ions bound to PP2B, the comparison of "all_Ca" to "all_Ca*" is inaccurate due to a difference in the molecular species included in these observables. A similar inaccuracy, related to the naming of the observables, explains the discrepancy between the time courses of the total number of cAMP observables (FIGURE 7). In contrast to the RB model, the multiple copies of cAMP bound to R2C2 are not included in this ODE model time course (compare FIGURE 8A).

As a complete list of molecular species in the RB model is not included in the specification, it is also unknown whether there are other trajectories of molecular species summed up in "all_Ca*". To obtain these, all molecular species containing Ca^{2+} in the RB model simulation were isolated from snapshot data ammounting to 24 compared to only 13 in the ODE model ("all_Ca"). These 13 species correspond in molecular composition to 18 of the 24 RB species sampled in total. The six absent species in the "all_Ca" observable are composed of an active form of PP2B containing 4 Ca^{2+} ions, either free or bound to phosphorylated CK1, or DARPP-32 in 4 different combinations of phosphorylation states. The number of species comprising "all_Ca" observable in the RB model is higher by five because the half-active form of PP2B (bound to two Ca^{2+} ions) in the RB model exists in 6 variants. Whilst in the ODE model, it is represented as a single species, named "PP2BinactiveCa2".

Now that we know the composition of molecular species comprising the "all_Ca*" observable feature, we can rerun the RB model with the same components as in "all_Ca" of the ODE model and compare their trajectories directly. To match this newly defined RB observable of "all_Ca*" to the original model, these 18 species were summed to obtain a single observable (FIGURE 10B). In comparison to the unaltered species composition (FIGURE 10A), the result shows that discrepancy between the ODE and RB observable trajectories have diminished. Knowing that there is a higher number of forms representing the half-active PP2B in the RB model, we can try to obtain a closer match between the "all_Ca" observables by superimposing only one of 6 trajectories of the RB model. FIGURE 10C shows that the match is close to perfect. It demonstrates that the differences between the "all_Ca" observables of the two models can be explained by the difference in the number of representations of the molecular species. As the six trajectories have the same dynamics and the same average levels of abundances, choosing one of them is arbitrary (FIGURE S1). Moreover, the distinction between locations of two Ca^{2+} ions on the numbered sites of PP2B is irrelevant since all four sites are functionally indistinguishable.

The above analysis demonstrates the flexibility of the RB modelling as a tool to explore complexes formed during the simulation. Molecular species defined in the ODE framework are fixed and definite, whereas, in the RB model, they are a subject of investigation. Snapshots of molecular

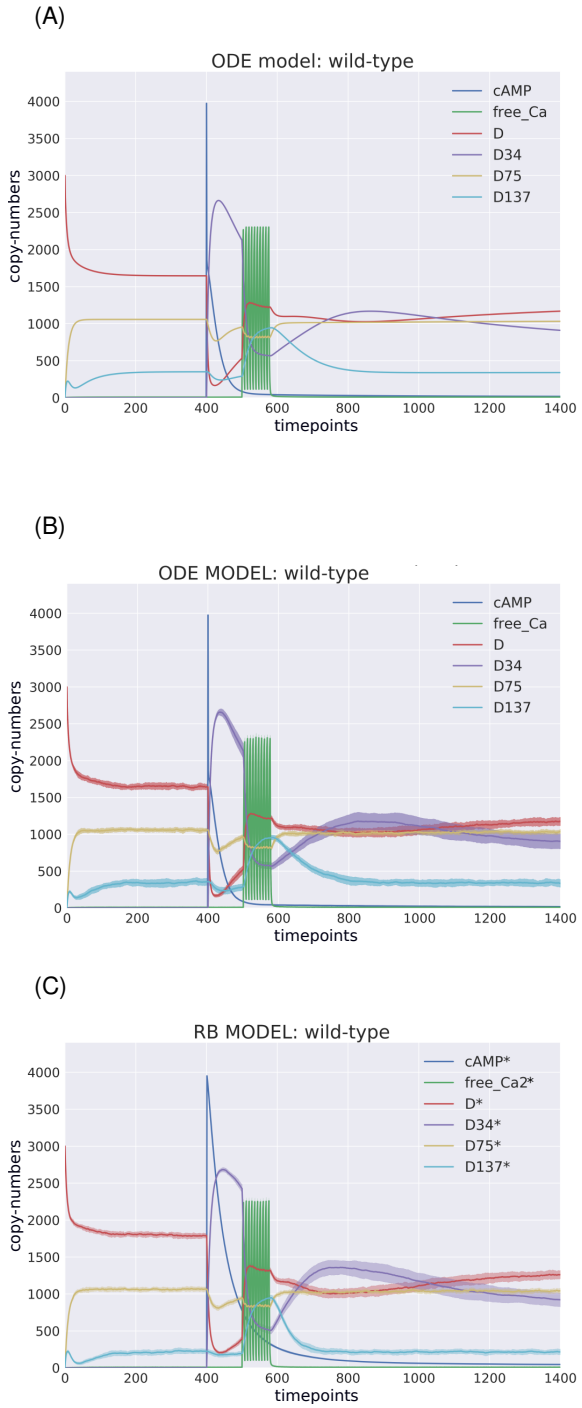

FIGURE 7: Time-courses of the ODE model for DARPP-32 isoforms triggered by a pulse of cAMP followed by a train of Ca2+ spikes obtained with (A) a deterministic solver, and (B) a stochastic simulation. Trajectories of the stochastic simulation were obtained from calculating mean value (*line*) and standard deviation (*shade*) based on 40 simulations. (C) RB model (stochastic simulation). Variable isoforms of DARPP-32: "D" - unphosphorylated; "D137" - Ser137 phosphorylated; "D75" - Thr75 phosphorylated; "D34" - Thr34 phosphorylated.

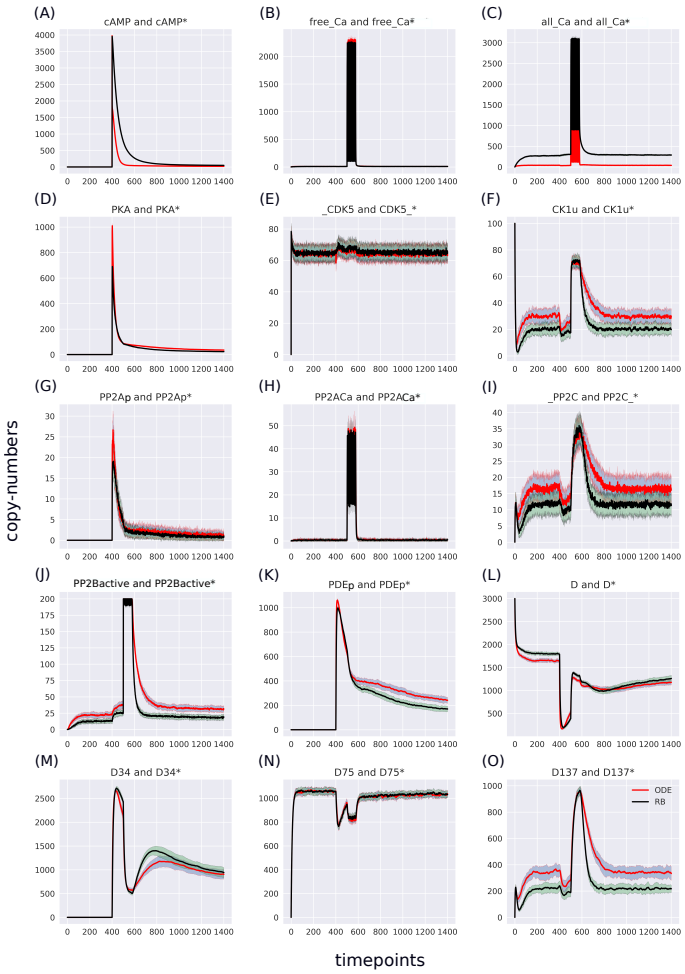

FIGURE 8: Overlaid time courses of the ODE stochastic and the RB stochastic in the baseline condition. Note that the scales on the y-axis are different to closely compare the traces of the observables. Trace colour: ODE (*red*), RB (*black*).

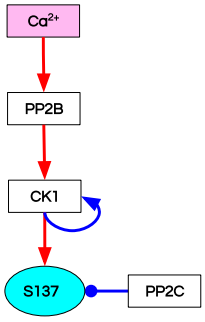

FIGURE 9: Reaction diagram of the observables showing the greatest divergence between the observables of the ODE and RB models. These observables are connected in a chain of mutually dependent activation reactions triggered by the influx of Ca^{2+}.

