

Intrinsic Noise Analyzer: A Software Package for the Exploration of Stochastic Biochemical Kinetics Using the System Size Expansion

Philipp Thomas^{1,2,3,9}, Hannes Matuschek^{4,9}, Ramon Grima^{1,2*}

1 School of Biological Sciences, University of Edinburgh, Edinburgh, United Kingdom, **2** SynthSys Edinburgh, University of Edinburgh, Edinburgh, United Kingdom, **3** Department of Physics, Humboldt University of Berlin, Berlin, Germany, **4** Institute of Physics and Astronomy, University of Potsdam, Potsdam, Germany

Abstract

The accepted stochastic descriptions of biochemical dynamics under well-mixed conditions are given by the Chemical Master Equation and the Stochastic Simulation Algorithm, which are equivalent. The latter is a Monte-Carlo method, which, despite enjoying broad availability in a large number of existing software packages, is computationally expensive due to the huge amounts of ensemble averaging required for obtaining accurate statistical information. The former is a set of coupled differential-difference equations for the probability of the system being in any one of the possible mesoscopic states; these equations are typically computationally intractable because of the inherently large state space. Here we introduce the software package intrinsic Noise Analyzer (iNA), which allows for systematic analysis of stochastic biochemical kinetics by means of van Kampen's system size expansion of the Chemical Master Equation. iNA is platform independent and supports the popular SBML format natively. The present implementation is the first to adopt a complementary approach that combines state-of-the-art analysis tools using the computer algebra system Ginac with traditional methods of stochastic simulation. iNA integrates two approximation methods based on the system size expansion, the Linear Noise Approximation and effective mesoscopic rate equations, which to-date have not been available to non-expert users, into an easy-to-use graphical user interface. In particular, the present methods allow for quick approximate analysis of time-dependent mean concentrations, variances, covariances and correlations coefficients, which typically outperforms stochastic simulations. These analytical tools are complemented by automated multi-core stochastic simulations with direct statistical evaluation and visualization. We showcase iNA's performance by using it to explore the stochastic properties of cooperative and non-cooperative enzyme kinetics and a gene network associated with circadian rhythms. The software iNA is freely available as executable binaries for Linux, MacOSX and Microsoft Windows, as well as the full source code under an open source license.

Citation: Thomas P, Matuschek H, Grima R (2012) Intrinsic Noise Analyzer: A Software Package for the Exploration of Stochastic Biochemical Kinetics Using the System Size Expansion. PLoS ONE 7(6): e38518. doi:10.1371/journal.pone.0038518

Editor: Jean Peccoud, Virginia Tech, United States of America

Received: April 12, 2012; **Accepted:** May 7, 2012; **Published:** June 12, 2012

Copyright: © 2012 Thomas et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The work was funded by SULSA (Scottish Universities Life Sciences Alliance) <http://www.sulsa.ac.uk/> and by SynthSys Edinburgh <http://www.csbe.ed.ac.uk/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Ramon.Grima@ed.ac.uk

⁹ These authors contributed equally to this work.

Introduction

Chemical kinetics is by its very nature stochastic. This stochasticity has several origins, chief among them being that spontaneous processes are responsible for the conformational changes which occur in unimolecular reactions while the process of bringing two molecules together to participate in a bimolecular reaction is Brownian motion [1]. This randomness is averaged out and hence non-apparent when the reactions under study involve a large number of molecules. This is the case of reactions occurring in test-tubes or in even larger systems. However inside cells, conditions are such that many species exist in low copy numbers [2]. The importance of stochasticity is particularly obvious in genetic regulatory networks since there are one or two copies of most genes per cell [3]. It is thus clear that stochastic modeling of intracellular networks is necessary to understand the complex biochemical processes underpinning a cell's response to both internal and external perturbations.

Current software implementations offer a broad range of stochastic modeling methods. Available packages can be divided into particle based descriptions and population based descriptions. Particle based methods adopt a microscopic approach that describes the movement of each individual reactant (non-solvent) molecule in space and time by means of Brownian dynamics. Popular software packages include the Greens Function Reaction-Diffusion algorithm [4], Smoldyn [5] and MCell [6]. Population based methods adopt a mesoscopic approach that retains the discreteness of reactants but does not need to simulate individual particle trajectory explicitly. This methodology, used by packages such as Smartcell [7] and MesoRD [8], is based on the reaction diffusion master equation [9,10]. The basic idea is to divide the reaction volume into smaller subvolumes, with reactions proceeding in each subvolume and molecules entering adjacent subvolumes by diffusion. Next one applies the well-mixed assumption to each subvolume (but not to the whole system) which implies that

we can ignore the positions and velocities of individual molecules inside each subvolume. The state of the system is then described by the number of molecules of each species in each subvolume, a description which is considerably reduced compared to that offered by particle based methods. This methodology relies on the knowledge of length scales over which the system is said to be spatially homogeneous [2].

A further reduced population description can be achieved by specifying to the situation in which the concentrations of interacting molecules are approximately spatially homogeneous over the entire reaction volume. Reaction kinetics is governed by two timescales: (i) the diffusion timescale, i.e., the time it takes for two molecules to meet each other and (ii) the reaction timescale, i.e., the time it takes for two molecules to react when they are in close proximity to each other. Concentration homogeneity over the entire compartment in which reactions occur, ensues when the reaction timescale is much larger than the diffusion timescale [2]. The large majority of available software packages, deterministic or stochastic, model this situation. Under such well-mixed conditions the Stochastic Simulation Algorithm (SSA) provides an accurate mesoscopic description of stochastic chemical dynamics. The SSA is a Monte Carlo technique by which one can simulate exact sample paths of the stochastic dynamics. The latter has been rigorously derived from microscopic physics by Gillespie for dilute well-mixed gases and solutions [11,12]. Over the past two decades, the popularization of the algorithm has led to its broad availability in many software packages (see Table 1). However in many situations of practical interest, the application of the SSA is computationally expensive mainly due to the two reasons: (i) whenever the fluctuations are large, e.g., the case of low copy numbers of molecules, a considerably large amount of ensemble averaging of the stochastic trajectories is needed to obtain statistically accurate results. (ii) the SSA simulates each reaction event explicitly which becomes computationally expensive whenever the copy number of at least one molecular species is large [1].

The chemical master equation (CME) is a mathematically equivalent and hence complementary description to the SSA [9]. The CME is a system of linear ordinary differential equations with an unbounded or a typically very large finite state space given by all combinations of copy numbers of the reactant molecules. Hence the advantage of the CME over the SSA is that it does not require any ensemble averaging and is not based on time-consuming simulation of individual reactions. However the CME does not lend itself easily to numerical or analytical computation, the reason being the large dimensionality of its state space. Hence to-date, software packages exploiting the utility of the CME have been scarce (see Table 1). Direct numerical integration of the CME is possible through the finite state projection method [13] which is implemented in the python package CmePy [14]. However, the state space grows exponentially with the number of species and hence these methods have limited applicability in biologically relevant situations. A different type of approach involves the calculation of the moments of the probability distribution solution of the CME by approximate means. Generally there exists an infinite hierarchy of coupled moment equations for reaction networks with bimolecular interactions. In order to make progress, a common method involves the truncation of the hierarchy by means of a moment-closure scheme. The software MomentClosure [15] implements the normal moment-closure approximation for mass action networks by setting all cumulants higher than a desired order to zero. A variety of alternative closures schemes are implemented in the package StochDynTools [16]. The advantage of these approaches is that they generally present quick ways to investigate the effects of noise without the need for averaging over many realizations of the stochastic process. However, these methods are based on ad hoc assumptions for the choice of the closure scheme and hence their accuracy and range of validity is often unknown.

In this article we introduce the software package, intrinsic Noise Analyzer (iNA), which enables a complementary approach using van Kampen's system size expansion (SSE) of the CME together

Table 1. Current software approaches for stochastic modeling.

Package	REs	Stochastic Simulation			CME			GUI	SBML	Ref
		Mean	Var	PDF	Mean	Var	PDF			
BioNetS	✓	✓	✓	✓	✓			✓	✓	[72]
Cain	✓	✓ _p	✓	✓	✓			✓	✓	[73,74]
CellMC		✓ _p							✓	[75]
Copasi	✓	✓			✓			✓	✓	[18]
Dizzy	✓	✓	✓					✓	✓	[76]
SimBiology	✓	✓ _p	✓	✓	✓			(2)	✓	[77]
StochKit		✓ _p	✓	✓	✓				(4)	[78]
StochPy	(5)	✓	✓	✓	✓			(3)	✓	[79]
CMEpy					✓ _{FSP}	✓	✓	✓	(3)	[14]
MomentClosure	✓				✓ _{MA}	✓	✓	(1)	✓	[15]
StochDynTools	✓				✓ _{MA}	✓	✓	(2)		[16]
iNA	✓	✓ _p	✓	✓	✓ _{SSE}	✓	✓	✓	✓	

✓_p implementation uses multi-core parallelism, ⁽¹⁾ Maple, ⁽²⁾ Matlab, ⁽³⁾ matplotlib, ⁽⁴⁾ using converter, ⁽⁵⁾ if used as plugin for PyScs [80], ✓_{FSP} method based on the Finite-State Projection algorithm, ✓_{MA} method based on moment approximation, ✓_{SSE} method based on system size expansion.

Existing software packages are divided in groups presenting implementations based on stochastic simulation or such based on the Chemical Master Equation. The software iNA combines both by using the system size expansion which has not been available in a software package yet. In particular, we also list whether the package allows for evaluation of mean concentrations, variances (Var) and the probability density function (PDF).

doi:10.1371/journal.pone.0038518.t001

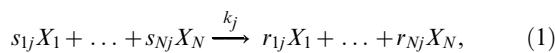
with traditional approaches of deterministic and stochastic simulation of chemical kinetics. The present implementation features the approximate computation of a selected number of moments of the probability distribution function solution of the CME by use of the Linear Noise Approximation (LNA) as well as the effective mesoscopic rate equations (EMREs). The advantage of these methods is that they are not based on ad-hoc assumptions like moment-closure approximations but rather they are based on the SSE which is a systematic expansion of the CME in powers of the inverse volume of the compartment in which reactions occur. The LNA provides the lowest order approximation to the second moments of the probability distribution (the variance and covariance of fluctuations) while the EMREs provide the first-order correction to the concentrations predicted by the deterministic rate equations (REs). iNA is the first software package to bridge the gap between deterministic simulation of chemical kinetics using rate equations, stochastic simulation of kinetics by the SSA and systematic analytic approximations of the CME. The novel combination of these complementary approaches makes iNA a valuable tool for the study of intrinsic noise in biological systems that has not yet been available to researchers and non-expert users in standalone software. The article is organized as follows. First we describe iNA's input specifications, the implemented methods and the features of the GUI. We then present the use of iNA to explore the dynamics of three models commonly encountered in biochemical kinetics: non-cooperative enzyme kinetics, cooperative enzyme kinetics and a gene network associated with circadian rhythms. Finally we discuss the design and implementation of the SBML parser and the methods, along with optimizations that increase the performance of the analysis. We complete the presentation by a derivation of the SSE-based methods at the heart of iNA.

The software is available as executable binaries for Linux (Fedora, Ubuntu, OpenSuse), MacOSX (10.5, 10.6, 10.7) and Microsoft Windows (7, XP) from <http://code.google.com/p/intrinsic-noise-analyzer>, as well as the full source code under the open source GPL2 license. The SBML files used in this article are available under the same URL.

Results

Input Format Specification

The general formulation of biochemical kinetics considers a number N of distinct chemical species confined in a mesoscopic volume of size Ω under well-mixed conditions. Species interact via R chemical reactions of the type



where j is the reaction index running from 1 to R , X_i denotes chemical species i , k_j is the reaction rate of the j^{th} reaction and s_{ij} and r_{ij} are the stoichiometric coefficients. We associate with each reaction a propensity function $\hat{a}_j(\vec{n}, \Omega)$ such that the probability for the j^{th} reaction to occur in the time interval $[t, t + dt)$ is given by $\hat{a}_j(\vec{n}, \Omega)dt$. The vector $\vec{n} = (n_1, \dots, n_N)^T$ denotes a mesoscopic state where n_i is the number of molecules of the i^{th} species. Note that our general formulation does not require all reactions to be necessarily elementary, i.e., unimolecular and bimolecular chemical reactions, but can also describe effective reactions. If the j^{th} reaction is elementary then its reaction rate k_j is a constant while if it is non-elementary the reaction rate is a function of the instantaneous concentrations, i.e., the elements of the vector \vec{n}/Ω .

This description of the biochemical reaction networks in terms of reactants, products, the associated stoichiometries and kinetic laws are part of the Systems Biology Markup Language (SBML) [17] which has become a standard representation of such networks. iNA natively supports SBML compatible with level 2 version 4 which makes the software versatile to work with models exchanged from other applications as Copasi [18], CellDesigner [19], SBML editor [20] or shorthand SMBL [21] as well as with some of the many models that are freely available in public databases [22]. The basic components of SBML parsed by iNA are definitions of units, compartments, species and reactions. iNA reads all SBML files that describe reaction networks as defined by the reaction scheme (1) along with the associated propensities. The former is obtained from SBML's reactant and product stoichiometry definitions while the latter is parsed from the "KineticLaw" construct. Species can be specified in terms of both amount (*mol*, molecule numbers and derived units) and concentration (*molar*, number concentrations and derived units).

The validation of models considered suitable for stochastic analysis requires the software to make several restrictions on SBML model definitions. Currently, events and explicit time dependent rate parameters are not supported. Also, reactions defined by the SBML specific reversible attribute cannot be validated for stochastic models due to ambiguities in the associated "KineticLaw". We refer users to the software package Copasi which allows for convenient conversion between SBML's reversible and irreversible reaction definitions. For consistency, the definition of a species must conform all species to be defined non-constant and free of algebraic constraints. The implementation of iNA assumes that such constraints arise naturally from the stoichiometry of the reaction network. Furthermore it is required for all parameters to be evaluated before runtime of simulations. Therefore the software does not allow parameters to be defined by ODEs or assignment rules. The software package gives the appropriate error messages in cases where one of the above specifications is not met.

Stochastic Simulation

Over the past two decades the SSA has enjoyed widespread popularity mainly because of the ease by which one can simulate stochastic reaction networks [23,24]. Given that the system is in state \vec{n} at time t , Gillespie proved using the laws of probability [25] that the probability per unit time for the j^{th} reaction to occur at time $t + \tau$ is

$$\pi(\tau, j | \vec{n}, t) = \hat{a}_j(\vec{n}, \Omega) \exp\left(-\sum_{i=1}^R \hat{a}_i(\vec{n}, \Omega)\tau\right). \quad (2)$$

The SSA generates a stochastic trajectory of the kinetics by sampling a reaction index j and a reaction time τ according to Eq. (2), followed by an update of the population size $n_i(t + \tau) = n_i(t) + S_{ij}$ for every species i . Note that the net change of the molecule number of species i by reaction j is given by the stoichiometric matrix $S_{ij} = r_{ij} - s_{ij}$. Despite its popularity, stochastic simulation has two major shortcomings. Firstly, simulations have to be carried out a significantly large number of times because of the considerable amount of independent realizations needed to obtain accurate statistical averages. Secondly, simulations can become quite slow when the population number of any molecular species is large [1]. Stochastic simulation is a basic component of the software iNA with support for simultaneous

simulation of independent realizations using shared memory parallelism of the OpenMP standard [26]. The software features two implementations of the SSA via the direct and the optimized direct method [25,27]. The output data is presented in terms of mean concentrations, variances and correlations as a function of time which allow for direct statistical interpretation.

