



HAL
open science

Necessary and sufficient conditions for protocell growth

Erwan Bigan, Loïc Paulevé, Jean-Marc Steyaert, Stéphane S. Douady

► **To cite this version:**

Erwan Bigan, Loïc Paulevé, Jean-Marc Steyaert, Stéphane S. Douady. Necessary and sufficient conditions for protocell growth. *Journal of Mathematical Biology*, 2016, 10.1007/s00285-016-0998-0. hal-01338156

HAL Id: hal-01338156

<https://hal.science/hal-01338156>

Submitted on 4 Jul 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Necessary and sufficient conditions for protocell growth

Erwan Bigan · Loïc Paulevé · Jean-Marc Steyaert · Stéphane Douady

Received: date / Accepted: date

Abstract We consider a generic protocell model consisting of any conservative chemical reaction network embedded within a membrane. The membrane results from the self-assembly of a membrane precursor and is semi-permeable to some nutrients. Nutrients are metabolized into all other species including the membrane precursor, and the membrane grows in area and the protocell in volume. Faithful replication through cell growth and division requires a doubling of both cell volume and surface area every division time (thus leading to a periodic surface area-to-volume ratio) and also requires periodic concentrations of the cell constituents. Building upon these basic considerations, we prove necessary and sufficient conditions pertaining to the chemical reaction network for such a regime to be met. A simple necessary condition is that every moiety must be fed. A stronger necessary condition implies that every siphon must be either fed, or connected to species outside the siphon through a pass reaction capable of transferring net positive mass into the siphon. And in the case of nutrient uptake through passive diffusion and of constant surface area-to-volume ratio, a sufficient condition for the existence of a fixed point is that every siphon be fed. These necessary and sufficient conditions hold for any chemical reaction kinetics, membrane parameters or nutrient flux diffusion constants.

Keywords Protocell · Chemical Reaction Network · Persistence · Siphons

Erwan Bigan
Laboratoire d'Informatique, École Polytechnique, Palaiseau, France, and
Laboratoire Matière et Systèmes Complexes, Université Paris Diderot, Paris, France
E-mail: bigan.erwan@orange.fr

Loïc Paulevé
Laboratoire de Recherche en Informatique, CNRS and Université Paris Sud, Orsay, France

Jean-Marc Steyaert
Laboratoire d'Informatique, École Polytechnique, Palaiseau, France

Stéphane Douady
Laboratoire Matière et Systèmes Complexes, Université Paris Diderot, Paris, France

Mathematics Subject Classification (2000) MSC 34A12 · 34C11 · 80A30 · 92B05 · 92C42

1 Introduction

Cellular growth and division relies upon harmonious growth whereby all cell constituents (cytoplasm and membrane) are synthesized at the same average rate (averaging over one cell cycle) so that at the end of a cycle one ends up with two daughter cells, each quasi-identical to the mother cell. This raises the following question: does such synchronization require some fine tuning of stoichiometric and/or kinetic parameters (with such fine tuning resulting from evolution in the case of modern evolved cells) or is it automatically achieved for any such choice? This question is of interest not only for the origins of life (how likely may have been the emergence of some protocell capable of stationary growth) or for the design of an artificial protocell, but also for future whole-cell models.

Existing whole-cell models do not answer this question, because they presume the existence of a stationary growth state, do not model the mechanistic steps leading to membrane growth, and rely upon some global phenomenological assumptions. The reason is not only the complexity of the system to be modeled, but also the lack of knowledge about many reaction kinetics. The nature and degree of such assumptions vary depending on the considered model. Constraint-based models such as Flux Balance Analysis (Orth et al 2010), or the model proposed by Molenaar et al 2009, assume that optimal cells have been selected over the course of evolution, such optimum being embodied in an objective function which is optimized under certain constraints. The reaction network stoichiometry constrains any stationary flux distribution (Orth et al 2010) or ribosome allocation (Molenaar et al 2009). An implicit assumption is that the optimal flux distribution or ribosome allocation can be achieved independently of any kinetic limitation. An explicit assumption is the choice of the objective function, which may be the biomass yield per consumed substrate (which requires knowledge of the cell composition) or the growth rate. In contrast, mechanistic whole-cell models do not assume cell optimality, and rely upon a kinetic description of the elementary chemical reactions and physical interactions. This typically results in a coupled set of Ordinary Differential Equations (ODEs) where concentrations of chemical species are variables of the system, and there exists one ODE for any such variable. ODEs also include dilution terms which depend on the growth rate (Pawłowski and Zielenkiewicz 2004). Yet, there is most often no mechanistic description of the elementary steps leading to membrane surface area and cell volume growth (such as membrane precursor synthesis and incorporation in the growing membrane). The growth rate is an additional phenomenological variable, which requires an additional assumption to make the model fully determined (e.g. constant density in Weiße et al 2015, or constant RNA polymerase concentration in Tadmor and Tlustý 2008).

Existing protocell models come closer to answering this question, but only partially. Because the system to be modeled is so much simpler than modern evolved cells, most protocell models include a fine-grained description of all constituents, and a mechanistic description of all chemical reactions, without any global phenomenological assumption. In particular, the growth rate is not a phenomenological parameter: growth in membrane surface area results from the kinetic incorporation of membrane precursors (lipids), and this leads to volume growth in a way which depends on the protocell shape. Also in contrast to whole-cell models, most protocell models do not presume the ability to grow and self-replicate, but attempt to explain when and why such cellular self-replication may occur. Even if emergent synchronization has been numerically observed in such models (characterized by the existence of a stationary growth state), it is unclear what grants such property to the system. Such emergent stationary growth was numerically observed in protocell models based upon large random autocatalytic Chemical Reaction Networks (CRNs) with specific stoichiometry (Kondo and Kaneko 2011; Himeoka and Kaneko 2014). This was also the case in our own previous work using random CRNs with arbitrary stoichiometry (Bigan et al 2015a). Varying the size of such networks, the ability of the cell to 'work' appeared to be determined by simple conditions pertaining only to the CRN stoichiometry and topology, and independent of the detailed kinetics.

The present work not only explains such numerical results but also gives formal answers to the above-mentioned emergent synchronization question. Using a protocell model relying upon any conservative CRN, it proves necessary and sufficient conditions for the existence of a stationary growth state. These conditions only depend on the CRN structure and are independent of the detailed kinetics.

1.1 Relation with Chemical Reaction Network Theory (CRNT)

Relating the existence of a stationary state (fixed point of a set of ODEs) to topological and stoichiometrical properties of a CRN independently of the kinetic details is precisely what Chemical Reaction Network Theory achieves (Feinberg 1979, 1995; Gunawardena 2003). Fundamental CRNT theorems relate such existence to the deficiency, which is an integer number only depending on the CRN structure and independent of the kinetics (Feinberg 1972). Existing theorems only apply to low deficiencies (0 or 1) and are intimately linked to the existence of complex-balanced equilibria (Horn 1972) for which a Lyapunov function can be constructed. However, in the case of the protocell model considered in this work, pseudo-reactions accounting for dilution of chemical species, membrane precursor incorporation and nutrient influx must be added to the embedded CRN. This typically results in large deficiencies¹ which makes existing theorems non-applicable.

¹ For the reference numerical protocell example given in Bigan et al 2015a, the deficiency is 9 (32) for the threshold (maximum-size) network, respectively (using the reference nutrient

The reason why specific results could nevertheless be obtained within the present work is the following: whereas general CRNT theorems are applicable to any CRN (conservative or not), a cellular model defines a very special class of non-conservative CRNs: starting from a conservative CRN representing the inner cell chemistry, specific non-conservative pseudo-reactions are added to account for dilution, membrane precursor incorporation (in the self-assembled membrane), and transmembrane fluxes. The resulting set of ODEs is reminiscent of that for a Continuous-Flow Stirred Tank Reactor (CFSTR), for which additional theorems giving necessary conditions for multistationarity have been proved (Schlosser and Feinberg 1994; Craciun and Feinberg 2005, 2006). Yet, there is a significant difference between the CFSTR and our protocell model: whereas the flow rate is fixed for a CFSTR, its counterpart here is the growth rate, which is a variable function of the protocell state. In the present work, we shall relate the existence of a stationary growth state to a lower bound for the growth rate, which is itself related to persistence of the membrane precursor (Lemmas 2 and 3). Important theoretical results relating persistence to siphons have previously been proved in CRNT (Angeli et al 2007), and we shall make use of such previous work and concepts to prove Theorem 3.

1.2 Relation with moieties and siphons, and with our own previous work

In organic chemistry, a moiety refers to a functional part of a molecule. For example, acyl or phosphate groups are moieties. These functional groups are typically left unchanged by the chemical reactions: when counting the total number of a given moiety, the same result is obtained for both the reactant and product side. In mathematical chemistry, the concept of moiety has been extended to include any elementary conserved quantity, and algorithms have been devised to determine all moieties (Schuster and Höfer 1991), which may not be as obviously visible as the two above simple examples. To each moiety corresponds a positive linear combination of concentrations that is kept constant by the chemical reactions.

In Petri nets, siphons refer to subsets of places that remain empty once they have lost all their tokens (Murata 1989). Applying this concept to chemical reaction networks, siphons are subsets of chemical species that remain forever absent, if already initially so (Angeli et al 2007). In other words, siphons are subsets of species the absence of which cannot be compensated by the chemical reactions.² In mathematical chemistry, siphons are closely related to the concept of persistence (Angeli et al 2007).

and membrane precursor combination). Consistently, numerical simulations on protocells based on random CRNs reveal that the stationary growth states are not complex-balanced.

² One intuitive way to grasp the meaning of siphons is the following: some CRNs are so strongly coupled that if one attempts to lower the concentration of one particular species by an appropriate 'sink' (e.g. incorporation into a structured membrane), then one 'sinks' all other concentrations as well. This is the case when the only siphon is the full set of species. But for CRNs with a weaker coupling, it is foreseeable to 'sink' some species while

Using these concepts, we have previously proved necessary and sufficient conditions for protocell stationary growth: (i) every moiety of the underlying CRN must be fed, or in other words, using the above example, some of the nutrients permeating through the membrane should carry an acyl or a phosphate group; and (ii) assuming a constant surface area-to-volume ratio, it suffices that every siphon of the underlying CRN contain the support of a moiety (that is fed).

We here extend these preliminary results by proving stronger necessary and weaker sufficient conditions:³ (i) every siphon must be either fed or connected to its complement by a specific kind of reaction, named pass reaction, that is capable of transferring mass into the siphon; and (ii) assuming a constant surface area-to-volume ratio, it suffices that every siphon be fed.

Our earlier proofs explained some of our previously reported numerical results using random conservative CRNs and a single diffusing nutrient (Bigan et al 2015a): no growth with two non-overlapping moieties; growth for any choice of nutrient and membrane precursor when the only siphon is the full set of species. The weaker sufficient condition also explains results in intermediate situations, for CRNs having a single moiety and a single siphon that is shorter than the full set of species: when the nutrient is chosen in this siphon, growth is always observed; else, growth is observed in some cases but not in others. And the stronger necessary condition helps understand what may make growth possible or not when the necessary conditions are met but the sufficient condition is not.

These conditions are illustrated with two examples (Section 6): a simple toy model, and a whole-cell model inspired by Molenaar et al 2009. And the applicability of these theoretical results to modern evolved cells or protocells is extensively discussed in Section 7.3.

1.3 Outline

This paper is organized as follows. Section 2 gives some definitions related to CRNs. Section 3 presents a mechanistic fine-grained protocell model and gives some related definitions. Section 4 proves an introductory proposition relating oscillations in concentrations of cell constituents to the cellular shape. Section 5 proves necessary and sufficient conditions for the existence of a stationary growth state. Section 6 illustrates these results with examples. Section 7 is devoted to discussion. And Section 8 gives a conclusion.

keeping other concentrations positive. A subset of species that can be 'sank' is a siphon that is shorter than the full set of species.

³ Condition \mathcal{A} is stronger (weaker) than condition \mathcal{B} if $\mathcal{A} \implies \mathcal{B}$ ($\mathcal{B} \implies \mathcal{A}$), respectively. Stronger necessary and weaker sufficient conditions are desirable for a finer delineation of 'working' protocells.

Notations

In this text, a vector \mathbf{v} is written in bold font. Its i^{th} coordinate is denoted v_i . The i^{th} coordinate of a vector carrying its own index \mathbf{v}_k is denoted $v_{k,i}$.

2 Definitions related to CRNs

Following standard practice in Chemical Reaction Network Theory (Érdi and Tóth 1989) a stoichiometry $N \times R$ matrix \mathbf{S} may be associated to any CRN where N is the total number of chemical species $\mathcal{S} = \{A_i\}_{i=1,\dots,N}$ participating in the CRN and R is the total number of reactions. The r^{th} column of \mathbf{S} is the reaction vector \mathbf{v}_r having as components $\{v_{r,i}\}_{i=1,\dots,N}$ where $v_{r,i}$ is the difference in stoichiometry of species A_i between products and reactants for the r^{th} reaction. For the CRN in a closed system, the time derivative of the $N \times 1$ concentration vector \mathbf{c} is given by $\dot{\mathbf{c}} = \mathbf{S}\mathbf{f}$ where \mathbf{f} is the $R \times 1$ rate vector.

Definition 1 A CRN is *conservative* if there exists a strictly positive $N \times 1$ vector \mathbf{m} such that $\mathbf{m}^T \mathbf{S} = 0$ where $(\cdot)^T$ denotes the transpose of (\cdot) . m_i is the molecular mass of species A_i .

Basically, a CRN is conservative if each chemical species may be assigned a positive mass such that mass conservation be guaranteed for every chemical reaction. There may exist multiple solutions to the set of mass conservation equations. This notion is explicitated through the definition below.

Definition 2 For a conservative CRN, the kernel of the transpose of \mathbf{S} denoted $\text{Ker}(\mathbf{S}^T)$ has dimension $\dim(\text{Ker}(\mathbf{S}^T)) = p \geq 1$, and the rank of \mathbf{S} is $\dim(\text{Im}(\mathbf{S})) = \dim(\text{Im}(\mathbf{S}^T))$ which is equal to $N - \dim(\text{Ker}(\mathbf{S}^T)) = N - p \leq N - 1$. The set of mass vectors compatible with mass conservation is $\{\mathbf{m} \mid m_i > 0, i = 1, \dots, N \text{ and } \mathbf{m}^T \mathbf{S} = 0\}$. This set constitutes a pointed convex cone having $p' \geq p$ generating vectors $\{\mathbf{b}_k\}_{k=1,\dots,p'}$, p of which are linearly independent and constitute a basis of $\text{Ker}(\mathbf{S}^T)$ (Schuster and Höfer 1991). We define as *moieties* the elements of such a set of generating vectors $\{\mathbf{b}_k\}_{k=1,\dots,p}$.

