



**HAL**  
open science

## Closure to efficient causation, computability and artificial life

M.L. María Luz Cárdenas, Juan-Carlos Letelier, Claudio Gutiérrez, Athel Cornish-Bowden, Jorge Soto-Andrade

► **To cite this version:**

M.L. María Luz Cárdenas, Juan-Carlos Letelier, Claudio Gutiérrez, Athel Cornish-Bowden, Jorge Soto-Andrade. Closure to efficient causation, computability and artificial life. *Journal of Theoretical Biology*, 2010, 263 (1), pp.79. 10.1016/j.jtbi.2009.11.010 . hal-00564468

**HAL Id: hal-00564468**

**<https://hal.science/hal-00564468>**

Submitted on 9 Feb 2011

**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.

## Author's Accepted Manuscript

Closure to efficient causation, computability and artificial life

María Luz Cárdenas, Juan-Carlos Letelier, Claudio Gutiérrez, Athel Cornish-Bowden, Jorge Soto-Andrade

PII: S0022-5193(09)00536-0  
DOI: doi:10.1016/j.jtbi.2009.11.010  
Reference: YJTBI5775



[www.elsevier.com/locate/jtbi](http://www.elsevier.com/locate/jtbi)

To appear in: *Journal of Theoretical Biology*

Received date: 19 May 2009  
Revised date: 6 November 2009  
Accepted date: 15 November 2009

Cite this article as: María Luz Cárdenas, Juan-Carlos Letelier, Claudio Gutiérrez, Athel Cornish-Bowden and Jorge Soto-Andrade, Closure to efficient causation, computability and artificial life, *Journal of Theoretical Biology*, doi:[10.1016/j.jtbi.2009.11.010](https://doi.org/10.1016/j.jtbi.2009.11.010)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Closure to efficient causation, computability and artificial life

María Luz Cárdenas<sup>a\*</sup>, Juan-Carlos Letelier<sup>b</sup>, Claudio Gutiérrez<sup>c</sup>,  
Athel Cornish-Bowden<sup>a</sup> and Jorge Soto-Andrade<sup>d</sup>

<sup>a</sup>Unité de Bioénergétique et Ingénierie des Protéines, Centre National de la Recherche Scientifique, 31 chemin Joseph-Aiguier, 13402 Marseille Cedex 20, France; <sup>b</sup>Departamento de Biología, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile; <sup>c</sup>Departamento de Ciencias de la Computación, Universidad de Chile, Blanco Encalada 2120, Santiago, Chile; <sup>d</sup>Departamento de Matemáticas, Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile

## Abstract

The major insight in Robert Rosen's view of a living organism as an  $(M, R)$ -system was the realization that an organism must be "closed to efficient causation", which means that the catalysts needed for its operation must be generated internally. This aspect is not controversial, but there has been confusion and misunderstanding about the logic Rosen used to achieve this closure. In addition, his corollary that an organism is not a mechanism and cannot have simulable models has led to much argument, most of it mathematical in nature and difficult to appreciate. Here we examine some of the mathematical arguments and clarify the conditions for closure.

## 1 Introduction

Studies of artificial life depend heavily on efforts to set up and simulate models of living organisms in the computer. According to Robert Rosen (1991), however, a living organism is not a machine, and it so cannot have a computer-simulable model. Not surprisingly, his conclusion has stimulated an intense argument among computer scientists, mathematicians and biologists (Landauer and Bellman, 2002; McMullin, 2004; Wells, 2006; Chu and Ho, 2006; 2007a,b; Louie, 2007; Wolkenhauer, 2007; Wolkenhauer and Hofmeyr, 2007; Mossio et al., 2009), because if it is valid it imposes a formidable barrier to modern theories of computation in a topic that is as central to our scientific endeavour as it is to the nature of living systems. It is important to emphasize at the outset that Rosen did not argue that artificial life was impossible<sup>1</sup>, but only that

\*The authors are listed in a random order, and correspondence may be addressed to any of them.

<sup>1</sup>His most widely read book, *Life Itself* (Rosen, 1991), is subtitled *A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*.

organisms are “closed to efficient causation” and that this essential property excludes any possibility of simulable models.

The numerous new papers cited above that deal with the issue of computability make it necessary to examine the controversy. We are convinced that full understanding of Rosen’s work requires study of more than just the well known closure diagram in *Life Itself* (Fig. 10C.6 of Rosen (1991); Fig. 1a below). His argument against computability requires a detailed analysis of metabolic closure as set out in a series of papers spanning 15 years (Rosen, 1958a,b; 1959; 1966; 1971; 1973). Progress in the matter of computability requires a thorough knowledge of the conceptual steps that Rosen took in this regard (for example the distinction that he made between simulation and modelling), and not just the summary encapsulated by the diagram.

A recent analysis in terms of  $\lambda$ -calculus and the theory of computer programming (Mossio et al., 2009) led to the opposite conclusion, that a system closed to efficient causation can certainly have computable models. The authors pointed out the apparent contradiction that autopoiesis (Maturana and Varela, 1980), which has strong underlying similarities with Rosen’s theory (Letelier et al., 2003), including closure to efficient causation, is claimed to have computable models (McMullin, 2004). Moreover, it is not obvious that the examples of  $(M, R)$ -systems that we have proposed (Letelier et al., 2005a; 2006; Cornish-Bowden et al., 2007; Cornish-Bowden and Cárdenas, 2007) cannot be simulated. Other criticisms of Rosen’s analysis also deserve to be answered, as they affect his insights in relation to closure, which we regard as essential for understanding living systems.

How does the existence, or not, of computable models of living organisms affect other aspects of the study of life? According to Chemero and Turvey (2008), “little rides on whether genuine artificial life is possible”, and we agree: the existence of simulable models of living organisms has less importance than Rosen’s essential insight, that organisms must be closed to efficient cause and hence metabolically closed. However, the argument about simulability will certainly continue: the work of many groups, including those attempting to develop life *in silico*, depends on the assumption that computer simulation of living systems is in principle possible, and any claims that it is not possible can expect to meet vehement opposition.

We believe that the way forward, both for getting a better understanding of life and for deciding whether it can have simulable models, will require Rosen’s abstract and mathematical ideas to be brought into much closer correspondence with biological reality. Future models need not only to reflect the mathematics accurately but they must also be biochemically reasonable.

## 2 Closure to efficient cause

Closure to efficient cause, illustrated in Fig. 1, is central to Rosen’s view of life, and we shall briefly resume what it means. Rosen drew the diagram, for example in Fig. 10C.6 of Rosen (1991), as in Fig. 1a. Cottam et al. (2007), have argued that the underlying logic is symmetrical and can be better illustrated with a symmetrical figure-of-eight layout, as in Fig. 1b. Their arrangement is visually appealing but it suggests a misleading parallelism between efficient and material causation, and in any case, as we note below, material causation *must* be asymmetrical if thermodynamic requirements

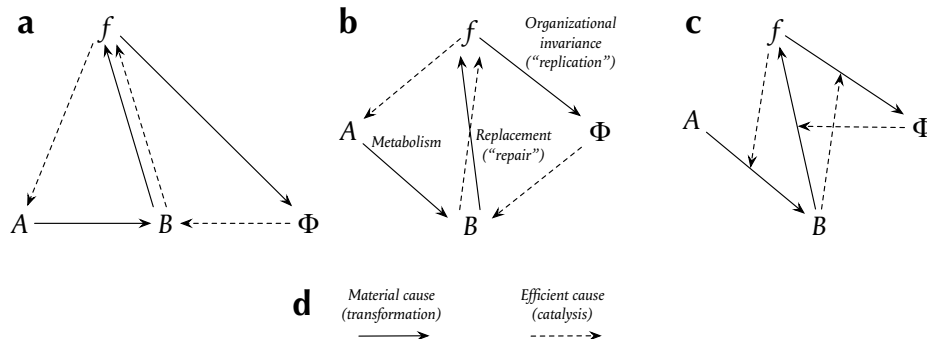


Fig. 1. The diagram is based on Fig. 10C.6 of Rosen (1991), shown (a) in the unsymmetrical way used by Rosen, (b) in the more symmetrical way suggested by Cottam et al. (2007), and (c) in a way suggested by Goudsmit (2007) in which the representation of efficient causes resembles normal practice in chemistry, with catalysts shown as acting on reactions rather than on substrates. (d) In all three variants full arrows represent material causation, or chemical transformation, whereas broken arrows show efficient causation, or catalysis. Rosen's terms "repair" and "replication" shown in parentheses for these last two processes are misleading, as they have *nothing* to do with the ordinary uses of these words in modern biochemistry, for example for DNA repair and DNA replication. Here we follow the terminology introduced previously (Letelier et al., 2006).

are to be satisfied. In the third layout (Fig. 1c), suggested by Goudsmit (2007), the distinction between efficient and material causation is represented in a way that is much closer to normal practice in chemistry and biochemistry. The three layouts do not imply three different schemes, but one scheme shown in three different but completely equivalent ways.