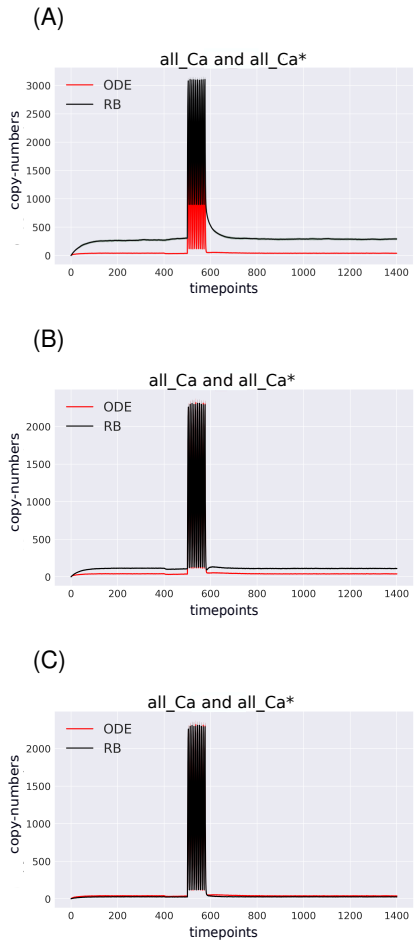

FIGURE 10: Comparison of variable compositions of molecular species containing Ca^{2+} ions tracked in the system in both models with (B) unaltered observables; (B) all molecular species containing Ca^{2+} ions selected as indicated by names in the original model and summed to obtain a single trace, where 13 molecular species in the ODE model are represented by 18 species in the RB model; (C) 13 molecular species of ODE model matched to 13 of RB model, where only 1 of 6 molecular species of inactive PP2B was selected.

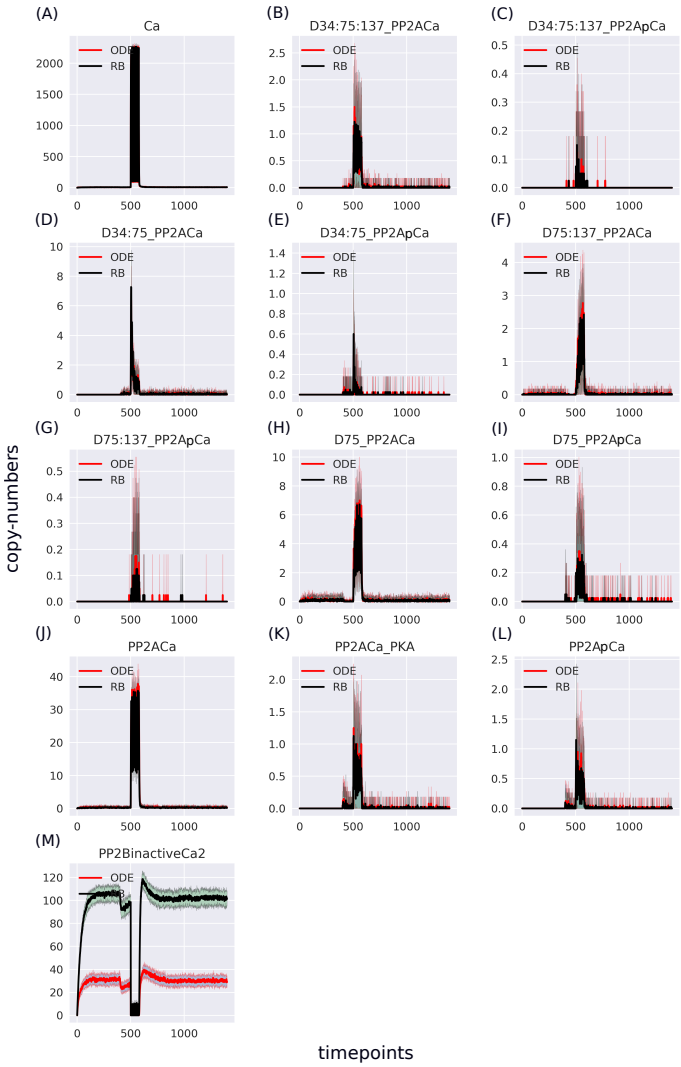

FIGURE 11: Traces of 13 pairs of molecular species containing Ca^{2+}, selected to match the ODE model. The largest disparity lies in the "PP2BinactiveCa2" variable - summation result of 6 entities representing an inactive form of PP2B in the RB model.

mixtures permit the determination of created species and their abundances. The results of the RB model can be easily dissected by tracking precisely defined observables of interest during the model run. In addition, the automated agglomeration of time courses into observables renders the RB modelling more advantageous compared to the ODE-based framework. This automation is less error-prone and independent of variable names.

4.4 *Rate constants of reactions formulated with "combinatorial-binding" notation should be increased to match ODE trajectories*

The largest discrepancy between trajectories can be observed in "PP2BinactiveCa2" (FIGURE 11M) that for the RB time course was obtained by summing 6 entities representing a half-active PP2B into one. If divided by 6, representing a single variant of half-active PP2B, the trajectory of the RB model becomes lower than the one of the ODE model (FIGURE S2M). These six forms of the half-active PP2B suggest that a better fit between the two models can be achieved by decreasing the constant rate of rules that represent the binding of Ca^2+ to free PP2B. However, this could bring the desired effect solely for the half-active PP2B ("PP2BinactiveCa2") but not for the dynamics of other coupled observables. In particular, the fully active PP2B ("PP2Bactive"), of which the "PP2BinactiveCa2" is an intermediate form. With the current parameter values, the fully active PP2B in the RB model is lower than in the ODE model (FIGURE 8J). Thus, the decrease in the rate constant of its intermediate form would lead to a further decrease in its copy number. Therefore, although there are more species of the half-active PP2B in the RB model, its fully active form has lower levels than the ODE model. To further examine this observation, we can return to the

comparison of model specifications (Table 2). The four reactions of PP2B activation are represented by 24 rules, explicit in site-specific detail that includes all combinations of positioning Ca^{2+} on four sites of PP2B. Therefore, the larger number of rules defining transition from inactive to active PP2B slows down this process in the rule representation. To support this hypothesis, we can review the second example that required a much larger number of rules, i.e. an activation of PKA. Figure 8D shows that the RB trajectory of the "PKA*" observable also reaches a much lower peak than its ODE counterpart. Accordingly, values of rate constants of rules that the number increased due to the "combinatorial binding" notation in the RB model should be increased to closely match the ones in the ODE model.

4.5 *RB language facilitates modifications of dynamic models*

Site-directed mutations The authors of the ODE model analysed it in two other conditions of site-directed mutagenesis affecting the Ser137 site. The same perturbations were applied to the RB model to establish if they would generate similar results. Figures 12 juxtapose simulation results of both models affected with constSer137 and Ser137Ala mutations. As exemplified by the six key observables, there is a close match in initial conditions and a general pattern of dynamics between the time courses of the two models. Therefore, we can conclude that the RB modelling allows emulating experimentally observed perturbations similarly to the ODE-based modelling.

Competitive and non-competitive site binding Two variants of the RB model with different binding site specifications are compared to test whether the dynamics of the model are affected when DARPP-32s binds multiple partners at once. The first specification is a competitive variant of the model with one binding site (oBS). In the second, a non-competitive variant, the partners bind simultaneously (three-binding-sites, tBS). The superimposed trajectories of two models (Figure S3) demonstrate that this modification has no effect on model dynamics. As the direct consequence of this modification is an increase in the size of the complexes to more than two proteins, it seems that larger complexes are rarely formed during the simulation. This interpretation is confirmed in a direct examination of species counts (Figure S4). As these three binding sites do not counter each other's binding properties, the lack of difference might be caused by the similarity in occupancy between a single site and all three sites together. The probability of a site being connected depends on copy numbers of reactants and the strength of binding affinities. Reactions in the model are classified as weak, with dissociation rates in the range of μM. Low-affinity bindings generally lead to lower levels of site occupancy. Moreover, the amount of DARPP-32 molecules exceeds the total counts of all its interactors. Therefore, with the current proportions of reactants, all three sites of DARPP-32 cannot be saturated to expose the difference in the binding capacity of DARPP-32. To expose the potential differences in dynamics between two binding scenarios, the site occupancy must also be modified. The simplest