The Chemical Master Equation and the System Size Expansion

An equivalent formulation for the stochastic reaction network described by Eq. (1) is the CME which can be derived from combinatorial arguments [9,10,28] and or from microphysics [11,12]. The CME gives the time-evolution equation for the probability $P(\vec{n},t)$ that the system is in a particular mesoscopic state $\vec{n}=(n_1,\dots,n_N)^T$,

$$\frac{\partial P(\vec{n},t)}{\partial t} = \sum_{j=1}^R \left(\prod_{i=1}^N E_i^{-S_{ij}} - 1 \right) \hat{a}_j(\vec{n},\Omega) P(\vec{n},t). \quad (3)$$

Here we have introduced the step operator $E_i^{-S_{ij}}$ which is defined by its action on a general function of molecular populations as $E_i^{-S_{ij}} g(n_1,\dots,n_i,\dots,n_N) = g(n_1,\dots,n_i - S_{ij},\dots,n_N)$ [9]. The CME is equivalent to a set of coupled differential-difference equations for each possible mesoscopic state, i.e., each combination of reactant molecule numbers. Typically, the large number of such states makes the CME intractable for numerical and analytical computation. The software package iNA uses an alternative approach based on van Kampen's SSE of the CME which is applicable whenever the dynamics of the reaction network is monostable [9]. In brief, the method constitutes a large volume expansion whose successive terms can be used to approximate the moments of the probability density function to any desired accuracy. Thereby it is implicit that whenever the reaction volume is large (or equivalently the molecular populations are large at constant concentration) the average concentrations can be approximated by the macroscopic REs,

$$\frac{\partial [\vec{X}]}{\partial t} = \underline{S} \vec{f}([\vec{X}]), \quad (4)$$

which are exactly the same as those used in deterministic models of biochemical kinetics. Note that $[\vec{X}]$ is the vector of macroscopic concentrations and \vec{f} is the macroscopic rate function vector, see Methods section. Note that matrices are underlined throughout the article.

The leading order term of the SSE is given by the LNA which has been the key tool in analytical studies of noise [9,29,30]. The merit of the method is that it provides a simple means of calculating the fluctuations about the concentration solution of the REs. In particular one is typically interested in the covariance of the time-dependent concentrations

$$\underline{\Sigma}(t) = \left\langle \left(\frac{\vec{n}(t)}{\Omega} - [\vec{X}(t)] \right) \left(\frac{\vec{n}(t)}{\Omega} - [\vec{X}(t)] \right)^T \right\rangle, \quad (5)$$

where the angled brackets denote the statistical average. Within the LNA the elements of the covariance matrix are determined by the time-dependent equation

$$\frac{\partial \underline{\Sigma}}{\partial t} = \underline{J} \underline{\Sigma} + \underline{\Sigma} \underline{J}^T + \Omega^{-1} \underline{D}. \quad (6)$$

Note that the matrix \underline{J} is the Jacobian which gives the extent by which small perturbations of the REs, Eq. (4), decay. The matrix \underline{D} is the diffusion matrix which quantifies the size of the perturbation due to intrinsic noise. Both matrices can be constructed from the stoichiometric coefficients and the macroscopic rate function vector \vec{f} . The diagonal elements of $\underline{\Sigma}$ are the variances and hence determine the standard deviation of concentration fluctuations by

$$\sigma_i(t) = \sqrt{\Sigma_{ii}(t)}. \quad (7)$$

The off-diagonal elements are the covariances which determine Pearson's correlation coefficients between the concentration fluctuations of species X_i and X_j ,

$$\text{corr}(X_i, X_j)(t) = \frac{\Sigma_{ij}(t)}{\sigma_i(t) \sigma_j(t)}. \quad (8)$$

Considering higher terms of the expansion one can obtain corrections to the REs which stems from a coupling of the mean concentrations to the higher order moments of the concentration fluctuations. These corrections have been calculated for networks composed of elementary reactions by Grima [31]. The new time-evolution equations which are obtained from this analysis are called effective mesoscopic rate equations (EMREs) and they are here extended for general reaction networks composed of elementary and non-elementary reaction steps (see *Methods* section). The EMREs are given by

$$\frac{\partial \langle \frac{\vec{n}}{\Omega} \rangle}{\partial t} = \frac{\partial}{\partial t} [\vec{X}] + \underline{J} \left(\langle \frac{\vec{n}}{\Omega} \rangle - [\vec{X}] \right) + \vec{\Delta}, \quad (9)$$

where $\vec{\Delta}$ is a vector which describes the coupling of the mean concentrations to the variance and covariance of fluctuations in the concentrations. It depends on the macroscopic concentrations and the covariance matrix which can be obtained by the solution of Eq. (4) and (6), respectively. We refer the interested reader to the section *Methods* for the definition of the vector $\vec{\Delta}$ together with an explicit derivation of the prescribed methods. EMREs have been shown to accurately describe the mean concentration over a wide range of copy numbers [31–35]. Generally the predictions of the REs and the EMREs agree only for reaction networks of unimolecular reactions. For more general reaction networks such as those involving bimolecular or non-elementary reactions, EMREs provide finite-copy number corrections to the REs. It is to be kept in mind that EMRE's provide meaningful results if and only if the predicted mean concentrations for all species are positive; in practice this means that the EMRE is valid for reaction volumes above a certain breakdown volume which is system specific (see [35] for further discussion).

The present software allows the computation of the SSE methods either in time dependent conditions or at steady state. The former is obtained by numerical integration of the set of coupled ordinary differential equations, Eqs. (4), (6) and (9). The

latter uses the same set of equations with the time derivative set to zero and reduces them to a set of simultaneous algebraic equations which typically can be solved with less computational effort and higher numerical accuracy. For the first time, all of the above methods are made available to a broad audience by the software package iNA.

Features of the GUI

The software iNA aims at ease of use of analytical approximations that facilitate the exploration of stochastic effects in biochemical reaction networks. We have therefore focused on a minimal GUI composed of a model tree with table and plot views. Analyses tasks can be easily accessed through wizards which guide the user through configuration. The GUI is divided into a menu bar and a main window. After a model has been loaded using the menu bar it will be available in the list of “Open models” on the left hand side. The latter list is hierarchically organized as follows

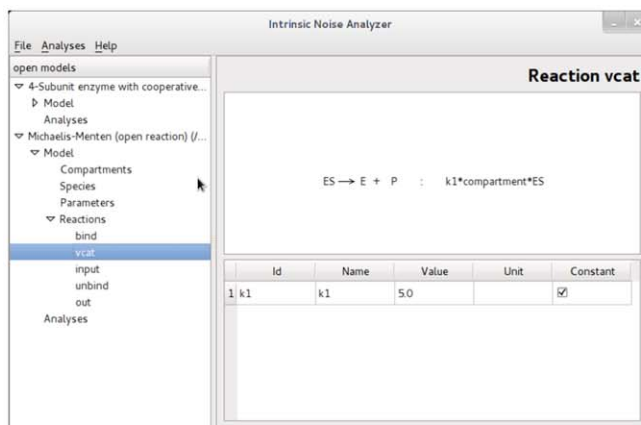
- *Model*, where the basic components of the model can be accessed. These include compartments, species, parameters and reactions; see Fig. 1a.
- *Analyses*, where the results of the Linear Noise Approximation, EMRE or SSA analyses can be accessed; see Fig. 1b.

The *Model* section contains a basic SBML reader for viewing the parameters that define the biological model and making sure that the SBML file has been parsed correctly. The *Analyses* section can be filled with items selected from the option *Analyses* in the menu bar. Currently there are three wizards available.

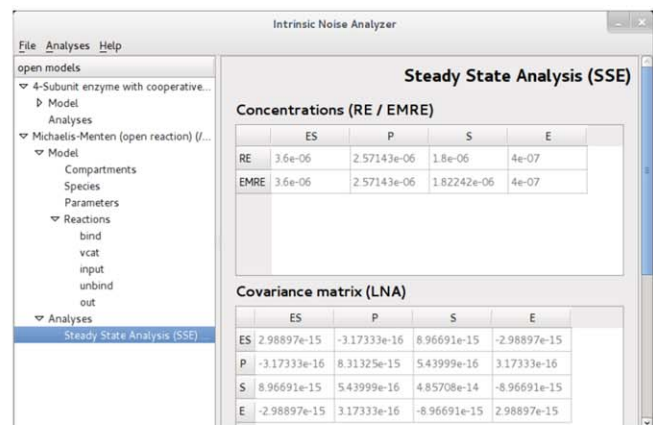
- a *Steady State Analysis (SSE)* wizard which customizes the computation of LNA, RE and EMRE in steady state conditions,
- a *Time Course Analysis (SSE)* wizard which customizes the computation of the LNA, RE and EMRE in time-dependent conditions,
- a *Stochastic Simulation Algorithm (SSA)* wizard guiding the initialization of stochastic simulations.

All wizards allow the user to select a subset of species to be analyzed, see Fig. 1c. The output of each analysis is simply a data table which can be saved to a text file or visualized by the predefined plot widgets.

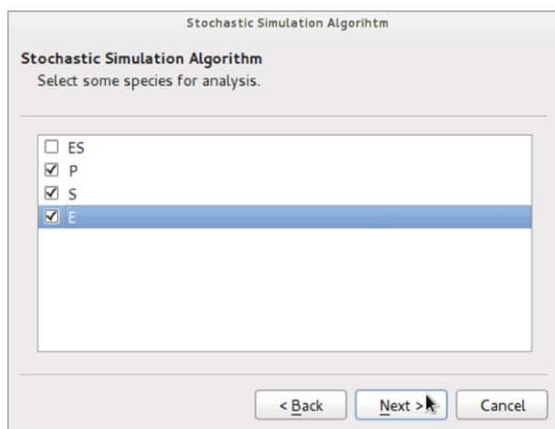
The *Steady State Analysis* is the most basic analysis provided by the program. It yields mean concentrations and the covariance matrix according to the LNA. Note that to the LNA level of approximation, the mean concentrations are the same as those obtained from solving the REs. The roots of the REs are computed



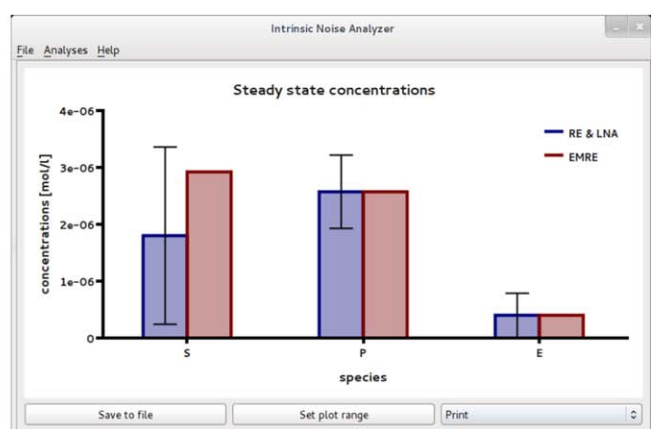
(a)



(b)



(c)



(d)

Figure 1. GUI of the software iNA. (a) *Model views* gives information on reactions, rate constants, propensities and species. (b) *Table views* provide the analysis results in an easy-to-read format. (c) iNA’s wizards allow for user friendly configuration of analyses. (d) *Plot views* visualize results in neat format. Note that SSE in (b) stands for system size expansion. doi:10.1371/journal.pone.0038518.g001

by iNA using the Newton-Raphson method with line search [36] and therefore it is required to specify precision and maximum number of iterations of the algorithm. iNA also outputs the mean concentration predictions according to EMREs [31]. These account for stochastic effects and hence are generally expected to be much closer to the true concentration prediction of the CME. Thereby one can obtain a quick estimate of the effect of noise on the reaction network. iNA offers convenient visualization of the outputs using the *Plot* option shown in Fig. 1d. This bar plot features two separate columns for the concentrations calculated by the REs and the EMREs of the individual species. The former is complemented with an error bar which indicates the standard deviation of the concentration fluctuations calculated using the LNA. The *Time Course Analysis* presents a wizard which is as simple to use as conventional integrators for deterministic REs by which one specifies the final time of integration, and the maximum relative and absolute errors. Therefore it is clear that all results obtained by the SSE should be checked for consistency with numerical integration carried out using smaller error estimates. The time course consists of theoretical estimates according to the deterministic REs and the corresponding fluctuations around it which have been estimated by the LNA at any point in time. At the same time the estimation of the mean concentrations using EMREs is shown and can be compared to its deterministic counterpart. The results can be accessed by a table view presenting the numerical data of the analysis or by selected plot views showing the mean concentrations and fluctuations computed by the REs and LNA, the correlation coefficients obtained from the LNA as a function of time or a comparison between the concentration predictions of EMREs and the REs.

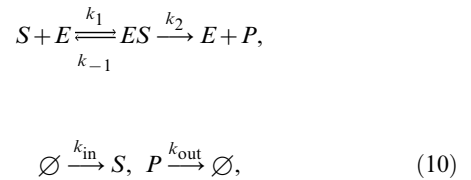
In order to validate the results, the *Stochastic Simulation* wizard offers the choice between two different implementations of the SSA (see section Design and Implementation) and allows the user to adjust the number of independent realizations that are used to calculate the statistical averages. This enables direct comparison of the simulation results with those obtained from the analysis using the SSE methods. All outputs are exportable to text files.

Applications

We showcase the utility of the present software by analyzing three models of biochemical kinetics: the Michaelis-Menten reaction [37], a multi-subunit enzyme with cooperative kinetics [38,39] and a gene network with negative feedback that has been proposed for circadian rhythms in *Drosophila* and *Neurospora* [40]. The SBML files that have been used in this section are listed in Table 2.

Michaelis-Menten Reaction with Substrate Input

The Michaelis-Menten reaction is a well studied example of biochemical kinetics. Over the past decade, stochastic models of the reaction have received considerable attention by means of analytical and stochastic simulation methods [23,24,32,41]. We here consider an embedded reaction mechanism that also accounts for flux conditions naturally found in living cells which comprises substrate input and product consumption reaction steps.



where S denotes the substrate species, E the free enzyme species, ES is the enzyme-substrate complex and P the product species. The k 's refer to the associated reaction rate constants. In what follows, we consider the same reaction occurring in two compartments characterized by two different length scales: the cellular scale which is the scale of large organelles such as mitochondria ($0.5 - 1 \mu m$) [42] and the scale of small sub-cellular compartments such as lipid rafts ($10 - 200 nm$) [43]. Two SBML files F1 and F2 (see Table 2) have been provided, one for each length scale. The rate constants are the same for both files and are shown in Table 3. These can also be conveniently accessed by the *Model view* of iNA, see Fig. 1a. Note that these rate constants are obtained from an experimental study of the enzyme Malate dehydrogenase [44].