Each version of Fig. 1 consists of three parts, each with a set of chemical transformations (material causation) and a set of catalysts (efficient causation). Metabolism is the complete set of chemical transformations  $A \rightarrow B$ , catalysed by a set of enzymes  $f$ . Replacement (or "repair" in Rosen's misleading terminology) is the resynthesis of the set of catalysts  $f$  as necessitated by degradation, wear and tear, and growth, catalysed by a replacement system  $\Phi$ . Organizational invariance ("replication" in Rosen's terminology) is the process that enables an organism to maintain this replacement system, as we discussed in more detail elsewhere (Letelier et al., 2006). Notice that it does not correspond to cell reproduction or DNA replication.<sup>2</sup>

It is obvious that enzymes must be synthesized from the products of metabolism and that this synthesis requires additional catalysts, so the arrows  $\Phi \dashrightarrow B \rightarrow f$  require no explanation, but it is less obvious why  $B$  should be the efficient cause of the replacement enzymes. In fact the efficient cause is not  $B$  but a function  $\beta$  that is related to  $B$  but is not *the same as*  $B$ . It might seem *a priori* that a separate set of

<sup>2</sup>This unfortunate terminology has misled some authors. For example, Mossio et al. (2009) wrote that " $\Phi$ , as 'replication function', may be associated with nucleic acids." There are actually two errors here, the first possibly typographical but the other probably caused by Rosen's terminology:  $\Phi$  is responsible for the "repair" function, not "replication", and "replication" in Rosen's sense is quite different from DNA replication, etc. Throughout the present paper we follow the terminology we introduced previously (Letelier et al., 2006), which we believe to be less confusing.



Fig. 2. Different criteria of closure. As there are various ways in which a system can be considered to be closed, it is important to distinguish between these. (a) Organisms are open systems in the thermodynamic sense, and they obtain the energy they need for all metabolic and motor functions by irreversibly converting nutrients into excreta; in the case of the blood-stream form of the parasite *Trypanosoma brucei* illustrated this is a matter of converting glucose into pyruvate. (b) On the other hand they are structurally closed, in the sense that different individuals, for example in the bacterial population illustrated, are separated from one another by physical barriers, and there is no doubt about where one ends and another begins. (c) Closure to efficient causation does not refer to structural closure, however, but to organizational closure: all catalysts needed for metabolism are themselves products of metabolism.

enzymes would be needed, but this would imply infinite regress, whereas organisms are not infinite. Rosen's device for avoiding this infinite regress was to interpret catalysis of the transformation  $f \rightarrow \Phi$  as a property  $\beta$  of the metabolic products  $B$ . We have discussed the implications of this elsewhere (Letelier et al., 2006), but it is important to emphasize at the outset that although  $\beta$  can be regarded as a *property* of  $B$  it is not the *same* as  $B$ ; this is a subtle point that continues to be widely misunderstood, and we return to the topic in Section 8.

In Fig. 1 all catalysts are synthesized internally; none is produced by any external agency. It is in this sense that the system is catalytically closed, or closed to efficient cause in Rosen's terms. There is no implication of material closure, however, and no conflict with the thermodynamic requirement that a living organism be open to the flux of matter, to allow metabolic energy to be extracted from food (Fig. 2a):  $A$  in Fig. 1 is understood to include molecules available from the environment, and likewise  $B$  to include molecules that are excreted, but the distinction between external and internal molecules is not explicit in the figure. In a third sense an organism is again closed, as there is always a physical separation (membrane, cell wall, skin, etc.) between one individual and another (Fig. 2b). This is important, and is emphasized in particular in the theory of autopoiesis (Maturana and Varela, 1980), but Fig. 1 is concerned with organizational closure (Fig. 2c), not with structural closure.

### 3 Analysis of closure in terms of hypersets

Hypersets are generalized sets in which the restriction that sets cannot be members of themselves is relaxed. This restriction was made at the beginning of the 20th century as a way of resolving Russell's paradox and the problems of ambiguity that arise when impredicative definitions are permitted, i.e. definitions that allow the entity being defined to participate in its own definition. However, the impredicativity that

is central to Rosen's view of an organism does not prevent it from being logically coherent (Kercel, 2007).

Chemero and Turvey (2006; 2008) have discussed the principle and usefulness of hypersets, and have shown how simple graphs of hypersets allow one to recognize if a system embodies circular definitions and is thus complex in Rosen's sense. Mossio et al. (2009) are critical of some of these authors' arguments — both in relation to their understanding of impredicativity and in the implications that impredicativity has for computability — but they do not address other serious problems with these papers, especially the misunderstanding of what catalysis is.

On the basis of their approach Chemero and Turvey (2006; 2008) conclude that catalytic closure does not necessarily mean closure to efficient cause. As this conclusion contradicts Rosen, it is important to understand that they arrived at it by means of an idiosyncratic definition of catalytic closure: "a system is catalytically closed just in case every product of the system is also a catalyst in the system" (Chemero and Turvey, 2006). They attributed this definition to Kauffman (1993), but in fact it inverts Kauffman's own definition: "Catalytic 'closure' must be achieved and maintained. That is, it must be the case that every member of the autocatalytic set has at least one of the possible last steps<sup>3</sup> in its formation catalyzed by some member of the set, and that connected sequences of catalyzed reactions lead from the maintained food set to all members of the autocatalytic set."

According to Kauffman, therefore, catalytic closure requires every catalyst to be a product of metabolism, whereas according to Chemero and Turvey it requires every product of metabolism to be a catalyst. Kauffman's definition agrees with related ideas in  $(M,R)$ -systems (Rosen, 1991) and autopoiesis (Maturana and Varela, 1980). As far as we are aware no one has previously suggested that it is a requirement for life that every metabolic product be a catalyst, and examination of real metabolism in real organisms reveals innumerable examples of products with no known functions as catalysts.

In their more recent paper Chemero and Turvey (2008) define catalytic closure in a more acceptable way, as we discuss below, so we need to examine whether the problems in the previous one (Chemero and Turvey, 2006) arose from an unfortunate choice of words, or whether they reflect a real failure of their analysis. This can be determined by ignoring their words and examining the series of reactions they used to illustrate them in their 2006 paper:



Their designation of A, B and C as catalysts in this system makes no sense if the pair of symbols on the left-hand side of each process is interpreted as a pair of co-reactants (see Fig. 3a). Instead,  $P + Q \rightarrow A$ , for example, apparently means a reaction  $P \rightarrow A$  catalysed by Q. If so, a more conventional and intelligible way of symbolizing the whole set of reactions is as follows:



<sup>3</sup>This unfortunate choice of wording could be taken to refer one of the steps near the end of a series of steps, in which case one might wonder why only one of these steps needs to be catalysed. In fact it refers to the possibility of parallel pathways to the same metabolite, and it requires that the final step of at least one of these must be catalysed.