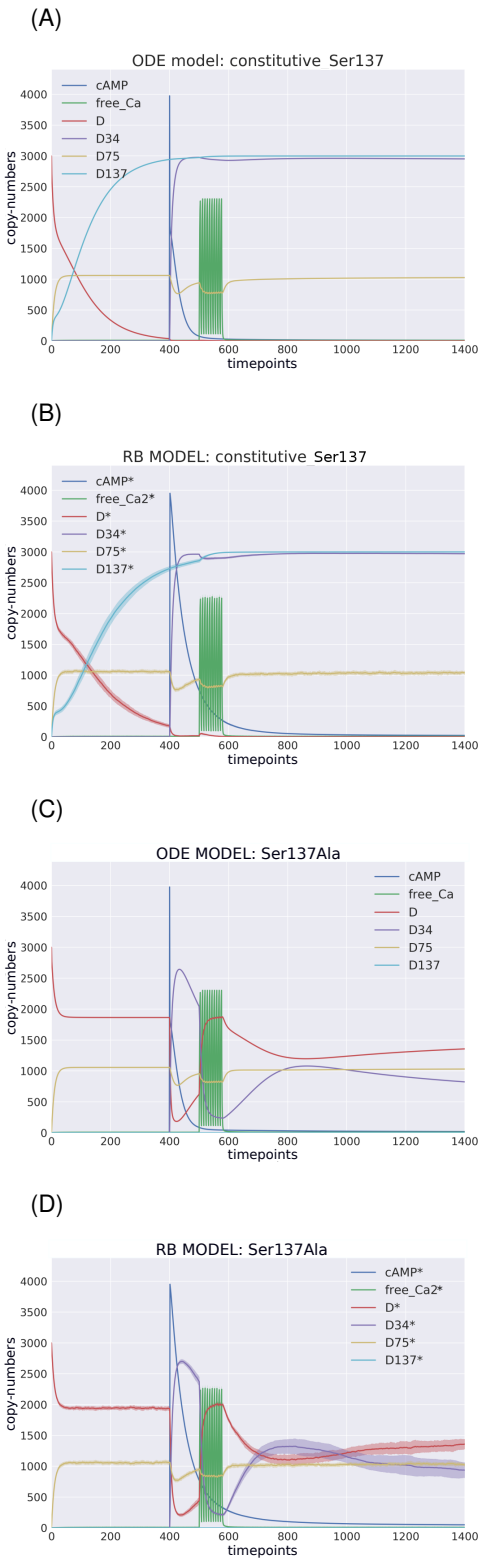

Figure 12: Comparison of the constitutive Ser137 mutation induced in (A) ODE model in deterministic setting; (B) RB model in stochastic setting; and the Ser137Ala mutation in (C) ODE model in deterministic setting; (D) RB model in stochastic setting; The same interference performed on rate constants of the two models caused similar dynamics.

test of this explanation could be performed by altering the size of the reactant pools. For instance, a significant decrease in DARPP-32 levels could increase the proportion of other reactants.

5 Discussion

ODE-based modelling is a classical and commonly used method for creating detailed dynamic models of biological systems [72–74]. It is frequently a point of reference and comparison to newly proposed modelling methods [14, 35, 39, 75, 76]. Nevertheless, modelling of signalling systems with ODE poses difficulties due to complexities underlying molecular interactions [41, 77]. To address these problems, RB modelling was developed. Although a large body of reviews discuss the advantages of RB modelling over ODE [11, 13, 17], a direct comparison of time courses of one molecular system encoded in the two modelling formalisms has not been presented before. The manual translation of this ODE model into any RB language was necessary despite the existence of a method [78] for automated translation of the SBML-format encoding ODE-based models to an RB model format that failed to perform this task. This paper presents results of encoding reactions underlying an ODE model to the RB language and a comparison of their specification and simulation results.

Effects of the RB framework on the model notation Encoding reactions into rules slightly reduced the size of model specification and increased counts of molecular species, which confirmed the well-known advantage of rule representation. Translation of reactions into rules involves the removal of non-influential contexts denoting states of reactants, such as phosphorylation or binding state. In consequence, this shortens and simplifies model representation.

Closer analysis of reaction subsets representing more general molecular mechanisms showed that reduction in the reaction number is only true for reactions occurring between the same reactants, describe the same transformation, and are parameterised with the same values of rate constants, but differ only concerning binding or internal state of reactants. In this type of reaction, the number of unique reaction rates was equal to the number of rules.

An increase in the number of rules representing reactions occurred where the same partner binds an agent at multiple sites. In such reactions, all possible positions and stages of the binding process had to be explicitly encoded. The "combinatorial binding" notation is not a general property of the RB language but characteristic of the Kappa syntax. In the BioNetGen Language (BNGL), an alternative RB framework, a rule is definable with sites named in the same way. It implies that a rule pattern defined for one of them applies to the others [67]. However, it is yet to be established how this simplified representation in BNGL would affect the trajectories of observables.

A shift in modelling focus with the use of the RB framework The process of encoding reactions into rules turns attention to questions such as how many binding partners can simultaneously bind a protein. The translation process has shown that information about interfaces of interacting proteins and their alternative states would considerably ease the process of model development by guiding decisions on agents' signatures. Therefore, data resources that could support RB modelling, such as a source of protein interaction interfaces, post-translational modifications (PTMs), and protein domains. For instance, proteins containing phosphatase catalytic domains are enzymes of dephosphorylation reactions [79]. However, such detailed information is not accessible for most molecular agents.

Discrepancies in dynamics between models Comparison of trajectories showed an agreement between models' dynamics, with some discrepancies. They mainly appeared due to a lack of precision in the variable names of the ODE model caused by differences in molecular species comprising tracked observables. When the simulation of the RB model was performed with observables exactly matching the ones in the ODE model, almost all paired trajectories fitted perfectly. The only problematic observable involved the "combinatorial binding" notation. Further analysis suggested that activation of proteins encoded with this notation is slower than in the ODE model. These specific discrepancies should be accounted for when reactions and rate constants of ODE models are reused to construct RB models. The difference in reaction speed between the models cannot be caused by the simulation paradigm since both were run with variants of Gillespie's SSA. This difference could simply reflect the difference in size of the rule representation compared to the reaction representation, as the more rules/reactions to activate an agent, the more events need to be executed to trigger that agent, thus slowing down its activation.

Facilitation of modification and reuse of dynamic models We used snapshots to demonstrate detailed exploration of the emerging molecular species during the simulation of an RB model. By overlaying trajectories of molecular species we enabled tracking of the source of differences between models. Emerging molecular species are relatively easy to examine and analyse. Conversely ODE models rely heavily on complete knowledge of the system and the arbitrary naming of variables. Retrieval of molecule counts, hidden in individual molecular species of the ODE model, would require further deconstruction of the reaction system and the arduous extension to much more complex models. The implicit difficulty of identifying all molecular entities arising from the molecular species in ODE models fundamentally impacts on their veracity.

The precise identification of molecules among species could be performed by parsing the SBML-based model encoding. However, exploration of the Fernandez et al. [1] model web page in the BioModels website[3], shows incomplete annotations of molecular species behind variable names, both concerning the actual counts of interactions

[3] `http://www.ebi.ac.uk/biomodels-main/BIOMD0000000153`

(e.g. Ca^{2+}) and their components. In this light, automated identification of molecules in the RB framework is particularly advantageous, as it offers a transparent framework with error-prone identification of molecules in a modelled system. This feature is particularly vital when the modelled system is composed of numerous molecules, and their particular states are to be analysed in detail.

The RB model was tested with two types of site-directed modifications, demonstrating the framework's flexibility to reproduce experimentally conducted perturbations. Though the binding site modification effectively changed the model reaction network, it did not affect the model response. Nevertheless, this intervention demonstrated the ease of performing such alterations within the RB framework. Additionally, the pattern notation improves model clarity and provides an intuitive representation of a model akin to a set of chemical reactions rather than equations, potentially improving the learning curve for a modeller-to-be.

Limiting factor of the simulation time The execution time of stochastic model simulations in comparison to deterministic ones has always been an issue addressed by multiple optimisation strategies [80]. It is no different in the case of the two models compared here. It takes almost 40min[4] to simulate this particular RB model with the KaSim simulator. The solution of the ODE model in the COPASI environment in the deterministic setting returns in an instance. The COPASI stochastic simulator produces results in no more than 15sec. Thus, reducing simulation times for RB models remains an important challenge for use cases in which speed is critical.

Further explorations of the DARPP-32 RB model There are two main routes for further exploration of the RB model. The first one is a modification of parameters defining different phases of combinatorially bound Ca^{2+} ions to PP2B and cAMP to R2C2. A particular task would be to identify factors by which the binding constants could be modified to counteract the many intermediate variants of these complexes and the lower copies of their activated final forms. It would be interesting to identify conditions under which we could observe a difference in model dynamics after adding binding sites to DARPP-32. As a starting point, the modification could be achieved by significantly decreasing the copy number of DARPP-32 compared to other interactors. As mentioned by Fernandez et al. [1], levels of DARPP-32 vary considerably, between μM to tens of μMs, in the striatum. With the greater availability of single-cell techniques for protein quantification [81] it would be worth establishing more precisely the range of DARPP-32 even at the resolution of a dendritic spine [82]. Estimating variability between cells could also be used to compare the varying levels of phosphorylated DARPP-32 at Thr34 that were observed in the stochastic simulations.

