Exact equilibrium-state solution of an intracellular complex formation model:
$kA \leftrightarrow P$ reaction in a small volume

Zoran Konkoli*
Chalmers University of Technology, Department of Microtechnology and Nanoscience—MC2
Bionano Systems Laboratory, SE-412 96 Gothenburg, Sweden
(Received 9 April 2010; published 26 October 2010)

A generic model of complex formation in small volumes was studied under the assumption of perfect mixing. Particles $A$ react in clusters, and each reaction converts $k$ $A$ particles into a $P$ particle. The back reaction is also allowed. The equilibrium state of the model is solved exactly. Fluctuations in product particle number are reduced by increasing the degree of cooperativity $k$. Three qualitatively distinct reactant fluctuation characteristics emerge.

DOI: 10.1103/PhysRevE.82.041922 PACS number(s): 87.15.A-, 05.40.−a, 64.60.an

I. BACKGROUND

Formation of macromolecular complexes is ubiquitous in living cells [1]. Already at a cell membrane, complex formation is an essential ingredient of receptor signalling activity [2]. A ligand binding signaling complex grows by aggregation of various proteins that regulate each other. Also, complex formation is crucial for various intracellular processes such as scaffolding, gene regulation, and various forms of post translational protein regulation. In addition, complex formation plays a role in regulating intracellular noise [3], especially in regulation of gene expression [4].

To gain some understanding of the equilibrium properties of the process of macromolecular complex formation, a relatively simple reaction model will be studied, where a multiparticle reaction forms a final product. Here and in the following, a reaction event where more than two reactants react simultaneously will be referred to as a multiparticle reaction. Likewise, any model that describes a multiparticle reaction will be referred to as a multiparticle reaction model. In principle, such models only differ in the number of reactant types that are binding and related stoichiometric coefficients (the number of required molecules). In this work a particular type of a multiparticle reaction model will be used. In the following, the scope of the validity of the model will be defined. The set of features that motivated the model will be laid out.

Complex formation pathways exhibit rather complex structure and involve combinations of random and cooperative binding processes [5]. Given a complex to be constructed, random binding describes the situation when constituents of the complex merge in an arbitrary order. In such a case series of reactions that form the complex are not unique. A sequence of reactions will be chosen at random, depending on the details of the underlaying motion of the reactants. In contrast, during cooperative binding, the series of binary reactions forms the complex in the well defined order. Each binding of a reactant decreases (increases) the dissociation (affinity) constant for the reaction that follows. Such thermodynamics favors one pathway that is the most likely to be followed. For example, cooperative processes occur during formation of a 30S ribosome [6], and of a cholesterol-sphingomyelin condensed complex [7]. In general, it is hard to draw a formal border between the random and cooperative binding, and it is an issue to distinguish between the two experimentally.

Both random and cooperative complex formation processes have one feature in common. Regardless whether a complex is formed in a random or in a cooperative fashion, a specific number of chemicals needs to be in place for the complex to form. For example, one molecule less than the required number, and the complex will not form. Furthermore, if all reactions that form the complex are much faster than other reactions in the cell, it is reasonable to expect that such a process can be described by using a multiparticle reaction model. The goal of this study is to understand which effects such “all or none” behavior bears on the noise characteristics of the kinetics of complex formation.

The simplicity of a multiparticle reaction model can be criticized in many ways. First, the probability that all required reactants meet at the same time and at the right place is rather small. A truly multiparticle reaction would be a rather rare event in the cell and would not survive the selection process during the evolution. It is clear that a multiparticle reaction model needs to be seen as an effective model that emerges as a special limit of a more realistic (serial) reaction model. Second, the question is whether the intuitive understanding of the validity of a multiparticle reaction model discussed above stands the rigorous mathematical tests. The issue is largely open but some work has been done to resolve it.