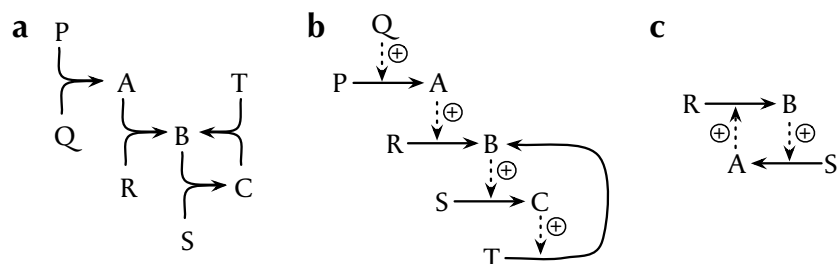


Fig. 3. Models of Chemero and Turvey. (a) A natural interpretation (but probably not the one intended) of the equations used by Chemero and Turvey (2006) to define their model. Note that the model contains no catalysis, and it violates the law of conservation of matter, because the  $B \rightleftharpoons C$  cycle is a bottomless pit into which all the reactants disappear, with nothing coming out. (b) What we take to be the intended meaning of the equations of Chemero and Turvey (2006). Full and broken arrows have the same meanings as in Fig. 1, and the plus signs simply emphasize that each broken arrow represents a catalytic effect, not an inhibition. (c) The more recent model of Chemero and Turvey (2008).

On this interpretation (Fig. 3b) their system satisfies their definition of catalytic closure (every product is a catalyst), but not that of Kauffman (1993), because Q is a catalyst but not a product of any reaction in the system. The fact that Q is not a system product is essential to their argument, as it leads them to an artificial distinction between catalytic closure and closure to efficient cause, and to say that “in general, catalytically closed systems are not closed to efficient causation.” Their system, however, does not satisfy any recognized criterion of catalytic closure (Rosen, 1991; Maturana and Varela, 1980; Kauffman, 1993), so it provides no support for their contention that their analysis “agrees with Chu and Ho (2006), who dispute Rosen’s purported proof that artificial life is impossible.” In any case, Rosen did not claim to have provided a “proof that artificial life is impossible”; on the contrary, as we noted in the Introduction, he made an explicit distinction between the existence of simulable models of life and the possibility of artificial life (Rosen, 1973).

The later paper (Chemero and Turvey, 2008) does not resolve the problems raised by the first. Much of the text and illustrations are identical, but the part that is of most concern to us here is different, albeit in a way that does not clarify their argument. They no longer state that every product of metabolism must be a catalyst, but instead that “when every product in a network of chemical reactions has a catalyst that is also a product in the network, the network of reactions is a collectively autocatalytic system”: this is closer to the definition given by Kauffman (1993). One may object that catalysts are usually associated with reactions rather than with products, but this is trivial compared with the objections that can be made to the previous statements. The new discussion refers to a scheme that fails to make it clear which version of the generalization is being illustrated, because now both catalysts are products, and both products are catalysts:



As may be seen by drawing it as in Fig. 3c, this model does represent a collectively



autocatalytic system, even though neither of the individual steps is autocatalytic.

Unfortunately the recent paper (Chemero and Turvey, 2008) makes no reference to the earlier one, and so it is not clear whether the changes are corrections or just different ways of expressing the same ideas. As it reaches essentially the same conclusion, expressed in the same words, that the analysis “agrees with Chu and Ho (2006), who dispute Rosen’s purported proof that artificial life is impossible,” it seems best regarded as a re-expression of the original argument and not as a corrected one.

Chemero and Turvey (2007) have a related paper that appeared between the two we have discussed. In it they define the following system of reactions:



and comment that “the second and third steps are autocatalytic. X reacts with B to produce Y and Y is a catalyst for production of more X”. Again, however, the meanings of the symbols and words are obscure and in addition it is difficult to make chemical sense of the individual reactions (especially  $2X + Y \rightarrow 3X$ ). As far as one can tell they have simply written a series of four equations without paying any attention to chemical plausibility. On the one hand the + sign does seem to be used in the conventional way to concatenate co-reactants; on the other hand the statement that Y is a catalyst is true only if the second and third steps are regarded as a single process, but not if they are treated separately, as the words seem to suggest. Unfortunately, therefore, this paper is only of limited help for clarifying the other two. There is, however, an important point illustrated by the idea of two or more reactions in a sequence as a single process. Once there is a reactant for one step regenerated by another step there is necessarily autocatalysis, as discussed by King (1977a) and as implied by Cornish-Bowden and Cárdenas (2007).

#### 4 Autocatalytic sets and autopoiesis

There is a fundamental difference between autocatalytic sets (Kauffman, 1993; 1986) and autopoiesis (Maturana and Varela, 1980), and it is necessary to understand this because of the implications of these approaches to the origin of life. For Kauffman an autocatalytic set is inevitably a *large* set, with, as a minimum, thousands of elements based on amino acids or RNA bases, because only large systems can have the statistical properties needed for closure to become virtually inevitable. For example, if any one molecule has a probability of  $10^{-9}$  of catalysing a particular ligation or cleavage step in the system,<sup>4</sup> the probability of catalytic closure is very low unless there are at least  $3 \times 10^8$  different kinds of molecules altogether. This number is greatly decreased, to about 18000, if there is the same probability of catalysing exchange reactions as well as ligation and cleavage, but it remains very large in comparison with the numbers of elements usually imagined to be necessary for a minimal autopoietic system: although Maturana and Varela may not have explicitly described their systems as small, most of their discussion implies that autopoietic systems can be small enough to be represented by models that are very small in Kauffman’s terms. They are thus much

<sup>4</sup>Orgel (2008) recently discussed the plausibility of this and other assumptions in Kauffman’s model. His analysis is interesting and important, but it does not bear directly on the point we are making here.

closer in spirit than Kauffman to most theories of the origin of life, and, indeed to current ideas on the number of distinct entities that would be needed for a minimal cell.

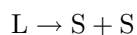
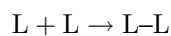
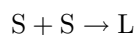
Although Kauffman (1993) concluded that autocatalytic sets had to be very large sets, it is important to realize that this conclusion followed from his wish to assess the likelihood that closure could arise out of purely chance properties of its component elements. It does not follow necessarily from the definition of an autocatalytic set, however, if one is not concerned with the statistical arguments, and, as we show later (Section 9), a very small set can satisfy the definition.

Where do  $(M,R)$ -systems (Rosen, 1991) fit into this spectrum? Rosen's discussion was always expressed in terms so abstract that there is no particular implication about how many elements an  $(M,R)$ -system needs to contain in order to be viable. In our earlier studies we followed Morán et al. (1996) in implicitly assuming that an  $(M,R)$ -system could be small enough to be represented by a model intelligible in chemical terms, but that may not be what Rosen intended. However, our model was designed to fix some of the ideas, and not to be a realistic model of an organism: in other words, just as each of Rosen's arrows represents a large number of parallel processes each of the arrows in our model should be taken as an oversimplified image of a complicated reality. In particular, proper energy management in an organism needs much more than an overall process that is assumed to be irreversible.

It is important to notice that although Kauffman's systems may be very large in terms of the number of different molecules they contain, they are rather *small* in terms of the number of different chemical types of molecule they contain. It is virtually certain that an organism cannot consist of just a mixture of peptides, or a mixture of nucleic acids.

## 5 Are autopoietic systems really computable?

As Letelier et al. (2003) pointed out, autopoietic systems share many features of  $(M,R)$ -systems and can be regarded as a subset of  $(M,R)$ -systems. They are thus by implication closed to efficient causation, and should inherit the property of not having computable models<sup>5</sup>. However, it has been claimed that autopoietic systems have been modelled (McMullin, 2004),<sup>6</sup> based on a simple simulation carried out by Varela et al. (1974) and later reconstructed in a more modern computational context by McMullin and Varela (1997). This model involves the following processes:



in which  $S$  is a molecule available from the environment and  $L-L \dots$  represents a linear chain of any length that can turn back on itself to form a closed cycle. The first

<sup>5</sup>The important distinction between simulation and modelling is discussed in section 7.