Moieties essentially correspond to positive linear combinations of chemical species concentrations that are left invariant by the chemical reactions.⁴ Moieties have non-negative but not necessarily strictly positive components: indeed, it can be proved that if a moiety has all its components strictly positive, then this is the only moiety, $p = 1$ (see proof in Appendix 1 of Bigan et al 2015a).

⁴ Moieties are non-negative basis vectors of the left-null space of the stoichiometry matrix. Any mass vector can be decomposed as a positive linear combination of moieties. In contrast, extreme pathways (Schilling et al 2000) or the closely related elementary flux modes (Schuster and Hilgetag 1994) are non-negative basis vectors of the right-null space of the stoichiometry matrix. Any stationary flux distribution can be decomposed as a positive linear combination of extreme pathways (or elementary flux modes).

Definition 3 The *support* of a moiety \mathbf{b} (resp. of a concentration vector \mathbf{c}) denoted as $\text{supp}(\mathbf{b})$ (resp. $\text{supp}(\mathbf{c})$) is the subset of chemical species A_i along which \mathbf{b} (resp. \mathbf{c}) has non-zero components.

Definition 4 For a CRN (conservative or not), a *siphon* Z is a subset of \mathcal{S} such that for every species A_i in Z and every reaction where A_i appears as product, then at least one of the reactant species also belongs to Z . A siphon is minimal if it does not contain strictly any other siphon.

The concept of siphon was introduced in Angeli et al 2007 to study persistence in CRNs. Siphons are essentially the sets of chemical species whose absence cannot be compensated by the chemical reactions.

Definition 5 Consider a conservative CRN having a siphon Z that is shorter than the full set of species \mathcal{S} , $\mathcal{S} \setminus Z \neq \emptyset$. Let \mathbf{b} be a moiety or positive linear combination of moieties of this CRN. Reordering species between $\mathcal{S} \setminus Z$ and Z , \mathbf{b} may be decomposed in block form as $\mathbf{b}^T = (\mathbf{b}_{\mathcal{S} \setminus Z}^T | \mathbf{b}_Z^T)$. A reaction characterized by its reaction vector \mathbf{v} is a *pass* reaction for \mathbf{b} if $\Delta b_Z = (\mathbf{0}_{\mathcal{S} \setminus Z}^T | \mathbf{b}_Z^T) \mathbf{v} > 0$, where $\mathbf{0}_{\mathcal{S} \setminus Z}$ is the null vector for the subset $\mathcal{S} \setminus Z$.

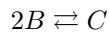
We have $(\mathbf{0}_{\mathcal{S} \setminus Z}^T | \mathbf{b}_Z^T) \mathbf{v} = -(\mathbf{b}_{\mathcal{S} \setminus Z}^T | \mathbf{0}_Z^T) \mathbf{v}$ because the CRN is conservative. The existence of a pass reaction for \mathbf{b} requires that $\text{supp}(\mathbf{b}) \cap Z \neq \emptyset$ and $\text{supp}(\mathbf{b}) \cap (\mathcal{S} \setminus Z) \neq \emptyset$.

Δb_Z is the total product weight minus the total reactant weight for the moiety \mathbf{b} , for species belonging to Z . A pass reaction for the moiety \mathbf{b} is a reaction resulting in a net positive weight transfer (for the moiety \mathbf{b}) from $\mathcal{S} \setminus Z$ into Z . It must involve species in both Z and $\mathcal{S} \setminus Z$. But not all such reactions need necessarily be pass reactions. All possible categories of reactions involving both Z and $\mathcal{S} \setminus Z$ are listed in Table 1 using the following notation: $\mathcal{S} \setminus Z + Z \rightarrow Z$ denotes any reaction such that the reactant complex contains species in $\mathcal{S} \setminus Z$ and in Z , and such that the product complex only contains species in Z . Attributes of such categories are also given: compatibility with the definition of a siphon, and compatibility with net positive weight transfer (for the moiety \mathbf{b}) from $\mathcal{S} \setminus Z$ into Z , i.e. $\Delta b_Z > 0$.

Pass reactions belong either to category $\mathcal{S} \setminus Z + Z \rightarrow \mathcal{S} \setminus Z + Z$ or to category $\mathcal{S} \setminus Z + Z \rightarrow Z$ (with the additional constraint $\Delta b_Z > 0$ in both cases).

The concept of pass reaction is an original contribution of the present work.

Illustration with a simple example Consider the following CRN consisting of two bidirectional reactions:



It can be easily verified that it is conservative with the only possible mass assignment (up to a multiplying factor) being $(m_A, m_B, m_C) = (1, 1, 2)$. The

Reaction category	Compatible with siphon	Compatible with net weight transfer (for \mathbf{b})
$\mathcal{S} \setminus Z \rightarrow Z$	✗	✓
$Z \rightarrow \mathcal{S} \setminus Z$	✓	✗
$\mathcal{S} \setminus Z + Z \rightarrow \mathcal{S} \setminus Z + Z$	✓	✓ if $\Delta b_Z > 0$
$\mathcal{S} \setminus Z \rightarrow \mathcal{S} \setminus Z + Z$	✗	✓
$\mathcal{S} \setminus Z + Z \rightarrow \mathcal{S} \setminus Z$	✓	✗
$Z \rightarrow \mathcal{S} \setminus Z + Z$	✓	✗
$\mathcal{S} \setminus Z + Z \rightarrow Z$	✓	✓ if $\Delta b_Z > 0$

Table 1 Categories of reactions involving species in both $\mathcal{S} \setminus Z$ and Z . The second column indicates whether reactions are compatible with the definition of a siphon or not. The third column indicates whether reactions are compatible with net positive weight transfer (for the moiety \mathbf{b}) from $\mathcal{S} \setminus Z$ into Z , i.e. with $\Delta b_Z > 0$.

mass vector \mathbf{m} is thus the only moiety. It can also be easily verified that $Z = \{B, C\}$ is the only minimal siphon, which is shorter than the full set of species $\mathcal{S} = \{A, B, C\}$, $\mathcal{S} \setminus Z = \{A\} \neq \emptyset$. The first forward reaction $A + B \rightarrow C$ is a pass reaction (for the mass) because the net mass transfer into Z is $\Delta m_Z = m_C - m_B = m_A > 0$.

3 Protocell model and related definitions

3.1 Assumptions

The following assumptions are made in the proposed protocell model:

1. Existence of some chemistry represented by a conservative CRN that is active at least inside the protocell, and that may be assumed either active or inactive outside the protocell (all proofs hold in both cases).
2. Self-assembly of one of the chemical species (membrane precursor A_{me}) in a structured membrane: the incorporation of A_{me} into the growing membrane is assumed to be kinetically controlled, with rate $F_{\text{output,me}}$ per unit area (the corresponding rate vector $\mathbf{F}_{\text{output}}$ has only one non-zero component: $F_{\text{output,me}}$, along A_{me}). The membrane is further characterized by the number of molecules per unit area N_{me} .
3. Membrane precursor incorporation kinetics: the membrane precursor incorporation rate per unit area $F_{\text{output,me}}$ is assumed to be a continuous monotonically increasing function of the concentration vector \mathbf{c} such that $F_{\text{output,me}}(\mathbf{c} = \mathbf{0}) = 0$ and $F_{\text{me}}(\mathbf{c}) > 0$ iff \mathbf{c} has non-zero components along a subset \mathcal{S}_{me} of \mathcal{S} . \mathcal{S}_{me} includes at least A_{me} (membrane precursor must be present for it to be incorporated in the membrane) and may also include other species (e.g. enzymes or other metabolites that may be required in case of catalyzed or active membrane precursor incorporation). This is a very mild assumption as it is verified by all foreseeable kinetics (mass-action, Michaelis-Menten, or active membrane precursor incorporation). Two situations may be envisioned for this incorporation: (i)

either it originates only from the inside of the protocell in which case the rate of incorporation is $F_{\text{output,me}}(\mathbf{c})$, or (ii) it originates from both sides of the membrane in which case the total rate of incorporation is $F_{\text{output,me}}(\mathbf{c}) + F_{\text{output,me}}(\mathbf{c}_{\text{out}})$. Case (i) corresponds to a situation where not all required species are present outside, $\mathcal{S}_{\text{me}} \not\subseteq \text{supp}(\mathbf{c}_{\text{out}})$, or there is some physical or chemical constraint preventing such an incorporation from the outside (e.g. polarity of the self-assembled membrane). Case (ii) corresponds to a non-polar membrane with all required species also present outside, $\mathcal{S}_{\text{me}} \subseteq \text{supp}(\mathbf{c}_{\text{out}})$.

4. Semi-permeability of self-assembled membrane to a subset \mathcal{S}_{nu} of \mathcal{S} . Nutrient uptake from the outside growth medium may result from any mechanism (e.g. passive diffusion or active transport). The resulting nutrient flux vector per unit area is $\mathbf{F}_{\text{input}} = \mathbf{F}_{\text{input}}(\mathbf{c}, \mathbf{c}_{\text{out}})$. It has non-zero components along the subset \mathcal{S}_{nu} that are functions of the inside \mathbf{c} and outside concentration vectors \mathbf{c}_{out} . For any species $A_i \in \mathcal{S}_{\text{nu}}$, $F_{\text{input},i} = 0$ if $c_{\text{out},i} = 0$ (nutrients must be present in the outside growth medium in order to flow inside).
5. Homogeneous concentrations: all chemical species are assumed to be homogeneously distributed and any intracellular or extracellular diffusion effect is neglected. This is a simplifying assumption as, e.g. inside the cell, one should expect nutrient $A_i \in \mathcal{S}_{\text{nu}}$ (resp. membrane precursor A_{me}) concentration to be highest (resp. lowest) near the membrane that acts as an effective source (resp. sink) for such chemical species. Similarly, any intracellular spatial organization is neglected.
6. Large outside growth medium volume compared to the protocell volume: so that the outside concentration vector \mathbf{c}_{out} remains constant even in the presence of a growing protocell. It is either at equilibrium ($\mathbf{S}\mathbf{f}(\mathbf{c}_{\text{out}}) = 0$ assuming the CRN is also active outside), or submitted to a continuous nutrient flow. This equilibrium or stationary \mathbf{c}_{out} is assumed to be reached starting with all and only those species in \mathcal{S}_{nu} in the initial state or in the input flow.

Figure 1 gives a schematic representation of the protocell model.

3.2 Ordinary Differential Equation (ODE) system

With the above assumptions, the ODE system governing the time evolution of concentrations is given by:

$$\dot{\mathbf{c}} = \mathbf{S}\mathbf{f} + \rho(\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}}) - \mu_{\text{inst}}\mathbf{c} \quad (1)$$

where \mathbf{c} is the $N \times 1$ inside concentration vector having as components the concentrations of the N different chemical species inside the protocell, \mathbf{S} is the $N \times R$ stoichiometry matrix associated with the CRN, $\mathbf{f} = \mathbf{f}(\mathbf{c})$ is the $R \times 1$ rate vector having as components the rates for each chemical reaction, $\rho = \mathcal{A}/\mathcal{V}$ is the surface area-to-volume ratio, $\mathbf{F}_{\text{input}} = \mathbf{F}_{\text{input}}(\mathbf{c}_{\text{out}}, \mathbf{c})$ is the $N \times 1$ nutrient

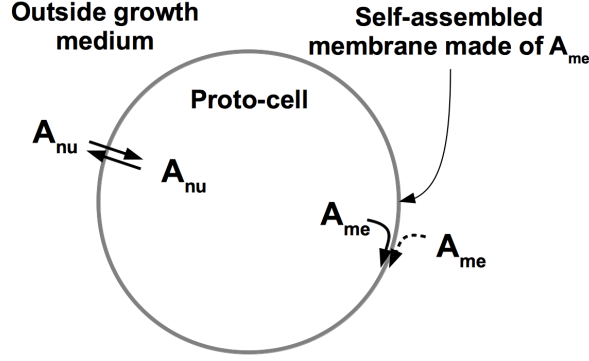


Fig. 1 Schematic of the protocell model. Some nutrient species ($A_{nu} \in \mathcal{S}_{nu}$) flow across the semi-permeable membrane (e.g. by passive diffusion or active transport). The membrane results from the self-assembly of the membrane precursor A_{me} . The membrane surface area grows as the result of A_{me} unidirectional incorporation in the self-assembled membrane.

flux vector per unit area with only non-zero components $F_{input,i}$ for $A_i \in \mathcal{S}_{nu}$, $\mathbf{F}_{output} = \mathbf{F}_{output}(\mathbf{c})$ is the $N \times 1$ membrane incorporation flux vector per unit area with only non-zero component $F_{output,me}(\mathbf{c})$ along A_{me} , and $\mu_{inst} \mathbf{c}$ represents the dilution factor with instantaneous growth rate defined as the instantaneous relative rate of change in volume, $\mu_{inst} = \dot{\mathcal{V}}/\mathcal{V}$.

By definition of ρ and μ_{inst} , we have:

$$\frac{\dot{\rho}}{\rho} = \frac{\dot{\mathcal{A}}}{\mathcal{A}} - \mu_{inst} \quad (2)$$

The membrane surface area \mathcal{A} increases as the result of A_{me} incorporation. If incorporation originates only from the inside of the protocell (case (i) in Assumption 3 above), as A_{me} gets incorporated into the growing membrane with rate (per unit area) $F_{output,me}$, the membrane area grows as:

$$\frac{\dot{\mathcal{A}}}{\mathcal{A}} = \frac{F_{output,me}(\mathbf{c})}{N_{me}} \quad (3)$$

If incorporation originates from both sides (case (ii) in Assumption 3 above), the above equation should be replaced by:

$$\frac{\dot{\mathcal{A}}}{\mathcal{A}} = \frac{F_{output,me}(\mathbf{c}) + F_{output,me}(\mathbf{c}_{out})}{N_{me}} \quad (4)$$

Equations 1, 2 and 3 or 4 do not suffice to make the ODE system autonomous: an additional relation is needed, which depends on the assumption made regarding the protocell shape and the mechanical properties of the membrane. We shall consider different cases to illustrate this point:

1. If, on the one hand, the membrane is so rigid that it is constrained to take a filament shape with fixed diameter (constant $\rho = \rho_0$) regardless of any difference in osmotic pressure across the membrane, then this additional relation is simply $\dot{\rho} = 0$ and the protocell is an autonomous dynamical system having N variables (the components of \mathbf{c}). The μ_{inst} trajectory is then simply given by Equation 2 (taken with $\dot{\rho} = 0$), and Equation 3 or 4.
2. If, on the other hand, the membrane is free to take any shape without any mechanical constraint, the cell shape (and corresponding ρ ratio) adjusts itself to balance the osmotic pressure across the membrane. Assuming the dynamics of such a balancing to be very fast, this corresponds to a constraint of the kind $\mathbf{u}^T \mathbf{c} = P$ where \mathbf{u} is the osmolarity vector such that u_i is the number of particles in solution when dissolving one molecule of A_i , and $P = \mathbf{u}^T \mathbf{c}_{\text{out}}$ is the constant outside growth medium osmolarity. The inside osmolarity is constant, $\mathbf{u}^T \dot{\mathbf{c}} = 0$. Replacing $\dot{\mathbf{c}}$ by Equation 1 and extracting μ_{inst} gives $\mu_{\text{inst}} = \mathbf{u}^T (\mathbf{S}\mathbf{f} + \rho(\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}})) / \mathbf{u}^T \mathbf{c}$. Feeding back this expression for μ_{inst} in Equations 1 and 2 gives an autonomous dynamical system having $N + 1$ variables (the components of \mathbf{c} , and ρ), with both concentrations and surface area-to-volume ratio trajectories solution of this extended set of ODEs. This is similar to the approach taken in Mavelli and Ruiz-Mirazo 2013 or Morgan et al 2004.
3. Between these two above extremes, a more sophisticated description of the mechanical properties of the membrane (as in Surovtsev et al 2009 or Božič and Svetina 2004) would result in a more complex relation and in an autonomous dynamical system having $N + 2$ variables (the components of \mathbf{c} , and, \mathcal{V} and \mathcal{A} , or equivalently ρ and μ_{inst}).