For example, a series of cooperativelike ligand binding models have been investigated rigorously (on mean field level) in [8]. Various ligand binding models have been compared with a multiparticle reaction model. The criteria for the validity of the multiparticlelike reaction model have been identified in a strict mathematical sense. In rough terms, given a pathway, the criteria identified in [8] state that the intermediate states should never accumulate significantly. If this is true, then a multiparticle model can be used. The criteria coincide with the definition of strongly positive cooperative binding given in [5]. Thus multiparticle reaction model should be a valid description of reactions that are strongly cooperative.
There are a number of examples where time scales for various reaction processes are separated. Few examples from [9] will be briefly discussed. Binding of a signal molecule to a transcription factor takes 1 ms. A transcription factor binds its DNA site in a second. A series of such events forms a transcription complex (the transcription factors, polymerase) that initiates gene expression which takes several minutes. The time scales of the processes just discussed are clearly separated. The build up of transcription factors is much faster from other processes that follow which happen at the minutes scale. Thus when modeling genetic networks, formation of transcription complexes is often described using multiparticle models [10].

Most of intracellular complex formation reactions are approximatively at an equilibrium. Reactions in cells are organized in a hierarchical manner where a series of fast reactions controls slow reactions [9]. For example, the time needed to change the concentration of a translated protein by 50% is roughly 1 h [9]. This time is much longer than characteristic times for protein regulation, transcription, translation, or signaling events. Thus the focus will be understanding equilibrium properties.

It will be assumed that the particles mix well, which greatly simplifies calculations. For example, the time for a protein to transverse a cell is rather short: 0.1 s in bacteria and 10 s in mammalian cells [9]. If the cell exploration time is longer than the complex formation time, the assumption is incorrect.

In addition, a closed reaction volume is assumed. Particles injection and removal are forbidden. Giving the mode of the operation of the living cell such restriction is rather severe since a closed system is biologically dead. Nevertheless, the assumption greatly facilitates finding of the exact solution, as will be shown.

For the reasons discussed so far the model bears some relevance on describing strongly cooperative complex formation processes but, in general, it should be viewed as a toy model of complex formation.

II. MODEL AND THE TECHNICAL SETUP

To describe complex formation processes the simplest possible multiparticle reaction model will be used. In strict mathematical terms the model is defined as follows. First, the reaction volume is assumed to be closed. Second, it will be assumed that there is only one type of reactant A. Third, it will be assumed that exactly k reactants are needed for a complex to form. The complex being formed by such reaction will be denoted by P. The focus will be on studying multiparticlelike reactions with k > 2. In practice, k can be very large. In the following, the symbol k will be referred to as the degree of cooperativity, which is closely related to the notion of the Hill coefficient [8]. Actually, the model studied in here is a simplified version of the Hill model of ligand binding [11]. Various versions of the Hill model are extensively used to describe cell dynamics in pharmacological modeling [12]. From now on the validity of the multiparticle reaction model will not be questioned further.

In standard chemical notation the multiparticle reaction model used in this work is defined as

\[ \lambda kA \rightleftharpoons P. \]

Variable k describes the degree of the cooperativity (the order of the reaction). The forward and the back-reaction rates are \( \lambda \) and \( \delta \), respectively. A and P particles will be counted and the symbols n and m denote their numbers, respectively. Please note that sufficient number of reactants needs to be present in the system (at least k) for the forward reaction to occur.

The system can be described by a master equation,

\[ \partial_t P(n,m,t) = \lambda \mathcal{O}_f + \delta \mathcal{O}_b. \]

Here and in the following the symbol \( \partial_n \) denotes the derivative \( \partial / \partial n \). \( P(n,m,t) \) is the occupancy probability for a configuration \((n,m)\). The terms describing the forward and the back reaction are given by

\[ \mathcal{O}_f = \binom{n + k}{k} P(n + k,m - 1,t) - \binom{n}{k} P(n,m,t), \]

\[ \mathcal{O}_b = (m + 1)P(n - k,m + 1,t) - mP(n,m,t). \]