<sup>6</sup>These claims appear to be accepted by Mossio et al. (2009) in their recent paper.

reaction is assumed to occur only in a specific part of the space (inside the boundary formed by a cyclic chain of L molecules, a two-dimensional representation of what in a three-dimensional system would be a membrane). This inevitably means that it is a *catalysed* reaction, because if it were spontaneous it would occur everywhere, not just inside the boundary. The simplest expression of the model did not give the expected results, because newly formed L–L... chains tended to clog the interior of the system instead of integrating themselves into the boundary molecule. This was solved in an *ad hoc* manner by introducing an unexplained inhibition to prevent premature bonding between L molecules.

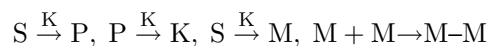
There are two points that need to be distinguished here: first whether the simulated systems were real autopoietic systems closed to efficient causation; second whether they were models in Rosen's sense, or simply simulations. So far as the first point is concerned, they are valuable as illustrations of how cell membranes can in principle result from sequential application of simple rules, but they were not simulations of autopoietic systems, because the catalyst needed for the essential chemical reaction was not generated by the system itself but was given from outside. McMullin (2004) recognized this, saying that "this appears to violate the demand for closure in the processes of production". However, it does not merely "appear" to violate this fundamental requirement; it does violate it. As the simulated system was not closed to efficient causation it tells us nothing about whether systems that are closed to efficient causation can have simulable models.

In his attempt to justify allowing exceptions to the requirement of closure, McMullin (2004) commented further that "there must surely be *some* exceptions allowed (specifically, covering the case of the S particles that are simply harvested from the environment)", in which the S particles correspond to the external molecules included in our simple model of an  $(M, R)$ -system (see Fig. 6 below). There is a confusion here between closure to material causation (which no one claims to be a property needed for life) and closure to efficient causation, an explicit property of  $(M, R)$ -systems, and at least an implicit one in autopoiesis.<sup>7</sup> However, it is a confusion that is easy to fall into, and deserves some explanation. It has been generally accepted since Schrödinger (1944) characterized living organisms as "feeding on negative entropy" that they are inevitably *open* systems, that maintain themselves indefinitely in states far from thermodynamic equilibrium by ingesting low-entropy food and excreting high-entropy products. This thermodynamic requirement does not in any way conflict with the organizational requirement that they be *closed* systems, because openness to material causation is not opposed to closure to efficient causation (see Fig. 2).

Referring to the failure of the autopoietic model to include regeneration of the catalyst, McMullin (2004) goes on to say that "more recent elaborations of this original model have specifically allowed for production of the [catalyst] particles (Breyer et al., 1998), so there is no *fundamental* difficulty here" (emphasis in the original). However, later in the paper he admits that "full achievement of self-reproduction is not reported" [in the simulations cited], so it is not clear how the argument is advanced. Although Breyer et al. (1998) did define a model with regeneration of the catalyst, which may

<sup>7</sup>In two recent papers Fernando and Rowe (2007; 2008) discuss both organizational closure and thermodynamic considerations. Although they do not confuse the two concepts, they also do not distinguish between them as explicitly as necessary to maintain them distinct.

be represented as follows:



the simulation results that they present do not appear to have been obtained with this model, but with a simpler one presented earlier in which there is no catalyst regeneration. Nonetheless, we do not doubt that with further elaboration of the computer code a genuine simulation of an autopoietic system could be achieved, and in this limited sense we do not contest McMullin's conclusion; however, it would still not constitute a model in Rosen's sense.

The model of Breyer et al. (1998) has a peculiarity that makes it difficult to regard it as a plausible biochemical model. As one may see, the catalyst  $K$  acts on a single substrate  $S$  to produce two different products ( $S \xrightarrow{K} P, S \xrightarrow{K} M$ ). This type of behaviour is common in catabolic processes, in which, for example, a proteolytic enzyme may act at various different sites in a protein substrate, releasing a different product in each case. However, we are concerned here with anabolic metabolism, for which we know of no such examples. The model gives the impression that the catalyst was included as an afterthought with no attempt to take biochemical reality into account.

Finally, it is worth noting that Zeleny (1995) has also denied that autopoiesis can be modelled in the sort of way that McMullin and others have attempted, writing as follows:

Approaches which sacrifice [the] essential individuality of components, like the statistical systems of differential equations used in the traditional systems sciences, cannot model autopoiesis. They are definitionally incapable of treating autopoietic systems as social systems.

## 6 $(M, R)$ -systems considered in terms of Cartesian closed categories

In developing their abstract cell model of a living organism, Wolkenhauer and Hofmeyr (2007) state, but do not prove, that the category needed for a mathematical model of a self-organizing cell must be Cartesian closed. In essence this means that the category behaves like the category of sets and mappings regarding the relationship between functions of two variables and functions of one variable. It is well known, indeed, that the graph of a function of two variables, such as  $f(x, y)$ , can be regarded in a global way as a "sheet" hovering over the  $(x, y)$  plane, or alternatively, as a family of slices, for example slices parallel to the  $y$  axis. In algebraic terms, a two-variable function

$$(x, y) \mapsto f(x, y)$$

may be described as a one-variable function:

$$x \mapsto f_x : y \mapsto f(x, y)$$

where  $f_x(y) = f(x, y)$ , whose values  $f_x$  are again one-variable functions, obtained by fixing the value  $x$  of the first variable as a parameter. The graph of each partial

function  $f_x$  appears then as a vertical “slice” of the graph of the two-variable function  $f$ , parallel to the  $y$  axis.

There are several similar statements in the paper about the necessity of Cartesian closure, but no proof, and in fact, it is not true: Cartesian closure may indeed be sufficient to allow formulation of a mathematical model of self-organization, as Wolkenhauer and Hofmeyr show by means of a rewording of Rosen’s own argument, but it is not necessary, and the arithmetical example that we described previously (Letelier et al., 2006) contradicts the claim that it is, as we discuss at the end of this section.

In the case relevant here, we are led to consider *concrete categories* whose objects are sets *endowed with a structure* and whose morphisms are simply “structure-preserving mappings”. For example, the objects could be vector spaces with linear maps as morphisms; or they could be groups, with group homomorphisms as morphisms; or they could be ordered sets with monotone mappings, and so on. For this sort of category to be Cartesian closed, it is necessary to know that the set  $A^B$  of morphisms from any object  $B$  to any object  $A$  may be endowed (in a “natural” way) with the same sort of structure as the other objects in the category and that all maps involved in the slicing procedure above are morphisms in the category. More precisely, if  $f : X \times Y \rightarrow Z$  is a morphism, then the maps  $f_x : Y \rightarrow Z$ , for all  $x \in X$  and the map  $x \mapsto f_x$ , from  $X$  to  $Z^Y$ , are morphisms in the category, and also vice versa.