While the proofs of Proposition 1 (Section 4) and of Theorems 1 and 2 (Section 5) hold in any such case and do not require to explicitly take into account such an additional complexity in the model, the proof of the sufficient condition given by Theorem 3 assumes a constant surface area-to-volume ratio (case 1 in the above enumeration).

3.3 Definitions related to protocells

Definition 6 A protocell is an autonomous ODE system such that:

1. Starting from positive initial conditions, it defines trajectories for the concentration vector \mathbf{c} and the cell geometry represented by the cell volume \mathcal{V} and the cell surface area \mathcal{A} (or equivalently, by the instantaneous growth rate μ_{inst} , and the surface area-to-volume ratio ρ).
2. It verifies Assumptions 1-6 of Section 3.1, and Equations 1, 2, and 3 or 4.

Starting from initial conditions $(\mathbf{c}_0, \mathcal{V}_0, \mathcal{A}_0)$, defining trajectories for $(\mathcal{V}, \mathcal{A})$ is equivalent to defining trajectories for $(\mu_{\text{inst}}, \rho)$ because the two may be deduced from one another through the relations $\mu_{\text{inst}} = \dot{\mathcal{V}}/\mathcal{V}$ and $\rho = \mathcal{A}/\mathcal{V}$.⁵

As mentioned at the end of Section 3.2 above, Equations 1, 2 and 3 or 4 do not suffice to make the ODE system autonomous: an additional relation is needed, which depends on the assumption made regarding the protocell shape and the mechanical properties of the membrane.

Definition 7 A protocell characterized by asymptotic trajectories for its constituent concentration vector $\mathbf{c}(t)$, its membrane surface area $\mathcal{A}(t)$ and its volume $\mathcal{V}(t)$ is said to be *working* if $\mathbf{c}(t)$ is periodic with periodicity T and if both $\mathcal{A}(t)$ and $\mathcal{V}(t)$ are doubled during a time period T .

The surface area-to-volume ratio $\rho(t) = \mathcal{A}(t)/\mathcal{V}(t)$ of a working protocell is thus also periodic because $\rho(t+T) = \mathcal{A}(t+T)/\mathcal{V}(t+T) = (2\mathcal{A}(t))/(2\mathcal{V}(t)) = \mathcal{A}(t)/\mathcal{V}(t) = \rho(t)$.⁶

Cell division is not taken into account in this work. The above definition only ensures that faithful replication through cell division be possible. The above definition also implicitly assumes that the global surface $\mathcal{A}(t)$ and volume $\mathcal{V}(t)$ (considering the entire lineage as a single system) and the concentrations $\mathbf{c}(t)$ are continuous functions of time (instantaneous variations through a bursting effect upon cell division are excluded).

Definition 8 The instantaneous growth rate μ_{inst} of a working protocell is defined as $\mu_{\text{inst}} = \dot{\mathcal{V}}/\mathcal{V}$ and its average value is denoted μ_{avg} .

μ_{inst} is also periodic because both the volume and its derivative double every time period T . We have $\mu_{\text{avg}} = (1/T) \int_0^T (\dot{\mathcal{V}}/\mathcal{V}) dt = \log(2)/T$. Although μ_{avg} is positive, μ_{inst} may not necessarily be so at all times. For example, a cell constriction could induce a negative instantaneous growth rate at some point during the cell cycle.

Definition 9 A protocell characterized by asymptotic trajectories for its constituent concentration vector $\mathbf{c}(t)$, its membrane surface area $\mathcal{A}(t)$ and its volume $\mathcal{V}(t)$ is said to be *stationary working* if $\mathbf{c}(t)$ is constant and if both $\mathcal{A}(t)$ and $\mathcal{V}(t)$ are exponentially increasing at the same constant strictly positive relative rate $\mu_{\text{inst}} = \mu_{\text{avg}} = \mu = \dot{\mathcal{V}}/\mathcal{V} = \dot{\mathcal{A}}/\mathcal{A}$.

A stationary working protocell is thus a particular example of a working protocell, with both $\mathcal{A}(t)$ and $\mathcal{V}(t)$ being doubled during a time period $T = \log(2)/\mu$. A stationary working protocell has a constant surface area-to-volume ratio because the relative rates of increase are equal for its surface area and its volume. An example of stationary working protocell is that of filamentous growth, with negligible cross-section compared to its surface area.

⁵ It is implicitly assumed that the $(\mathcal{V}, \mathcal{A})$ trajectories are such that at any instant, there is enough membrane surface area \mathcal{A} to accommodate the volume \mathcal{V} , $\mathcal{A} \geq \sqrt[3]{36\pi\mathcal{V}}$. Else, the protocell may burst. See Mavelli and Ruiz-Mirazo 2013.

⁶ It can also be constant, which is a peculiar periodic function.

Definition 10 Consider a protocell characterized by its embedded conservative CRN, and by its nutrient flux vector $\mathbf{F}_{\text{input}} = \mathbf{F}_{\text{input}}(\mathbf{c}, \mathbf{c}_{\text{out}})$ having nonzero components along a subset \mathcal{S}_{nu} of \mathcal{S} (Assumption 4). A subset Y of the full set of species \mathcal{S} is *fed* if $Y \cap \mathcal{S}_{\text{nu}} \neq \emptyset$.

Y is fed if it contains some of the species that can cross the membrane. Note that Assumption 6 ensures that all species that can cross the membrane are present in the outside growth medium.

This notion shall be applied to the supports of moieties and to siphons. In the case of moieties, we shall say for simplicity that a moiety \mathbf{b} is fed if its support $\text{supp}(\mathbf{b})$ is fed.

4 A constant concentration vector requires a constant surface area-to-volume ratio

As a preamble to necessary and sufficient conditions proved in the next section, we present here a basic introductory proposition.

Proposition 1 *If a working protocell has a constant concentration vector trajectory, then its surface area-to-volume ratio trajectory is also constant, i.e. the protocell is working stationary.*

It is reminded that by virtue of Definition 6, a protocell verifies Assumptions 1-6 of Section 3.1, and Equations 1, 2, and 3 or 4.

Proof Multiplying both sides of Equation 1 by the transpose of the mass vector \mathbf{m}^T and using Equation 2 to replace the instantaneous growth rate μ_{inst} gives:

$$\dot{D} = \rho \mathbf{m}^T (\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}}) - \left(\frac{\dot{\mathcal{A}}}{\mathcal{A}} - \frac{\dot{\rho}}{\rho} \right) D \quad (5)$$

where $D = \mathbf{m}^T \mathbf{c}$ is the density inside the protocell and $\dot{\mathcal{A}}/\mathcal{A}$ is given by Equation 3 or 4. A constant concentration vector \mathbf{c} results in a constant density, $\dot{D} = 0$, constant nutrient input flux $\mathbf{F}_{\text{input}}$ and membrane precursor incorporation output flux $\mathbf{F}_{\text{output}}$ vectors, and a constant $\dot{\mathcal{A}}/\mathcal{A}$ (owing to Equations 3 or 4). The left hand side of the above equation equals zero and the right hand side only contains constant elements, except potentially ρ and $\dot{\rho}$. This makes ρ solution of a differential equation of the form:

$$\frac{\dot{\rho}}{\rho} = \alpha - \beta \rho \quad (6)$$

with:

$$\alpha = \frac{\dot{\mathcal{A}}}{\mathcal{A}} \quad (7)$$

and:

$$\beta = \frac{\mathbf{m}^T(\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}})}{D} \quad (8)$$

The differential equation 6 can be solved analytically, which results in:

$$\rho(t) = \frac{\alpha}{\beta + \gamma e^{-\alpha t}} \quad (9)$$

where $\gamma = (\alpha/\rho_0) - \beta$ and $\rho_0 = \rho(t=0)$ is the initial condition. $\rho(t)$ is a monotonic function that cannot be periodic unless $\gamma = 0$, i.e. $\rho_0 = \alpha/\beta$ in which case $\rho(t)$ remains constant. \square

In particular, this means that if for a working protocell ρ is periodic but not constant (e.g. because of certain geometrical or mechanical constraints), then the concentration of at least some of its constituents is also periodic but not constant. Model examples of such situations have previously been described (Morgan et al 2004; Surovstev et al 2007; Surovtsev et al 2009).

There have also been numerous experimental reports of various concentrations oscillating across the cell cycle in eukaryotic cells (see Busa and Nuccitelli 1984, Wittmann et al 2005 and references therein). It should be noted that such cells tend to have a spherical shape at the beginning of a cell cycle, and that with spherical growth the surface area-to-volume ratio ρ decreases as the sphere radius increases. The required periodicity of ρ implies its increase at some later point during the cell cycle (which may typically result from cell constriction). Proposition 1 suggests that, besides additional regulation mechanisms, variations of ρ across the cell cycle should contribute to such oscillations in concentrations. This stresses the importance of taking cell geometry into account in whole-cell modeling.

5 Necessary and sufficient conditions for a working protocell

5.1 First necessary condition

Theorem 1 *For a protocell to be working, every moiety \mathbf{b} (of the embedded conservative CRN) the support of which contains any species in \mathcal{S}_{me} , $\text{supp}(\mathbf{b}) \cap \mathcal{S}_{me} \neq \emptyset$, must be fed. This is necessary to ensure persistence of species in \mathcal{S}_{me} and a positive growth rate for the protocell ODE system. Further, if all species (not just those in \mathcal{S}_{me}) are to be present in the stationary growth state, then every moiety (not just those such that $\text{supp}(\mathbf{b}) \cap \mathcal{S}_{me} \neq \emptyset$) must be fed.*

It is reminded that by virtue of Definition 6, a protocell verifies Assumptions 1-6 of Section 3.1, and Equations 1, 2, and 3 or 4.

The following lemma is useful to prove this theorem:

Lemma 1 *If a nonnegative function of the concentration vector of a working protocell, $y(t) = \psi(\mathbf{c}(t)) \geq 0$ is such that $\dot{y} \leq -\mu_{inst} \times y$, then $y \rightarrow 0$ if $t \rightarrow +\infty$.*

This is obvious if $\mu_{\text{inst}} = \mu_{\text{avg}} > 0$ is a constant (as is the case for a stationary working protocell) which results in an exponentially decreasing y . It also holds in the most general case where μ_{inst} varies (and may even turn negative at some point during the cell cycle). The proof is given below.

Proof Replacing $y(t)$ by its asymptotic trajectory $y_{\text{as}}(t)$, if $y_{\text{as}} \neq 0$ were possible, we would have:

$$\frac{\dot{y}_{\text{as}}}{y_{\text{as}}} \leq -\mu_{\text{inst}} \quad (10)$$

Integrating this inequality over any time period T would give:

$$\int_{t_1}^{t_1+T} \frac{\dot{y}_{\text{as}}}{y_{\text{as}}} dt = \log\left(\frac{y_{\text{as}}(t_1+T)}{y_{\text{as}}(t_1)}\right) \leq - \int_{t_1}^{t_1+T} \mu_{\text{inst}} dt = \log\left(\frac{V(t_1)}{V(t_1+T)}\right) \quad (11)$$

where t_1 is an arbitrary time. The above inequality is equivalent to:

$$\frac{y_{\text{as}}(t_1+T)}{y_{\text{as}}(t_1)} \leq \frac{V(t_1)}{V(t_1+T)} \quad (12)$$

y being a function of the concentration vector \mathbf{c} that itself converges asymptotically towards a periodic trajectory, $y_{\text{as}}(t)$ is also periodic and the left-hand side of the above inequality is $LHS = 1$. The protocell volume doubling every time period T , the right-hand side is $RHS = 1/2$. This is contradictory with the above inequality. We must therefore have $y_{\text{as}} = 0$. \square

We shall now prove Theorem 1.

Proof Multiplying both sides of the above ODE system (given by Equation 1) by \mathbf{b}^T where \mathbf{b} is any of the p moieties $\{\mathbf{b}_k\}_{k=1,\dots,p}$, gives the ODE governing the time evolution of the quantity $\mathbf{b}^T \mathbf{c}$:

$$\mathbf{b}^T \dot{\mathbf{c}} = \rho \mathbf{b}^T (\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}}) - \mu_{\text{inst}} \mathbf{b}^T \mathbf{c} \quad (13)$$

If there exists a moiety \mathbf{b} that is not fed, then $\text{supp}(\mathbf{b}) \cap \mathcal{S}_{\text{nu}} = \emptyset$. As $\mathbf{F}_{\text{input}}$ only has nonzero components along \mathcal{S}_{nu} , we have $\mathbf{b}^T \mathbf{F}_{\text{input}} = 0$. $\mathbf{b}^T \dot{\mathbf{c}}$ may then be bounded as follows:

$$\mathbf{b}^T \dot{\mathbf{c}} \leq -\mu_{\text{inst}} \mathbf{b}^T \mathbf{c} \quad (14)$$

Applying Lemma 1 with $y = \mathbf{b}^T \mathbf{c}$ results in $\mathbf{b}^T \mathbf{c} \rightarrow 0$ with $t \rightarrow +\infty$. All species in $\text{supp}(\mathbf{b})$ asymptotically disappear.