Extending the model to match evolving knowledge on the DARPP-32 interaction network So far, only early signalling events of DARPP-32 have been modelled, localised mainly in the cytosol. Thus, the Ser102 site was omitted in our model because there is no evidence that it can be affected by DA or Glu signalling [83, 84]. However, a recent study by Nishi et al. [85] suggests that Glu can decrease the effect of DA signalling (phosphorylation of three other sites) by dephosphorylating DARPP-32 at Ser102 causing the accumulation of DARPP-32 in the nucleus. The Ser102 site regulates nuclear transportation of DARPP-32 [86] representing late signalling events. Stipanovich et al. [86] have shown that nuclear accumulation of DARPP-32 is promoted by drugs of abuse; the inclusion of Ser102 in future models could be valuable to further explore this. In addition, the results of the recent study by Nishi et al. [85] could be used to calibrate phosphorylation of the four phosphorylation sites as they were measured under the same conditions.

Need for formal prioritisation methods of emerging molecular species RB modelling offers tools for the dissection of emerging molecular species during simulations. As the number of such molecular species increases it becomes increasingly difficult to identify which model components are of particular importance to the system. There is great potential to exploit commonly used methods for model exploration, such as sensitivity analysis, to identify critical model features including parameters and model output variables. It would be advantageous to support the modeller's assumptions with automated methods to prioritise model outputs for downstream analysis and greater insight into the underlying biological systems.

6 Conclusions

Dynamic molecular modelling has become increasingly important in uncovering and integrating our dispersed knowledge of molecular mechanisms. Choosing the best formalism meet this challenge is a difficult task. In this work, we have presented a detailed and systematic comparison of two major formal approaches to quantitative modelling. We demonstrated the advantages and disadvantages of RB modelling to the preeminent ODE approach. We found RB modelling to be a more detailed and flexible way to represent biological molecular systems, enabling exploration of individual molecular entities, model extension, and future reuse. These conclusions confirm the huge potential of the RB formalism and hopefully will embolden future exploration and research in this topic.

7 Code accessibility

The code reproducing the figures and the encoding of the Kappa models can be found on github : `https://github.com/ewysocka/rb_vs_ode_model_of_darpp-32`.

8 Acknowledgments

The project was jointly founded by UCB PHARMA S.A., acting through its UK subsidiary UCB CELLTECH, and the University of Edinburgh. The authors would like to thank

[4]The total CPU time measured with the "`time`" command on Linux OS

Andrew Miller and Evangelia Petsalaki whose suggestions have contributed to the improvement of this manuscript.

9 Conflict of interest

The authors do not declare any conflict of interest.

List of Acronyms

Ca^{2+} calcium ions

cAMP cyclic adenosine monophosphate

DA dopamine

DARPP-32 dopamine- and cAMP-regulated neuronal phosphoprotein with molecular weight 32 kDa

Glu glutamate

MSPN medium spiny projection neurons

ODE ordinary differential equation

PTM post-translational modification

RB rule-based

SBML Systems Biology Markup Language

Ser102 Serine 102

Ser137 Serine 137

SSA Stochastic Simulation Algorithm

Thr34 Threonine 34

Thr75 Threonine 75

List of Figures

1 The first step in building a kinetic model of molecular interactions is the definition of *chemical reactions* between molecules. These reactions describe chemical transitions between states of one reactant catalysed by the other. An arrow denotes the direction of the transformation and a doubled arrow reaction reversibility. A set of chemical reactions is denoted with a *reaction diagram*. Quantitative evaluation of model behaviour is achieved by converting coupled chemical reactions (simultaneously solved) to a set of ODEs. Each rate equation expresses the change of concentration of a single molecular species over time, formulated with reaction rates that directly take part in the creation and elimination of this species [7]. On the right-hand side of the figure, a rate equation for X_{00} is defined with 4 rate laws of mass action. The law of mass action states that the speed of reaction is proportional to the product of the concentrations of the reactants [18, p.141]. Each reaction rate is weighted by a *rate constant*, specific to each reaction. Positive and negative signs of reaction rates denote the direction of arrows pointing respectively towards and away of X_{00} [7]. Example adapted from Hlavacek et al. [17]. 2

2 Main concepts and elements of the Kappa language (description in the main text). 3

3 Reaction diagrams representing different aspects of the DARPP-32 network included in the ODE model by Fernandez et al. [1] Nodes: DARPP-32 *(cyan)*, second messengers *(magenta)*, kinases/phosphatases *(white)*. Edges: inhibition reactions *(blue)*, activation reactions *(red)*. 4

4 Approach to comparison of ODE and RB modelling frameworks. 4

5 Approach to comparisons of time-series. 5

6 Two scenarios of binding patterns illustrate the implications of changing the number of binding sites. In the non-competitive case (A), the phosphorylation sites can bind to their respective kinases and phosphatases. In the other scenario (B), only one site can connect. This leads to competition between the interactors. 5

7 Time-courses of the ODE model for DARPP-32 isoforms triggered by a pulse of cAMP followed by a train of Ca2+ spikes obtained with (A) a deterministic solver, and (B) a stochastic simulation. Trajectories of the stochastic simulation were obtained from calculating mean value (*line*) and standard deviation (*shade*) based on 40 simulations. (C) RB model (stochastic simulation). Variable isoforms of DARPP-32: "D" - unphosphorylated; "D137" - Ser137 phosphorylated; "D75" - Thr75 phosphorylated; "D34" - Thr34 phosphorylated. 8

8 Overlaid time courses of the ODE stochastic and the RB stochastic in the baseline condition. Note that the scales on the y-axis are different to closely compare the traces of the observables. Trace colour: ODE (*red*), RB (*black*). 8

9 Reaction diagram of the observables showing the greatest divergence between the observables of the ODE and RB models. These observables are connected in a chain of mutually dependent activation reactions triggered by the influx of Ca^{2+}. 8

10 Comparison of variable compositions of molecular species containing Ca^{2+} ions tracked in the system in both models with (B) unaltered observables; (B) all molecular species containing Ca^{2+} ions selected as indicated by names in the original model and summed to obtain a single trace, where 13 molecular species in the ODE model are represented by 18 species in the RB model; (C) 13 molecular species of ODE model matched to 13 of RB model, where only 1 of 6 molecular species of inactive PP2B was selected. 9

11 Traces of 13 pairs of molecular species containing Ca^{2+}, selected to match the ODE model. The largest disparity lies in the "PP2BinactiveCa2" variable - summation result of 6 entities representing an inactive form of PP2B in the RB model. 9

12 Comparison of the constitutive Ser137 mutation induced in (A) ODE model in deterministic setting; (B) RB model in stochastic setting; and the Ser137Ala mutation in (C) ODE model in deterministic setting; (D) RB model in stochastic setting; The same interference performed on rate constants of the two models caused similar dynamics. 10

S1 A half-active PP2B is a complex composed of PP2B and two Ca^{2+} ions. The RB model simulation generates six different molecular species that represent this complex due to a combinatorial binding of Ca^{2+} ions to 4 identical sites of PP2B. The plot shows superimposed trajectories of these six variants of the half-active PP2B. Neither of these six trajectories differentiates itself from others by a pattern of dynamics nor an average level of abundances. 21

S2 Separated molecular species containing Ca^{2+}, selected as in the original model. The ^textsf'PP2BinactiveCa2" trajectory of the RB model was obtained by selecting one of 6 entities representing the inactive form of PP2B in the RB model. There is still a discrepancy between the models, but the trajectory is lower for the RB model. 22

S3 Comparison of two variants of the RB model in which the agent representing DARPP-32 had one binding site (oBS, red trace) and three binding sites (tBS, black trace). Overlaid trajectories of corresponding agents demonstrate no effect of this modification on the model trajectories. 23

S4 Overlay of change in species counts overtime for two models where DARPP-32 can bind at a single site (oBS) and three sites (tBS) (A). The species set size is similar in both model variants. The change in the number of unique species counts is aligned with trajectories of the stimuli (B). The dynamics of complex formation are dictated by the pattern of stimulus introduction, as the greatest differences between oBS and tBS occur during stimulus application. 24

List of Tables

1 The specification of a model can be divided into components. The total number of elements in each component is shown for both the ODE model and the RB model. 7

2 The list of reactions in the Fernandez et al. [1] publication was divided into components based on more general molecular processes represented by subsets of reactions, such as phosphorylation or activation. We can closely examine reaction-rule relation by comparing models by components. The table shows the number of reaction rules versus reaction instances and a unique number of rates per model component. It is noticeable that the reduction in the number of reaction instances due to the translation of reactions into Kappa language occurred in only two model components (1. & 8.), while in two others, it resulted in an expansion of the rule number (5. & 6.). 7

S1 Names of RB observables and corresponding names of ODE observables with definitions. To obtain observable ODEs, the time series of the corresponding molecular species are summed based on their names. 21

References

[1] É. Fernandez, R. Schiappa, J. A. Girault, and N. Le Novère. DARPP-32 is a robust integrator of dopamine and glutamate signals. *PLoS Computational Biology*, 2 (12):1619–1633, 2006.