The problem with insistence on Cartesian closure is that it inevitably excludes functions that ought to be acceptable. For example, let  $f$  be the familiar dot product  $(x, y) \mapsto x \cdot y$  in which  $x, y$  may be real numbers, or vectors in the plane or in 3-dimensional space: the partial mappings  $f_x$  are indeed linear<sup>8</sup>, and so is the mapping  $x \mapsto f_x$ , because of the distributive law for the product with respect to addition of numbers. So the slicing maps

$$x \mapsto f_x$$

with  $f_x(y) = f(x, y)$ , are linear, even though  $f : (x, y) \mapsto x \cdot y$  is not linear but bilinear, i.e.  $f(x, y)$  is linear in  $y$  if  $x$  is fixed, and is also linear in  $x$  if  $y$  is fixed. Here we notice that the “natural” — and useful — property for the global function  $f$  of two variables is bilinearity, not linearity. Similar consideration apply to abelian groups, the category of our arithmetical example. Even though the category is not Cartesian closed, the map  $(m, n) \mapsto mn$ , where  $m$  and  $n$  are integers modulo 12, or “hours”, is bi-additive, and  $f$ ,  $\Phi$  and  $\beta$  (in Rosen’s symbols) are all additive functions, i.e. they are morphisms of the category of abelian groups. It is in this sense that our simple mathematical example (Letelier et al., 2006) is a counter-example to the assertion by Wolkenhauer and Hofmeyr (2007) that the associated category of a mathematical model of a self-organizing cell must be Cartesian closed.

---

<sup>8</sup>We use “linear” here in the mathematical sense: a function  $g$  is linear if  $g(u + v) = g(u) + g(v)$ . In physics  $g : u \mapsto 5 + u$  for  $u \in V = \mathbb{R}$  would be regarded as linear, though in mathematics it is called an affine function.

## 7 Simulation and modelling

Simulating an organism and creating a model of an organism may appear to be the same thing, and so it is important for discussing Rosen's work to emphasize that he attached clearly different meanings to these two ideas. There are two key issues in the notion of simulation, the first of which is the possibility of developing by means of a computer program a sequence of steps that behave identically to the phenomenon to be simulated. In general this simulation could not give any information about the structure of the phenomenon<sup>9</sup>. In computer science one would say that the phenomenon of interest is simulable by a Turing machine (the most general form of computing today). However, there is a second issue that is just as important, which is to be able to obtain information about the structure of the phenomenon from the simulation. This amounts to deciding which variables and parameters of the phenomenon to keep and record, how much time the simulation will run, etc. These variables and parameters can then give us the structural information we seek. For Rosen the first of these is "simulation" whereas the second is "modelling", and he considers that for understanding the structure of phenomena the interesting things are models. His attitude is, of course, that of a scientist rather than of an engineer,<sup>10</sup> as he regards understanding things as more important than predicting how they will behave. The notion of closure under efficient causation and his famous diagram (see Fig. 1 below) are attempts to present a model of life in this sense, not just a simulation of life.

The distinction between modelling and simulation can be clearly explained by a diagram (Fig. 4). The boxes  $F_1$  and  $F_2$  represent (formal) systems, and the arrows  $\phi$  and  $\mu$  indicate the notion of "entailment" in each of these systems. Then, the system  $F_2$  *simulates* the system  $F_1$  if there are encodings  $\alpha$  (intuitively the notion of measurement or encoding) and  $\beta$  (intuitively prediction or decoding) and an "inferential machinery"  $\mu$  such that the diagram commutes, that is,

$$\phi = \beta \circ \mu \circ \alpha. \quad (1)$$

Note that if the arrows  $\alpha$  and  $\beta$  are computable and the entailment relation  $\mu$  in  $F_2$  is computable, the predictions are computable, by combining their respective results. This is exactly what happens with the idea of simulation in the sciences today.

However, the equation does not tell us much about the structure of  $\mu$ . What is this structure, and how does it relate to  $\phi$ ? This is not at all the concern of simulation, because without any knowledge of this we can "predict" certain phenomena of  $F_1$  using it. A deeper level of understanding of the phenomena  $F_1$  occurs when we

<sup>9</sup>As a simple and familiar example, methods of numerical analysis such as orthogonal polynomials are used in engineering applications to simulate with as much precision as desired the behaviour of a mathematical function that has an unknown analytical form but for which numerical values are known at many points. These methods are very valuable for making predictions about the numerical values of the function at additional points, but they tell us nothing about the real functional relationship. In Rosen's terminology, therefore, orthogonal polynomials can simulate a phenomenon as precisely as desired, but they do not model it.

<sup>10</sup>The distinction may be further illustrated by recalling the controversy two decades ago over whether the kinetics of a multienzyme system were best represented in terms of equations that offered insight into the real nature of the system (Kacser and Burns, 1973) or by ones that were effective for predicting its behaviour under a range of conditions (Savageau, 1976). See Cornish-Bowden (1989).

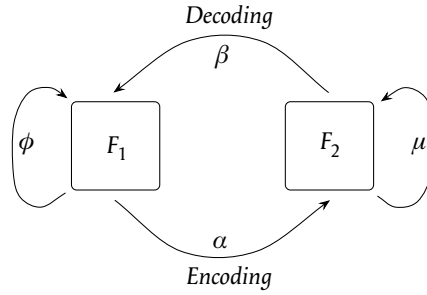


Fig. 4. Simulation. The diagram illustrates the relationship between two formalisms  $F_1$  and  $F_2$ , and is based on Fig. 3F.2 of Rosen (1991) and diagram (1) in section 4.9 of Louie (2009).

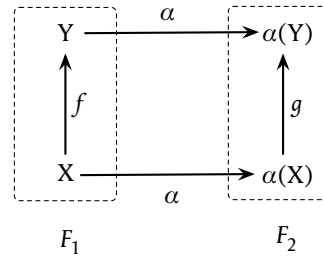


Fig. 5. Modelling. The diagram in Fig. 4 is shown for a deeper level of understanding. The system of  $F_2$  is a model (and not just a simulation) of that of  $F_1$  if the entailment structure  $g$  in  $F_2$  is related to the entailment structure  $f$  in  $F_1$  by a known function, so  $g = \alpha(f)$ . The figure is based on diagram (3) in section 4.9 of Louie (2009).

can “deduce” in some sense the structure of the arrow  $\mu$  from the arrow  $\phi$  and the encodings. This is what Rosen (1991) calls a *model*, and Louie (2009) represents it as in Fig. 5. In the notations of that diagram, a *model* has the more stringent condition that the entailment notion in the formalism  $F_2$  is represented as a function (the key point being that the structure of that function is *known*). Louie represents that relation by the equation  $g = \alpha(f)$ .

The question is now whether Rosen’s model can be simulated by a computer. Note that the computability of the function  $\phi$  does not imply that the relation  $\mu$  is computable, that is that  $\alpha$  applied over processes ( $\alpha(f)$ ) is computable. Even if we assume that there is a mathematical expression of his ideas that is simulable by a computer, this says nothing about whether we can get information about the structure of the phenomenon simulated, for example if it stabilizes, if it terminates, etc. This is what Rosen claims when stating that no model of closure to efficient causation could be Turing-simulable.

There are several formalisms for specifying mathematical models, such as logic, category theory and so on, and different authors and research groups have chosen different models for formalizing Rosen’s ideas. He himself sometimes used frameworks similar to algebraic notations and sometimes category theory, but when discussing computability he did not chose any particular framework. There are various equiv-

alent formalisms for this, including Turing machines, recursive functions, universal grammars,  $\lambda$ -calculus, rewrite systems and so on. All of them model the same class of functions, namely *recursive functions*.

The question therefore arises naturally of which one is best suited for modelling Rosen's ideas on computability. Although he expressed his ideas in this area by directly addressing the language of recursion theory, or of "sequential machines", the use of functions and the notion of efficient causation brings to mind the ideas behind  $\lambda$ -calculus, namely the formalization of the notion of function and evaluation<sup>11</sup>. This approach was suggested by Fontana and Buss (1996) a decade ago, and is actively pursued in a recent paper of Mossio et al. (2009). The use of rewrite rules is another attractive approach, which has been formalized in the area of artificial chemistry by, for example, Dittrich et al. (2001).

It is important to keep in mind that all these formalisms are equivalent, in the sense that there are translations among them, and all represent the same mathematical object. That is why we think it is better to keep the discussion in the biological area at the abstract level of computability, using the formalisms that better fit the particular problems to be addressed.