Michaelis-Menten kinetics on the cellular scale. We here consider the reaction scheme (10) to take place in a compartment of volume $\Omega = 0.5 fl$ (femtoliters) which corresponds roughly to the size of a bacterium or a large organelle [45]. We now analyze the steady-state stochastic properties of the reaction by means of the LNA and the EMRE.

Table 3. Rate constants for Michaelis-Menten kinetics with substrate input (SBML files F1 and F2).

k_{in}	$1.8 \times 10^{-5} M/s$	k_{out}	$7s^{-1}$
k_1	$5 \times 10^7 (Ms)^{-1}$	k_{-1}	$5s^{-1}$
k_2	$5s^{-1}$		

doi:10.1371/journal.pone.0038518.t003

Table 2. SBML model definition files that have been used in this article.

F1	enzymekinetics1.xml	Enzyme with Michaelis-Menten kinetics, $\Omega = 0.5 fl$.
F2	enzymekinetics2.xml	Enzyme with Michaelis-Menten kinetics, $\Omega = 0.01 fl$
F3	coopkinetics1.xml	Multi-subunit enzyme with cooperative kinetics, $\Omega = 1 fl$
F4	coopkinetics2.xml	Multi-subunit enzyme with cooperative kinetics, $\Omega = 0.1 fl$
F5	coremodel1.xml	Circadian clock model, $\Omega = 2 fl$, weak negative feedback
F6	coremodel2.xml	Circadian clock model, $\Omega = 0.2 fl$, weak negative feedback
F7	coremodel3.xml	Circadian clock model, $\Omega = 0.2 fl$, strong negative feedback

All files and the software iNA are available from the URL <http://code.google.com/p/intrinsic-noise-analyzer>.
doi:10.1371/journal.pone.0038518.t002

The SBML file F1 specifies an initial condition of about 1200 enzyme molecules in non-complex form, which corresponds to a total enzyme concentration of $4 \mu\text{M}$ (micromolar); the substrate and product concentrations are initially zero. The model definition can be opened in iNA to perform the *Steady State Analysis*. The output generated is a table view which is shown in Fig. 1b that can also be exported to a text file as well as visualized by a plot (see Fig. 2a). We obtained values for the substrate and product concentrations to be $1.8 \mu\text{M}$ and $2.6 \mu\text{M}$ according to the REs. The table view gives further information on the intrinsic fluctuations in terms of the covariance matrix in steady state. We have computed the coefficient of variation (CV), a non-dimensional quantity which measures the inverse signal-to-noise ratio. The CV of species X is defined as $CV(X) = \sigma_X/[X]$, where σ_X is the standard deviation defined by Eq. (7). We obtain $CV(S) = 0.122$ and $CV(P) = 0.035$. Note that the EMRE concentrations obtained from the *Steady State Analysis* of iNA are very close to those predicted by the REs. This suggests that the LNA predictions of CV are accurate.

We verified the accuracy of the LNA predictions by obtaining 1000 independent realizations using the *Stochastic Simulation* wizard; the result is shown in Fig. 2b. The figure shows that the concentrations have reached steady state after about 5s. We then exported the data to a file and averaged over the output for $t \geq 5s$. The resulting mean concentrations are given by $[S] = 1.8 \mu\text{M}$ and $[P] = 2.6 \mu\text{M}$ with mean variances of $5.0 \times 10^{-2} (\mu\text{M})^2$ and $8.2 \times 10^{-3} (\mu\text{M})^2$, respectively. These values correspond to CV of about 0.124 and 0.035 which are in excellent agreement with the predictions of the LNA computed by iNA.

Michaelis-Menten kinetics in a small intracellular compartment. Next we study the same reaction in a compartment of reduced volume $\Omega = 0.01fl$ which roughly corresponds to a spherical volume of diameter $270nm$. The enzyme concentrations correspond to a total copy number of only 24 molecules. In such a small compartment, the average substrate concentrations can be greatly enhanced by intrinsic noise. The *Steady State Analysis* performed by iNA is summarized in Fig. 3(a). The substrate concentration is $1.8 \mu\text{M}$ according to the RE and $2.9 \mu\text{M}$ according to the EMRE. This implies a noise-induced enhancement of the concentration by about 60%. Note for all other concentrations – those of enzyme, complex and product – the predictions of the RE and EMRE theory agree exactly and thus

are independent of the volume of the compartment. These phenomena have been found earlier by Grima [32] using the SSE and are well reproduced by the software iNA.

We verify these predictions by stochastic simulation; the result is shown in Fig. 3b. We can now compare this result to the time course obtained from the REs and the LNA, see Fig. 3c. A comparison of the latter two subfigures reveals that the approach to steady state as obtained from SSA simulations is significantly slower than the one predicted by the macroscopic REs. Note also that the variance of the substrate concentration of the SSA is considerably larger than that predicted by the LNA. However we see that the time course of the stochastic simulation in Fig. 3b is well reproduced by the EMRE, see Fig. 3d. We exported the data from stochastic simulation to a text file and performed a time average over concentrations for which $t \geq 10s$. The results are $3.0 \mu\text{M}$ and $2.6 \mu\text{M}$ for substrate and product concentrations, respectively, which are in excellent agreement with the predictions of the EMREs. We have also estimated the coefficient of variation by stochastic simulations to be $CV(S) = 0.78$. Note that in this case the LNA predicts a value of 0.86 which overestimates the size of intrinsic noise by about 10%.

Comparing Figs. 3b and 3c, shows that the relative magnitudes of the concentrations of substrate and product are reversed when noise is taken into account. In contrast, a comparison of Figs. 2a and 2b, shows that the relative magnitudes of the concentrations of substrate and product remain the same even when noise is taken into account. This indicates that a noise-induced concentration inversion effect occurs below some critical compartment volume, a phenomenon which has recently been described by Ramaswamy et al. for the trimerization reaction with bursty input and for a two gene circuit with negative feedback [35]. In the section *Methods*, we use EMREs to calculate the theoretical value for the critical volume at which the concentrations of substrate and product become equal. Above the critical volume, the product concentration is larger than that of the substrate and hence the REs are qualitatively correct. Below the critical volume, the substrate concentration is larger than that of the product and REs are then qualitatively incorrect. The critical volume is calculated to be $0.0146fl$ for the enzyme Malate dehydrogenase considered here. Note that the compartment volume used in this example ($0.01fl$) is smaller than the critical volume whilst the volume used in the previous example ($0.5fl$) is much larger. This explains why the concentration inversion effect is only observed in the current example.

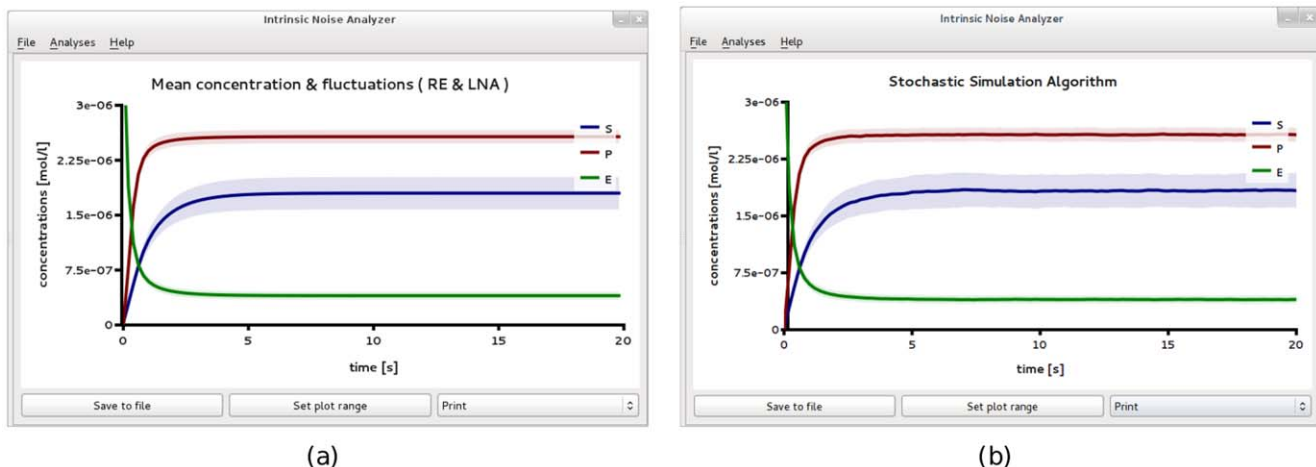


Figure 2. Michaelis-Menten kinetics in a large compartment of volume $\Omega = 0.5fl$. Plots showing the results of *Linear Noise Approximation* (a) and *Stochastic simulation* using an ensemble of independent 1,000 realizations (b). The two are in excellent agreement. Both figures have been obtained from iNA's analysis of the SBML file F1. doi:10.1371/journal.pone.0038518.g002

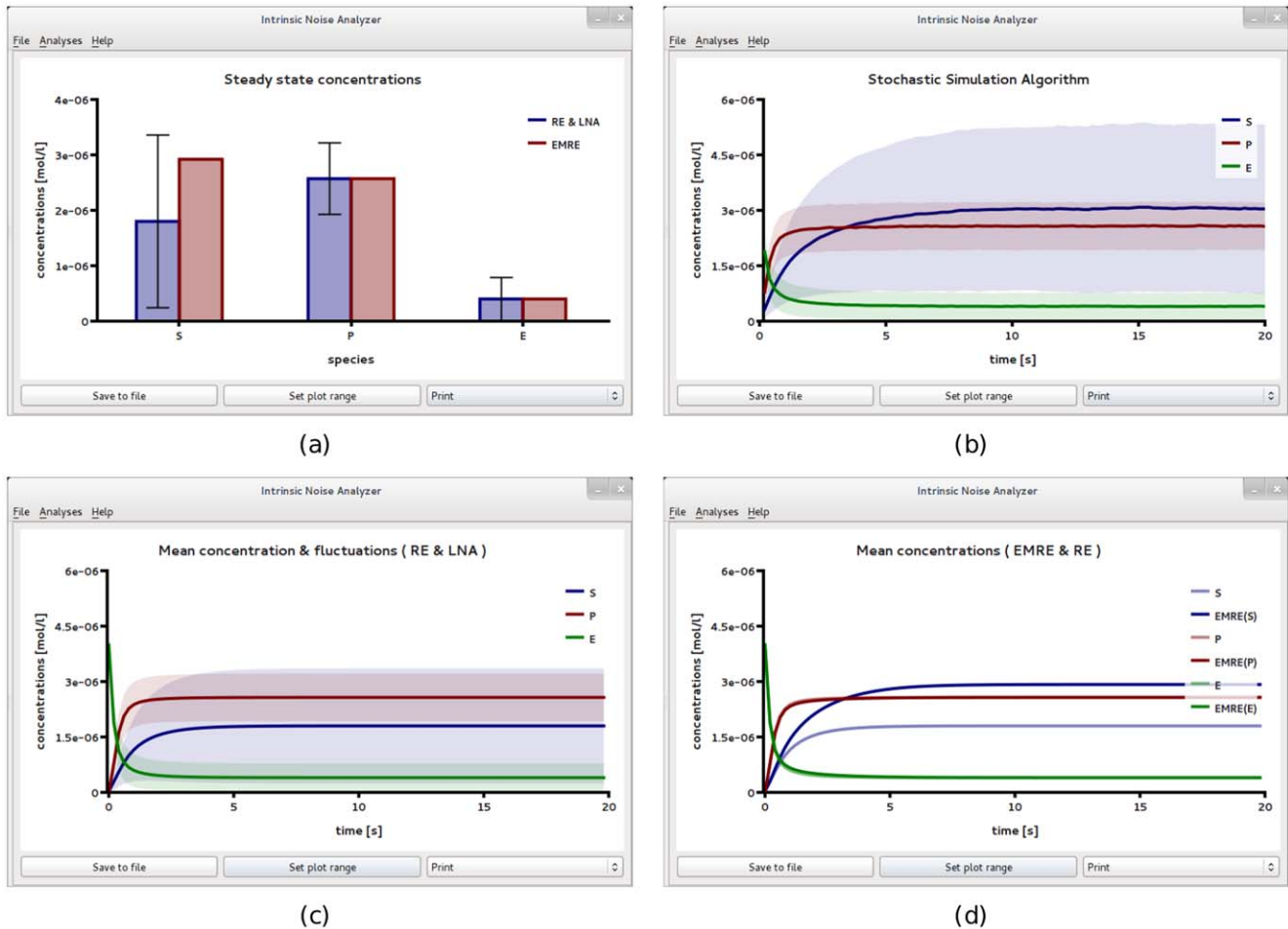
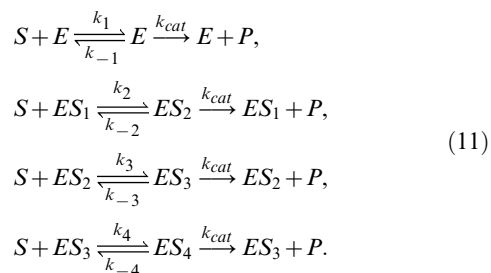


Figure 3. Michaelis-Menten kinetics in a small compartment volume $\Omega = 0.01 \text{ fl}$. (a) Plot of iNA's *Steady State Analysis* shows amplified EMRE substrate concentrations in comparison with those predicted by the REs. This conclusion is supported by stochastic simulations (b) using an ensemble size of 10,000 realizations which are in excellent agreement with steady state and time course predictions shown in (a) and (d), respectively. (c) For comparison we show the result of the LNA time course which fails to accurately predict both the mean substrate concentrations and the variance of fluctuations about them. The figures have been obtained from iNA's analysis of the SBML file F2. doi:10.1371/journal.pone.0038518.g003

Cooperative Enzyme Kinetics

Apart from Michaelis-Menten kinetics, it is often found that many enzymes exhibit positive cooperativity in substrate binding. Such behavior is a kinetic signature of enzymes with multiple interdependent binding sites. As a consequence, experimental ligand binding curves show a sigmoidal dependence on the abundance of specific substrates [37,46]. Tyson [38] used REs to study the deterministic cooperative kinetics of a two subunit enzyme catalyzed system (an enzyme with two binding sites). Here we extend the model to involve four subunits and study its stochastic properties using iNA. The reaction scheme reads



We enforce the condition $K_{M1} \gg K_{M2} \gg K_{M3} \gg K_{M4}$ where $K_{Mi} = (k_{-i} + k_{cat})/k_i$ is the Michaelis-Menten constant of the i^{th} enzyme reaction. This condition ensures that the binding of a substrate molecule to the enzyme molecule with n sites already occupied by other substrate molecules occurs quicker than substrate binding to an enzyme molecule with $n-1$ sites occupied; hence the phenomenon of positive cooperativity. In order to account for flux conditions that are found naturally in living cells, we also include substrate input and product removal reactions



We have written two SBML model files, F3 and F4, that describe the chemical reactions with an initial total enzyme concentration of $1 \mu M$ inside two different compartment volumes (1 fl and 0.1 fl). The rate constants are the same for both compartments and can be found in Table 4. The choice of rate constants leads to Michaelis-Menten constants which are in the physiological range [47] and which guarantee positive cooperativity: $K_{M1} = 20 \mu M$, $K_{M2} = 2 \mu M$, $K_{M3} = 0.2 \mu M$ and $K_{M4} = 0.02 \mu M$.