[2] L. Xie, X. Ge, H. Tan, L. Xie, Y. Zhang, T. Hart, X. Yang, and P. E. Bourne. Towards Structural Systems Pharmacology to Study Complex Diseases and Personalized Medicine. *PLoS Computational Biology*, 10(5):e1003554, 2014.

[3] J. Li and U. R. Mansmann. A molecular signaling map and its application. *Cellular Signalling*, 26(12): 2834–2842, 2014.

[4] L. Jia, J. Piña-Crespo, and Y. Li. Restoring wnt/β-catenin signaling is a promising therapeutic strategy for alzheimer's disease. *Molecular Brain*, 12(1), 2019.

[5] N. D. Volkow and M. Boyle. Neuroscience of addiction: Relevance to prevention and treatment. *The American journal of psychiatry*, 2018.

[6] H. Kitano. Systems Biology: A Brief Overview. *Science*, 295(5560), 2002.

[7] J. C. Sible and J. J. Tyson. Mathematical modeling as a tool for investigating cell cycle control networks. *Methods (San Diego, Calif.)*, 41(2):238–47, 2007.

[8] A. Dräger, N. Hassis, J. Supper, A. Schröder, and A. Zell. SBMLsqueezer: a CellDesigner plug-in to generate kinetic rate equations for biochemical networks. *BMC systems biology*, 2(1):39, 2008.

[9] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22(24):3067–3074, 2006.

[10] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, A. S. Perelson, and B. Goldstein. The complexity of complexes in signal transduction. *Biotechnology and bioengineering*, 84(7):783–94, 2003.

[11] V. Danos. *Rule-Based Modelling of Cellular Signalling*, volume 4703. Springer Berlin Heidelberg, Berlin, Heidelberg, 2007.

[12] L. A. Chylek, E. C. Stites, R. G. Posner, and W. S. Hlavacek. Innovations of the Rule-Based Modeling Approach. In A. Prokop and B. Csukás, editors, *Systems Biology: Integrative Biology and Simulation Tools*, pages 273–300. Springer Netherlands, Dordrecht, 2013.

[13] L. A. Chylek, L. A. Harris, C.-S. Tung, J. R. Faeder, C. F. Lopez, and W. S. Hlavacek. Rule-based modeling: a computational approach for studying biomolecular site dynamics in cell signaling systems. *Wiley interdisciplinary reviews. Systems biology and medicine*, 6(1): 13–36, 2014.

[14] L. A. Chylek, L. A. Harris, J. R. Faeder, and W. S. Hlavacek. Modeling for (physical) biologists: an introduction to the rule-based approach. *Physical biology*, 12(4):045007, 2015.

[15] G. Antunes, A. C. Roque, and F. M. Simoes-de Souza. Stochastic Induction of Long-Term Potentiation and Long-Term Depression. *Nature Scientific Reports*, 6(1): 30899, 2016.

[16] A. Köhler, J. Krivine, and J. Vidmar. A Rule-Based Model of Base Excision Repair. In *CMSB 2014: Computational Methods in Systems Biology*, pages 173–195. Springer, Cham, 2014.

[17] W. S. Hlavacek, J. R. Faeder, M. L. Blinov, R. G. Posner, M. Hucka, and W. Fontana. Rules for modeling signal-transduction systems. *Science's STKE : signal transduction knowledge environment*, 2006(344):re6, 2006.

[18] E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach. *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley-VCH, 2005.

[19] D. Wilkinson. *Stochastic Modelling for Systems Biology*. Chapman & Hall/CRC Mathematical & Computational Biology. Taylor & Francis, 2006.

[20] D. J. Wilkinson. Stochastic modelling for quantitative description of heterogeneous biological systems. *Nature reviews. Genetics*, 10(2):122–33, 2009.

[21] D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.

[22] Y. Chen, F. Cunningham, D. Rios, W. M. McLaren, J. Smith, B. Pritchard, G. M. Spudich, S. Brent, E. Kulesha, P. Marin-Garcia, D. Smedley, E. Birney, and P. Flicek. Ensembl variation resources. *BMC Genomics*, 11(1):293, 2010.

[23] J. E. Ladbury and S. T. Arold. Noise in cellular signaling pathways: causes and effects. *Trends in biochemical sciences*, 37(5):173–8, 2012.

[24] D. T. Gillespie. Exact Stochastic Simulation of Coupled Chemical Reactions. *Journal of Physical Chemistry*, 81 (1):2340–2361, 1977.

[25] P. Seshacharyulu, M. P. Ponnusamy, D. Haridas, M. Jain, A. K. Ganti, and S. K. Batra. Targeting the EGFR signaling pathway in cancer therapy. *Expert opinion on therapeutic targets*, 16(1):15–31, 2012.

[26] J. Chen, F. Zeng, S. J. Forrester, S. Eguchi, M.-Z. Zhang, and R. C. Harris. Expression and Function of the Epidermal Growth Factor Receptor in Physiology and Disease. *Physiological Reviews*, 96(3):1025–1069, 2016.

[27] B. J. Mayer, M. L. Blinov, and L. M. Loew. Molecular machines or pleiomorphic ensembles: signaling complexes revisited. *Journal of biology*, 8(9):81, 2009.

[28] R. Suderman and E. J. Deeds. Machines vs. Ensembles: Effective MAPK Signaling through Heterogeneous Sets of Protein Complexes. *PLoS Computational Biology*, 9(10):e1003278, 2013.

[29] E. Bartocci and P. Lió. Computational Modeling, Formal Analysis, and Tools for Systems Biology. *PLoS Computational Biology*, 12(1):e1004591, 2016.

[30] D. Machado, R. S. Costa, M. Rocha, E. C. Ferreira, B. Tidor, and I. Rocha. Modeling formalisms in Systems Biology. *AMB Express, Springer Open Journal*, 1 (45), 2011.

[31] N. Tenazinha and S. Vinga. A Survey on Methods for Modeling and Analyzing Integrated Biological Networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(4):943–958, 2011.

[32] Z. Ji, K. Yan, W. Li, H. Hu, and X. Zhu. Mathematical and Computational Modeling in Complex Biological Systems. *BioMed Research International*, 2017:1–16, 2017.

[33] N. Le Novère. Quantitative and logic modelling of molecular and gene networks. *Nature Reviews Genetics*, 16(3):146–158, 2015.

[34] J. Baeten. A brief history of process algebra. *Theoretical Computer Science*, 335(2-3):131–146, 2005.

[35] F. Ciocchetta and J. Hillston. Process algebras in systems biology. In M. Bernardo, P. Degano, and G. Zavattaro, editors, *Formal Methods for Computational Systems Biology*, pages 265–312, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[36] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the pi-calculus process algebra. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing*, pages 459–70, 2001.

[37] M. L. Guerriero, C. Priami, and A. Romanel. Modeling static biological compartments with beta-binders. In H. Anai, K. Horimoto, and T. Kutsia, editors, *Algebraic Biology*, pages 247–261, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg.

[38] L. Dematté, R. Larcher, A. Palmisano, C. Priami, and A. Romanel. Programming Biology in BlenX. In S. Choi, editor, *Systems Biology for Signaling Networks*, chapter 31, pages 777–820. Springer-Verlag New York, 1 edition, 2010.

[39] V. Danos and C. Laneve. Formal Molecular Biology. *Theoretical Computer Science*, 325:69–110, 2004.

[40] B. Liu and P. S. Thiagarajan. Modeling and analysis of biopathways dynamics. *Journal of bioinformatics and computational biology*, 10(4), 2012.

[41] M. I. Stefan, T. M. Bartol, T. J. Sejnowski, and M. B. Kennedy. Multi-state Modeling of Biomolecules. *PLoS Computational Biology*, 10(9):e1003844, 2014.