The important question here is to define what is meant by the terms *computable* and *not computable* in the context of Rosen's ideas. He himself never used these terms in his book *Life Itself* (Rosen, 1991), referring instead to machines, mechanisms and the notion of *simulation* (by a machine). On p. 192 of his book he explains this important idea as follows:

Thus, the word "simulable" becomes synonymous with "evaluable by a Turing machine". In the picturesque language of Turing machines, this means the following: if  $f$  is simulable, then there is a Turing machine  $T$  such that, for any word  $w$  in the domain of  $f$ , suitably inscribed on an input tape to  $T$ , and for a suitably chosen initial state of  $T$ , the machine will halt after a finite number of steps, with  $f(w)$  on its output tape.

Notice in particular Rosen's words "the machine will halt after a finite number of steps", as this is the key to some of the argument about whether his conclusions are correct. Obviously, if one drops the condition of halting after a finite number of steps from the definition of computability then one may arrive at a different conclusion from his.

## 8 Circularity and computer programming

### 8.1 Rosen's analysis

The conclusion of Mossio et al. (2009) that  $(M, R)$ -systems can have computable models is based on an analysis of the fundamental equations of  $(M, R)$ -systems in terms of the theory of computer programming, specifically in terms of  $\lambda$ -calculus. Their analysis omits an essential part of the argument, however, and arrives in consequence at a result that we contest. As we discussed previously (Letelier et al., 2006) the summary

<sup>11</sup>Some authors use the word "application" for what we are calling evaluation, a term we find more appropriate.



of Rosen's system shown in Fig. 1 can be expressed in mathematically much more rigorous terms by a series of mappings,

$$A \xrightarrow{f} B \xrightarrow{\Phi} \text{Map}(A, B) \xrightarrow{\beta} \text{Map}(B, \text{Map}(A, B))$$

and three equations, the first of which expresses the idea that all of metabolism has an operator  $f$  as its efficient cause:

$$a \xrightarrow{f} b, \quad \text{i.e. } f(a) = b$$

In biological terms  $f$  can be regarded as the set of enzymes needed to catalyse the reactions. The second equation shows that the efficient cause of the replacement of enzymes necessitated by degradation, wear and tear, and growth is another operator  $\Phi$ :

$$b \xrightarrow{\Phi} f, \quad \text{i.e. } \Phi(b) = f$$

with the products of metabolism as material cause. To avoid infinite regress, maintenance of  $\Phi$  needs to be achieved *without introducing another layer of causation*, and Rosen's suggestion was that the operator  $\beta$  needed for maintaining  $\Phi$  could be obtained by solving the equation  $\Phi(b) = f$  for  $\Phi$ . Indeed, if we assume that for each possible  $f$  there is only one  $\Phi$  that satisfies this equation, the operator  $\beta$  is then defined simply by the condition

$$f \xrightarrow{\beta} \Phi, \quad \text{i.e. } \beta(f) = \Phi$$

In more abstract mathematical terms we may look upon the value  $\Phi(b)$  of  $\Phi$  at  $b$  as the result of applying the operator  $ev_b$ , or "evaluation at  $b$ ", to the function  $\Phi$ :

$$\Phi(b) = ev_b(\Phi)$$

In these terms we have  $ev_b(\Phi) = f$ , and since the operator  $ev_b$  is by hypothesis invertible, we get

$$\Phi = (ev_b)^{-1}(f)$$

So, finally our operator  $\beta$  is none other than the inverse  $(ev_b)^{-1}$  of the "evaluation at  $b$ " operator  $ev_b$ .

The notion that  $b$  can be identified with  $\beta$  needs to be treated with care. As we have just explained, invertibility of the evaluation operator (a very demanding condition in normal situations) implies that knowledge of  $b$  should allow  $\beta$  to be obtained by the following series of steps: first,  $b$  is used to define a function  $ev_b$ , and if this evaluation at  $b$  is invertible we have that  $ev_b$  can be used to calculate  $(ev_b)^{-1}$ , a far from trivial task in general. However, if the conditions are fulfilled there is a bijection (one-to-one correspondence) between  $b$  and  $(ev_b)^{-1} = \beta$  and we can write  $\beta = P(b)$ , in which  $P$  is a complex operator with the property that

$$P(b) = P(\hat{b}) \text{ implies that } b = \hat{b}$$

It is therefore important to realize that "identification" does not signify "equality" in this context. That is why we (Letelier et al., 2006) twice referred to identification but not to equality. Nonetheless, it is unfortunate that we used the words "with  $\beta$  equivalent to  $b$ " to mean "with a one-to-one relationship between  $\beta$  and  $b$ " after equation 14 in that paper, as this was to invite exactly the sort of misunderstanding that we were seeking to avoid. Perhaps the clearest statement would be that  $b$  can be identified, via a one-to-one mapping, with  $\beta$ .

## 8.2 Expression in terms of $\lambda$ -calculus

After correctly noting that  $\Phi = \beta(f)$ , Mossio et al. (2009) continue by stating that “Rosen makes the crucial observation that the infinite regress can be avoided by introducing a circularity:  $\beta$  can be identified with  $B$ , which is already produced by the system”<sup>12</sup>, and then write  $\Phi = B(f)$ . However, although Rosen’s own diagram (cf. Fig. 1) could be taken to mean that  $\beta \equiv B$  it is clear from our analysis and from Rosen’s own papers that this is not what he meant.

Writing their three equations in  $\lambda$ -calculus notation, they become as follows:

$$(fA) = B$$

$$(\Phi B) = f$$

$$(Bf) = \Phi$$

They then substituted the first and third equation into the second to get

$$((fA)f)(fA) = f$$

This is a fixed-point equation meaning that  $f$  is a fixed point of

$$G = \lambda x.((xA)x(xA))$$

As the classical Curry fixed-point operator  $Y$  of  $\lambda$ -calculus provides a fixed point  $YM$  for any term  $M$ , in particular for our  $G$ , they conclude that they can take

$$f = YG$$

to obtain a solution for Rosen’s equations.

Notice that this result implies that the entire system, not just  $B$  but also  $f$  and  $\Phi$ , is fully determined by  $A$  alone. This seems biologically bizarre, and is not in any way implicit in Fig. 1. More important, the misidentification  $\beta \equiv B$  undermines the whole approach, because without this it is not possible to arrive at a simple expression of closure in terms of  $\lambda$ -calculus, which provides no convenient way to represent the inverse of a term like  $B$ .

## 8.3 Self-referential equations

In the simple arithmetical example discussed below (Section 8.7), metabolic states are just numbers modulo 12 and all admissible mappings are just scaling maps, and  $\beta$  is the scaling map given by multiplication by  $b^{-1}$ , not by  $b$ , and so the fixed point equation for  $f$  above turns out to be no more than a tautology,  $f = f$ , in which  $f$  is indeed a fixed point, but of the identity mapping!

This result is trivial, of course, so it is important to point out that the underlying behaviour is not trivial. As we discussed previously (Letelier et al., 2005b), the fundamental property of living organisms can be represented by a self-referential equation, also discussed by Mossio et al. (2009), of the following form:

$$f(f) = f$$

<sup>12</sup>Note that Mossio et al. (2009) use  $B$  for our  $b$ , which we (and Rosen) regard as an element of  $B$ .

Written as a mathematical equation it appears puzzling, but the biological property that it represents is more understandable:

$$\text{molecules}(\text{molecules}) = \text{molecules}$$

or

$$\text{metabolism}(\text{metabolism}) = \text{metabolism}$$

Thus metabolism can be regarded as a function that acts on metabolism to regenerate metabolism.