If $\text{supp}(\mathbf{b}) \cap \mathcal{S}_{\text{me}} \neq \emptyset$, then there exists at least one species in \mathcal{S}_{me} that asymptotically disappears. The membrane area can no longer grow from the inside. Neither can it grow from the outside because the constant \mathbf{c}_{out} is assumed to be reached starting with only species in \mathcal{S}_{nu} in the initial state (or in the constant input flow through the growth medium), see Assumption 6. Any species in the support $\text{supp}(\mathbf{b})$ of any non-fed moiety \mathbf{b} , $\text{supp}(\mathbf{b}) \cap \mathcal{S}_{\text{nu}} = \emptyset$, remains absent outside, $\text{supp}(\mathbf{b}) \cap \text{supp}(\mathbf{c}_{\text{out}}) = \emptyset$, because moieties correspond

to conserved quantities. This results in a non-working protocell because the membrane surface area can neither grow from the inside nor from the outside ($\dot{\mathcal{A}} \rightarrow 0$ because of Equation 3 or 4), whereas \mathcal{A} should double every time period T if the protocell were working). \square

5.2 A stronger necessary condition

Theorem 2 *For a protocell to be working, for every siphon Z (of the embedded conservative CRN) containing any species in \mathcal{S}_{me} , $Z \cap \mathcal{S}_{me} \neq \emptyset$, and every moiety \mathbf{b} the support of which intersects with $Z \cap \mathcal{S}_{me}$, $\text{supp}(\mathbf{b}) \cap (Z \cap \mathcal{S}_{me}) \neq \emptyset$, Z must be either fed or connected to its complement $\mathcal{S} \setminus Z$ through a pass reaction for \mathbf{b} . This is necessary to ensure persistence of species in \mathcal{S}_{me} and a positive growth rate for the protocell ODE system. Further, if persistence of all chemical species is required (not just those in \mathcal{S}_{me}), then for every siphon Z (not just those intersecting with \mathcal{S}_{me}) and every moiety \mathbf{b} the support of which intersects with Z , $\text{supp}(\mathbf{b}) \cap Z \neq \emptyset$, Z must be either fed or connected to its complement $\mathcal{S} \setminus Z$ through a pass reaction for \mathbf{b} .*

It is reminded that by virtue of Definition 6, a protocell verifies Assumptions 1-6 of Section 3.1, and Equations 1, 2, and 3 or 4.

Applying this condition to the mass vector \mathbf{m} , every siphon Z should either be fed or connected to its complement $\mathcal{S} \setminus Z$ through a pass reaction for the mass.

Proof Consider a siphon Z intersecting with the support of a moiety \mathbf{b} . Assume Z is neither fed, nor connected to $\mathcal{S} \setminus Z$ by any pass reaction. As Z is not fed, $\mathcal{S} \setminus Z \neq \emptyset$ because there is at least one nutrient species (see Assumption 6).

We can rearrange species and reaction indices as follows: first species in $\mathcal{S} \setminus Z$ followed by those in Z (as in Definition 5); first reactions involving only species in $\mathcal{S} \setminus Z$, followed by those involving species in both $\mathcal{S} \setminus Z$ and Z , and last those involving only species in Z . The stoichiometry matrix then takes the following block form:

$$\mathbf{S} = \left(\begin{array}{c|c|c} \mathbf{S}_{\mathcal{S} \setminus Z} & & 0 \\ \hline & \mathbf{S}_{\mathcal{S} \setminus Z, Z} & \\ \hline 0 & & \mathbf{S}_Z \end{array} \right)$$

where $\mathbf{S}_{\mathcal{S} \setminus Z}$ (resp. \mathbf{S}_Z) is the stoichiometry matrix subblock for reactions only involving species in $\mathcal{S} \setminus Z$ (resp. in Z) and $\mathbf{S}_{\mathcal{S} \setminus Z, Z}$ is the stoichiometry matrix subblock for reactions connecting $\mathcal{S} \setminus Z$ and Z . With this rearrangement, the transposes of the moiety \mathbf{b} and of the reaction flux vector take the following block form:

$$\mathbf{b}^T = (\mathbf{b}_{\mathcal{S} \setminus Z}^T | \mathbf{b}_Z^T)$$

and,

$$\mathbf{f}^T = (\mathbf{f}_{\mathcal{S}\setminus Z}^T | \mathbf{f}_{\mathcal{S}\setminus Z, Z}^T | \mathbf{f}_Z^T)$$

Decomposing the full ODE system (given by Equation 1) into two ODE subsystems for concentrations of species belonging to $\mathcal{S}\setminus Z$ and Z , respectively, and multiplying these subsets by $\mathbf{b}_{\mathcal{S}\setminus Z}^T$ and \mathbf{b}_Z^T , respectively, gives:

$$\begin{aligned} \mathbf{b}_{\mathcal{S}\setminus Z}^T \dot{\mathbf{c}}_{\mathcal{S}\setminus Z} &= \rho \mathbf{b}^T \mathbf{F}_{\text{input}} + \mathbf{b}_{\mathcal{S}\setminus Z}^T \mathbf{S}_{\mathcal{S}\setminus Z} \mathbf{f}_{\mathcal{S}\setminus Z} \\ &\quad + (\mathbf{b}_{\mathcal{S}\setminus Z}^T | \mathbf{0}_Z^T) \mathbf{S}_{\mathcal{S}\setminus Z, Z} \mathbf{f}_{\mathcal{S}\setminus Z, Z} - \mu_{\text{inst}} \mathbf{b}_{\mathcal{S}\setminus Z}^T \mathbf{c}_{\mathcal{S}\setminus Z} \end{aligned} \quad (15)$$

$$\begin{aligned} \mathbf{b}_Z^T \dot{\mathbf{c}}_Z &= (\mathbf{0}_{\mathcal{S}\setminus Z}^T | \mathbf{b}_Z^T) \mathbf{S}_{\mathcal{S}\setminus Z, Z} \mathbf{f}_{\mathcal{S}\setminus Z, Z} + \mathbf{b}_Z^T \mathbf{S}_Z \mathbf{f}_Z - \rho \mathbf{b}^T \mathbf{F}_{\text{output}} - \mu_{\text{inst}} \mathbf{b}_Z^T \mathbf{c}_Z \end{aligned} \quad (16)$$

We have $\mathbf{b}_{\mathcal{S}\setminus Z}^T \mathbf{S}_{\mathcal{S}\setminus Z} = 0$ and $\mathbf{b}_Z^T \mathbf{S}_Z = 0$ because the CRN is conservative. The two above equations may be rewritten as:

$$\dot{B}_{\mathcal{S}\setminus Z} = \rho \mathbf{b}^T \mathbf{F}_{\text{input}} - \sum_{r \in \{\mathcal{S}\setminus Z, Z\}} \Delta b_{Z,r} \times f_r - \mu_{\text{inst}} B_{\mathcal{S}\setminus Z} \quad (17)$$

$$\dot{B}_Z = \sum_{r \in \{\mathcal{S}\setminus Z, Z\}} \Delta b_{Z,r} \times f_r - \rho \mathbf{b}^T \mathbf{F}_{\text{output}} - \mu_{\text{inst}} B_Z \quad (18)$$

where $B_{\mathcal{S}\setminus Z} = \mathbf{b}_{\mathcal{S}\setminus Z}^T \mathbf{c}_{\mathcal{S}\setminus Z}$ (resp. $B_Z = \mathbf{b}_Z^T \mathbf{c}_Z$) is the contribution of species in $\mathcal{S}\setminus Z$ (resp. Z) to the total weight for moiety \mathbf{b} , $B = \mathbf{b}^T \mathbf{c}$, r is the index spanning the set $\{\mathcal{S}\setminus Z, Z\}$ of reactions connecting Z to $\mathcal{S}\setminus Z$, $\Delta b_{Z,r} = (\mathbf{0}_{\mathcal{S}\setminus Z}^T | \mathbf{b}_Z^T) \mathbf{v}_r = -(\mathbf{b}_{\mathcal{S}\setminus Z}^T | \mathbf{0}_Z^T) \mathbf{v}_r$ is the net weight for moiety \mathbf{b} being transferred from $\mathcal{S}\setminus Z$ into Z by reaction indexed by r (i.e. total product weight minus total reactant weight for species belonging to Z), with reaction vector \mathbf{v}_r . As there is no pass reaction connecting $\mathcal{S}\setminus Z$ to Z , $\Delta b_{Z,r} \leq 0, \forall r \in \{\mathcal{S}\setminus Z, Z\}$, and:

$$\dot{B}_Z \leq -\mu_{\text{inst}} B_Z \quad (19)$$

In the above Equations 17 and 18 we have implicitly assumed that $A_{me} \in Z$. If $A_{me} \in \mathcal{S}\setminus Z$ instead, the term $-\rho \mathbf{b}^T \mathbf{F}_{\text{output}}$ should be in Equation 17 instead of 18. The above inequality 19 holds in both cases.

Applying Lemma 1 with $y = B_Z$ results in $B_Z \rightarrow 0$ with $t \rightarrow +\infty$. All species in $\text{supp}(\mathbf{b}) \cap Z$ asymptotically disappear, with the same consequences as in the proof of Theorem 1.

If $\text{supp}(\mathbf{b}) \cap (Z \cap \mathcal{S}_{\text{me}}) \neq \emptyset$, then there exists at least one species in \mathcal{S}_{me} that asymptotically disappears. The membrane area can no longer grow from the inside. Neither can it grow from the outside because the constant \mathbf{c}_{out} is assumed to be reached starting with only species in \mathcal{S}_{nu} in the initial state (or in the constant input flow through the growth medium), see Assumption 6. Any species in any non-fed siphon Z , $Z \cap \mathcal{S}_{\text{nu}} = \emptyset$, remains absent outside,

$Z \cap \text{supp}(\mathbf{c}_{\text{out}}) = \emptyset$, because siphons are subsets of species whose absence cannot be compensated by chemical reactions. This results in a non-working protocell because the membrane surface area can neither grow from the inside nor from the outside ($\dot{\mathcal{A}} \rightarrow 0$ because of Equation 3 or 4), whereas \mathcal{A} should double every time period T if the protocell were working). \square

This second necessary condition is stronger than the first one because of the following proposition.

Proposition 2 *For a conservative CRN, the support of any moiety is a siphon.*

Proof For any moiety \mathbf{b} and any reaction in a CRN, we shall denote by product \mathbf{b} -weight (resp. reactant \mathbf{b} -weight) the sum of the components of \mathbf{b} along product species (resp. reactant species) weighted by the corresponding stoichiometric coefficients. Conservativity ensures that the \mathbf{b} -weight is conserved on either side of the reaction, i.e. the product \mathbf{b} -weight equals the reactant \mathbf{b} -weight.

If any species A_i in $\text{supp}(\mathbf{b})$ appears on the product side of a reaction, then the corresponding product \mathbf{b} -weight is strictly positive, and so is the corresponding reactant \mathbf{b} -weight. A strictly positive reactant \mathbf{b} -weight implies that at least one of the reactant species belongs to $\text{supp}(\mathbf{b})$, which matches the definition of a siphon. \square

Assume the second necessary condition is verified. For any moiety \mathbf{b} , the above proposition states that $Z = \text{supp}(\mathbf{b})$ is a siphon. Either Z is fed, or connected to its complement $\mathcal{S} \setminus Z$ by a pass reaction for \mathbf{b} . The existence of such a pass reaction would require $\text{supp}(\mathbf{b}) \cap (\mathcal{S} \setminus Z) \neq \emptyset$ (see note after Definition 5), but this is impossible because $Z = \text{supp}(\mathbf{b})$. This proves that $Z = \text{supp}(\mathbf{b})$ must be fed. The first condition is then also verified. This proves that the second condition is stronger than the first one.

5.3 Sufficient condition

The sufficient condition presented in this section only holds in the case where nutrient uptake results from passive diffusion and where the surface area-to-volume ratio $\rho = \rho_0$ remains constant.

From Equation 2, the growth rate is given by $\mu_{\text{inst}} = \dot{\mathcal{A}}/\mathcal{A}$ which is itself given by Equation 3 or 4 depending on whether membrane precursor incorporation originates only from the inside or from both sides of the membrane.

Denoting the effective diffusion constant of species $A_i \in \mathcal{S}_{\text{nu}}$ as \mathcal{D}_i , each of the non-zero components $F_{\text{input},i}$, $A_i \in \mathcal{S}_{\text{nu}}$, of the nutrient flux vector per unit area $\mathbf{F}_{\text{input}}$ can be expressed as:

$$F_{\text{input},i} = \mathcal{D}_i(c_{\text{out},i} - c_i) \quad \forall i \in \mathcal{S}_{\text{nu}} \quad (20)$$

We shall denote $\mathbf{F}_{\text{input,max}}$ the maximum input flux vector having as non-zero components $F_{\text{input,max},i} = \mathcal{D}_i c_{\text{out},i}$, $\forall i \in \mathcal{S}_{\text{nu}}$.

Theorem 3 Consider a protocell for which:

1. Nutrient uptake results from passive diffusion
2. The surface area-to-volume ratio ρ remains constant

If every siphon Z (of the embedded conservative CRN) containing any species in \mathcal{S}_{me} is fed, then there exists a fixed point with all species in \mathcal{S}_{me} being persistent and with positive growth rate for the protocell ODE system. Further, if every siphon (not just those containing any species in \mathcal{S}_{me}) is fed, then all chemical species are persistent and present in the fixed point.

It is reminded that by virtue of Definition 6, a protocell verifies Assumptions 1-6 of Section 3.1, and Equations 1, 2, and 3 or 4.

In essence, this theorem suggests that a sufficient condition for a protocell to be working is that every siphon of the embedded conservative CRN is fed, provided the additional assumptions of nutrient uptake through passive diffusion and of constant $\rho = \rho_0$ are made. This is not exactly so because the existence of a fixed point does not guarantee its stability (working stationary protocell) or its oscillatory behavior (working protocell). More complex bounded trajectories cannot be excluded, although numerical analyses on randomly generated CRNs suggest that this fixed point is actually stable and unique, but this remains a conjecture at this stage.^{7,8}

As for the above necessary conditions, this sufficient condition only assumes mass conservation of the embedded CRN and holds for any kinetics including that for membrane precursor incorporation.

The main steps of the proof are summarized below, each step corresponding to a distinct lemma:

1. Any autonomous dynamical system with bounded trajectories has a fixed point.⁹ In particular, any reaction system with bounded density has a fixed point. This can be deduced from Brouwer's fixed point theorem (Wei 1962) (Lemma 2).
2. If A_{me} is persistent (i.e. c_{me} has a lower bound), then the cytoplasmic density has an upper bound. This results from directly bounding Equation 13 taken for $\mathbf{b} = \mathbf{m}$. The chemical system trajectory is thus bounded (Lemma 3).
3. If every siphon (of the embedded conservative CRN) containing A_{me} is fed, then A_{me} is persistent. This is because if A_{me} were to disappear, then so would all other species in at least one of the siphons containing A_{me} , and this is contradictory with this siphon being fed (Lemma 4).

⁷ We were unable to construct a Lyapunov function for this protocell model. As mentioned in Section 1.1, the Lyapunov function used in existing deficiency-based CRNT theorems requires complex-balanced equilibria. Consistently, numerical simulations show that protocell stationary growth states are not complex-balanced (see Footnote 1).

⁸ The fact that complex bounded trajectories cannot be excluded (in which case the protocell might not be working) is the reason why Definitions 6 and 7 are kept distinct.

⁹ This fixed point may depend on the initial conditions. A typical example (considered in Wei 1962) is a closed conservative chemical reaction system: both the trajectory upper bound and the equilibrium point depend on the initial conditions (which determine the total mass in the system). See also Footnote 10.

We now enunciate and prove these lemmas.