Table 4. Rate constants for model of cooperative enzyme kinetics with substrate input (SBML files F3 and F4).

k_{in}	$9 \mu M/s$	k_{out}	$10s^{-1}$
k_1	$1(\mu Ms)^{-1}$	k_{-1}	$10s^{-1}$
k_2	$10(\mu Ms)^{-1}$	k_{-2}	$10s^{-1}$
k_3	$1 \times 10^2(\mu Ms)^{-1}$	k_{-3}	$10s^{-1}$
k_4	$1 \times 10^3(\mu Ms)^{-1}$	k_{-4}	$10s^{-1}$
k_{cat}	$10s^{-1}$		

doi:10.1371/journal.pone.0038518.t004

Cooperative enzyme kinetics at the cellular scale. An enzyme concentration of $1 \mu M$ realized in a cellular volume of $1fl$ implies a copy number of about 600 enzyme molecules per cell. The *Steady State Analysis* of iNA enables one to obtain a quick overview of the noise characteristics of the reaction network in steady state conditions. Comparing the prediction of the REs and EMREs, as summarized in Fig. 4a, we find excellent agreement at this length scale. Since the average concentrations are well

captured by the macroscopic REs we can investigate the fluctuations around them using the LNA. The size of the error bars in Fig. 4a show that as for the Michaelis-Menten reaction, the largest fluctuations in steady-state conditions occur for the substrate species. The mean substrate concentration as obtained from the *table view* is $1.1 \mu M$ with a CV of about 0.14. This implies fluctuations of roughly 90 substrate molecules showing that molecular fluctuations in enzyme kinetics can be very significant even at the cellular scale.

We now use iNA to calculate the transient correlations in the dynamics. Note that the presence of correlations is a distinct feature of the stochastic description. The correlation coefficient as defined by Eq. (8) can be computed by iNA using the LNA, as well as from stochastic simulations, see Fig. 4 b,c. The two are in excellent agreement. The fluctuations of free enzyme and substrate concentrations become anti-correlated as the steady state is approached; this is expected and simply due to the fact that they bind to each other. However, what is more interesting is that the transient correlations can exhibit a complex biphasic behavior, alternating between positively and negatively correlated states as a function of time. Damped oscillations in the correlation coefficients can be traced to the presence of a pair of complex conjugate

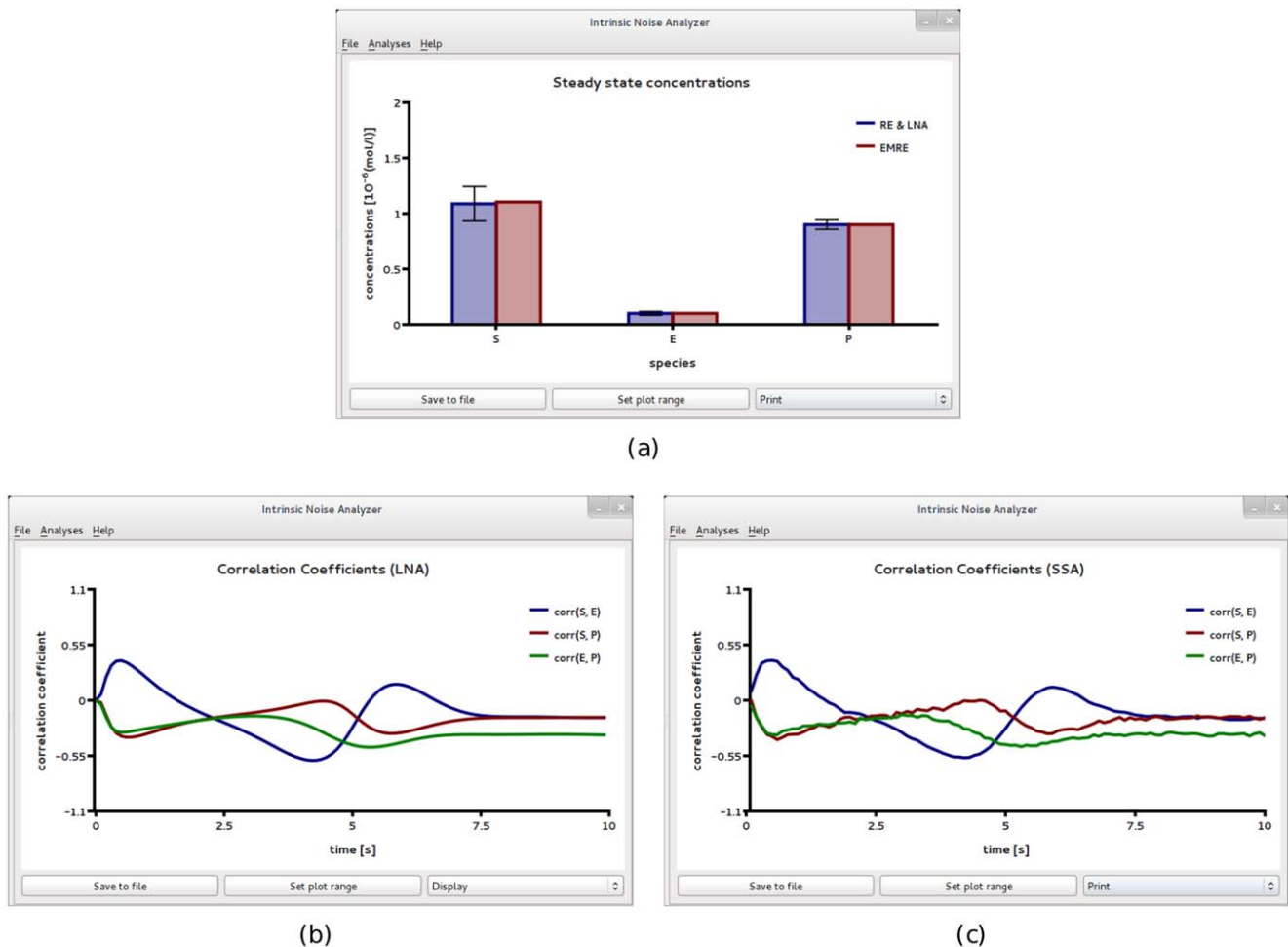


Figure 4. Cooperative enzyme kinetics in a large compartment of volume $\Omega = 1fl$. (a) Plot of the *Steady State Analysis* which shows good agreement of RE and EMRE results. (b) Estimation of time dependent correlation coefficients using the *Linear Noise Approximation* which are found to be in good agreement with those calculating from ensemble averaging of 5,000 stochastic simulations (c). Note that the oscillatory behavior in the correlation coefficients indicates the presence of noise induced oscillations. The figures were obtained from iNA's analysis of the SBML file F3. doi:10.1371/journal.pone.0038518.g004

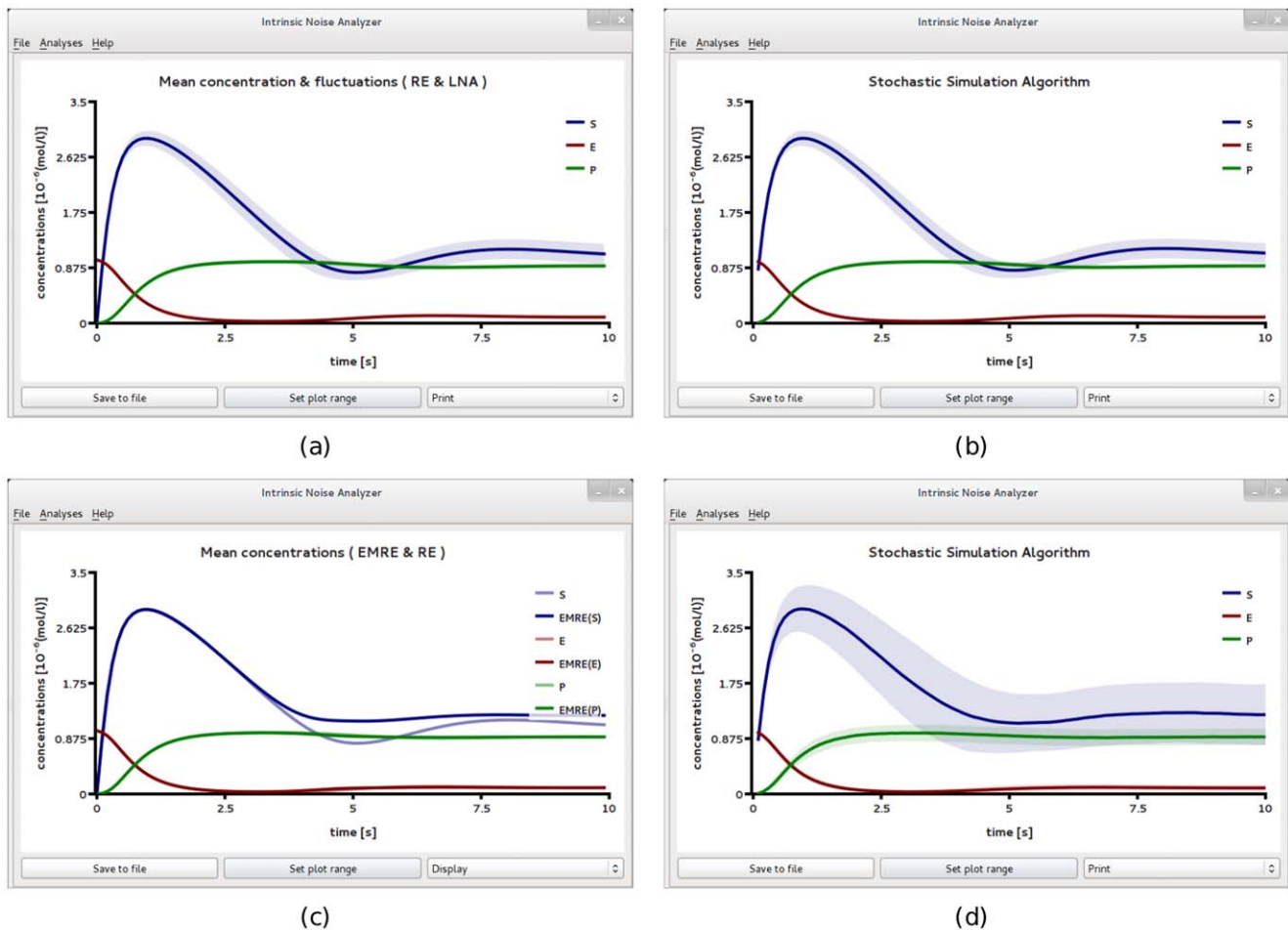


Figure 5. Transient dynamics of cooperative enzyme kinetics in large and small compartments. (a) shows damped oscillations in the mean concentrations of the RE and the fluctuations about them at the cellular scale $\Omega = 1/l$ (large compartment) which are in excellent agreement with the ensemble averaged data obtained from 5,000 stochastic simulations (b). In (c) we show how the EMRE predicts that noise destroys the oscillations when the reaction is taking place inside a small intracellular compartment of size $\Omega = 0.1/l$. This is well reproduced by stochastic simulations using 10,000 realizations (d). Figure (a) and (b) have been obtained from iNA's analysis of SBML file F3 while (c) and (d) were obtained from analysis of SBML file F4. doi:10.1371/journal.pone.0038518.g005

eigenvalues of the Jacobian of the REs. The latter have been connected with the presence of noise-induced oscillations that are observed in single realizations of the stochastic dynamics [48]. Similar oscillations have also been found for enzymes with negative cooperativity, i.e., in the presence of an inhibitor, and have been related to the presence of damped oscillations in the corresponding deterministic model [49]. Indeed the *Time Course Analysis* carried out by iNA shows transient damped oscillations in the mean concentrations, see Fig. 5a. We have also verified this transient dependence against stochastic simulations which are shown in Fig. 5b and found to be in good agreement with the results of the LNA.

Cooperative enzyme kinetics in a small intracellular compartment. Many enzymes operate in small compartments. Hence for physiological concentrations, the enzyme copy numbers can be quite small. In the second SBML model file, F4, we have reduced the compartmental volume to $0.1/l$ which at constant total enzyme concentration implies a copy number of only 60 enzyme molecules. The time course of the mean concentrations can easily be investigated by means of the EMRE implemented by iNA. The result is shown in Fig. 5c. We observe that the time course is in good agreement with the macroscopic REs only for

free enzyme and product concentrations. In contrast, we find that the EMRE predicts the damped oscillations in the substrate concentration to die out quicker than predicted by the REs. This inherent dephasing of individual realizations (compared to the previous case in which the volume was $1/l$) can be traced back to increased noise due to smaller copy numbers at the subcellular scale. We have qualitatively verified this effect by ensemble averaging 10,000 stochastic simulation realizations, see Fig. 5d.

Gene Regulatory Network with a Negative Feedback Loop

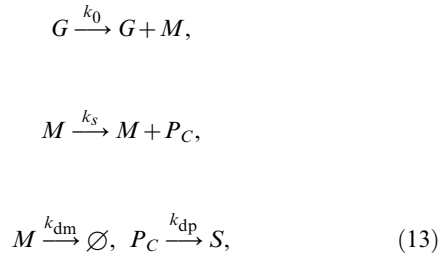
Many genes are represented by only a single copy inside living cells. It is also a fact that many important regulatory molecules exist in low copy numbers. Hence the biochemical process of gene expression is inherently stochastic which also implies that there is a significant amount of noise in the protein levels.

A remarkable property of gene expression is the emergence of cellular rhythms which are commonly attributed to a negative feedback loop in which clock proteins inhibit the expression of their own gene [50–52]. There is an ongoing debate whether circadian rhythms (a class of cellular rhythms) observed in constant

darkness are self-sustained or noise-induced [53]. The latter, which can be analyzed by means of the SSE, is considered here.

We consider the case where cell-cell coupling and cell-to-cell variability are negligible. Then a single SSA realization of the stochastic dynamics of a circadian clock circuit models single cell circadian rhythms of clock expression while the average over independent realizations models the ensemble rhythm at the population level.