#### 8.4 Alternative mode of closure

Mossio et al. (2009) also discuss an alternative derivation of their result from the version of his argument that Rosen (1991) gave on pages 238–241 of his book (i.e. before reaching the more definitive analysis that he gave in pages 248–252). In effect, he considered that the version of closure embodied in Fig. 1a is not the only possible version, and one could also achieve closure in which a single term is the efficient cause of two different objects. In the context of their discussion we agree with Mossio et al. (2009) in their footnote 12 that “it would have been preferable for Rosen to demonstrate his theorem on the diagram representing canonical  $(M, R)$ -systems; but in the event, nothing hangs on this.” There are of course many ways of closing the diagram, but arguing about which of these is best would be like the sterile theological arguments over whether including the word *filioque* in the Nicene Creed is heretical or not. Nonetheless, there is a serious point here: as we showed (Cornish-Bowden et al., 2007), multifunctionality is an absolutely vital component of any system closed to efficient cause, because closure cannot be achieved if each molecule fulfils one role and one role only. Thus the playing of multiple roles is not in itself an objection to the approach Rosen used in pages 238–241 of his book, though Mossio et al. (2009) described it as “rather peculiar”. In fact, more and more experimental examples of multifunctionality are being reported (Tipton et al., 2003; Sriram et al., 2005; Gancedo and Flores, 2008).

#### 8.5 Recursive cycle of programs

Mossio et al. (2009) continued by relating these ideas to the theory of computer programming, in which one can set up a chain of programs such that a program  $\Phi$  produces as output a program  $f$ , which produces as output a third program  $B$ , which has output  $\Phi$ . It will be evident, however, that this discussion, although interesting, incorporates the misidentification  $\beta \equiv B$ , and thus does not invalidate Rosen’s proof that an organism cannot have a simulable model. The non-identity of  $\beta$  with  $b$  is, however, essential for understanding Rosen’s idea, and so we shall return to this point below.

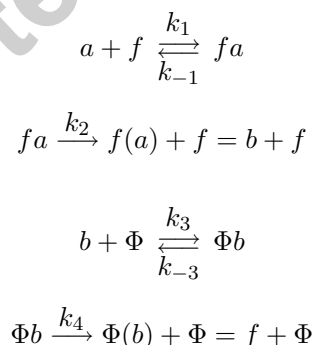
Even without the problem of the misidentification of  $\beta$  with  $b$ , the argument would still not necessarily invalidate the proof, because Rosen insisted on the original definition of *computability* used by Turing (1936), including the condition that the simulation program must terminate after a finite number of steps. However, although Mossio et al. (2009) evoke programs that can be written with finite strings of signs, they do

not show that the cycle of such programs is guaranteed to halt after a finite number of operations. Their definition of computability is therefore weaker than Rosen's and does not affect his proof. Mossio et al. (2009) recognize this, of course, but they consider the halting condition to be something "imposed by Rosen" rather than inherent in the definition of computability. In other words they do not necessarily disagree with Rosen, or, more recently, with Louie (2007), about the facts, but about whether halting after a finite number of steps is essential to the definition.

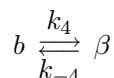
## 8.6 Rosen's analysis revisited

As the identification  $\beta \equiv b$  (where  $b \subset B$ ) is common in many analyses of Rosen's work, it is necessary to return once more to his own argument in order to see that this is an oversimplification that he avoided. He presented his idea in many publications, with different notations and different degrees of rigour, and one cannot escape the necessity to read these in depth to obtain a full understanding of his thinking. His paper of 1966 (Rosen, 1966) is especially important, and the summary in his book (Rosen, 1991) is not a complete account. Nonetheless, as it is the account most familiar to most readers we point out that although his text on pp. 248–252 is very clear<sup>13</sup> it needs to be read in the knowledge that the three figures on the same pages (Figs. 10C.3, 10C.5 and 10C.6) are unclear, first because they encourage the misidentification between  $\beta$  and  $B$  (in the figures) or with  $b$  (in the text), though the text makes it clear that that was not intended, and because the definition of the types of arrows used is inverted (silently, and presumably in error) between Figs. 10C.5 and 10C.6.

To clarify the relationship between  $\beta$  and  $b$  even further we need to refer to another paper in the series (Rosen, 1971). There Rosen made a radical departure from his previous purely algebraic approach, and formulated the problem in a way more familiar to kinetically minded biochemists, with the following equations:



Later in the paper he supplemented these with a further set of equations:



<sup>13</sup>Here Rosen used the symbol  $\hat{b}$  for the evaluation at  $b$ , which is more often symbolized (as above) as  $ev_b$ .

$$\beta + f \begin{array}{c} \xrightarrow{k_5} \\ \xleftarrow{k_{-5}} \end{array} \beta f$$

$$\beta f \xrightarrow{k_6} \Phi + \beta$$

As may be seen, especially in the equation  $b \rightleftharpoons \beta$  in the second set, Rosen regarded  $b$  and  $\beta$  as different entities, albeit related and interconvertible.

### 8.7 Arithmetical example of an $(M, R)$ -system

A major difficulty in accepting Rosen's analysis arises from a difficulty in accepting that the evaluation map for functions on sets can be invertible, as this is an unusual property of such a map, and as Rosen's device for achieving closure to efficient causation depends on its possible existence it can only be correct if the property is possible.

Our main point is that metabolisms are represented by quite special mappings  $f$  from  $A$  to  $B$ . We follow here the notations introduced by Rosen, who was never fully clear about the kind of mappings represented by his  $f$ , sometimes even writing  $H(A, B)$  to denote the set of all such mappings  $f$  from  $A$  to  $B$ ; this implied that this set could be a proper subset of the set  $Map(A, B)$  of all possible mappings from  $A$  to  $B$ .

In our simple arithmetical example (Letelier et al., 2006) we took  $A$  and  $B$  to be the integers modulo 12, which can be interpreted as "hours":  $0, 1, 2 \dots 11$ , with  $12 \equiv 0$ . Hours may be added modulo 12; for example,  $9 + 6 = 15 = 3$  modulo 12. We then take  $H(A, B)$  to consist of all mappings that preserve addition, which are simply "scaling mappings", i.e. multiplication by a constant number modulo 12 (which can be recovered by evaluating the scaling mapping at 1). It follows that evaluation at 1 is invertible, because from any number  $k$  modulo 12 we can recover the scaling mapping given by multiplication by  $k$ , whose value at 1 is of course  $k$ .

This example is not biological, but the essential point is that it disposes of any suggestion that Rosen's conclusion is impossible on the mathematical grounds that it requires an impossible inversion. Study of this example led us to argue that the evaluation maps needed for Rosen's argument must be applied to a class of restricted, structure-preserving, or "admissible", maps, a tiny sub-set of the universe of possible mappings from  $A$  to  $B$  (Letelier et al., 2006). The smallness of this subset is visible even in an example as simple as the one we described, for which there are  $12^{12}$ , i.e. about  $8.9 \times 10^{12}$ , possible functions from  $A$  to  $B$  to which the evaluation map at  $b$  may be applied, but only 12, or  $1.3 \times 10^{-10}\%$ , of these are "structure preserving", i.e. scaling maps given by multiplication by a constant number  $k$  modulo 12. On this set  $H(A, B)$  with 12 elements, the evaluation maps at 1, 5, 7 and 11 are invertible. Evaluation maps at 2, 3, 4, 6, 8, 9, 10 are not one-to-one and therefore not invertible: scaling by 3 and scaling by 9 take the same value  $6 = 18$  modulo 12 at 2, for instance.