Lemma 2 *Consider a protocell for which the surface area-to-volume ratio ρ remains constant. If the cytoplasmic density $D = \mathbf{m}^T \mathbf{c}$ remains bounded for the entire protocell trajectory given by the autonomous differential equation $\dot{\mathbf{c}} = g(\mathbf{c})$ with $g(\mathbf{c}) = \mathbf{S}\mathbf{f} + \rho_0(\mathbf{F}_{\text{input}} - \mathbf{F}_{\text{output}}) - \mu_{\text{inst}}\mathbf{c}$, then there exists a stationary point \mathbf{c}_{st} such that $g(\mathbf{c}_{\text{st}}) = 0$.*

Proof The proof results from a generalization of Brouwer's fixed point theorem to dynamical systems, and was given in Wei 1962. This result is not specific to chemical systems and holds for any continuous autonomous dynamical system $\dot{\mathbf{x}} = g(\mathbf{x})$ having a semiflow. A more restrictive statement of this result, based on a similar proof as that given in Wei 1962 can also be found in Basener et al 2006 and Richeson et al 2002: if a subset K of the entire space of all possible \mathbf{x} is non-empty, convex and forward-invariant for the flow of g , then there exists a fixed point \mathbf{x}_0 such that $g(\mathbf{x}_0) = 0$.¹⁰ \square

Lemma 3 *Consider a protocell for which:*

1. *Nutrient uptake results from passive diffusion*
2. *The surface area-to-volume ratio ρ remains constant*

If all concentrations for chemical species in \mathcal{S}_{me} admit strictly positive lower bounds, then for any moiety or positive linear combination of moieties \mathbf{b} , the corresponding quantity $\mathbf{b}^T \mathbf{c}$ has an upper bound. In particular, $D = \mathbf{m}^T \mathbf{c}$ is bounded.

Proof If all concentrations for chemical species in \mathcal{S}_{me} have strictly positive lower bounds, then so has the membrane precursor incorporation rate $F_{\text{output}} \geq F_{\text{output, min}} > 0$ because of Assumption 3. With the assumption of constant $\rho = \rho_0$ the growth rate is $\mu_{\text{inst}}(t) = \mathcal{A}/\mathcal{V}$ which is given by Equation 3 or 4. Whether given by any of these two equations, $\mu_{\text{inst}}(t)$ has a strictly positive lower bound $\mu_{\text{inst}}(t) \geq \mu_{\text{min}} = F_{\text{output, min}}/N_{\text{me}} > 0$, and the time evolution of $\mathbf{b}^T \mathbf{c}(t)$ can be bounded as follows:

$$\mathbf{b}^T \dot{\mathbf{c}} \leq \rho_0 \mathbf{b}^T \mathbf{F}_{\text{input, max}} - \mu_{\text{min}}(B_{\text{me}} + \mathbf{b}^T \mathbf{c}) \quad (21)$$

where $\mathbf{F}_{\text{input, max}}$ is the maximum nutrient flux vector and $B_{\text{me}} = \rho_0 \times b_{\text{me}} \times N_{\text{me}}$. The above equation may be rewritten in a more compact form:

$$\dot{B} \leq \xi - \mu_{\text{min}} B \quad (22)$$

where $B = \mathbf{b}^T \mathbf{c}$ and $\xi = \rho_0 \mathbf{b}^T \mathbf{F}_{\text{input, max}} - \mu_{\text{min}} B_{\text{me}}$. This is equivalent to:

$$\dot{B} + \mu_{\text{min}} B \leq \xi \quad (23)$$

¹⁰ The original proof in Wei 1962 holds even if the trajectory bound depends on the initial conditions. Whereas the statement in Basener et al 2006 and Richeson et al 2002 relies upon the existence of a convex and forward-invariant set, which corresponds to a trajectory bound independent of the initial conditions (after a sufficient time).

Multiplying both sides by $\exp(\mu_{\min}t)$ gives:

$$(\dot{B} + \mu_{\min}B) \exp(\mu_{\min}t) \leq \xi \exp(\mu_{\min}t) \quad (24)$$

or equivalently:

$$(B \exp(\mu_{\min}t)) \leq \xi \exp(\mu_{\min}t) \quad (25)$$

Integrating from $t' = 0$ until $t' = t$ gives:

$$B(t) \exp(\mu_{\min}t) - B_0 \leq \xi \frac{\exp(\mu_{\min}t) - 1}{\mu_{\min}} \quad (26)$$

where $B_0 = B(t = 0)$. Multiplying both sides by $\exp(-\mu_{\min}t)$ and rearranging terms gives:

$$B(t) \leq B_0 \exp(-\mu_{\min}t) + \xi \frac{1 - \exp(-\mu_{\min}t)}{\mu_{\min}} \quad (27)$$

The right-hand side is a bounded function, which proves that $B(t)$ has an upper bound. Applying this to the particular case of the mass, $\mathbf{b} = \mathbf{m}$, shows that if all concentrations for chemical species in \mathcal{S}_{me} have strictly positive lower bounds, then the cytoplasmic density D is bounded. \square

Lemma 4 *Consider a protocell for which:*

1. *Nutrient uptake results from passive diffusion*
2. *The surface area-to-volume ratio ρ remains constant*

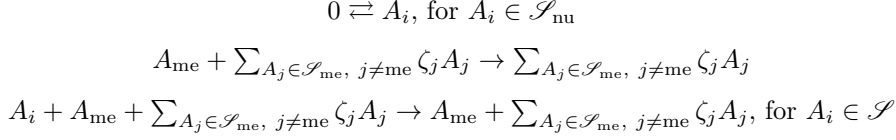
If every siphon Z (of the embedded conservative CRN) containing any species in \mathcal{S}_{me} is fed, then the concentrations trajectories of all species in \mathcal{S}_{me} have strictly positive lower bounds when starting from strictly positive initial concentrations (i.e. are persistent).

The proof will make use of a previous result that relates persistence to siphons in CRNs. It was first proved in Angeli et al (2007) for conservative CRNs and later extended to non-conservative CRNs in Angeli et al (2011). It requires only mild assumptions on reaction kinetics (these should be nonnegative, continuous, and monotonic functions of concentrations, with null reaction rate if any reactant concentration is null). The key result is Theorem 4 in Angeli et al (2011) which states that any extinction set is a siphon.¹¹

We shall exploit this result by applying it to an extended non-conservative CRN the dynamics of which are fully described by the ODE system given by Equation 1, in the case of nutrient flux through passive diffusion and in the

¹¹ An extinction set is defined as follows: if for some particular set of strictly positive initial conditions, the corresponding ω -limit set of a bounded concentration vector trajectory has zero concentration for some non-empty subset Z of all species, then Z is an extinction set. Note that this definition as well as the statement of Theorem 4 in Angeli et al (2011) calls for bounded concentration vector trajectories. However, its proof appears not to rely upon this assumption and only requires continuity (of trajectories and kinetics), which is the reason why we use it here in this extended form.

case of constant surface area-to-volume ratio. This extended non-conservative CRN is obtained by adding to the original embedded conservative CRN, the following pseudo-reactions accounting for nutrient flux across the membrane, membrane precursor incorporation, and dilution:



The nutrient flux (bidirectional) pseudo-reaction has forward (constant) rate equal to $\rho_0 F_{\text{input}, \max, i}$ for $A_i \in \mathcal{S}_{\text{nu}}$ and reverse rate equal to $\rho_0 \mathcal{D}_i c_i$. The rate of the (unidirectional) membrane precursor incorporation is $\rho_0 F_{\text{output}}(\{c_j\}_{j \in \mathcal{S}_{\text{me}}})$. It only consumes one molecule of A_{me} but may require the presence of additional species A_j with stoichiometric coefficient ζ_j (e.g. catalysts) in case \mathcal{S}_{me} contains more than just A_{me} . And the rate of the (unidirectional) dilution for species A_i is $\mu_{\text{inst}} c_i = (F_{\text{output}}(\{c_j\}_{j \in \mathcal{S}_{\text{me}}})/N_{\text{me}})c_i$ in case membrane precursor incorporation originates only from inside. If incorporation also occurs from the outside, this can be taken into account through additional pseudo-reactions $A_i \rightarrow 0$ with rate $(F_{\text{output}}(\mathbf{c}_{\text{out}})/N_{\text{me}})c_i$ for $A_i \in \mathcal{S}$. Pseudo-reactions accounting for membrane precursor incorporation and dilution and their rates have only been given for the sake of completeness, but only the nutrient flux forward pseudo-reaction is key to our proof of Lemma 4, as will be made clear from the following lemma.

Lemma 5 *If a non-conservative CRN contains a pseudo-reaction of the kind $0 \rightarrow A_i$, then no siphon may contain A_i .*

Proof By definition of a siphon, if A_i were to belong to a siphon Z , then there would at least be one species on the reactant side of the reaction $0 \rightarrow A_i$ that would also belong to Z . But this is impossible because no such species exists. \square

In the particular case of nutrient influx by diffusion and of constant surface area-to-volume ratio, if the subset $\{A_i\}$ is fed, then A_i cannot go extinct, whatever other reactions in the CRN may consume A_i .

We are now in a position to prove Lemma 4.

Proof Assume one of the species in \mathcal{S}_{me} were not persistent. Then this species would belong to a siphon Z of the extended non-conservative CRN. By definition of a siphon, Z would also be a siphon of the embedded conservative CRN, that is made of a subset of all reactions in the extended non-conservative CRN. And at least one of the nutrient species A_i in \mathcal{S}_{nu} would also belong to the siphon Z because every siphon of the embedded conservative CRN is fed. But by virtue of Lemma 5 the associated nutrient flux forward pseudo-reaction $0 \rightarrow A_i$, $A_i \in \mathcal{S}_{\text{nu}}$, prevents the existence of any such siphon (of the extended non-conservative CRN). This proves that all species in \mathcal{S}_{me} are persistent. \square

We now combine Lemmas 2, 3 and 4 to prove Theorem 3.

Proof If every siphon (of the embedded conservative CRN) containing any species in \mathcal{S}_{me} is fed, Lemma 4 ensures that all species in \mathcal{S}_{me} are persistent. Applying Lemma 3 to the particular case of the mass ensures that if all species in \mathcal{S}_{me} are persistent, then the cytoplasmic density D is bounded. And Lemma 2 ensures that if D is bounded, the protocell dynamical system has a stationary point. Persistence of all species in \mathcal{S}_{me} further ensures that this stationary point corresponds to a positive growth rate.

Further, if every siphon of the embedded conservative CRN (not just those containing any species in \mathcal{S}_{me}) is fed with some nutrient flux, then Lemma 4 ensures that all chemical species are persistent and present in the fixed point. \square

The proof only ensures the existence of a fixed point and a bounded trajectory, but does not guarantee its unicity nor its stability.

Summary of mathematical results and assumptions

Table 2 summarizes mathematical results proved in the preceding two sections, along with required assumptions.

		Result statement	Assumptions
Necessary conditions	Proposition 1	$\dot{c} = 0 \implies \dot{\rho} = 0$	– Assumptions 1-6 in Section 3.1
	Theorem 1	Each moiety must be fed	
	Theorem 2	Each siphon must be fed or connected to its complement via a pass reaction	
Sufficient condition	Theorem 3	Each siphon is fed	– Assumptions 1-6 in Section 3.1 – Fixed-diameter filament shape ($\rho = \rho_0$ constant) – Nutrient influx by diffusion

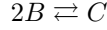
Table 2 Summary of results in Sections 4 and 5 and required assumptions. The statement of Theorem 2 has been simplified by only applying it to the particular case where $\mathbf{b} = \mathbf{m}$.

6 Illustration with examples

6.1 Simple example

Consider the same CRN example as given at the end of Section 2:





We recall that it is conservative with $(m_A, m_B, m_C) = (1, 1, 2)$ as single moiety, and with $\{B, C\}$ as only minimal siphon.

Let us endow this CRN with mass-action kinetics by letting $k_{1,2}^+$ (resp. $k_{1,2}^-$) be the forward (resp. reverse) kinetic constants for the first and second reaction.

Consider a protocell embedding such a CRN with C chosen as membrane precursor (with no other species required for membrane precursor incorporation). The corresponding rate is assumed to also follow mass-action kinetics with constant K_{me} per unit area, $F_{\text{output}} = K_{\text{me}}c_C$. As in the previous sections, N_{me} is the number of molecules per unit area in the structured membrane.

Let us assume that A is the single nutrient flowing across the membrane through passive diffusion with \mathcal{D} as effective diffusion constant, and that the protocell has a constant surface area-to-volume ratio $\rho = \rho_0$. Let us also assume that A is the only species present in the outside growth medium so that membrane precursor incorporation only occurs from the inside. The first necessary condition (Theorem 1) is met because the support of the only moiety is the full set of species which is fed. The second necessary condition (Theorem 2) is also met because $A + B \rightarrow C$ is a pass reaction (the net mass transfer into Z is $m_C - m_B = m_A > 0$). But the sufficient condition given in Theorem 3 is not met because the minimal siphon $\{B, C\}$ is not fed.

Consistently, we shall show analytically that for certain kinetic and membrane parameters, the protocell ODE system cannot have any fixed point.

The protocell ODE system is given by:

$$\dot{c}_A = \rho_0 \mathcal{D}(c_{\text{out},A} - c_A) - k_1^+ c_A c_B + k_1^- c_C - \mu_{\text{inst}} c_A \quad (28)$$

$$\dot{c}_B = -k_1^+ c_A c_B + k_1^- c_C - 2k_2^+ c_B^2 + 2k_2^- c_C - \mu_{\text{inst}} c_B \quad (29)$$

$$\dot{c}_C = k_1^+ c_A c_B - k_1^- c_C + k_2^+ c_B^2 - k_2^- c_C - \rho_0 K_{\text{me}} c_C - \mu_{\text{inst}} c_C \quad (30)$$

Looking for a fixed point, we set all time derivatives to zero. Adding Equations 29 and 30 with time derivatives set to zero gives:

$$-k_2^+ c_B^2 + k_2^- c_C - \mu_{\text{inst}} c_B - \rho_0 K_{\text{me}} c_C - \mu_{\text{inst}} c_C = 0 \quad (31)$$

Rearranging terms, we obtain:

$$k_2^+ c_B^2 + \mu_{\text{inst}}(c_B + c_C) = (k_2^- - \rho_0 K_{\text{me}}) c_C \quad (32)$$

The *LHS* of the above equation is positive and the *RHS* is positive only if $k_2^- > \rho_0 K_{\text{me}}$. This shows that the existence of a fixed point is excluded within a certain range of kinetic and membrane parameters.

Choosing B as nutrient instead of A , the sufficient condition becomes met which ensures the existence of a fixed point. Analytically proving the existence

of such a fixed point is quite challenging even with such a simple network, and this illustrates the power of Theorem 3.

The fact that the sufficient condition is not met does not preclude the existence of a fixed point. Indeed, Appendix A shows a variation of this simple example where nutrient influx results from active transport instead of passive diffusion: the nutrient A is only present outside and is transformed into C across the membrane under the activation of B . Appendix A proves analytically that such a protocell may or may not exhibit a fixed point (other than the degenerate zero concentration vector corresponding to an empty protocell) depending on kinetic and membrane parameters.