An ensemble average of an observable \( \mathcal{O}(n,m) \) is defined as \( \langle \mathcal{O}(n,m) \rangle = \sum_{n,m} \mathcal{O}(n,m)P(n,m,t) \). The sum cannot be evaluated easily since it is hard to diagonalize the master equation analytically. Instead, the exact equations of motion will be derived for a special set of observables which define \( (\mu, \nu) \)-point density functions as

\[ \rho_{\mu,\nu}(t) = \frac{n!m!}{(n - \mu)!(m - \nu)!}. \]

A few examples of such functions are \( \rho_{1,0} = \langle n \rangle \) and \( \rho_{0,1} = \langle m \rangle \) (average particle numbers), \( \rho_{2,0} = \langle (n - 1) \rangle \) and \( \rho_{0,2} = \langle (m - 1) \rangle \) (fluctuations, noise), or \( \rho_{1,1} = \langle nm \rangle \) (the correlation between A and P particle numbers). The goal is to investigate behavior of \( \rho_{\mu,\nu}(t) \) for large times.

III. EXACT STATIONARY STATE SOLUTION

By using standard techniques, the problem can be converted to a quantum field theory with a Hamiltonian

\[ \hat{H} = \left[ (\hat{a}^\dagger)^{\lambda} - \hat{p}^\dagger \right] \left[ \frac{\lambda}{k!} \hat{a}^{k} - \hat{p} \right], \]

where \( \hat{a} \) (\( \hat{a}^\dagger \)) and \( \hat{p} \) (\( \hat{p}^\dagger \)) denote the annihilation (creation) operators for A and P particles, respectively. The techniques for performing the derivations have been reviewed in [13]. The density functions can be evaluated as

\[ \rho_{\mu,\nu}(t) = \langle 1 | \hat{a}^{\nu} \hat{p}^{\nu} | \phi(t) \rangle, \]

where \( | 1 \rangle \) is the left eigenvector of \( \hat{a}^\dagger \) and \( \hat{p}^\dagger \) with the eigenvalue 1, and \( | \phi(t) \rangle \) satisfies \( \partial_t | \phi(t) \rangle = -\hat{H} | \phi(t) \rangle \). The equations of motion can be obtained from \( \partial_t \rho_{\mu,\nu} = -\{1[\hat{a}^{\nu} \hat{p}^{\nu}, \hat{H}] | \phi(t) \rangle \} \) where the square brackets denote a commutator. Using Wick’s theorem one can derive
The form of coefficients $C_{k,r}$ is irrelevant for the discussion that follows. The equations are an instance of the BBGKY hierarchy.

The equations that describe the equilibrium condition $\partial_t \rho_{\mu,v}(t)=0$ are given by $\gamma \rho_{\mu,v} = \rho_{\mu,v+1}$ where $\rho_{\mu,v} = \lim_{\Delta t \to 0} \rho_{\mu,v}(t)$, $\gamma = \lambda(V)$, and $\mu, v = 0, 1, 2, \ldots$. Possible to prove by using induction. The equations can be written in a more convenient form as

$$\rho_{\mu,v} = \gamma^\nu \rho_{\mu+k\nu,0}. \quad (9)$$

A density function with $\nu \neq 0$ can be obtained from the corresponding function with $\nu = 0$. Equation (9) does not fully specify the equilibrium density functions. The question is how to compute the functions with $\nu \neq 0$.

The density functions will be computed for pure states. A pure state contains a well defined number of $A$ and $P$ particles at $t=0$ being $n_0$ and $m_0$, respectively. For an arbitrary initial condition one only needs to average over the initial particle distribution. For the pure states, $n+m$ is conserved in time: $\langle (n+m)(f(m,n)) \rangle = 0$, and $\langle n(m+n) \rangle = \eta$. The conservation laws enforce the following form for the density functions

$$[a_0 - (\mu + k \nu)] \rho_{\mu,v} = \rho_{\mu,v+1} + k \rho_{\mu,v+1}. \quad (10)$$