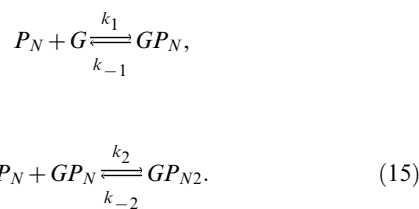
The genetic network under study is a variant of the core model for a circadian clock which has been considered by Gonze and Goldbeter in Ref. [40]. At the heart of the model is a set of effective reactions that describe transcription and translation



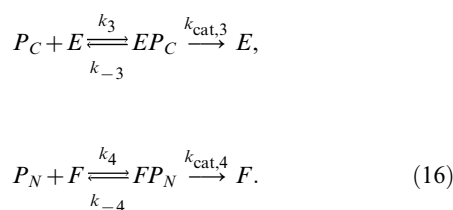
which is common to many simple models of gene expression [54,55]. The above involves the clock gene, G , the transcribed mRNA, M and the cytosolic clock protein, P_C . Note that we have also taken into account mRNA degradation with rate k_{dm} as well as the consumption of protein with rate k_{dp} in a different pathway involving S which is not explicitly considered here. The negative feedback loop arises from transport of the cytosolic clock protein into the nucleus (the nuclear protein is represented by P_N) [56]



where it binds to promoter regions of DNA while inhibiting the expression of its own gene



Furthermore we assume that degradation occurs via the following enzymatic processes



The constraint that the gene copy number is fixed to one implies that the RE model for this system depends on the compartment

volume. Taking into account discreteness, e.g., calculating the EMRE corrections, generates a volume dependence on top of this pre-existing volume dependence. To clearly distinguish the volume corrections due to the EMRE, we follow [35] (see its Supplementary Information) and scale the rates such that we eliminate the volume dependence of the REs. This ensures that the effective rates for transcription and DNA binding are the same for different reaction volumes. The rescaling is as follows: $k_0 = \Omega N_A k_0'$, $k_1 = \Omega N_A k_1'$, $k_{-1} = \Omega N_A k_{-1}'$, $k_2 = \Omega N_A k_2'$ and $k_{-2} = \Omega N_A k_{-2}'$. We further impose cooperative binding by the choice of rate constants $k_2 = 10k_1$ and $k_{-2} = k_{-1}$. The specific values of the rate constants used in this example can be found in Table 5.

Two SBML files F5 and F6 were created to describe the biochemical reactions in compartments of size 2ℓ and 0.2ℓ respectively. Figures Fig. 6 and 7 show the results obtained by analyzing file F5 using iNA. At this length scale, the population level descriptions of the mean concentrations and of the variances of fluctuations as given by the REs and the LNA respectively, match very well those obtained from stochastic simulations (compare Fig. 6a and b). In Fig 6c we show a single SSA trajectory which corresponds to single cell level data. Note that while at the population level, one can observe synchronous damped oscillations, at the single cell level oscillations are not evident. In Fig. 7 we show the results of a more detailed test on the accuracy of the LNA. We compare the correlation coefficients estimated from the LNA and stochastic simulations and find good agreement between the two. We observe that the fluctuations of the cytosolic and nuclear protein species are strongly correlated with a clear hierarchy of the correlation coefficients over the whole time course, i.e., $\text{corr}(P_C, P_N) > \text{corr}(M, P_C) > \text{corr}(M, P_N)$. Moreover, we see that over a short time interval M and P_N are anti-correlated due to the onset of the repression. The minima in the correlations at some point in time suggests that repression generally affects the fluctuations of all three species.

It is commonly believed that on average transient responses observed on the cell population level can be accurately described by deterministic REs [57]. This is contrary to the theory of EMREs which predicts that for nonlinear reactions considered in mesoscopic volumes, the mean concentration prediction of the macroscopic REs does not agree with that of the CME. In other words, if the circuit inside each cell is characterized by low copy numbers of interacting molecules then the mean concentrations at the population level (obtained using the SSA) will show deviations from the predictions of the REs. These deviations are too small to observe in Fig. 6. Hence we used iNA to analyze the SBML model

Table 5. Rate constants for circadian clock model (SBML files F5 and F6. File F7 has the same constants except that k_1 is multiplied by a factor of 100.).

k_{in}	$5d^{-1}$	k_{out}	$5d^{-1}$
k_0'	$500d^{-1}$	k_0	$\Omega N_A k_0'$
k_1'	$0.5d^{-1}$	k_1	$\Omega N_A k_1'$
k_{-1}'	$0.5d^{-1}$	k_{-1}	$\Omega N_A k_{-1}'$
k_2	$10k_1$	k_{-2}	k_{-1}
k_3	$0.5(\mu Md)^{-1}$	k_{-3}	$0.5d^{-1}$
k_4	$10(\mu Md)^{-1}$	k_{-4}	$5d^{-1}$
$k_{cat,3}$	$0.5d^{-1}$	$k_{cat,4}$	$10d^{-1}$
k_s	$5d^{-1}$	k_{dm}, k_{dp}	$5d^{-1}$

Note that the time unit is days (d).

doi:10.1371/journal.pone.0038518.t005

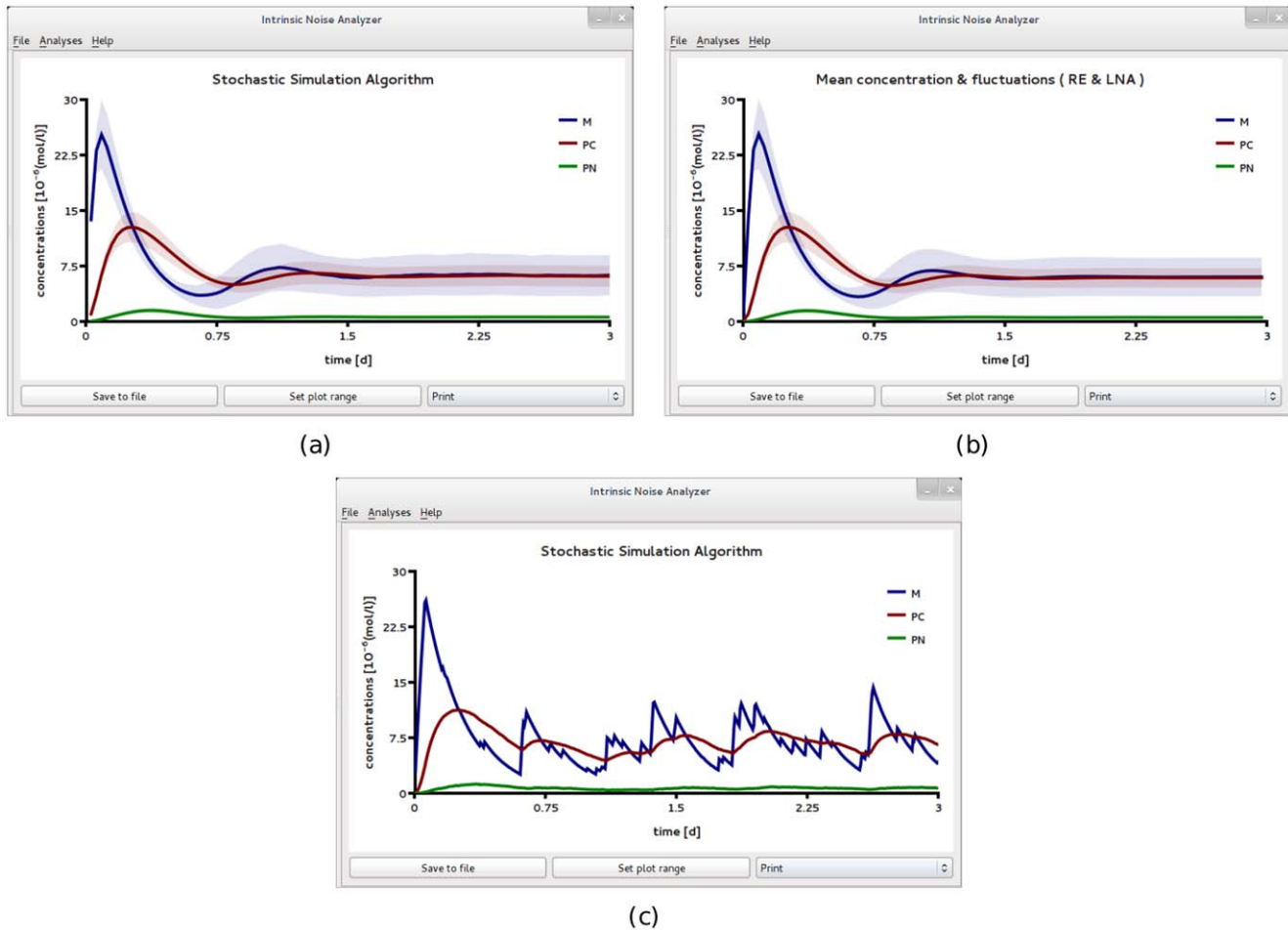


Figure 6. A circadian clock model realized in a large compartment of volume $\Omega = 2f/l$. (a) Mean concentrations and variance of fluctuations obtained by ensemble averaging 3,000 stochastic realizations. This corresponds to averages calculated over a population of an equal number of uncoupled identical cells, each having a circadian clock inside. The latter time course is well reproduced by the *Linear Noise Approximation* shown in (b). While the population level concentrations display damped oscillations, individual stochastic simulation realizations (which correspond to individual cell data) reveal no obvious periodicities, (c). All figures have been obtained by analyzing SBML file F5 with iNA. doi:10.1371/journal.pone.0038518.g006

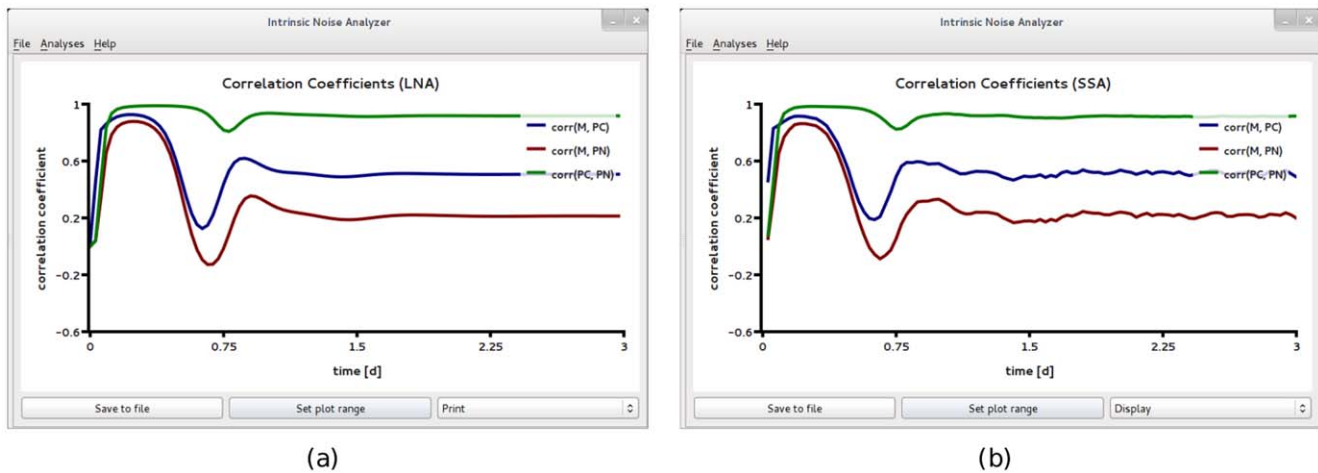


Figure 7. Transients in the correlations of concentration fluctuations for a circadian clock model realized in a large compartment of volume $\Omega = 2f/l$. Comparison of the time course of correlations obtained from the *Linear Noise Approximation* (a) and from ensemble averaging stochastic simulations of 3,000 independent cells (b). The two are in good agreement. The minima are a signature of repression due to the negative feedback loop. Both figures have been obtained by analyzing SBML file F5 with iNA. doi:10.1371/journal.pone.0038518.g007

definition F6 where the reaction volume is 10 times smaller than that in F5. Figs. 8a and b show the population level *Time Course Analysis* provided by iNA, according to the SSA, EMREs and the REs. Note that the EMRE predictions no longer agree with those of the macroscopic REs. In fact, we find that on the cell population level, the influence of intrinsic noise reduces the damping of synchronous oscillations compared to the large copy number case considered in Fig. 6b. A comparison of Figs. 8a and b shows that these predictions are supported by stochastic simulations. Reduced damping of the oscillations can be interpreted as increased coherence of individual stochastic realizations on the single cell level. We have also carried out single cell simulations, see Fig. 8c, which qualitatively support this conclusion. In fact, we find that the protein concentrations in a single cell exhibit sustained oscillations at a period of about one day which is in agreement with the period of synchronous damped oscillations at the population level, see Fig. 8. Note also the shift to lower frequency in the damped oscillations. Similar noise-induced shifts have been observed by other authors [58] using stochastic simulations.

To investigate this effect further, we increased the repression rate constant k_1 by a factor of 100, see SBML file F7 in Table 2

with a volume $\Omega=0.2fl$. Interestingly, we find that here the EMRE predicts synchronous damped oscillations even when they are absent in the REs, see Fig. 9a. Hence, these oscillations are purely induced by noise since they cannot be observed in the corresponding deterministic model. We have verified this effect against stochastic simulations which are in qualitative agreement as shown in Fig. 9b. Note that although the EMRE prediction is qualitatively correct in predicting oscillations where they are not captured by the REs, the phase and frequency of the oscillations differs significantly from the stochastic simulations at the cell population level, as shown in Fig. 9b. Higher-order corrections to the mean concentrations than those of the EMRE would be needed to correctly capture such details. The theory of such corrections has already been worked out [59] but they are not presently implemented in iNA. While at the cell population level, we have noise-induced damped oscillations, at the single cell level we observe persistent periodic rhythms in the mRNA and protein concentrations (see Fig. 9c) which are considerably more regular and conspicuous compared to the case of weak negative feedback in Fig. 8. Similar differences between the population level and cell level circadian dynamics have been experimentally observed [60].