[42] J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana. Internal coarse-graining of molecular systems. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6453–8, 2009.

[43] J. Feret, V. Danos, J. Krivine, R. Harmer, and W. Fontana. Internal coarse-graining of molecular systems (Supporting Information). *Proceedings of the National Academy of Sciences of the United States of America*, 106(16):6453–8, 2009.

[44] J. Krivine, V. Danos, and A. Benecke. Modelling Epigenetic Information Maintenance: A Kappa Tutorial. In A. Bouajjani and O. Maler, editors, *Computer Aided Verification*, pages 17–32, Berlin, Heidelberg, 2009. Springer-Verlag.

[45] J. Yang and W. S. Hlavacek. The efficiency of reactant site sampling in network-free simulation of rule-based models for biochemical systems. *Physical biology*, 8(5): 055009, 2011.

[46] N. Juty, R. Ali, M. Glont, S. Keating, N. Rodriguez, M. J. Swat, S. M. Wimalaratne, H. Hermjakob, N. Le Novère, C. Laibe, and V. Chelliah. BioModels: Content, Features, Functionality, and Use. *CPT: pharmacometrics & systems pharmacology*, 4(2):e3, 2015.

[47] T. Manninen, K. Hituri, E. Toivari, and M.-L. Linne. Modeling signal transduction leading to synaptic plasticity: evaluation and comparison of five models. *EURASIP journal on bioinformatics & systems biology*, 2011:797250, 2011.

[48] A. G. Nair, U. S. Bhalla, J. H. Kotaleski, and J. J. Saucerman. Role of DARPP32 and ARPP21 in the Emergence of Temporal Constraints on Striatal Calcium and Dopamine Integration. *PLOS Computational Biology PLoS Comput Biol*, 12(9), 2016.

[49] T. Nakano, T. Doi, J. Yoshimoto, and K. Doya. A kinetic model of dopamine- and calcium-dependent striatal synaptic plasticity. *PLoS Computational Biology*, 6(2):e1000670, 2010.

[50] M. Mattioni and N. Le Novère. Integration of Biochemical and Electrical Signaling-Multiscale Model of the Medium Spiny Neuron of the Striatum. *PLoS ONE*, 8(7):e66811, 2013.

[51] J. W. Bales, H. Q. Yan, X. Ma, Y. Li, R. Samarasinghe, and C. E. Dixon. The dopamine and cAMP regulated phosphoprotein, 32kDa (DARPP-32) signaling pathway: A novel therapeutic target in traumatic brain injury. *Experimental Neurology*, 229(2):300–307, 2011.

[52] J. Kim, I. S. Ryu, S. Y. Seo, and E. S. Choe. Activation of Protein Kinases and Phosphatases Coupled to Glutamate Receptors Regulates the Phosphorylation State of DARPP32 at Threonine 75 After Repeated Exposure to Cocaine in the Rat Dorsal Striatum in a Ca2+-Dependent Manner. *The international journal of neuropsychopharmacology / official scientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)*, 18(12):992–999, 2015.

[53] I. Buesa, Z. Aira, J. J. Azkue, J. Li, X. Zhang, and X. Chen. Regulation of Nociceptive Plasticity Threshold and DARPP-32 Phosphorylation in Spinal Dorsal Horn Neurons by Convergent Dopamine and Glutamate Inputs. *PLOS ONE*, 11(9):e0162416, 2016.

[54] NCBI Resource Coordinators. Database resources of the national center for biotechnology information. *Nucleic Acids Research*, 44(D1):D7, 2016.

[55] J. A. Marsh, B. Dancheck, M. J. Ragusa, M. Allaire, J. D. Forman-Kay, and W. Peti. Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators. *Structure (London, England : 1993)*, 18(9):1094–103, 2010.

[56] R. J. Beninger and T. V. Gerdjikov. Dopamine-Glutamate Interactions in Reward-Related Incentive Learning. In *Dopamine and Glutamate in Psychiatric Disorders*, pages 319–354. Humana Press, Totowa, NJ, 2005.

[57] K. Cho, M. H. Cho, J. H. Seo, J. Peak, K. H. Kong, S. Y. Yoon, and D. H. Kim. Calpain-mediated cleavage of DARPP-32 in Alzheimer's disease. *Aging Cell*, 14(5): 878–886, 2015.

[58] S. D. Philibin, A. Hernandez, D. W. Self, and J. A. Bibb. Striatal signal transduction and drug addiction. *Frontiers in Neuroanatomy*, 5(60), 2011.

[59] Y. Kunii, T. M. Hyde, T. Ye, C. Li, B. Kolachana, D. Dickinson, D. R. Weinberger, J. E. Kleinman, and B. K. Lipska. Revisiting DARPP-32 in postmortem human brain: changes in schizophrenia and bipolar disorder and genetic associations with t-DARPP-32 expression. *Molecular Psychiatry*, 19:192–199, 2014.

[60] H. Wang, M. Farhan, J. Xu, P. Lazarovici, W. Zheng, H. Wang, M. Farhan, J. Xu, P. Lazarovici, W. Zheng, H. Wang, M. Farhan, J. Xu, P. Lazarovici, and W. Zheng. The involvement of DARPP-32 in the pathophysiology of schizophrenia. *Oncotarget*, 5(0), 2015.

[61] M. Yger and J.-A. Girault. DARPP-32, Jack of All Trades... Master of Which? *Frontiers in Behavioral Neuroscience*, 5(56), 2011.

[62] M. Lindskog, M. Kim, M. A. Wikström, K. T. Blackwell, and J. H. Kotaleski. Transient calcium and dopamine increase PKA activity and DARPP-32 phosphorylation. *PLoS computational biology*, 2(9):e119, 2006.

[63] O. Gutierrez-Arenas, O. Eriksson, and J. H. Kotaleski. Segregation and crosstalk of D1 receptor-mediated activation of ERK in striatal medium spiny neurons upon acute administration of psychostimulants. *PLoS computational biology*, 10(1):e1003445, 2014.

[64] A. G. Nair, O. Gutierrez-Arenas, O. Eriksson, A. Jauhiainen, K. T. Blackwell, and J. H. Kotaleski. Modeling intracellular signaling underlying striatal function in health and disease. *Progress in Molecular Biology and Translational Science*, 123:277–304, 2014.

[65] J. Feret and J. Krivine. KaSim3 reference manual, 2012.

[66] V. Danos, H. Koeppl, and J. Wilson-Kanamori. Cooperative assembly systems. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 6937 LNCS:1–20, 2011.

[67] J. A. P. Sekar and J. R. Faeder. Rule-based modeling of signal transduction: a primer. *Methods in molecular biology (Clifton, N.J.)*, 880:139–218, 2012.

[68] B. Dancheck, A. C. Nairn, and W. Peti. Detailed structural characterization of unbound protein phosphatase 1 inhibitors. *Biochemistry*, 47(47):12346–56, 2008.

[69] L. Mollica, L. M. Bessa, X. Hanoulle, M. R. Jensen, M. Blackledge, and R. Schneider. Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment. *Frontiers in molecular biosciences*, 3:52, 2016.

[70] O. Engmann, A. Giralt, N. Gervasi, L. Marion-Poll, L. Gasmi, O. Filhol, M. R. Picciotto, D. Gilligan, P. Greengard, A. C. Nairn, D. Hervé, and J.-A. Girault. DARPP-32 interaction with adducin may mediate rapid environmental effects on striatal neurons. *Nature communications*, 6:10099, 2015.

[71] M. S. Choy, R. Page, and W. Peti. Regulation of protein phosphatase 1 by intrinsically disordered proteins. *Biochemical Society transactions*, 40(5):969–74, 2012.

[72] A. J. Lotka. Analytical note on certain rhythmic relations in organic systems. *Proceedings of the National Academy of Sciences*, 6(7):410–415, 1920.

[73] A. L. Hodgkin and A. F. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544, 1952.

[74] J. Lisman. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. *Proceedings of the National Academy of Sciences of the United States of America*, 86(23):9574–8, 1989.

[75] M. K. Morris, J. Saez-Rodriguez, P. K. Sorger, and D. A. Lauffenburger. Logic-based models for the analysis of cell signaling networks. *Biochemistry*, 49(15):3216–24, 2010.

[76] C. Chaouiya. Petri net modelling of biological networks. *Briefings in bioinformatics*, 8(4):210–9, 2007.

[77] B. N. Kholodenko. Cell-signalling dynamics in time and space. *Nature reviews. Molecular cell biology*, 7(3): 165–76, 2006.

[78] J.-J. Tapia and J. R. Faeder. The Atomizer: Extracting Implicit Molecular Structure from Reaction Network Models. *Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics (BCB'13)*, pages 726–727, 2013.