Equations (9) and (10) can be combined to give the relations for the density functions with $\nu = 0$: $\rho_{\mu,0} = \rho_{\mu,1} = \gamma \rho_{\mu,v}$. To solve the recurrence relations it is useful to introduce $\rho_{\mu} = \rho_{\mu,0}$. This gives $\rho_{\mu} = \rho_{\mu-1} + k \gamma \rho_{\mu} \eta^\mu$ for $\mu = 0, 1, 2, \ldots$. The normalization condition $\rho_{\mu,0} = 1$ leads to $\rho_{\mu} = 1$, and using Wick’s theorem results in $\rho_{\mu,v} = \rho_{\mu,v+1} = \mu + k \nu > \mu$, and $\rho_{\mu,v} = \rho_{\mu,v+1}$ for $\nu < 0$. The recurrence relations can be solved by mapping them to a differential equation:

$$\rho_{\mu,v} = \frac{a_0!}{v!} \Phi_{\mu,v}(x)^x \Phi_{\mu,v}(y). \quad (11)$$

where $\Phi_{\mu,v}(x) = x^v \Phi(x), \Phi(x) = \exp(x + \eta x^2)$. Finally, combining Eqs. (9) and (11) gives the exact explicit expression for stationary (equilibrium) state density functions

$$\rho_{\mu} = \gamma^\nu \frac{a_0!}{[a_0 - (\mu + k \nu)]!} \Phi_{\mu}(x)^x \Phi_{\mu}(y). \quad (12)$$

where the superscript on the density function symbol emphasizes the fact that the result holds for the pure states. The derivatives of $\Phi(x)$ can be readily computed through the Taylor expansion around $x=0$ which gives

$$\Phi_{\mu,v} = \sum_{i=0}^{\min(\mu,v)} \frac{\alpha!}{[\alpha - k]!} \gamma^\nu. \quad (13)$$

The symbol $[i/j]$ indicates the integer part of the ratio of the two integers $i$ and $j$. An equilibrium density can be expressed as a ratio of two polynomials in $\gamma$. The density functions for an arbitrary initial state can be calculated by averaging Eq. (12) over an initial probability distribution. Since the goal is to understand noise properties of biochemical reactions in individual cells the usage of the pure states will be implicit and the superscript $a_0$ will be omitted from $\rho_{\mu,v}$.IV. COMBINATORIAL NOISE CONTROL AND FINE TUNING OF NOISE CHARACTERISTICS

The most convenient way to characterize the fluctuations in particle numbers is to introduce the pair correlation functions $\chi_{AA}$, $\chi_{PP}$, and $\chi_{AP}$ for $AA$, $PP$, and $AP$ pairs, respectively, as

$$\chi_{AA} = \frac{\rho_{2,0}}{\rho_{1,1}^2} = 1 + \frac{Q_A}{\langle n \rangle}, \quad (14)$$

$$\chi_{PP} = \frac{\rho_{0,2}}{\rho_{0,1}^{\prime 2}} = 1 + \frac{Q_P}{\langle m \rangle}, \quad (15)$$

$$\chi_{AP} = \frac{\rho_{1,1}^{\prime}}{\rho_{0,1}^{\prime 2}} = \frac{\langle nm \rangle}{\langle n \rangle \langle m \rangle}. \quad (16)$$

The standard deviations are defined in the usual way as $\sigma_A^2 = \langle (n^2) \rangle - \langle n \rangle^2$ and $\sigma_P^2 = \langle (m^2) \rangle - \langle m \rangle^2$, and $Q_A = (\sigma_A^2 - \langle n \rangle^2)/\langle n \rangle$ and $Q_P = (\sigma_P^2 - \langle m \rangle^2)/\langle m \rangle$.

The two-point correlation functions are related to Mandel’s $Q$ parameter [14] used in optics to describe deviations from the Poisson distribution $[Q_A$ and $Q_P$ in Eqs. (14) and (15)]. Let $X=A, P$ and consider an arbitrary particle number distribution. When $\chi_{XX} \approx 1$ ($\chi_{XX} \approx 0$) the particle number distribution is Poisson like. The regime $\chi_{XX} < 1$ ($\chi_{XX} > 1$) points to better (worse) noise control relative to the Poisson case $\chi_{XX}=1$ and will be referred to as a sub-(supra-)Poissonian regime. For $\chi_{AP} < 1$, 1, A and P particles are uncorrelated.