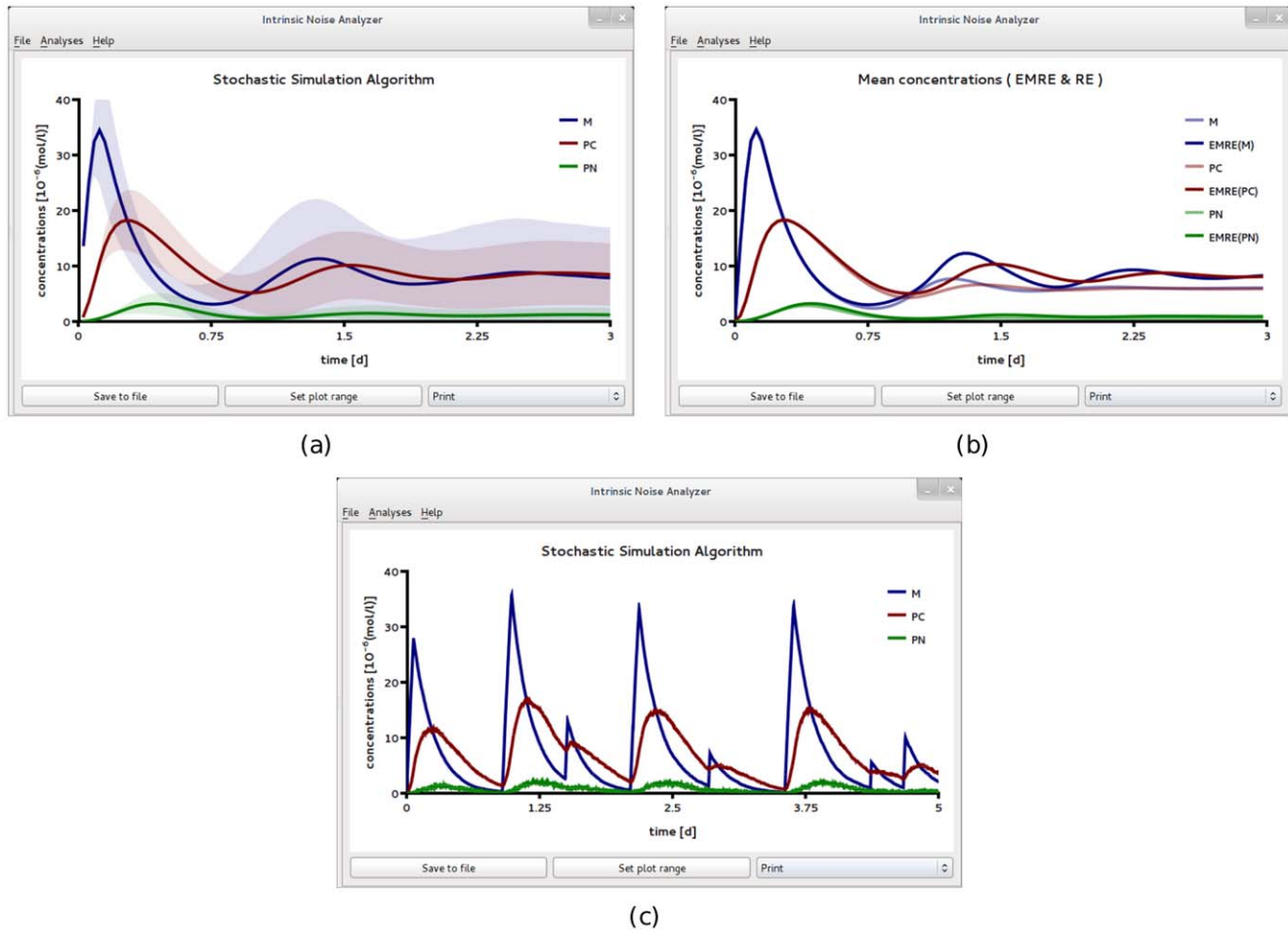


Figure 8. A circadian clock model realized in a small compartment of volume $\Omega=0.2fl$ with weak negative feedback. Time course analysis obtained from stochastic simulations of an ensemble of 50,000 independent cells (a) and from the REs and EMREs (b). The EMRE analysis shows synchronous damped oscillations which are amplified compared to the macroscopic REs. This suggests that noise increases the coherence at the single cell level compared to the case $\Omega=2fl$, see Fig. 6. This is also supported by single cell stochastic simulations which show bursty noise-induced oscillations of mRNA, M , and more regular oscillations for the cytosolic and nuclear protein concentrations, P_c and P_n , respectively. All figures have been obtained by analyzing SBML file F6 with iNA.
 doi:10.1371/journal.pone.0038518.g008

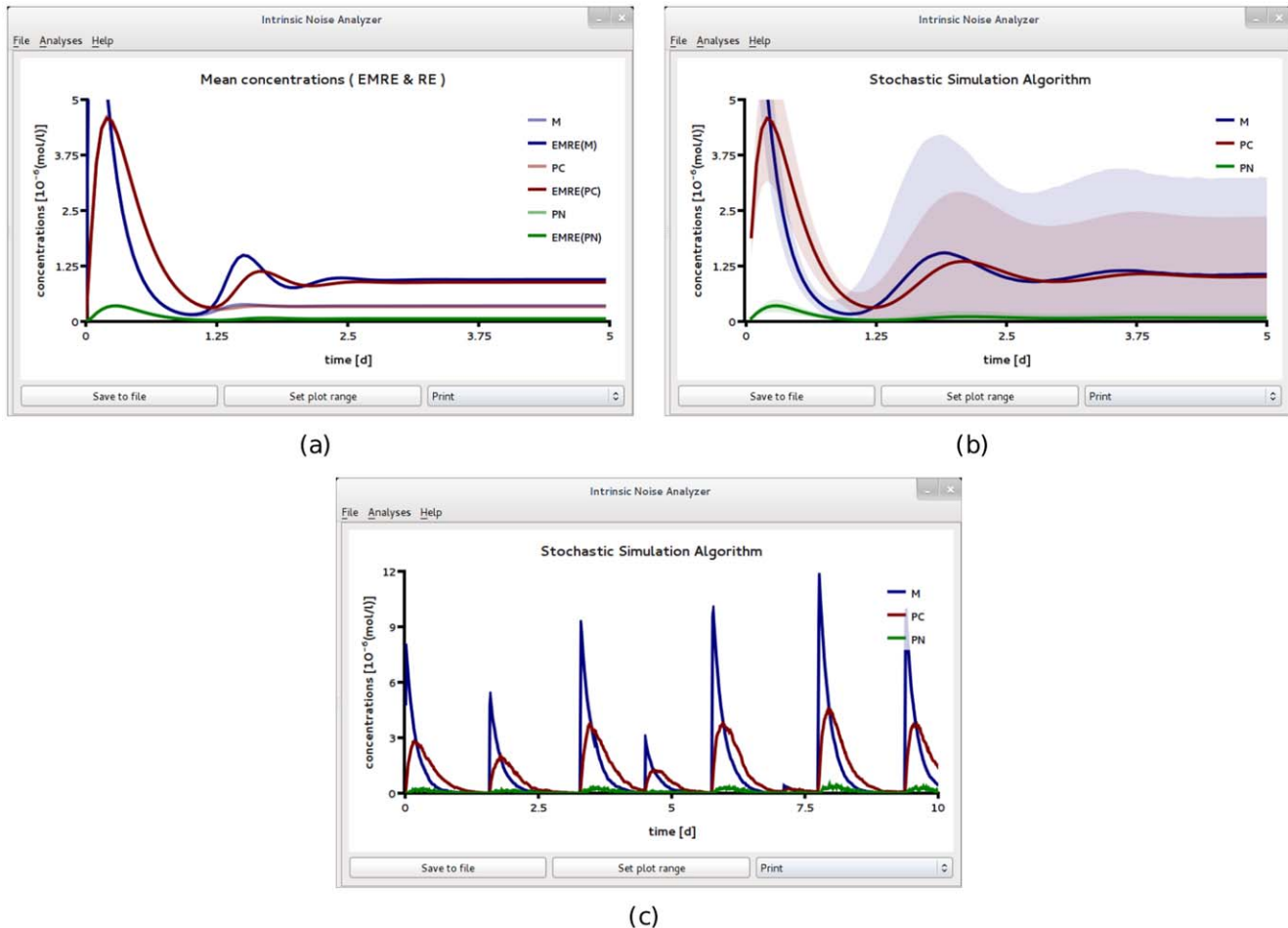


Figure 9. A circadian clock model realized in a small compartment of volume $\Omega = 0.2/l$ and with strong negative feedback. This rate constant k_1 which controls the strength of transcriptional repression is made a hundred times larger than that used for Fig. 8. In (a) we compare the predictions of the REs with those of the EMREs. The EMREs predict damped oscillations at the population level which are missed by the REs. The presence of these noise-induced synchronous oscillations is qualitatively confirmed by stochastic simulations of 100,000 independent cells (b). However, the phase of the damped oscillation is quantitatively different from the EMREs. In particular, panel (a) shows damped oscillations with a period of about 1 day (d) while the ones obtained from stochastic simulations in panel (b) have a significantly longer period. We have also carried out stochastic simulations at the single cell level (c) which show sustained bursty oscillations with a period of about 1.5 days in the mRNA, M , and cytosolic protein, P_c , concentrations. All figures have been obtained by analyzing SBML file F7 with iNA. doi:10.1371/journal.pone.0038518.g009

We conclude by noting that the fact that the RE model could not capture the population level dynamics suggests that the common practice of extracting rate constants by fitting population level data to RE model predictions (see for example [61]) may lead to erroneous results whenever the single cell dynamics are considerably affected by intrinsic noise.

Design and Implementation

The conceptual design of iNA consists of three layers: model interpretation, computation and GUI which are shown in Fig. 10. The input is a SBML model definition file from which an interpreter of the mathematical model representation is set up. The computational layer of iNA can be accessed by the user in two ways. One route is direct stochastic simulation of the model. This procedure has to be carried out repeatedly in order to obtain sufficiently accurate statistical averages. The second route is to obtain an approximate analytical solution for the mean concentrations and the variance and covariance of concentration fluctuations of the CME. For the latter we employ the Linear

Noise Approximation and the EMRE method which rely on van Kampen's SSE. The output of the computation is visualized by table and plot views implemented in the GUI. iNA has been implemented in C++ and makes use of the cross-platform library Qt (Nokia, Inc) which yields comparable graphical interfaces on different platforms from a single source code.

The SBML parser. The model file is parsed using the library libSBML [62] from which the interpreter is set up. The latter constructs a mathematical representation of the model which is suitable for the use of the computer algebra system Ginac [63]. Such representation comprises the list of species, the stoichiometric matrix and the set of propensities characterizing the reaction network. Though symbolic representations of SBML are available [64], these libraries are limited to the representations of models using deterministic REs. The implementation presented here is the first to adopt this approach for the computation of the SSE of the CME, which goes beyond the validity of the deterministic REs. Thereby it fills the gap between analytical approximations and stochastic simulation of biochemical kinetics.

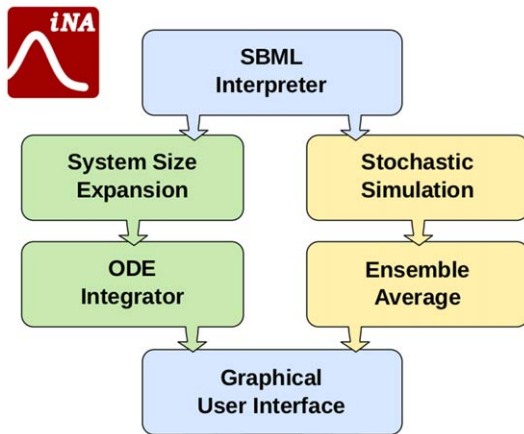


Figure 10. Schematic illustrating the architecture of the software iNA. The software reads an SBML file, sets up an internal mathematical representation of the biochemical reaction scheme, and computes the mean concentrations and the variances and covariance of concentration fluctuations as a function of time by stochastic simulations and by approximations based on the system size expansion. The results are then output via tables and plots implemented in the GUI.

doi:10.1371/journal.pone.0038518.g010

SSA and methods based on the SSE. The computational layer offers two methods for the stochastic simulation: the direct method developed by Gillespie [25] and the more recently introduced optimized direct method [27]. It is our experience that the latter typically speeds up simulations by a factor of two. The computation of the LNA and EMRE requires the series expansion of the CME in powers of the inverse square root of the volume of the compartment (see the section *Methods*). We make use of the library Ginac, by means of which we can compute derivatives analytically and obtain series expansions of the mathematical model representation instead of resorting to numerical approximations. The computations are heavily based on basic linear algebra, for which we resort to the C++ template library Eigen [65] that has been customized to work with symbolic representations of Ginac. In order to allow for unconstrained numerical integration, it is also necessary to identify the corresponding conservation laws of the reaction network which is done using the full-pivoting LU decomposition provided by Eigen. This algorithm has been described in Ref. [66] and has been shown to give reliable results. The numerical integration of the ordinary differential equations describing REs, the LNA and EMREs is performed using explicit Runge-Kutta algorithms (RK4, RK45, Dopri54) or implicit Rosenbrock method (4/5th order) for stiff problems [36,67,68].

Optimizations. Generally the number of simultaneous equations that need to be solved by the software is given by N for the REs, $(1/2)N(N+1)+N$ for the LNA and $(1/2)N(N+1)+2N$ for the EMREs, where N is the number of species (see the section *Methods*). Hence the complexity of the SSE based methods grows at least quadratically with the number of species depending on the algorithm of integration employed. In contrast, the performance of the SSA is limited first by the number of reactions over the time course, which is proportional to the size of the propensity and second by the large number of realizations required to obtain accurate statistical averages. We have addressed

the latter issue by supplying our stochastic simulators with easy-to-use OpenMP parallelism which is available on many platforms and can be accessed through the *Stochastic Simulation* wizard. The remaining bottlenecks concern the performance of expression evaluation which is at the heart of both analytical and stochastic simulation methods. In order to allow for efficient evaluation of the required expressions, we have set up a bytecode interpreter, a concept which is common to dynamic programming languages. Bytecode interpreters increase the execution performance by reducing the code complexity and allow for very efficient implementations while at the same time maintaining the portability of code across many platforms. The expressions computed by iNA are first compiled into a bytecode representation which is then reduced by fast peep-hole optimizations [69] and finally evaluated using an efficient interpreter implementing stack-machines. On single-core architectures we observed a speed up by a factor of 10–20 compared to the use of conventional methods for both stochastic simulations and SSE based methods. On multi-core architectures the performance of the SSE methods could be significantly improved by the use multiple stacks enabling parallel expression evaluation.

Methods

General Formulation of the SSE

In the section *The Chemical Master Equation and the System Size Expansion* we have introduced the CME, Eq. (3). The latter is typically intractable for analytical solution. In fact, exact solutions exist only for a handful of cases. However, the dynamics can be investigated systematically by means of van Kampen's SSE [9,70]. The starting point of the analysis is the so called Van Kampen ansatz

$$\frac{\vec{n}}{\Omega} = [\vec{X}] + \Omega^{-1/2}\vec{\epsilon}, \quad (17)$$

by which one separates the instantaneous concentration into a deterministic part given by the solution $[\vec{X}]$ of the macroscopic REs, Eq. (4), and the fluctuations around it. The change of variable causes the probability distribution of molecular populations $P(\vec{n},t)$ to be transformed into a new probability distribution of fluctuations $\Pi(\vec{\epsilon},t)$. In particular, the time derivative transforms as [9]

$$\frac{\partial P(\vec{n},t)}{\partial t} = \frac{\partial \Pi(\vec{\epsilon},t)}{\partial t} - \Omega^{1/2} \sum_i \frac{\partial [X_i]}{\partial t} \frac{\partial \Pi(\vec{\epsilon},t)}{\partial \epsilon_i}, \quad (18)$$

which takes into account the change in probability along the deterministic trajectory of $[X_i]$.