[79] F. Sacco, L. Perfetto, L. Castagnoli, and G. Cesareni. The human phosphatase interactome: An intricate family portrait. *FEBS letters*, 586(17):2732–9, 2012.

[80] D. T. Gillespie. Stochastic simulation of chemical kinetics. *Annual review of physical chemistry*, 58:35–55, 2007.

[81] C. A. Lo, I. Kays, F. Emran, T. J. Lin, V. Cvetkovska, and B. E. Chen. Quantification of Protein Levels in Single Living Cells. *Cell Reports*, 13(11):2634–2644, 2015.

[82] N. Otmakhov and J. Lisman. Measuring CaMKII concentration in dendritic spines. *Journal of neuroscience methods*, 203(1):106–14, 2012.

[83] J. A. Girault, H. C. J. Hemmings, K. R. Williams, A. C. Nairn, and P. Greengard. Phosphorylation of DARPP-32, a Dopamine- and cAMP-regulated Phosphoprotein, by Casein Kinase II. *Journal of Biological Chemistry*, 264(36):21748–21759, 1989.

[84] P. Svenningsson, A. Nishi, G. Fisone, J.-A. Girault, A. C. Nairn, and P. Greengard. DARPP-32: an integrator of neurotransmission. *Annual review of pharmacology and toxicology*, 44:269–96, 2004.

[85] A. Nishi, M. Matamales, V. Musante, E. Valjent, M. Kuroiwa, Y. Kitahara, H. Rebholz, P. Greengard, J.-A. Girault, and A. C. Nairn. Glutamate Counteracts Dopamine/PKA Signaling via Dephosphorylation of DARPP-32 Ser-97 and Alteration of Its Cytonuclear Distribution. *The Journal of biological chemistry*, 292(4): 1462–1476, 2017.

[86] A. Stipanovich, E. Valjent, M. Matamales, A. Nishi, J.-H. Ahn, M. Maroteaux, J. Bertran-Gonzalez, K. Brami-Cherrier, H. Enslen, A.-G. Corbillé, O. Filhol, A. C. Nairn, P. Greengard, D. Hervé, and J.-A. Girault. A phosphatase cascade by which rewarding stimuli control nucleosomal response. *Nature*, 453(7197):879–884, 2008.

Supplementary Material: Comparison of rule- and ordinary differential equation-based dynamic model of DARPP-32 signalling network

RB	ODE	Definition
cAMP*	cAMP	cAMP binding unspecified
free_Ca*	free_Ca	Ca^{2+} unbound
all_Ca*	all_Ca	Ca^{2+} binding unspecified
PKA*	PKA	PKA binding unspecified
CDK5_*	_CDK5	CDK5 bound
CK1u*	CK1u	CK1 unphosphorylated, binding unspecified
PP2Ap*	PP2Ap	PP2A phosphorylated, all bindings unspecified
PP2ACa*	PP2ACa	PP2A bound to Ca^{2+}, phosphorylation and other bindings unspecified
PP2C_*	_PP2C	PP2C bound
PP2Bactive*	PP2Bactive	PP2B active, binding unspecified
PDEp*	PDEp	PDE phosphorylated, binding unspecified
D*	D	DARPP-32 unphosphorylated at all sites, binding unspecified
D34*	D34	DARPP-32 phoshophorylated at Thr34 with unspecified binding, other sites' internal states and binding unspecified
D75*	D75	DARPP-32 phoshophorylated at Thr75 with unspecified binding, other sites' internal states and binding unspecified
D137*	D137	DARPP-32 phoshophorylated at Ser137 with unspecified binding, other sites' internal states and binding unspecified

Table S1: Names of RB observables and corresponding names of ODE observables with definitions. To obtain observable ODEs, the time series of the corresponding molecular species are summed based on their names.

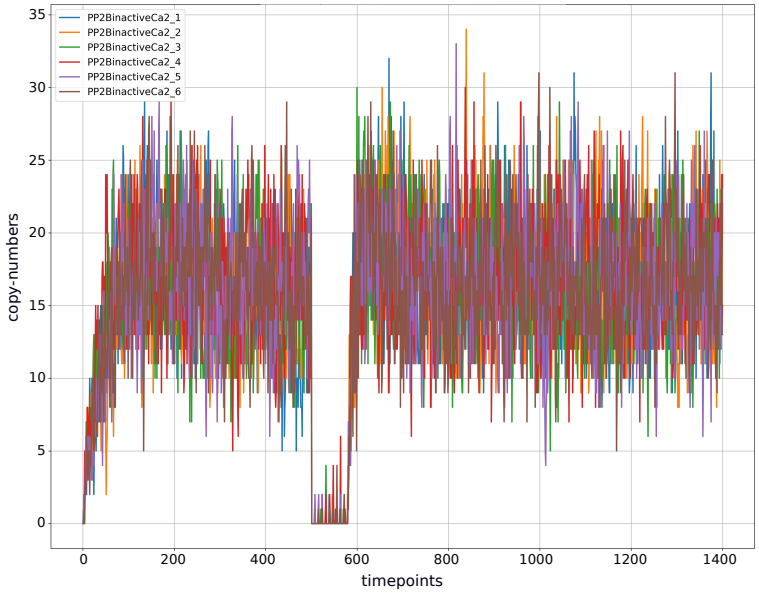

Figure S1: A half-active PP2B is a complex composed of PP2B and two Ca^{2+} ions. The RB model simulation generates six different molecular species that represent this complex due to a combinatorial binding of Ca^{2+} ions to 4 identical sites of PP2B. The plot shows superimposed trajectories of these six variants of the half-active PP2B. Neither of these six trajectories differentiates itself from others by a pattern of dynamics nor an average level of abundances.