The correlation functions have been investigated for nontrivial cases $a_0 \geq k$. When $a_0 < k$ the reactions are shut down due to the insufficient number of particles. The numbers of $A$ and $P$ particles should be negatively correlated. A large number of $P$ particles corresponds to the situation where there are very few $A$ particles and vice versa. The expectation $\chi_{AP} < 1$ was confirmed by the numerical analysis of a few cases (not shown).

The correlation functions for $AA$ and $PP$ pairs have more interesting behavior. In the $\gamma \rightarrow 0$ limit both correlation functions are smaller than one. When the back-reaction dominates there is a stochastic focusing (sub-Poissonian behavior); the $A$ ($P$) particle number distributions gets compressed toward large (small) $m (n)$ values. The distributions are very narrow.

Inspection of the $\gamma \rightarrow \infty$ limit reveals a rich set of noise characteristics. From Eqs. (12), (14), and (15) one concludes that $\chi_{AA} \propto \gamma^2$ and $\chi_{PP} \propto \gamma^2$ with $\gamma = [a_0/k] + [a_0-k/k]$ and $\theta = [a_0/k] + [a_0-k/k]$. To perform the integer arithmetic it is useful to use $a_0 = nk + \epsilon$ ($n$ and $\epsilon$ are integers). The variable $\epsilon$ specifies how many $A$ particles are left when all $A$ particles that are initially in the system are forced to react. Based on the value of $\epsilon$ several regimes can be identified.
For $\varepsilon = 0$ and $\gamma \rightarrow \infty$, $\chi_{AA} \approx \gamma$ while the PP correlation function saturates to a value less than one. When all $A$ particles can react, the fluctuations in $A$ ($P$) particle numbers are supra-(sub-)Poissonian. Such behavior emerges since the forward reaction is favored and the system spends most of the time in the states with small (large) numbers of $A$ ($P$) particles. Fluctuations from such states dramatically increase the number of A particles from zero to multiples of $k$ which leads to the supra-Poissonian $A$ particle number distribution. The $P$ particle number distribution is compressed strongly toward large $m$ values and becomes sub-Poissonian. When $\varepsilon = 1$ the system frequently visits the state with one $A$ particle (and maximum number of $P$ particles) and the particle number distributions become sub-Poissonian.

Figure 1 indicates that findings from the analysis of the limiting cases extend to the intermediate $\gamma$ values. The curves exhibit rather rich behavior in the intermediate regime. For $\varepsilon = 0$ the amount of noise increases as $\gamma$ becomes larger. With $\varepsilon = 1$ there is the interval in $\gamma$ where $\chi_{AA} > 1$, opening a possibility of noise control by tailoring $\gamma$. Other curves for $\varepsilon = 2, \ldots, k-1$ tend to stay more or less below one, except for the short region of small $\gamma$ values. Please note that for larger $\varepsilon$ values the scenarios repeat periodically. For example, $\varepsilon - k$ (curve not shown) would result in the similar behavior as for the $\varepsilon = 0$ case. Please note that such insights were made possible by having an exact solution at hand. Any approximate treatment would likely miss such behavior.

Figure 2 shows that larger complexes lead to the better noise control of the product molecules. For a fixed $a_0$ increase in the cooperativity degree $k$ leads to the decrease in $\chi_{PP}$ for all values of (especially small) $\gamma$. Due to the $\varepsilon$-effects discussed previously, the dependence of $\chi_{AA}$ correlation functions on $k$ is much more erratic (not shown). Figure 3 shows that the successive increases in $k$ make the $P$ particle number distribution increasingly sub-Poissonian. For a given system, an increase in $k$ has several effects. First, it reduces the lifetimes $\tau_{mn} \sim \Lambda^{2k} + \delta n$ of small $m$ (large $n$) states; $\langle \tau_n \rangle < \langle \tau_{n+1} \rangle$ for $k < n/2$. For the large $n$ states the increase in $k$ extends the number of the clusters that can react. Also, the jumps toward large $m$ (small $n$) states become more frequent. At the same time, the lifetime of large $m$ (small $n$) states increases since the number of clusters that can react decreases; $\langle \tau_n \rangle > \langle \tau_{n+1} \rangle$ for $k > n/2$. These effects combine to give the behavior shown in Fig. 3.