The expansion of the step operator. The core of the expansion now follows from the observation that the step operator can be written as the Taylor series $E_i^{-S_{ij}} g(n_1, \dots, n_i, \dots, n_N) = e^{-\partial/\partial n_i S_{ij}} g(n_1, \dots, n_i, \dots, n_N)$ which upon the change of variable, Eq. (17), can be expanded in terms of the inverse square root of the volume Ω .

$$\prod_{i=1}^N E_i^{-S_{ij}} - 1 = -\Omega^{-1/2} \sum_{i=1}^N S_{ij} \partial_i$$

$$\begin{aligned}
 & + \frac{\Omega^{-1}}{2} \sum_{i,k=1}^N S_{ij} S_{kj} \partial_i \partial_k - \frac{\Omega^{-3/2}}{6} \sum_{i,k,r=1}^N S_{ij} S_{kj} S_{kr} \partial_i \partial_k \partial_r \\
 & + O(\Omega^{-2}),
 \end{aligned} \tag{19}$$

where $\partial_i \equiv \partial / \partial \epsilon_i$.

The expansion of the propensity. We now turn to the expansion of the propensity which is done in two steps. Note that from here on we use the convention that twice repeated Greek indices have to be summed over 1 to N. First we expand the propensities in powers of epsilon. The result is

$$\begin{aligned}
 \hat{a}_j(\vec{n}, \Omega) &= a_j([X], \Omega) + \Omega^{-1/2} \epsilon_\alpha \partial_{[X_\alpha]} a_j([\vec{X}], \Omega) \\
 & + \frac{1}{2} \Omega^{-1} \epsilon_\alpha \epsilon_\beta \partial_{[X_\alpha]} \partial_{[X_\beta]} a_j([\vec{X}], \Omega) + O(\Omega^{-3/2}).
 \end{aligned} \tag{20}$$

Note that $a_j([X], \Omega)$ is simply the microscopic propensity evaluated at the deterministic concentrations. As shown below the latter depends explicitly on Ω and can be expanded into a finite series of the form

$$a_j([X], \Omega) = \Omega f_j^{(0)}([X]) + \Omega^{-1} f_j^{(1)}([X]) + \Omega^{-2} f_j^{(2)}([X]) + \dots \tag{21}$$

The first term denoted by $f_j^{(0)}([X])$ is typically associated with the macroscopic rate function of the j -th reaction. Combining Eq. (20) and (21) we find

$$\begin{aligned}
 \hat{a}_j(\vec{n}, \Omega) &= \Omega \left(f_j^{(0)}([\vec{X}]) + \Omega^{-1/2} \epsilon_\alpha \partial_{[X_\alpha]} f_j^{(0)}([\vec{X}]) \right. \\
 & \left. + \frac{1}{2} \Omega^{-1} \epsilon_\alpha \epsilon_\beta \partial_{[X_\alpha]} \partial_{[X_\beta]} f_j^{(0)}([\vec{X}]) + \Omega^{-1} f_j^{(1)}([\vec{X}]) + O(\Omega^{-3/2}) \right).
 \end{aligned} \tag{22}$$

In this way we obtain an expansion whose coefficients are independent of Ω . We will illustrate the expansion for the case of mass action kinetics and the case of non-elementary reactions. The former case allows to write the propensity in the generic form

$$\hat{a}_j(\vec{n}, \Omega) = \Omega k_j \prod_{z=1}^N \Omega^{-s_{zj}} (n_z)_{s_{zj}}, \tag{23}$$

where $(n_z)_{s_{zj}}$ is the falling factorial $n_z(n_z - 1) \dots (n_z - s_{zj} + 1)$. The microscopic propensity becomes

$$a_j([X], \Omega) = \Omega k_j \prod_{z=1}^N [X_z] ([X_z] - \Omega^{-1}) \dots ([X_z] - \Omega^{-1}(s_{zj} - 1)), \tag{24}$$

and hence by collecting terms of order Ω^0 and Ω^1 it follows that

$$f_j^{(0)}([\vec{X}]) = k_j \prod_{z=1}^N [X_z]^{s_{zj}},$$

$$f_j^{(1)}([\vec{X}]) = - \frac{k_j}{2} \prod_{z=1}^N [X_z]^{s_{zj}-1} s_{zj} (s_{zj} - 1) \tag{25}$$

For non-elementary reactions we consider the Michaelis-Menten example with reaction propensity $\Omega v_m n_S / (n_S + \Omega K_M)$ as has been suggested by Rao [71]. By expressing the particle number in terms of concentrations one observes that by construction macroscopic and microscopic propensities agree, i.e., we have $a_j([\vec{X}], \Omega) = \Omega f_j([\vec{X}])$. It follows that the macroscopic rate function becomes $f_j^{(0)} = f_j([\vec{X}]) = v_m [S] / ([S] + K_M)$ while $f_j^{(n)} = 0$ for all $n > 0$. The above expansions are computed automatically by iNA and do not require any further user input. Note however that in the situation where the user defined propensity cannot be expanded in the prescribed way the software simply assumes that $f_j^{(0)}([\vec{X}]) = a_j(\vec{n}, \Omega) / \Omega$ at constant Ω . By doing so the correctness of the numerical values used in the analysis is ensured.

The expansion of the CME. The expansion obtained in this way can be written as

$$\begin{aligned}
 & \frac{\partial \Pi(\vec{\epsilon}, t)}{\partial t} - \Omega^{1/2} \frac{\partial [X_\alpha]}{\partial t} \frac{\partial \Pi(\vec{\epsilon}, t)}{\partial \epsilon_\alpha} = \\
 & \left(\Omega^{1/2} \partial_\alpha D_\alpha + \Omega^0 \mathcal{L}^{(0)} + \Omega^{-1/2} \mathcal{L}^{(1)} + O(\Omega^{-1}) \right) \Pi(\vec{\epsilon}, t),
 \end{aligned} \tag{26}$$

where the individual terms in the expansion are obtained using (3) together with the expansions (19), (20) and (21) and grouping terms in descending powers of Ω .

$$\mathcal{L}^{(0)} = - \partial_\alpha J_\alpha^\beta + \frac{1}{2} \partial_\alpha \partial_\beta D_{\alpha\beta}, \tag{27}$$

$$\begin{aligned}
 \mathcal{L}^{(1)} &= - \partial_\alpha D_\alpha^{(1)} - \frac{1}{2!} \partial_\alpha J_\alpha^{\beta\gamma} q_\beta q_\gamma + \frac{1}{2!} \partial_\alpha \partial_\beta J_\alpha^\gamma q_\beta q_\gamma \\
 & - \frac{1}{3!} \partial_\alpha \partial_\beta \partial_\gamma D_{\alpha\beta\gamma}.
 \end{aligned} \tag{28}$$

Note that Eq. (28) corrects Eq. 14 in Ref. [31] which is missing the third term. However, this does not affect the result. Here we have defined the SSE coefficients given by

$$D_{ij..r}^{(n)} = \sum_{k=1}^R S_{ik} S_{jk} \dots S_{rk} f_k^{(n)}([\vec{X}]),$$

$$J_{ij..r}^{(n)st..z} = \frac{\partial}{\partial [X_s]} \frac{\partial}{\partial [X_t]} \dots \frac{\partial}{\partial [X_z]} D_{ij..r}^{(n)}. \tag{29}$$

Note that we adopt the convention to omit the index for $n = 0$, e.g., $D_\alpha^{(0)} \equiv D_\alpha$. We can now equate the terms of order $\Omega^{1/2}$ which appear on the left and right hand side of Eq. (26):

$$\frac{\partial [X_i]}{\partial t} = D_i = \sum_{k=1}^R S_{ik} f_k([\vec{X}]), \tag{30}$$

which are the macroscopic REs.

The linear noise approximation. We now proceed by constructing equations for the moments of the $\vec{\epsilon}$ variables. We follow the derivation presented in Ref. [59] and expand the probability distribution of fluctuations $\Pi(\vec{\epsilon}, t)$ in terms of the inverse square root of the volume

$$\Pi(\vec{\epsilon}, t) = \sum_{j=0}^{\infty} \Pi_j(\vec{\epsilon}, t) \Omega^{-j/2}. \tag{31}$$

In consequence there exists an equivalent expansion of the moments

$$\langle \epsilon_k \epsilon_l \dots \epsilon_m \rangle = \sum_{j=0}^{\infty} \langle \epsilon_k \epsilon_l \dots \epsilon_m \rangle_j \Omega^{-j/2}, \tag{32}$$

where the following definition has been used

$$\langle \epsilon_k \epsilon_l \dots \epsilon_m \rangle_j = \int \epsilon_k \epsilon_l \dots \epsilon_m \Pi_j(\vec{\epsilon}, t) d\vec{\epsilon}. \tag{33}$$

Using the expansion of the probability density, Eq. (31), together with Eq. (26) and (27) we find after equating all terms of order Ω^0 :

$$\begin{aligned} \frac{\partial}{\partial t} \Pi_0(\vec{\epsilon}, t) &= \mathcal{L}^{(0)} \Pi_0(\vec{\epsilon}, t) \\ &= \left(-\partial_x J_x^\beta \epsilon_\beta + \frac{1}{2} \partial_x \partial_\beta D_{x\beta} \right) \Pi_0(\vec{\epsilon}, t). \end{aligned} \tag{34}$$

The result is the Linear Noise Approximation which yields a linear Fokker-Planck equation. Its solution is well known to be a multivariate Gaussian distribution [9,29]. The time evolution equations of the first two moments are obtained by multiplying the latter by ϵ_i and $\epsilon_i \epsilon_j$ respectively and performing the integration over $\vec{\epsilon}$:

$$\frac{\partial}{\partial t} [\epsilon_i]_0 = J_i^\alpha [\epsilon_\alpha]_0, \tag{35}$$

$$\frac{\partial}{\partial t} [\epsilon_i \epsilon_j]_0 = J_i^\alpha [\epsilon_\alpha \epsilon_j]_0 + [\epsilon_i \epsilon_\alpha]_0 J_j^\alpha + D_{ij}. \tag{36}$$

Note that in the case of the deterministic initial conditions we have $\vec{n}/\Omega = [\vec{X}]$ initially and hence all moments of $\vec{\epsilon}$ given by Eq. (32) are zero for $t = 0$. We can also deduce by inspection of Eq. (35) that in this case we have $[\vec{\epsilon}]_0 = 0$ for all times. In order to relate the moments back to the moments of the concentration variables we use Eq. (32) and (17) to find that mean concentrations and covariance matrix are given by

$$\left\langle \frac{n_i}{\Omega} \right\rangle = [X_i] \tag{37}$$

$$\begin{aligned} \Sigma_{ij} &= \left\langle \left(\frac{n_i}{\Omega} - \left\langle \frac{n_i}{\Omega} \right\rangle \right) \left(\frac{n_j}{\Omega} - \left\langle \frac{n_j}{\Omega} \right\rangle \right) \right\rangle \\ &= \Omega^{-1} [\epsilon_i \epsilon_j]_0 \end{aligned} \tag{38}$$

Using the definition of the covariance matrix Σ_{ij} together with Eq. (36) we find that it satisfies the time dependent matrix equation

$$\frac{\partial}{\partial t} \underline{\Sigma} = \underline{J} \underline{\Sigma} + \underline{\Sigma} \underline{J}^T + \Omega^{-1} \underline{D}, \tag{39}$$

where the Jacobian and diffusion matrix are given by $(\underline{J})_{ij} = J_i^j$ and $(\underline{D})_{ij} = D_{ij}$, respectively [29]. It follows that to order Ω^0 the average concentrations are correctly described by the macroscopic REs. The size of fluctuations about the average are then given by the matrix Σ as obtained from the solution of Eq. (39).

Effective Mesoscopic Rate Equations. In this section we will use higher order corrections to derive a set of effective mesoscopic rate equations, the EMREs, that are expected to be closer to the true concentrations as predicted by the CME. We therefore utilize the system size expansion, Eq. (27) together with Eq. (31) and equate terms of the order $\Omega^{-1/2}$. The result is given by

$$\frac{\partial}{\partial t} \Pi_1(\vec{\epsilon}, t) = \mathcal{L}^{(0)} \Pi_1(\vec{\epsilon}, t) + \mathcal{L}^{(1)} \Pi_0(\vec{\epsilon}, t). \tag{40}$$

Multiplying the above equation by ϵ_i and performing the integration over $\int d\vec{\epsilon}$ we find

$$\frac{\partial}{\partial t} [\epsilon_i]_1 = J_i^\alpha [\epsilon_\alpha]_1 + \frac{1}{2} J_i^\alpha \beta [\epsilon_\alpha \epsilon_\beta]_0 + D_i^{(1)}. \tag{41}$$

We can use the above result together with Eq. (17) and (32) to find an equation for the average concentration accurate to order $O(\Omega^{-3/2})$:

$$\left\langle \frac{\vec{n}}{\Omega} \right\rangle = [\vec{X}] + \Omega^{-1} [\vec{\epsilon}]_1 + O(\Omega^{-3/2}). \tag{42}$$

Hence we conclude that the difference between the true concentrations predicted by the CME and those of the macroscopic RES evolves as

$$\frac{\partial}{\partial t} \left(\left\langle \frac{\vec{n}}{\Omega} \right\rangle - [\vec{X}] \right) = \underline{J} \left(\left\langle \frac{\vec{n}}{\Omega} \right\rangle - [\vec{X}] \right) + \vec{\Delta}, \tag{43}$$

where we have defined the vector $\vec{\Delta}$ with components

$$\Delta_i = \frac{1}{2} J_i^{\alpha\beta} \Sigma_{\alpha\beta} + \Omega^{-1} D_i^{(1)}. \quad (44)$$

Note that in the case of deterministic initial conditions this result enjoys an increased accuracy of order $O(\Omega^{-2})$ as has been shown in Ref. [59]. Note that by virtue of Eq. (17) we see that each nonzero solution of the above equation implies a correction to the concentrations as macroscopic REs. Necessary conditions for this to be true depend on the explicit form of the propensities such as propensities with nonlinear dependence on the molecular populations for which $J_i^{\alpha\beta} = \partial_x \partial_\beta S_{i;\gamma} f_\gamma(\vec{X}) \neq 0$ or propensities for which microscopic and macroscopic rate functions do not agree, i.e., $D_i^{(1)} = S_{i;\alpha} f_\alpha^{(1)}(\vec{X}) \neq 0$. Note that for a network composed entirely of elementary reactions $D_i^{(1)}$ is non-zero only for species which dimerize because of the special form of the propensity which is given by $k(n/\Omega)(n-1)$. It can be shown that in this case $D_i^{(1)} = -(1/2)G_i^\alpha[X_\alpha]$ with $G_i^\alpha \equiv J_i^{\alpha\alpha}$ and hence

$$\Delta_i = \frac{1}{2} \left(J_i^{\alpha\beta} \Sigma_{\alpha\beta} - \Omega^{-1} G_i^\alpha [X_\alpha] \right), \quad (45)$$

which agrees with the result given by Grima for networks composed of elementary reactions [31].