6.2 Whole-cell model inspired by Molenaar et al 2009

In this mechanistic, coarse-grained whole-cell model, metabolites, proteins and chemical reactions are the same as in Molenaar et al 2009. The original model was a constraint-based model which presumed the existence of a stationary growth state, and which was used to gain insight into specific characteristics of the cell. Instead, the same coarse-grained description is used here in a fully mechanistic model, in order to determine whether a stationary growth state exists for any choice of kinetic parameters or not.

Embedded CRN The embedded CRN consists of two metabolic reactions:



and of four protein synthesis reactions:



There are seven participating chemical species: three metabolites (intracellular substrate S , precursor P , and lipid L) and four proteins (ribosome R catalyzing the transformation of P into any protein, enzyme E_1 catalyzing the metabolization of S into P , enzyme E_2 catalyzing the metabolization of P into L , and transmembrane transporter T). n_i is the number of precursors P required to synthesize a protein indexed by i ($i = 1, 2, R$ or T , for E_1, E_2, R or T , respectively).

In the following, c_X is the concentration of species X for $X = S, P, L, R$ or T , and c_1 (c_2) is the concentration of E_1 (E_2), respectively. All the above reactions are endowed with Michaelis-Menten kinetics with kinetic coefficient

k_{cat_i} and saturation concentration K_{m_i}), where i is the protein index, $i = 1, 2$ or R (the relation between transmembrane transporter T and substrate influx is more complex than a simple Michaelis-Menten relation and is considered below).

The reaction rates for the two metabolic reactions (indexed by their reaction product, P and L , respectively) are given by:

$$\nu_P = \frac{k_{\text{cat}_1} c_S c_1}{K_{m_1} + c_S} \quad (39)$$

$$\nu_L = \frac{k_{\text{cat}_2} c_P c_2}{K_{m_2} + c_P} \quad (40)$$

And the rate of synthesis of a protein indexed by i ($i = 1, 2, R$ or T) is given by:

$$\nu_i = \frac{k_{\text{cat}_R} n_i c_P c_R}{K_{m_R} + n_1 c_P + n_2 c_P + n_R c_P + n_T c_P} = \frac{k_{\text{cat}_R} n_i c_P c_R}{K_{m_R} + n_{\text{tot}} c_P} \quad (41)$$

where $n_{\text{tot}} = n_1 + n_2 + n_R + n_T$. The above formulation reflects the distribution of the total ribosome concentration c_R : the fraction of ribosome devoted to synthesis of a protein indexed by i (i.e. bound to the corresponding mRNA) is $\phi_i = n_i c_P / (K_{m_R} + n_{\text{tot}} c_P)$ while the fraction of free ribosomes is $\phi_{\text{free}} = K_{m_R} / (K_{m_R} + n_{\text{tot}} c_P)$. We thus have $\nu_i = \phi_i k_{\text{cat}_R} c_R$. Such a formulation is consistent with the mechanistic model of Scott et al 2010, as well as with that of Weiße et al 2015 with the additional simplifying assumption that all genes are transcribed at the same rate in the present case.

Membrane growth L and T self-assemble into a structured membrane. Both contribute to the total surface membrane surface area \mathcal{A} :

$$\mathcal{A} = \mathcal{A}_L + \mathcal{A}_T \quad (42)$$

where \mathcal{A}_L (\mathcal{A}_T) is the contribution of L (T) to the membrane surface area, respectively. We call α the fraction of total membrane surface area occupied by T :

$$\alpha = \frac{\mathcal{A}_T}{\mathcal{A}} \quad (43)$$

Assuming the incorporation of both L and T is proportional to \mathcal{A}_L ¹², we have:

$$\frac{\dot{\mathcal{A}}_L}{\mathcal{A}_L} = \frac{F_{\text{output},L}}{N_L} \quad (44)$$

and:

¹² L or T can only incorporate among already self-assembled L . As in Molenaar et al 2009, this excludes a growing membrane composed only of T but the corresponding process is modeled mechanistically here, instead of being imposed as an external constraint.

$$\frac{\dot{\mathcal{A}}_T}{\mathcal{A}_L} = \frac{F_{\text{output},T}}{N_T} \quad (45)$$

where $F_{\text{output},L}$ ($F_{\text{output},T}$) is the rate of incorporation of L (T) in the growing membrane, and N_L (N_T) is the number of molecules of L (T) per unit area in the self-assembled membrane, respectively.¹³

As in the previous example, $F_{\text{output},L}$ ($F_{\text{output},T}$) is assumed to follow mass-action kinetics with constant K_L (K_T) per unit area, $F_{\text{output},L} = K_L c_L$ ($F_{\text{output},T} = K_T c_T$), respectively.

Assuming for simplicity a fixed-diameter filament shape ($\rho = \rho_0$), the growth rate is given by $\mu = \dot{\mathcal{A}}/\mathcal{A}$. Adding Equations 44 and 45 and multiplying the resulting sum by $\mathcal{A}_L/\mathcal{A} = 1 - \alpha$ gives:

$$\mu = \frac{\dot{\mathcal{A}}}{\mathcal{A}} = (1 - \alpha) \left(\frac{F_{\text{output},L}}{N_L} + \frac{F_{\text{output},T}}{N_T} \right) \quad (46)$$

The time derivative of α is obtained by deriving Equation 43:

$$\dot{\alpha} = \frac{\dot{\mathcal{A}}_T}{\mathcal{A}} - \frac{\mathcal{A}_T \dot{\mathcal{A}}}{\mathcal{A}^2} = \frac{\mathcal{A}_L \dot{\mathcal{A}}_T}{\mathcal{A} \mathcal{A}_L} - \frac{\mathcal{A}_T \mathcal{A}_L \dot{\mathcal{A}}}{\mathcal{A}^2 \mathcal{A}_L} \quad (47)$$

which, using the relations $\alpha = \mathcal{A}_T/\mathcal{A}$ and $1 - \alpha = \mathcal{A}_L/\mathcal{A}$, as well as Equations 44 and 45, gives:

$$\dot{\alpha} = (1 - \alpha) \left(\frac{F_{\text{output},T}}{N_T} - \alpha \left(\frac{F_{\text{output},L}}{N_L} + \frac{F_{\text{output},T}}{N_T} \right) \right) \quad (48)$$

Nutrient influx The nutrient influx per unit area, $F_{\text{input},S}$, is assumed proportional to the fraction of total membrane surface area occupied by T , which is α :

$$F_{\text{input},S} = \alpha \phi_{\text{input},S} \quad (49)$$

where $\phi_{\text{input},S}$ is the nutrient influx rate per unit surface area occupied by T (or equivalently, $\phi_{\text{input},S}/N_T$ is the nutrient influx rate per self-assembled molecule of transporter). We shall consider two cases for $\phi_{\text{input},S}$:

1. Facilitated diffusion (as with the GLUT glucose transporter family in mammalian cells, see Olson and Pessin 1996):

$$\phi_{\text{input},S} = \frac{\mathcal{D}(c_{\text{out},S} - c_S)}{1 + \left| \frac{c_{\text{out},S} - c_S}{K_{m,S}} \right|} \quad (50)$$

where $c_{\text{out},S}$ is the substrate concentration in the outside growth medium and \mathcal{D} is an effective diffusion constant.

¹³ Or equivalently, $1/N_L$ ($1/N_T$) is the membrane surface area per self-assembled molecule of L (T), respectively.

2. Active transport (as with the PTS glucose transporter in *E. coli*, see Jahreis et al 2008): in the most general formulation, $\phi_{\text{input},S} = \phi_{\text{input},S}(c_{\text{out},S}, c_{\text{act}})$ which should be an increasing and saturating function of the outside nutrient concentration $c_{\text{out},S}$ and of an inside activator concentration c_{act} (see also discussion in Section 7.2). Consistent with the original model of Molenaar et al 2009 which has no explicit energy currency metabolite, we arbitrarily assume that c_{act} is constant, which gives:¹⁴

$$\phi_{\text{input},S} = \frac{K_S c_{\text{out},S}}{1 + \frac{c_{\text{out},S}}{K_{m_S}}} \quad (51)$$

Applicability of necessary and sufficient conditions The model described here is different from the one described in Section 3 because of the composite nature of the membrane, and because nutrient flows only through the transporter T . It can be easily seen that the general proofs of the two necessary conditions are still valid, even with a composite membrane and selective nutrient flow. Let us verify whether these conditions are met or not for the present example.

The embedded CRN is conservative and has a single moiety because there is only one (up to a multiplying factor) solution to the mass conservation problem: masses of all metabolites, S , P and L , must be equal, $m_S = m_P = m_L = m$; and the mass m_i of any protein indexed by i is given by $m_i = n_i \times m$. The first necessary condition is thus met with a single nutrient influx of S .

The embedded CRN has three minimal siphons: $Z_1 = \{S\}$, $Z_2 = \{R\}$ and $Z_3 = \{E_1, P\}$. Z_1 is fed. Z_2 is not but $n_R P + R \rightarrow 2R$ is a pass reaction for Z_2 , with $\Delta m_{Z_2} = m_R$ as net mass inflow into Z_2 . Z_3 is also not fed but $S + E_1 \rightarrow P + E_1$ is a pass reaction for Z_3 , with $\Delta m_{Z_3} = m_P$. The second necessary condition is met because every siphon is either fed or connected to its complement through a pass reaction.

But it can also be easily seen that the proof of the sufficient condition is no longer valid with a selective nutrient flow only through T (it would still be valid with a composite membrane such that nutrient could flow through any of its constituents). This is because (i) the α multiplicative term in Equation 49 might vanish even if $\phi_{\text{input},S}$ were to meet the condition that each siphon be fed; and (ii) Equations 46 and 48 cannot be easily associated to pseudo-reactions, even introducing new pseudo-species.¹⁵

The mathematical results of the main text do not enable us to make any firm statement regarding the existence of a stationary growth state. Yet, Appendix B proves that a stationary growth state exists for any parameter set. This shows that when necessary conditions are met but no sufficient condi-

¹⁴ As explained in the next discussion section, the transmembrane protein T is distinct from the activator molecule which is a high-energy metabolite, down-converted to a lower-energy molecule in the active transport process.

¹⁵ Even if the proof of the sufficient condition still held, this sufficient condition (which grants the existence of a stationary growth state if every siphon is fed) would obviously not even be met because neither Z_2 nor Z_3 are fed.

tion is applicable, one cannot draw any firm conclusions without a detailed analysis.

Such a mechanistic coarse-grained full-cell model could be analyzed further to re-derive some of the already established bacterial growth laws (Scott et al 2010; Weiße et al 2015). But the point we wish to make is different: Appendix A presents a simple protocell example for which the necessary conditions are met but the sufficient condition is not, and for which a stationary growth state only exists within a certain range of parameters; this second example is a coarse-grained whole-cell model for which the necessary conditions are met but the sufficient condition is not even applicable. Yet, a stationary growth state exists for any parameter set. This shows that when necessary conditions are met but no sufficient condition is applicable, one cannot draw any firm conclusions without a detailed analysis.

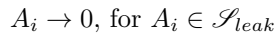
7 Discussion

7.1 Production, degradation, and leakage of chemical species

The proposed protocell model relies upon any conservative CRN. This also encompasses any situation where elementary building blocks (e.g. nucleotides) may be degraded or produced from simpler constituents inside the protocell.

The proved necessary and sufficient conditions even hold in the case where some species may leak out of the membrane. We shall assume that $\mathcal{S}_{\text{leak}}$, $\mathcal{S}_{\text{leak}} \cap \mathcal{S}_{\text{nu}} = \emptyset$, is the set of leaking species. With leakage, any conserved quantity $\mathbf{b}^T \mathbf{c}$ decreases at a faster rate, which implies that inequalities 14 and 19 still hold, and so do the necessary conditions.

Likewise, the density upper bound granted by Lemma 3 in case of persistence of any species in \mathcal{S}_{me} also holds. The only outstanding question is whether persistence of any species in \mathcal{S}_{me} (in case every siphon containing any species in \mathcal{S}_{me} is fed) still holds. The following pseudo-reactions must be added to the non-conservative extended CRN to account for leakage:



By definition of a siphon, such pseudo-reactions have no impact on siphons, because no species appears on the product side. This proves that Lemma 4 also holds with leakage of some species, and so does the sufficient condition.

In writing the above pseudo-reaction, we have implicitly assumed that leakage occurs through a passive mechanism such as diffusion, and that the leaking species are absent from the outside growth medium. If the leaking species are also present in the outside growth medium, they should be handled as any other nutrient species. The sufficient condition grants the existence of a fixed point, but does not tell us which species concentration settles at a higher or lower inside vs. outside concentration. In such a stationary growth regime, we can simply observe that stationary concentrations settle to values such that

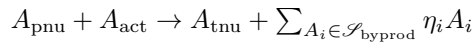
as a whole there is a net positive material influx to ensure a positive growth rate, i.e. the rate of mass nutrient influx exceeds the rate of mass leakage.

This proves that the same necessary conditions hold even with leakage of any kind, and that the sufficient condition holds even with leakage by passive diffusion.

7.2 Active transport vs. passive diffusion

The necessary conditions of Theorems 1 and 2 hold for any type of nutrient influx (e.g. passive diffusion or active transport). But the sufficient condition of Theorem 3 only holds in the case where nutrient flux results from passive diffusion (and for a constant surface area-to-volume ratio).

Considering active transport instead of passive diffusion, and assuming a single nutrient species for simplicity, a third species participating in the CRN should be assigned a specific role, beyond the membrane precursor and nutrient: the activator A_{act} , the consumption of which would be required to import the primary nutrient A_{pnu} from the outside growth medium inside the protocell. The corresponding reaction could be modeled as a unidirectional surfacic reaction occurring across the membrane:



where A_{tnu} is the transformed nutrient under the action of A_{act} and where $A_i \in \mathcal{S}_{\text{byprod}}$ are additional potential byproduct species that may be produced with stoichiometry coefficient η_i .

The original primary nutrient A_{pnu} (e.g. glucose) may never be present inside the protocell (unless it can also diffuse across the membrane) while A_{tnu} is the transformed nutrient inside the protocell (e.g. glucophosphate). This nutrient input reaction may thus be the only reaction in which the nutrient species A_{pnu} participates.¹⁶ In that case the full set of species excluding the primary nutrient, $\mathcal{S} \setminus \{A_{\text{pnu}}\}$, is obviously a siphon Z . The nutrient input reaction is a pass reaction for Z but this siphon is not fed (i.e. there is no species in Z crossing the membrane). This means persistence of A_{me} can no longer be ensured. In fact, the zero concentration vector (all species inside the protocell going extinct) is a fixed point of the ODE system, which is not the case with passive diffusion because there is still a constant strictly positive nutrient flux $F_{\text{input,max,nu}} = \mathcal{D}c_{\text{out,nu}}$ when all concentrations are null.