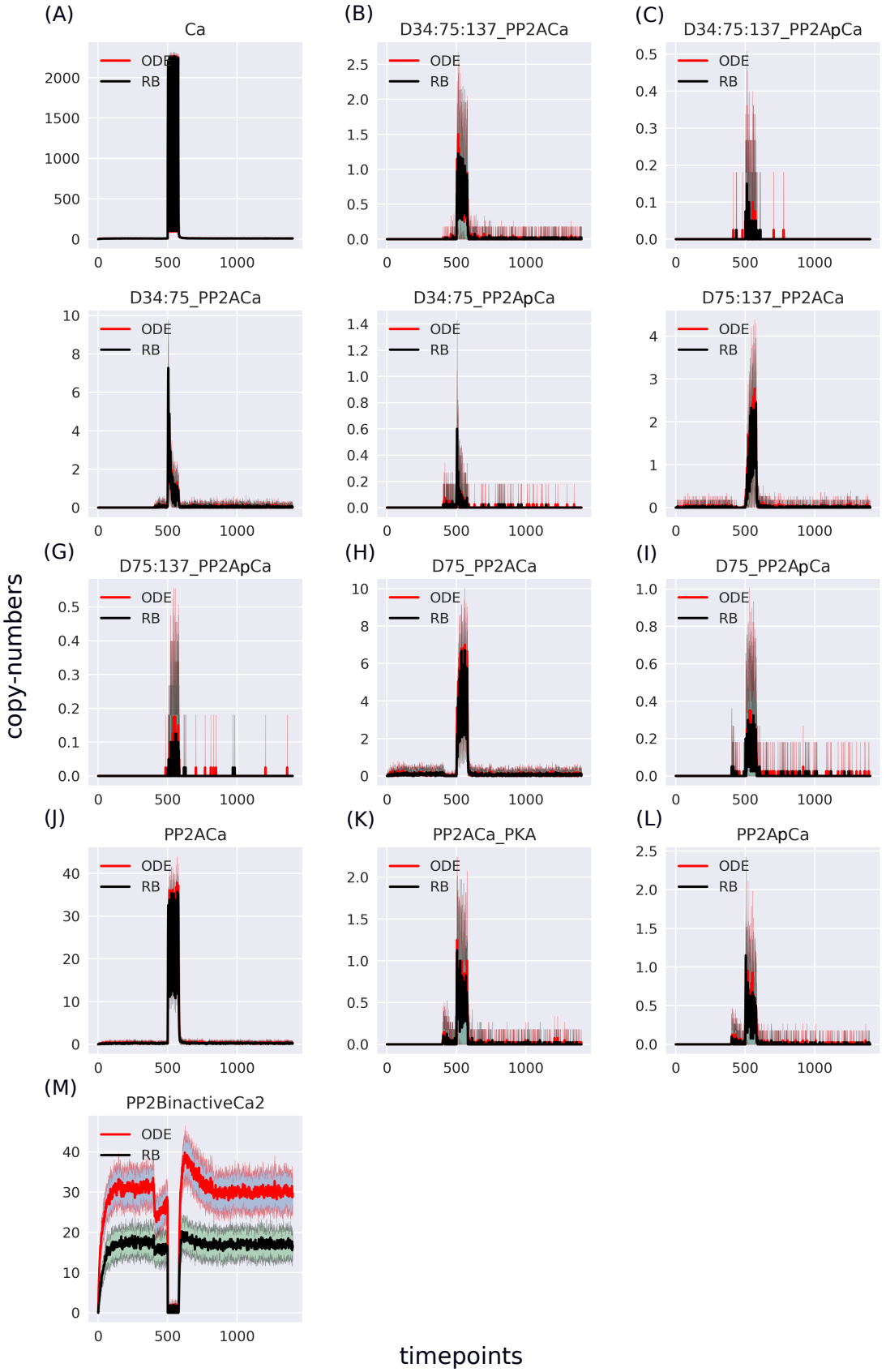

FIGURE S2: Separated molecular species containing Ca^{2+}, selected as in the original model. The ^textsf‘PP2BinactiveCa2” trajectory of the RB model was obtained by selecting one of 6 entities representing the inactive form of PP2B in the RB model. There is still a discrepancy between the models, but the trajectory is lower for the RB model.

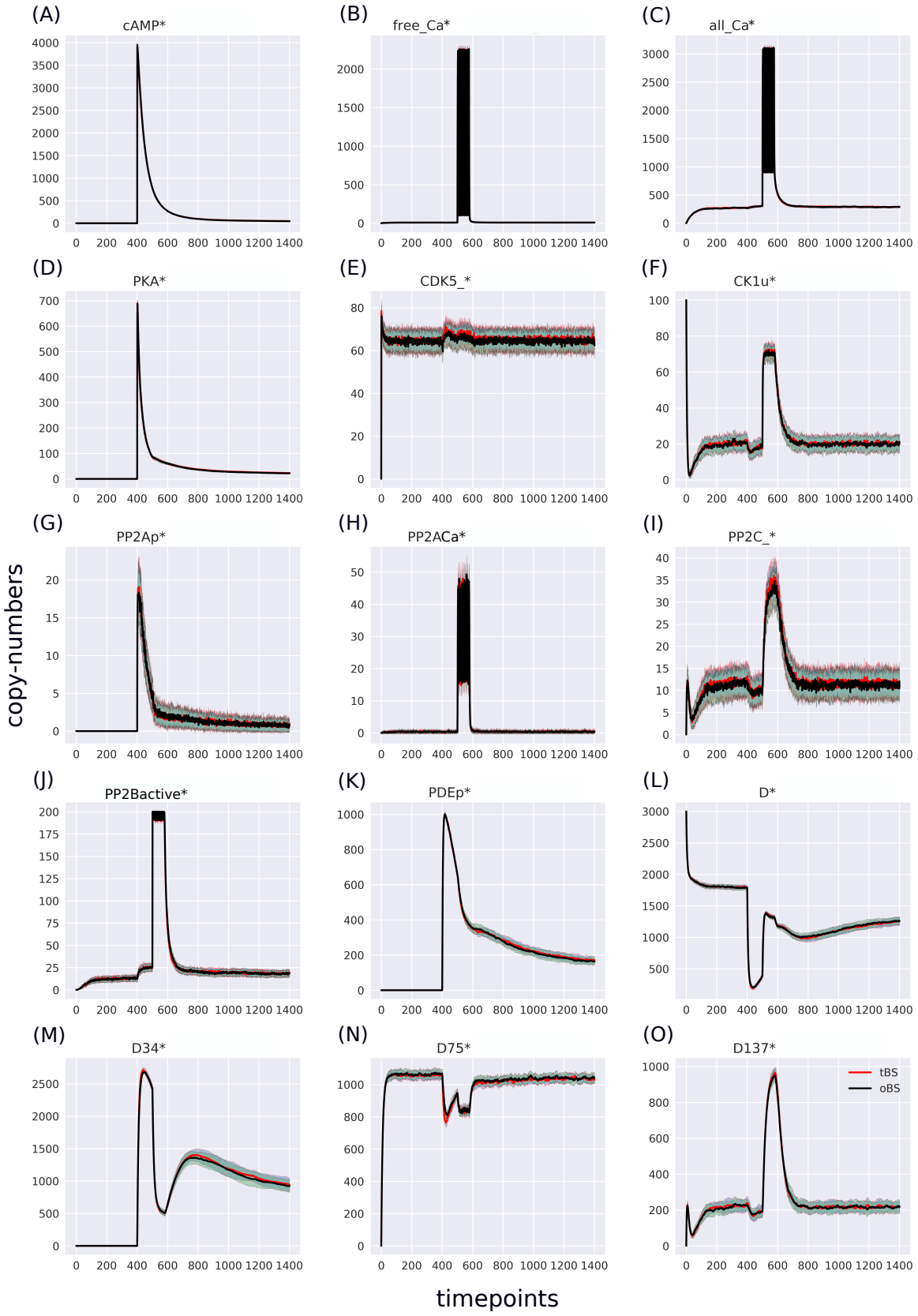

Figure S3: Comparison of two variants of the RB model in which the agent representing DARPP-32 had one binding site (oBS, red trace) and three binding sites (tBS, black trace). Overlaid trajectories of corresponding agents demonstrate no effect of this modification on the model trajectories.

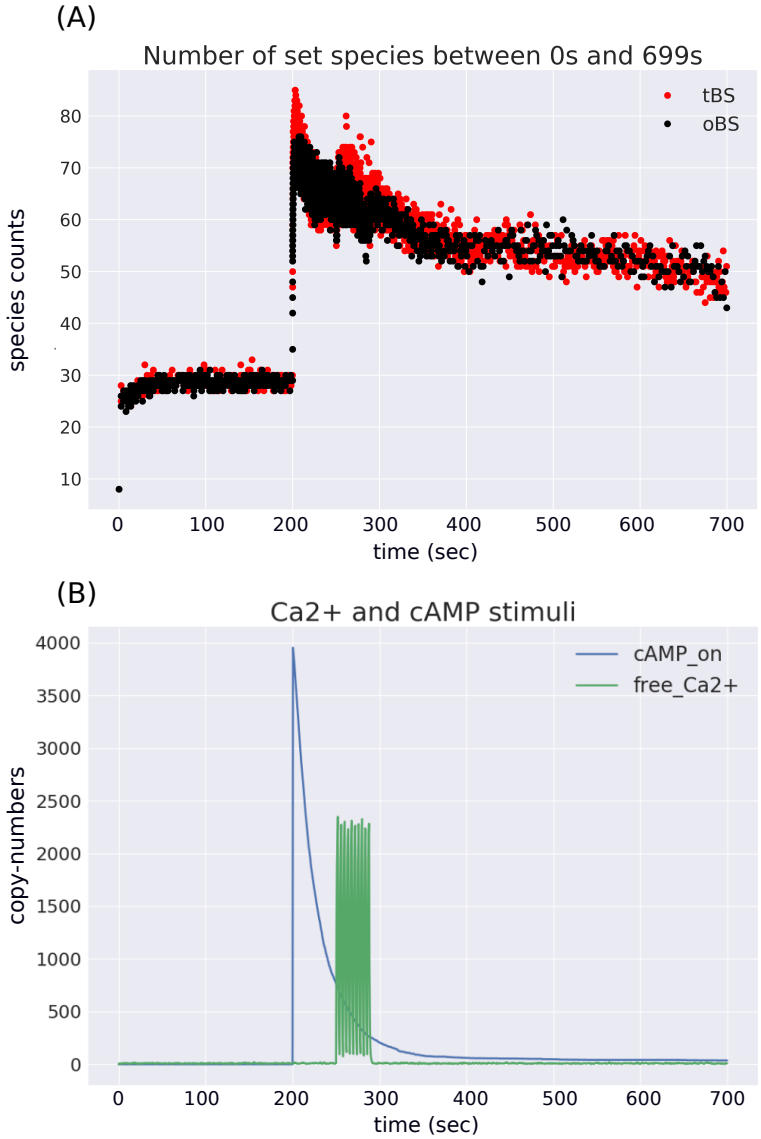

FIGURE S4: Overlay of change in species counts overtime for two models where DARPP-32 can bind at a single site (oBS) and three sites (tBS) (A). The species set size is similar in both model variants. The change in the number of unique species counts is aligned with trajectories of the stimuli (B). The dynamics of complex formation are dictated by the pattern of stimulus introduction, as the greatest differences between oBS and tBS occur during stimulus application.