Concentration Inversion Effect for the Michaelis-Menten Reaction

Recently, Ramaswamy et al. [35] have reported a noise-induced concentration inversion effect using the EMREs. The authors consider independent realizations of the same chemical system in compartments of different volumes. They find that for monostable reaction networks with bimolecular reactions, when a species is more more abundant than another for large compartmental volumes, i.e., the regime where the REs are valid, then the reverse occurs for an identical system realized in compartmental volumes below a critical value, i.e., the regime where the EMREs give a more accurate description of the true mean concentrations.

We here derive the equation for the critical volume for this discreteness-induced concentration inversion of substrate and product in the Michaelis-Menten scheme (10). The macroscopic REs are given by

$$\frac{d}{dt} [S] = k_{\text{in}} - k_1 [E][S] + k_{-1} [ES],$$

$$\frac{d}{dt} [ES] = k_1 [E][S] - k_{-1} [ES] - k_2 [ES],$$

$$\frac{d}{dt} [P] = k_2 [ES] - k_{\text{out}} [S], \quad (46)$$

where $[E] = [E_T] - [ES]$ by conservation of total enzyme concentration. Under steady state conditions the above equations can be solved by setting the time derivatives to zero which yields $[S] = K_M \alpha / (1 - \alpha)$, $[P] = k_{\text{in}} / k_{\text{out}}$ and $[ES] = [E_T] \alpha$ where $\alpha = k_{\text{in}} / (k_2 [E_T])$ is measuring the fractional enzyme saturation. The EMREs of the reaction have been first obtained by Grima in Ref. [32] where it was found that the REs are only accurate for describing the enzyme and product concentrations in mesoscopic volumes, i.e., we have $\langle n_C / \Omega \rangle = [C]$ and $\langle n_P / \Omega \rangle = [P]$. However they underestimate the mesoscopic substrate concentration. The critical volume under which the concentrations of substrate and product concentrations are equal is obtained from the condition

$$\left\langle \frac{n_S}{\Omega} \right\rangle = \left\langle \frac{n_P}{\Omega} \right\rangle, \quad (47)$$

which implies the same equality in terms of molecule numbers. The finite volume correction to the macroscopic REs as given in Ref. [33], Eq. 29, leads to the condition

$$[P] = [S] + \frac{1}{\Omega_{\text{crit}}} \frac{\alpha^2 K_M}{(1 - \alpha)(K_M + [E_T](1 - \alpha)^2)}. \quad (48)$$

Solving the above equation for Ω_{crit} we find

$$\Omega_{\text{crit}} = \alpha^2 \left(\frac{k_{\text{in}}}{k_{\text{out}}} (1 - \alpha) - K_M \alpha \right)^{-1} \left(1 + \frac{[E_T]}{K_M} (1 - \alpha^2) \right)^{-1}. \quad (49)$$

Using the rate constants for the Malate dehydrogenase enzyme summarized in Table 3 we find a critical volume of $\Omega_{\text{crit}} = 1.46 \times 10^{-17} l$.

Acknowledgments

It is a pleasure to thank Andrea Y. Weisse for clarifying some aspects of the time-dependent correlation coefficients as well as Ricardo Honorato-Zimmer for helping to compile the software for the Windows operating system.

Author Contributions

Conceived and designed the experiments: PT HM RG. Performed the experiments: PT HM RG. Analyzed the data: PT HM RG. Contributed reagents/materials/analysis tools: PT HM RG. Wrote the paper: PT HM RG.

References

- Gillespie D (2007) Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry 58: 35–55.
- Grima R, Schnell S (2008) Modelling reaction kinetics inside cells. Essays in Biochemistry 45: 41.
- Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. Nature 467: 167–173.
- van Zon J, Ten Wolde P (2005) Greens-function reaction dynamics: A particle-based approach for simulating biochemical networks in time and space. The Journal of Chemical Physics 123: 234910.
- Andrews S, Addy N, Brent R, Arkin A (2010) Detailed simulations of cell biology with smoldyn 2.1. PLoS computational biology 6: e1000705.
- Kerr R, Bartol T, Kaminsky B, Dittrich M, Chang J, et al. (2008) Fast monte carlo simulation methods for biological reaction-diffusion systems in solution and on surfaces. SIAM Journal on Scientific Computing 30: 3126.
- Ander M, Beltrao P, Di Ventura B, Ferkinghoff-Borg J, Foglierini M, et al. (2004) Smartcell, a framework to simulate cellular processes that combines stochastic approximation with diffusion and localisation: analysis of simple networks. Systems biology 1: 129.
- Hattne J, Fange D, Elf J (2005) Stochastic reaction-diffusion simulation with MesoRD. Bioinformatics 21: 2923–2924.
- van Kampen N (2007) Stochastic processes in physics and chemistry. North-Holland Personal Library, 3rd edition.

10. Gardiner C (2007) Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences. Springer, 3rd edition.
11. Gillespie D (1992) A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications* 188: 404–425.
12. Gillespie D (2009) A diffusional bimolecular propensity function. *The Journal of Chemical Physics* 131: 164109.
13. Munsky B, Khammash M (2006) The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics* 124: 044104.
14. Cmepey v0.3 documentation. <http://fcostin.github.com/cmepey/>. Accessed: 10 Feb 2012.
15. Gillespie C (2009) Moment-closure approximations for mass-action models. *Systems Biology, IET* 3: 52–58.
16. Hespanha J (2008) Moment closure for biochemical networks. In: *Communications, Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on. IEEE*, 142–147.
17. Hucka M, Finney A, Bornstein B, Keating S, Shapiro B, et al. (2004) Evolving a lingua franca and associated software infrastructure for computational systems biology: the systems biology markup language (sbml) project. *Systems Biology* 1: 41–53.
18. Mendes P, Hoops S, Sahle S, Gauges R, Dada J, et al. (2009) Computational modeling of biochemical networks using copasi. *Methods in Molecular Biology* 500: 17–59.
19. Funahashi A, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, et al. (2008) CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proceedings of the IEEE* 96: 1254–1265.
20. Nicolas R, Donizelli M, Le Novère N (2007) Sbmleditor: effective creation of models in the systems biology markup language (sbml). *BMC Bioinformatics* 8: 79.
21. Gillespie C, Wilkinson D, Proctor C, Shanley D, Boys R, et al. (2006) Tools for the sbml community. *Bioinformatics* 22: 628–629.
22. Li C, Donizelli M, Rodriguez N, Dharuri H, Endler L, et al. (2010) BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology* 4: 92.
23. Higham D (2008) Modeling and simulating chemical reactions. *SIAM Review, Education Section* 50: 347–368.
24. Turner T, Schnell S, Burrage K (2004) Stochastic approaches for modelling in vivo reactions. *Computational Biology and Chemistry* 28: 165–178.
25. Gillespie D (1977) Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry* 81: 2340–2361.
26. The OpenMP API specification for parallel programming. <http://openmp.org/wp/openmp-specifications/>. Accessed: 10 Feb 2012.
27. Cao Y, Li H, Petzold L (2004) Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *The Journal of Chemical Physics* 121: 4059.
28. McQuarrie D (1967) Stochastic approach to chemical kinetics. *Journal of Applied Probability* 4: 413–478.
29. Elf J, Ehrenberg M (2003) Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research* 13: 2475–2484.
30. Hayot F, Jayaprakash C (2004) The linear noise approximation for molecular fluctuations within cells. *Physical Biology* 1: 205.
31. Grima R (2010) An effective rate equation approach to reaction kinetics in small volumes: Theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of Chemical Physics* 133: 035101.
32. Grima R (2009) Noise-induced breakdown of the Michaelis-Menten equation in steady-state conditions. *Physical Review Letters* 102: 218103.
33. Grima R (2009) Investigating the robustness of the classical enzyme kinetic equations in small intracellular compartments. *BMC Systems Biology* 3: 101.
34. Thomas P, Straube A, Grima R (2010) Stochastic theory of large-scale enzyme-reaction networks: Finite copy number corrections to rate equation models. *The Journal of Chemical Physics* 133: 195101.
35. Ramaswamy R, Gonzalez-Segredo N, Sbalzarini IF, Grima R (2012) Discreteness-induced concentration inversion in mesoscopic chemical systems. *Nature Communications* 3: 779.
36. Press W, Teukolsky S, Vetterling W, Flannery B (2007) Numerical recipes: the art of scientific computing. Cambridge University Press, 3rd edition.
37. Fersht A (1999) Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. WH Freeman.
38. Tyson J (2002) Biochemical oscillations. *Computational Cell Biology* : 230–260.
39. Shou M, Mei Q, Ettore Jr M, Dai R, Baillie T, et al. (1999) Sigmoidal kinetic model for two co-operative substrate-binding sites in a cytochrome p450 3a4 active site: an example of the metabolism of diazepam and its derivatives. *Biochemical Journal* 340: 845.
40. Gonze D, Goldbeter A (2006) Circadian rhythms and molecular noise. *Chaos* 16: 26110.
41. De Ronde W, Daniels B, Mugler A, Sinityn N, Nemenman I (2009) Mesoscopic statistical properties of multistep enzyme-mediated reactions. *Systems Biology, IET* 3: 429–437.
42. Tam Z, Cai Y, Gunawan R (2010) Elucidating cytochrome c release from mitochondria: Insights from an in silico three-dimensional model. *Biophysical Journal* 99: 3155–3163.
43. Pike L (2009) The challenge of lipid rafts. *Journal of Lipid Research* 50: S323–S328.
44. Lodola A, Shore J, Parker D, Holbrook J (1978) Malate dehydrogenase of the cytosol. A kinetic investigation of the reaction mechanism and a comparison with lactate dehydrogenase. *Biochemical Journal* 175: 987.
45. Luby-Phelps K (1999) Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. *International Review of Cytology* 192: 189–221.
46. Hutzler J, Tracy T (2002) Atypical kinetic profiles in drug metabolism reactions. *Drug metabolism and disposition* 30: 355–362.
47. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, et al. (2011) The moderately efficient enzyme: evolutionary and physico-chemical trends shaping enzyme parameters. *Biochemistry* 50: 4402.
48. McKane A, Nagy J, Newman T, Stefanini M (2007) Amplified biochemical oscillations in cellular systems. *Journal of Statistical Physics* 128: 165–191.
49. Davis K, Roussel M (2006) Optimal observability of sustained stochastic competitive inhibition oscillations at organellar volumes. *FEBS Journal* 273: 84–95.
50. Aronson B, Johnson K, Loros J, Dunlap J (1994) Negative feedback defining a circadian clock: autoregulation of the clock gene frequency. *Science* 263: 1578.
51. Goldbeter A (2002) Computational approaches to cellular rhythms. *Nature* 420: 238–245.
52. Sato T, Yamada R, Ukai H, Baggs J, Miraglia L, et al. (2006) Feedback repression is required for mammalian circadian clock function. *Nature Genetics* 38: 312–319.
53. Westermark P, Welsh D, Okamura H, Herzel H (2009) Quantification of circadian rhythms in single cells. *PLoS Computational Biology* 5: e1000580.
54. Thattai M, Van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences* 98: 8614.
55. Shahrezaei V, Swain P (2008) The stochastic nature of biochemical networks. *Current opinion in biotechnology* 19: 369–374.
56. Luo C, Loros J, Dunlap J (1998) Nuclear localization is required for function of the essential clock protein FRQ. *The EMBO Journal* 17: 1228–1235.
57. Mettetal J, Muzzey D, Pedraza J, Ozbudak E, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. *Proceedings of the National Academy of Sciences of the United States of America* 103: 7304.
58. Ramaswamy R, Sbalzarini I (2011) Intrinsic noise alters the frequency spectrum of mesoscopic oscillatory chemical reaction systems. *Scientific Reports* 1.
59. Grima R, Thomas P, Straube A (2011) How accurate are the nonlinear chemical fokker-planck and chemical langevin equations? *The Journal of Chemical Physics* 135: 084103.
60. Welsh D, Yoo S, Liu A, Takahashi J, Kay S (2004) Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Current Biology* 14: 2289–2295.
61. Locke J, Millar A, Turner M (2005) Modelling genetic networks with noisy and varied experimental data: the circadian clock in arabidopsis thaliana. *Journal of Theoretical Biology* 234: 383–393.
62. Bornstein B, Keating S, Jouraku A, Hucka M (2008) Libsbml: an api library for sbml. *Bioinformatics* 24: 880.
63. Bauer C, Frink A, Kreckel R (2002) Introduction to the ginac framework for computation within the c++ programming language. *Journal of Symbolic Computation* 33: 1–12.
64. Machné R, Finney A, Müller S, Lu J, Widder S, et al. (2006) The sbml ode solver library: a native api for symbolic and fast numerical analysis of reaction networks. *Bioinformatics* 22: 1406.
65. Guennebaud G, Jacob B (2010) Eigen v3. <http://eigen.tuxfamily.org>.
66. Vallabhajosyula R, Chickarmane V, Sauro H (2006) Conservation analysis of large biochemical networks. *Bioinformatics* 22: 346–353.
67. Fehlberg E (1969) Low-order classical runge-kutta formulas with stepsize control and their application to some heat transfer problems. Technical Report 315, National Aeronautics and Space Administration.
68. Shampine LF (1982) Implementation of rosenbrock methods. *ACM Transactions on Mathematical Software* 8: 93–113.
69. Aho A, Lam M, Sethi R, Ullman J (2007) Compilers: principles, techniques, and tools. Pearson/Addison Wesley.
70. Van Kampen N (1976) The expansion of the master equation. *Advances in Chemical Physics* : 245–309.
71. Rao C, Arkin A (2003) Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm. *The Journal of Chemical Physics* 118: 4999.
72. Adalsteinsson D, McMillen D, Elston T (2004) Biochemical network stochastic simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMC Bioinformatics* 5: 24.
73. Mauch S (2009). Cain: Stochastic simulations for chemical kinetics. <http://cain.sourceforge.net>.
74. Mauch S, Stalzer M (2011) Efficient formulations for exact stochastic simulation of chemical systems. *Computational Biology and Bioinformatics, IEEE/ACM Transactions on* 8: 27–35.
75. Caulfield E, Hellander A (2010) Cellmc – a multiplatform model compiler for the cell broadband engine and x86. *Bioinformatics* 26: 426–428.
76. Ramsey S, Orrell D, Bolouri H (2005) Dizzy: stochastic simulation of large-scale genetic regulatory networks. *Journal of Bioinformatics and Computational Biology* 3: 415–436.
77. Liu Q, Qi X, Fan S (2010) Simulating bioreaction processes based on simbiology. *Computer Applications and Software* 27: 212–214.

78. Sanft K, Wu S, Roh M, Fu J, Lim R, et al. (2011) Stochkit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics* 27: 2457–2458.
79. StochPy user guide v0.1. <http://stompy.sourceforge.net/html/userguide.html>. Accessed: 10 Feb 2012.
80. Olivier B, Rohwer J, Hofmeyr J (2005) Modelling cellular systems with PySCeS. *Bioinformatics* 21: 560–561.