¹⁶ The activator molecule is distinct from the transmembrane protein introduced in the example of Section 6.2, which acts as a channel for A_{pnu} and which was not included in the generic model of Section 3. In the case of the PTS glucose transport system for *E. coli*, this activator is phosphoenolpyruvate (*pep*) which is down-converted to pyruvate (*pyr*) while having at the same time glucose (*glu*) converted to glucophosphate (*g6p*) (Jahreis et al 2008). The corresponding transport reaction is $glu + pep \rightarrow g6p + pyr$, which only occurs across the membrane so that only glucophosphate (and not glucose) is present inside the cell. This activator molecule should also not be confused with a co-factor, which is typically left unchanged by a biochemical reaction. The activator molecule should rather be seen as an energy currency fueling the active transport process.

The fact that the sufficient condition is not met does not preclude the existence of a fixed point, as shown by the examples of Section 6 and Appendices A and B.

7.3 Applicability to modern evolved cells or to protocells

Similar to other CRNT results, the proved necessary and sufficient conditions only rely upon stoichiometric and topological characteristics of the embedded CRN, and are independent of the reaction kinetics. That living systems might 'work' for any parameter set is counterintuitive and calls for a specific discussion.

In actual chemical reaction networks, kinetics are constrained by thermodynamics (Ederer and Gilles 2007). A CRN in a closed system must verify detailed balance at thermodynamic equilibrium, which induces constraints such as the Wegscheider condition on the choice of kinetic parameters. As proofs hold for any conservative CRN, they also hold for the particular class of thermodynamically-consistent CRNs. Yet, there are several reasons why living systems might only 'work' within specific parameter ranges.

1. A first reason is that even if the sufficient condition is met, Theorem 3 only ensures a strictly positive growth rate, which could be so low that in practice this would correspond to no growth. Indeed, endowing random conservative CRNs with thermodynamically-consistent kinetics, numerical simulations show that a significant growth rate is only achieved when the nutrient (membrane precursor) Gibbs energy per unit molecular weight is high (low), respectively (Bigan et al 2015a).
2. A second reason is that non-fed siphons might be quite widespread:
 - (a) With respect to modern evolved cells, biochemical reaction networks exhibit non-fed siphons. Besides the above-discussed active transport case, the subset of species $Z = \{R\}$ consisting of only the ribosome R always constitute a siphon, as illustrated by the coarse-grained mechanistic whole-cell model presented in Appendix B. This is because the reactions where R appears on the reactant side are: (i) either protein synthesis reactions of the kind $P + R \rightarrow E + R$ where P is a lumped term designating all amino-acid precursors and E is the synthesized protein; or (ii) a ribosome synthesis reaction of the kind $P + R \rightarrow 2R$. $Z = \{R\}$ verifies the definition of a siphon, and this siphon is not fed because ribosomes are not imported from the outside growth medium but only synthesized inside.
 - (b) With respect to protocells, non-fed siphons may be key to the effective chemical insulation between the inside and the outside of the protocell. Nutrients could remain stable in the outside growth medium and be metabolized into all protocell constituents inside the protocell. This suggests that the set of all cell constituents excluding the nutrient would be a siphon, which would be empty in the outside growth medium. We have proposed that such chemical architectures, combined with a pass

reaction catalyzed by the membrane, might have led to the emergence of active transport (which provides the ultimate insulation between the inside and the outside) by an evolutionary scenario (Bigan et al 2015c). This raises the question of what may be needed to make a living system 'work' when necessary conditions are met, but the sufficient condition is not. Equation 18 shows that the pass reaction should transfer mass into the siphon at a sufficient rate for the siphon not to go empty (persistence of species in the non-fed siphon ensures a bounded density and the existence of a fixed point). This rate is expected to increase on (i) increasing the kinetic coefficient of the pass reaction, (ii) increasing the standard Gibbs energy drop across the pass reaction if it is reversible (so as to keep sufficiently low the rate at which this reverse reaction removes matter from the siphon), or (iii) increasing the synthesis rate of other species in Z appearing on the reactant side of the pass reaction (which play the same role as A_{act} in the case of active transport). These three factors should contribute to making the system 'work'.

This is consistent with the 'working' parameter range identified for the simple example analyzed in Appendix A: the active transport pass reaction is unidirectional (corresponding to an infinite standard Gibbs energy drop), and the existence of a stationary growth state is granted provided the kinetic coefficient (k_2^-) of the reaction synthesizing the species in Z appearing on the reactant side of the pass reaction (B), is sufficiently high. It is also consistent with the example of Section 6.2 and Appendix B: the pass reactions are also unidirectional, and the existence of a stationary growth state is granted for any parameter set.

It should be noted that multiple pass reactions may have to be crossed to reach a non-fed siphon starting from a nutrient: combining the above ribosome siphon example ($Z = \{R\}$) with nutrient influx by active transport, would require at least two pass reactions to be crossed: the nutrient influx pseudo-reaction; and the ribosome synthesis reaction, $P + R \rightarrow 2R$.

3. A third and last reason is that the assumption of a fixed-diameter filament shape might not necessarily hold.¹⁷ Although filamentation is a widely observed growth mode (Jensen and Woolfolk 1985; Dickinson 2008), vesicles most often take a spherical shape (Stano and Luisi 2010). Besides, even for rod-shaped bacteria growing by elongation, the rod diameter is not necessarily constant: it has been observed to increase on increasing nutrient quality and growth rate for some Gram-negative bacteria (Schaechter et al 1958). This suggests a complex interplay between osmotic pressure, metabolism and the mechanical properties of the membrane. Further the-

¹⁷ Such an assumption does not automatically grant synchronization of the rates of synthesis of membrane and cytoplasmic constituents: a counterexample is given by the simple example of Appendix A, for which the existence of a fixed point is only granted within some parameter range.

oretical work is needed to investigate protocell behavior with a refined modeling of the membrane taking into account its mechanical properties.¹⁸

On the other hand, and in spite of these caveats, results presented in this work also suggest that, even if living systems might not 'work' for any parameter set, they may well do so over very broad ranges of parameters. In particular for protocells, synchronization between the rate of synthesis of membrane and cytoplasmic constituents may not require any fine tuning of kinetic or membrane parameters, but may simply emerge as a system-level property.¹⁹ This result is expected to be relevant for the origins of life and the artificial life research communities, because it shows that the emergence of cellular self-replication is more likely than one might think at first sight. The results presented in this work are only a first step towards a finer delineation of minimal conditions for cellular self-replication. In particular, more theoretical work is required towards identifying the conditions under which pass reactions can transfer mass into siphons at a sufficient rate to make the cell 'work'. The present results should also stimulate the further development of fully mechanistic whole-cell models (i.e. including a mechanistic description of elementary processes leading to cellular growth, starting with membrane surface area growth). The conclusions of such models should have a lesser dependence on global phenomenological assumptions (such as constant density in Weiße et al 2015, or constant RNA polymerase concentration in Tadmor and Tlusty 2008).

8 Conclusion

We have proposed a generic protocell model consisting of any conservative chemical reaction network embedded within a membrane resulting from the self-assembly of one the chemical species participating in the reaction network. The membrane is assumed to be semi-permeable to some other chemical species (nutrients) flowing across the membrane. With these assumptions, we have proved that a necessary condition for the protocell to grow and divide periodically is that every moiety must be fed, and that a stronger necessary condition implies that every siphon must be connected to its complement through a pass reaction resulting in net positive mass transfer into the siphon. With the additional assumptions of constant surface area-to-volume ratio and of nutrient flow by passive diffusion, we have proved that a sufficient condition for the existence of a stationary growth state is that every siphon be fed.

¹⁸ Since the original submission of the present work, we have carried out additional numerical and theoretical work on protocells having a membrane that is so flexible that the osmotic pressure is quasi-instantaneously balanced across the membrane. Filamentation was found to be an emerging growth mode in such a case, but with a filament diameter that depends on the environmental conditions as well as on the membrane and chemical reaction network parameters (Bigan et al 2015d).

¹⁹ This statement only applies to protocells. Some modern evolved cells have acquired complex regulatory mechanisms to regulate their size or shape, which are presumably fine-tuned. An example is the osmotic stress pathway and osmoadaptation in yeast (Hohmann 2002).

These conditions hold for any kinetics, membrane parameters or diffusion constants. This work is a first step towards a finer delineation of chemical reaction networks compatible with cellular growth. It shows that cellular growth may simply emerge as a system-level property without requiring any fine tuning of kinetic or membrane parameters. We hope this work will further stimulate the development of formal fully mechanistic fine-grained whole-cell models, which might eventually lead to the identification of a new generation of 'growth laws' without requiring any global phenomenological assumption.

Acknowledgements The authors would like to thank Pierre Legrain, Laurent Schwartz and Pierre Plateau for stimulating discussions.

References

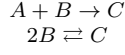
- Angeli D, De Leenheer P, Sontag ED (2007) A Petri net approach to the study of persistence in chemical reaction networks. *Mathematical Biosciences* 210(2):598–618
- Angeli D, De Leenheer P, Sontag ED (2011) Persistence results for chemical reaction networks with time-dependent kinetics and no global conservation laws. *SIAM Journal on Applied Mathematics* 71(1):128–146
- Basener W, Brooks BP, Ross D (2006) The Brouwer fixed point theorem applied to rumour transmission. *Applied Mathematics Letters* 19(8):841–842
- Bigan E, Steyaert JM, Douady S (2015a) Minimal conditions for protocell stationary growth. *Artificial Life* 21(2):166–192
- Bigan E, Steyaert JM, Douady S (2015b) On necessary and sufficient conditions for proto-cell stationary growth. *Electronic Notes in Theoretical Computer Science* 316:3–15
- Bigan E, Steyaert JM, Douady S (2015c) Chemical schemes for maintaining different compositions across a semi-permeable membrane with application to proto-cells. *Origins of Life and Evolution of Biospheres* 45(4):439–454
- Bigan E, Steyaert JM, Douady S (2015d) Filamentation as a primitive growth mode? *Physical Biology* 12(6):066024
- Božič B, Svetina S (2004) A relationship between membrane properties forms the basis of a selectivity mechanism for vesicle self-reproduction. *European Biophysics Journal* 33(7):565–571
- Busa W, Nuccitelli R (1984) Metabolic regulation via intracellular pH. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 246(4):R409–R438
- Craciun G, Feinberg M (2005) Multiple equilibria in complex chemical reaction networks: I. the injectivity property. *SIAM Journal on Applied Mathematics* 65(5):1526–1546
- Craciun G, Feinberg M (2006) Multiple equilibria in complex chemical reaction networks: II. the species-reaction graph. *SIAM Journal on Applied Mathematics* 66(4):1321–1338

- Dickinson J (2008) Filament formation in *Saccharomyces cerevisiae* – a review. *Folia Microbiologica* 53(1):3–14
- Ederer M, Gilles E (2007) Thermodynamically feasible kinetic models of reaction networks. *Biophysical Journal* 92(6):1846–1857
- Érdi P, Tóth J (1989) *Mathematical models of chemical reactions: theory and applications of deterministic and stochastic models*. Manchester University Press
- Feinberg M (1972) Complex balancing in general kinetic systems. *Archive for Rational Mechanics and Analysis* 49(3):187–194
- Feinberg M (1979) *Lectures on chemical reaction networks*. Notes of lectures given at the Mathematics Research Center, University of Wisconsin, Madison, WI. Available for download at <https://crnt.osu.edu/LecturesOnReactionNetworks>
- Feinberg M (1995) The existence and uniqueness of steady states for a class of chemical reaction networks. *Archive for Rational Mechanics and Analysis* 132(4):311–370
- Gunawardena J (2003) *Chemical reaction network theory for in-silico biologists*. Notes available for download at <http://vcp.med.harvard.edu/papers/crnt.pdf>
- Himeoka Y, Kaneko K (2014) Entropy production of a steady-growth cell with catalytic reactions. *Physical Review E* 90(4):042,714
- Hohmann S (2002) Osmotic stress signaling and osmoadaptation in yeasts. *Microbiology and Molecular Biology Reviews* 66(2):300–372
- Horn F (1972) Necessary and sufficient conditions for complex balancing in chemical kinetics. *Archive for Rational Mechanics and Analysis* 49(3):172–186
- Jahreis K, Pimentel-Schmitt EF, Brückner R, Titgemeyer F (2008) Ins and outs of glucose transport systems in eubacteria. *FEMS Microbiology Reviews* 32(6):891–907
- Jensen RH, Woolfolk CA (1985) Formation of filaments by *Pseudomonas putida*. *Applied and Environmental Microbiology* 50(2):364–372
- Kondo Y, Kaneko K (2011) Growth states of catalytic reaction networks exhibiting energy metabolism. *Physical Review E* 84(1):011,927
- Mavelli F, Ruiz-Mirazo K (2013) Theoretical conditions for the stationary reproduction of model protocells. *Integrative Biology* 5(2):324–341
- Molenaar D, van Berlo R, de Ridder D, Teusink B (2009) Shifts in growth strategies reflect tradeoffs in cellular economics. *Molecular Systems Biology* 5(1):323
- Morgan JJ, Surovtsev IV, Lindahl PA (2004) A framework for whole-cell mathematical modeling. *Journal of Theoretical Biology* 231(4):581–596
- Murata T (1989) Petri nets: properties, analysis and applications. *Proceedings of the IEEE* 77(4):541–580
- Olson AL, Pessin JE (1996) Structure, function, and regulation of the mammalian facilitative glucose transporter gene family. *Annual Review of Nutrition* 16(1):235–256

- Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? *Nature Biotechnology* 28(3):245–248
- Pawłowski PH, Zielenkiewicz P (2004) Biochemical kinetics in changing volumes. *Acta Biochimica Polonica* 51:231–243
- Richeson D, Wiseman J, et al (2002) A fixed point theorem for bounded dynamical systems. *Illinois Journal of Mathematics* 46(2):491–495
- Schaechter M, Maaløe O, Kjeldgaard NO (1958) Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *Microbiology* 19(3):592–606
- Schilling CH, Letscher D, Palsson BØ (2000) Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of Theoretical Biology* 203(3):229–248
- Schlosser PM, Feinberg M (1994) A theory of multiple steady states in isothermal homogeneous CFSTRs with many reactions. *Chemical Engineering Science* 49(11):1749–1767
- Schuster S, Hilgetag C (1994) On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems* 2(02):165–182
- Schuster S, Höfer T (1991) Determining all extreme semi-positive conservation relations in chemical reaction systems: a test criterion for conservativity. *Journal of the Chemical Society, Faraday Transactions* 87(16):2561–2566
- Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: origins and consequences. *Science* 330(6007):1099–1102
- Stano P, Luisi PL (2010) Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chemical Communications* 46(21):3639–3653
- Surovstev IV, Morgan JJ, Lindahl PA (2007) Whole-cell modeling framework in which biochemical dynamics impact aspects of cellular geometry. *Journal of Theoretical Biology* 244(1):154–166
- Surovtsev IV, Zhang Z, Lindahl PA, Morgan JJ (2009) Mathematical modeling of a minimal protocell with coordinated growth and division. *Journal of Theoretical Biology* 260(3):422–429
- Tadmor AD, Tlusty T (2008) A coarse-grained biophysical model of *E. coli* and its application to perturbation of the rRNA operon copy number. *PLoS Computational Biology* 4(4):e1000,038
- Wei J (1962) Axiomatic treatment of chemical reaction systems. *The Journal of Chemical Physics* 36(6):1578–1584
- Weiß AY, Oyarzún DA, Danos V, Swain PS (2015) Mechanistic links between cellular trade-offs, gene expression, and growth. *Proceedings of the National Academy of Sciences* 112(9):E1038–E1047
- Wittmann C, Hans M, Van Winden WA, Ras C, Heijnen JJ (2005) Dynamics of intracellular metabolites of glycolysis and TCA cycle during cell-cycle-related oscillation in *Saccharomyces cerevisiae*. *Biotechnology and bioengineering* 89(7):839–847

A Analytical treatment of a simple example protocell with nutrient influx by active transport

Consider the same simple CRN example endowed with mass-action kinetics as described in Section 6, but where the first reaction is assumed unidirectional and only occurs across the membrane:



A is the primary nutrient present in the outside growth medium, $c_{\text{out},A} > 0$ and is absent from the protocell, C is the transformed nutrient under activation by B . B and C are the only species present inside. C also plays the role of membrane precursor with incorporation kinetic coefficient K_{me} per unit area. We shall assume a constant surface area-to-volume ratio ρ_0 so that the surfacic kinetic coefficient of the first surfacic reaction can be transformed in a volumic kinetic coefficient, k_1^+ . As in Section 6, the kinetic coefficients for the second reaction are k_2^\pm .