V. CONCLUSIONS

It has been recognized that macromolecular complex formation plays a role in controlling the fluctuations in particle numbers at the single cell level [3]. This study points to several ways of how macromolecular complexes can control intracellular noise.

First, it was shown that a multiparticel-like reaction can reduce the fluctuations in a product particle number. This functionality relies on a simple combinatorial principle and is likely generic. The principle only involves the well known property of binomial coefficients. Accordingly, the sole reason for the existence of a multiparticel-like complex formation process in a pathway might be to reduce the fluctuations in the number of product molecules.

Second, the characteristics of the fluctuations in reactant ($A$) particle number can be controlled by the ratio of the forward and the backward reaction rates (inverse of the dissociation constant). The particle number distribution can be sub- or supra-Poissonian depending on the $\gamma$ value. Rather rich dependence on $\gamma$ has been identified, where various regimes are crossed as the value of $\gamma$ changes.

FIG. 1. For a fixed reaction order $k$ the complex formation reaction has the ability to perform simple integer like arithmetic on $a_0 = \eta k + \varepsilon$ (all variables integers). The three classes of noise characteristics emerge depending on whether $\varepsilon = 0$ (full line), $\varepsilon = 1$ (dashed), and $\varepsilon = 2, \ldots, k-1$ (other lines that saturate to constant values). The plot was made with $k = 4$, $\eta = 2$ and $\varepsilon = 0, 1, 2, 3$.

FIG. 2. For a fixed $a_0$ (20 used in the plot), increase in the reaction order reduces the fluctuations in $P$ particle number; $k = 3$ (full line), $k = 4$ (dashed), $k = 5$ (dotted), $k = 6$ (dash-dot), $k = 7$ (dash-dot-dot), $k = 8$ (dash-dash-dot). The values for the standard deviation were scaled with the average particle numbers (the noise reduction is nontrivial).

FIG. 3. The asymptotic (equilibrium) distribution of $P$ particle number (same parameters as in Fig. 2). Increase in the cooperativity degree $k$ makes the distribution increasingly sub-Poissonian; $k = 3$ (full line), $k = 4$ (dashed), and $k = 7$ (dotted). See the text for the explanation of such behavior.
Third, the reactants noise characteristics strongly depend on how many reactants are left once the maximum number of the complexes is formed. The behavior will get weakened when the system is made open but it might persist if the number of particles does not fluctuate more than \( k \) which could be maintained by other processes in the cell. In the opposite case when fluctuations in the reactant particle number are larger that the cooperativity degree the behavior will not persist. It remains to be seen whether the case where fluctuations in the reactant particle number are larger than the average number of reactants in a reaction could be maintained by other processes in the cell.

From a technical point of view the work introduces a methodology that might be helpful for dealing with the BBGKY hierarchy of equations used in various fields (ranging from quantum physics, to study of density fluctuations in the universe, and spanning over social science and finance). There is an eternal problem of closing the (infinite) hierarchy of equations.

For example, it is hard to describe diffusion controlled reactions in such a way [15]. The standard closures provide qualitative results for systems with large number of particles [15]. Describing a low particle limit is much harder, especially if conservation laws are present [16]. The derivation of the exact result provides an insight of how to use a conservation law in a constructive way. The conservation laws enforced recurrence relation between the density functions. In general, such relations must not be invalidated by a closure. Other technical tricks can be extracted from the calculation but they will not be discussed in here.

ACKNOWLEDGMENTS

The financial support from Chalmers Biocenter and internal MC2 grant for strategic development is greatly acknowledged. The author would like to thank Aldo Jesorka for useful discussions.