The protocell ODE system is given by:

$$\dot{c}_B = -k_1^+ c_{\text{out},A} c_B - 2k_2^+ c_B^2 + 2k_2^- c_C - \mu_{\text{inst}} c_B \quad (52)$$

$$\dot{c}_C = k_1^+ c_{\text{out},A} c_B + k_2^+ c_B^2 - k_2^- c_C - \rho_0 K_{\text{me}} c_C - \mu_{\text{inst}} c_C \quad (53)$$

with the growth rate μ_{inst} given by:

$$\mu_{\text{inst}} = \frac{K_{\text{me}} c_C}{N_{\text{me}}} \quad (54)$$

Looking for a fixed point, we set all time derivatives to zero. Adding Equations 52 and 53 with time derivatives set to zero results in the same equation as Equation 32 which we restate here:

$$k_2^+ c_B^2 + \mu_{\text{inst}} (c_B + c_C) = (k_2^- - \rho_0 K_{\text{me}}) c_C \quad (55)$$

Replacing the growth rate μ_{inst} by its expression given in Equation 54 gives a quadratic equation that uniquely determines c_B as a function of c_C provided $k_2^- > \rho_0 K_{\text{me}}$ and $c_C < N_{\text{me}}(\frac{k_2^-}{K_{\text{me}}} - \rho_0)$:

$$c_B = \phi_1(c_C) = \frac{-K_{\text{me}} c_C + \sqrt{(K_{\text{me}} c_C)^2 + 4k_2^+ c_C N_{\text{me}} ((k_2^- - \rho_0 K_{\text{me}}) N_{\text{me}} - K_{\text{me}} c_C)}}{2k_2^+ N_{\text{me}}} \quad (56)$$

It can be verified that $\phi_1(c_C) = 0$ for $c_C = 0$ and for $c_C = N_{\text{me}}(\frac{k_2^-}{K_{\text{me}}} - \rho_0)$, that it is strictly positive for c_C in between those two zeros, and that its derivative is infinite at $c_C = 0$.

Adding Equation 52 to Equation 53 multiplied by two (thus obtaining the time derivative of the system density) and setting all time derivatives to zero, we obtain:

$$k_1^+ c_{\text{out},A} c_B - 2\rho_0 K_{\text{me}} c_C - \mu_{\text{inst}} (c_B + 2c_C) = 0 \quad (57)$$

Replacing the growth rate μ_{inst} by its expression given in Equation 54 and rearranging terms gives the following expression for c_B as a function of c_C :

$$c_B = \phi_2(c_C) = \frac{2K_{\text{me}} c_C (\rho_0 N_{\text{me}} + c_C)}{k_1^+ c_{\text{out},A} N_{\text{me}} - K_{\text{me}} c_C} \quad (58)$$

It can be verified that $\phi_2(c_C) = 0$ for $c_C = 0$, that $\phi_2(c_C) \rightarrow +\infty$ with $c_C \rightarrow \frac{k_1^+ c_{\text{out},A} N_{\text{me}}}{K_{\text{me}}}$, that it takes strictly positive values for c_C in between, and that its derivative is finite at $c_C = 0$.

Because at $c_C = 0$, the derivative of ϕ_1 is infinite and that of ϕ_2 is finite, $\phi_1(c_C) > \phi_2(c_C)$ in the neighbourhood of $c_C = 0$. Besides, $\phi_2(c_C) > \phi_1(c_C)$ in the neighbourhood of some strictly positive c_C value because of the following considerations:

1. If $k_1^+ c_{\text{out},A} > k_2^- - \rho_0 K_{\text{me}}$, then $\phi_2(c_C) > \phi_1(c_C) = 0$ at $c_C = N_{\text{me}}(\frac{k_2^-}{K_{\text{me}}} - \rho_0)$.
2. Else if $k_1^+ c_{\text{out},A} \leq k_2^- - \rho_0 K_{\text{me}}$, then $\phi_2(c_C) \rightarrow +\infty$ and $\phi_1(c_C) \rightarrow \phi_1(N_{\text{me}}(\frac{k_2^-}{K_{\text{me}}} - \rho_0))$ (which is finite) when $c_C \rightarrow N_{\text{me}}(\frac{k_2^-}{K_{\text{me}}} - \rho_0)$.

This implies that $c_B = \phi_1(c_C)$ and $c_B = \phi_2(c_C)$ intersect for some strictly positive set of concentrations (c_{B0}, c_{C0}) .

We have thus proved the existence of a fixed point provided $k_2^- > \rho_0 K_{\text{me}}$. This simple example illustrates the fact that in the case of active transport, the protocell may or may not exhibit a fixed point (other than the degenerate zero concentration vector corresponding to an empty protocell) depending on kinetic and membrane parameters.

B Analytical treatment of a mechanistic, coarse-grained whole-cell model inspired by Molenaar et al 2009

This appendix proves analytically that the whole-cell model described in Section 6.2 exhibits a stationary growth state for any parameter set.

B.1 ODE system

The ODE system governing the cell dynamics is composed of:

1. Three differential equations for metabolites:

$$\dot{c}_S = \rho_0 \alpha \phi_{\text{input},S} - \nu_P - \mu c_S \quad (59)$$

$$\dot{c}_P = \nu_P - \nu_L - n_1 \nu_1 - n_2 \nu_2 - n_R \nu_R - n_T \nu_T - \mu c_P \quad (60)$$

$$\dot{c}_L = \nu_L - \rho_0 F_{\text{output},L} - \mu c_L \quad (61)$$

2. Four differential equations for proteins:

$$\dot{c}_i = \nu_i - \mu c_i \quad (62)$$

where $i = 1, 2$ or R , and:

$$\dot{c}_T = \nu_T - \rho_0 F_{\text{output},T} - \mu c_T \quad (63)$$

3. One differential equation for α (Equation 48)

B.2 Expressing all concentrations as functions of α and of the normalized precursor concentration $p = c_P/K_{m_R}$

Looking for a stationary state, we set all time derivatives to zero. Combining Equations 62 and 41, both taken for $i = R$, gives:

$$\mu = \frac{k_{\text{cat}_R} n_{RP}}{1 + n_{\text{tot}} p} \quad (64)$$

where $p = c_P/K_{m_R}$ is the normalized precursor concentration. Combining Equations 62 and 41 for $j = 1, 2$ and R together gives:

$$c_i = \frac{n_i}{n_R} c_R \quad (65)$$

where $i = 1$ or 2 .

Combining Equations 63, 62 (taken for $i = R$) and 41 (taken for $i = R$ and T) gives:

$$c_R = \frac{n_R}{n_T} c_T \left(1 + \rho_0 \frac{K_T}{\mu}\right) \quad (66)$$

Equation 48 for α gives (excluding the degenerate case $\alpha = 1$ for which $\mu = 0$):

$$\alpha = \frac{\frac{F_{\text{output},T}}{N_T}}{\frac{F_{\text{output},L}}{N_L} + \frac{F_{\text{output},T}}{N_T}} \quad (67)$$

or equivalently:

$$1 - \alpha = \frac{\frac{F_{\text{output},L}}{N_L}}{\frac{F_{\text{output},L}}{N_L} + \frac{F_{\text{output},T}}{N_T}} \quad (68)$$

We are now in a position to express all concentrations (except c_S) as a function of two unknowns, α and the normalized precursor concentration p . For simplicity, we shall keep μ in these expressions, knowing that it is a simple function of p given by Equation 64. Combining the above equation with Equation 46, gives:

$$\mu = \frac{F_{\text{output},L}}{N_L} \quad (69)$$

and:

$$\frac{\alpha}{1-\alpha}\mu = \frac{F_{\text{output},T}}{N_T} \quad (70)$$

From which c_L and c_T are extracted:

$$c_L = \mu \frac{N_L}{K_L} \quad (71)$$

$$c_T = \mu \frac{\alpha}{1-\alpha} \frac{N_T}{K_T} \quad (72)$$

Feeding Equation 72 in Equation 66, and feeding the result in Equation 65, gives:

$$c_i = \frac{\alpha}{1-\alpha} \frac{n_i}{n_T} N_T \left(\rho_0 + \frac{\mu}{K_T} \right) \quad (73)$$

where $i = 1, 2$ or R . Finally, we recall that, by definition, $c_P = p \times K_{m_R}$.

B.3 Relating α to p

Setting the time derivative of c_L (given by Equation 61) to zero, using Equation 40, and rearranging terms gives:

$$\frac{k_{\text{cat}_2} c_P c_2}{K_{m_2} + c_P} = K_L \left(\rho_0 + \frac{\mu}{K_L} \right) c_L \quad (74)$$

Defining $\theta = K_{m_R}/K_{m_2}$, using Equations 71 (for c_L), 73 (for c_2) and 64 (for μ), and rearranging terms gives:

$$\frac{1-\alpha}{\alpha} = \frac{\frac{k_{\text{cat}_2} \theta}{1+\theta p} n_2 N_T \rho_0 + \frac{\mu}{K_T}}{\frac{k_{\text{cat}_R} n_R}{1+n_{\text{tot}} p} n_T N_L \rho_0 + \frac{\mu}{K_L}} \quad (75)$$

Equation 75 defines a monotonic univocal relation between α and p , such that $0 < \alpha(p=0) = \alpha_0$, $\alpha(p=+\infty) = \alpha_\infty < 1$. α_0 may be lower or greater than α_∞ depending on the specific choice of parameters.

B.4 Determining two independent relations of the kind $c_S = \phi_{1,2}(p)$

Setting the time derivative of c_P (Equation 60) to zero, dividing by c_R and rearranging terms, gives:

$$\frac{k_{\text{cat}_1} c_S}{K_{m_1} + c_S} \frac{n_1}{n_T} = \frac{k_{\text{cat}_2} \theta p}{1+\theta p} \frac{n_2}{n_T} + \mu \frac{(n_1^2 + n_2^2 + n_R^2 + n_T^2) n_R}{n_T} + \mu \frac{p K_{m_R}}{N_T \left(\rho_0 + \frac{\mu}{K_T} \right)} \frac{1-\alpha}{\alpha} \quad (76)$$

Combined with Equation 75, this defines a function $c_S = \phi_1(p)$ over the interval $[0; p_{\text{max},1}[$, where $p_{\text{max},1}$ is the value for which the right-hand side (*RHS*) of Equation 76 verifies $\text{RHS}(p_{\text{max},1}) = k_{\text{cat}_1} n_1/n_T$. This function is such that $\phi_1(p=0) = 0$ (and this is the only zero of ϕ_1), and $\phi_1(p) \rightarrow +\infty$ if $p \rightarrow p_{\text{max},1}$.

Setting the sum of the time derivatives of c_S , c_P and c_L (sum of Equations 59, 60 and 61) to zero and rearranging terms gives:

$$\begin{aligned} \rho_0 \alpha \phi_{\text{input},S} = \mu c_S + \mu \frac{(n_1^2 + n_2^2 + n_R^2 + n_T^2)}{n_T} N_T \left(\rho_0 + \frac{\mu}{K_T} \right) \frac{\alpha}{1 - \alpha} \\ + \mu p K_{m_R} + \mu \left(\rho_0 + \frac{\mu}{K_L} \right) N_L \end{aligned} \quad (77)$$

Combined with Equation 75, this defines a function $c_S = \phi_2(p)$. Considering in turn the two different cases for nutrient influx:

1. Facilitated diffusion: $\phi_{\text{input},S}$ is the function of $c_{\text{out},S}$ and of c_S given by Equation 50. If $p = 0$, then the *RHS* of Equation 77 is null, and the *LHS* is also null iff $c_S = c_{\text{out},S}$. We thus have $\phi_2(p = 0) = c_{\text{out},S}$. If $p \rightarrow +\infty$, then *RHS* $\rightarrow +\infty$ while the *LHS* cannot exceed a maximum value $\rho_0 \alpha_\infty \mathcal{D} c_{\text{out},S}$ which is reached if $c_S = 0$. Therefore, there must exist $p_{\text{max},2}$ such that $\phi_2(p_{\text{max},2}) = 0$.
2. Active transport, $\phi_{\text{input},S}$ is only a function of $c_{\text{out},S}$ and is thus a constant with respect to the internal state variables. It can be verified that $\phi_2(p)$ is a decreasing function of p such that $\phi_2(p = 0) = +\infty$ and $\phi_2(p = p_{\text{max},2}) = 0$ where $p_{\text{max},2}$ is defined by letting $c_S = 0$ in Equation 77.

In any of the two above cases, we have $\phi_2(p = 0) > \phi_1(p = 0)$ and $\phi_2(p = p_{\text{max}}) < \phi_1(p = p_{\text{max}})$ where $p_{\text{max}} = \min\{p_{\text{max},1}, p_{\text{max},2}\}$. This implies that $\phi_1(p)$ and $\phi_2(p)$ thus intersect for a finite p_{st} which fully defines a stationary growth state.

In the second above case, it is to be expected that if active transport had been more rigorously modeled by taking into account the activator introduced in Section 7.2 (equivalent to an energy currency, which would be consumed or down-converted by the active transport process), a stationary growth state might only exist provided the rate of synthesis of this energy currency is sufficient. This would be a situation similar to that described by the simple example of Appendix A.