At this point we need to refer back to our earlier criticism (Letelier et al., 2006) of the arguments by Landauer and Bellman (2002) that Rosen's analysis could not work. We incorrectly implied that these authors considered Rosen's argument to be impossible, whereas in reality (see Section 2.5 of their paper) they considered it too trivial to support his claim that an organism was not a mechanism. However, their

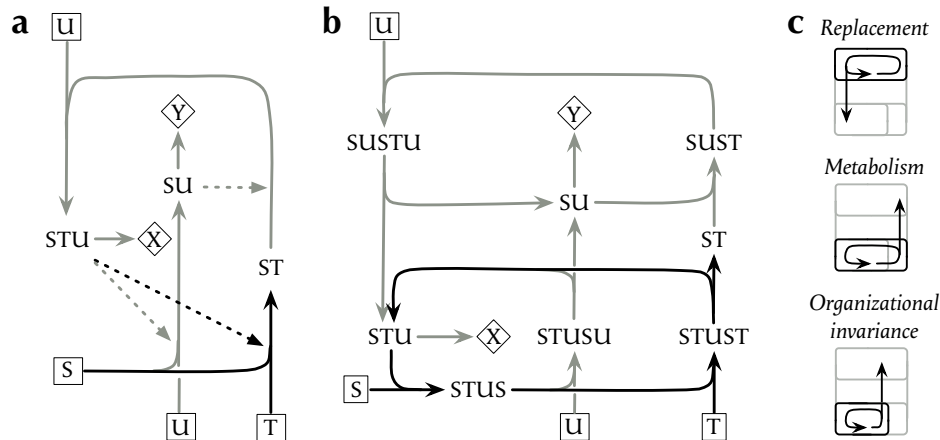


Fig. 6. A model of an  $(M, R)$ -system. This is the biological model suggested previously, (a) with reaction cycles compressed into catalysed reactions (Letelier et al., 2006), with catalysts shown as acting on reactions rather than on substrates (cf. Goudsmit, 2007), (b) with reaction cycles shown explicitly (Cornish-Bowden et al., 2007; Cornish-Bowden and Cárdenas, 2007), and (c) with an interpretation of the cycles in terms of metabolism, replacement and organizational invariance. The metabolites shown inside squares (input) or diamonds (output) are considered to be “external” and to have fixed concentrations. Fig. 3b of Cornish-Bowden et al. (2007), should have been very similar to the form shown in (b), but it was printed incorrectly, with three intermediates misidentified, STUST as SUST, STUSU as SUSU, and SUSTU also as SUSU (with the result that the names SUST and SUSU occurred twice each). The arrows shown in black highlight the part of the model that was considered to be “metabolism” during its construction, i.e.  $S + T \rightarrow ST$  catalysed by STU. Note, however, that this cannot be deduced simply from the structure of the network.

own argument depended on a simple example that was indeed too trivial, as it was based on a two-member set that could be regarded as its own inverse.

## 9 A simple metabolic model interpreted in terms of theories of closure

### 9.1 $(M, R)$ -systems

Although there has been a considerable resurgence of interest in Rosen’s view of organisms in recent years, a large part of the discussion has focussed on his diagram in Fig. 10C.6 of his book (Rosen, 1991), corresponding to Fig. 1a of the present paper. However, as we have emphasized in the Introduction, full understanding of Rosen’s work cannot be obtained from a single diagram, and we, following an idea of Morán et al. (1996), have been exploring the characteristics of a very small metabolic network in the hope of revealing in a simple system how the concepts embodied in the symbols  $f$ ,  $\Phi$  and  $\beta$  can be mapped, not into  $\lambda$ -calculus but into concepts more familiar to biologists.

The small network illustrated in Fig. 6 represents a minimal  $(M, R)$ -system. It was first given in outline (Letelier et al., 2006), and later in more detail (Cornish-Bowden

et al., 2007; Cornish-Bowden and Cárdenas, 2007)<sup>14</sup>.

Fig. 6 looks quite different from Fig. 1, and it is natural therefore to enquire what is the relationship between the two. In naive terms, this is illustrated in Fig. 7. In this figure we have tried to establish a clear correspondence between Figs. 1 and 6, in effect superimposing a specific biochemical network on Rosen's conceptual world. Despite the dangers of trying to encapsulate a complex theory in a simple example, it may allow the abstract ideas to be made more concrete, and several points emerge from study of this example:

1. The input in Fig. 6 ( $\{S, T, U\}$ ) corresponds to  $A$  in Fig. 1.
2. The products of the metabolic network ( $\{ST, SU, STU, X, Y\}$ ) correspond to  $B$  in Fig. 1.
3. The metabolic part of the model ( $S + T \rightarrow ST$ ) is only a subset of the complete set of reactions.
4. Some components fulfil multiple functions, for example  $STU$  embodies both  $f$  and  $\beta$ . We believe that multifunctionality is essential for achieving closure (Cornish-Bowden et al., 2007; Cornish-Bowden and Cárdenas, 2007), and for avoiding the combinatorial explosion that would otherwise occur, if every new enzyme required a new set of enzymes to maintain it. The particular example of it seen here, however, is a consequence of the smallness of the model, and should not be taken to suggest that  $f$  and  $\beta$  coincide in general.
5. Every function ( $f$ ,  $\Phi$  or  $\beta$ ) is embodied by some molecule produced by the system.
6. The molecules embodying  $B$  ( $\{ST, SU, STU, X, Y\}$ ) and  $\beta$  ( $STU$ ) are not the same. This point is crucial.
7. The set  $B$  is different from  $A$ , and contains elements that account for  $f$ ,  $\Phi$  and  $\beta$ .
8. The system is an  $(M, R)$ -system, but it is not an  $(M, R)$ -system with organizational invariance. As discussed (Letelier et al., 2006), the set of biochemical reactions accepts more than one assignment for  $\Phi$ ; in other words the particular assignment shown is not the only possible one, and so the required property of invertibility is violated.

In relation to this last point, however, note that this is an artificial example in which we, as its inventors, had a free choice in ascribing catalytic properties to whatever molecules we wished. In a real system, in contrast, there is little or no choice, because the reactions that the constituent molecules can undergo are circumscribed by their structures and the laws of chemistry. No more choice is involved than there is for a pigeon when it uses its eyes for seeing and its ears for hearing, rather than the other way around. Expanding this type of micro-metabolic example, and rooting it more

<sup>14</sup>Unfortunately Fig. 3b of Cornish-Bowden et al. (2007) was printed incorrectly, as detailed in the legend to Fig. 6. The corresponding illustration in Cornish-Bowden and Cárdenas (2007), Fig. 1b therein, was printed correctly.

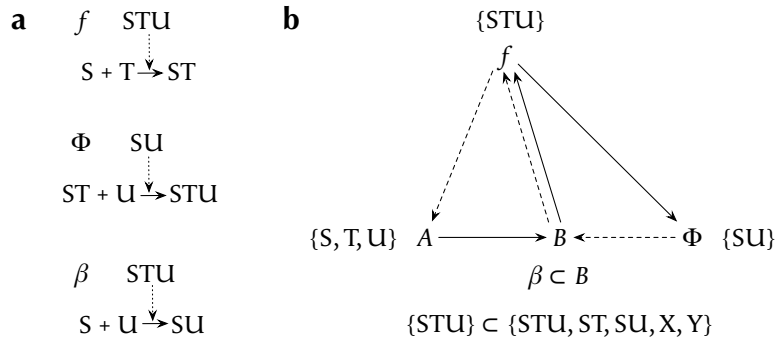


Fig. 7. An attempt to interpret the model of Fig. 6 in terms of the closure diagram (Fig. 1). (a) The three functions  $f$ ,  $\Phi$  and  $\beta$  are associated with two molecules (one of them used twice). (b) The closure diagram shown in Fig. 1a is labelled with the molecules in Fig. 6. The correspondence is at best approximate, and should not be taken too precisely. In particular, the fact that  $\{\text{STU}\}$  is identified both with  $f$  and with  $\beta$  is not a general property but is just a consequence of the smallness of the model. Nonetheless, some sort of multifunctionality is essential for closure, so although at least some molecules must have multiple functions the choice of STU in this example was arbitrary.

firmly in chemical possibilities that are known from the study of real metabolism, should eventually allow us to arrive at the smallest real metabolic system (which perhaps will turn out to be very large indeed) in which the condition of invertibility ( $\Phi(b) = f$ ) has one and only one solution, making  $\beta$  a well defined entity.

The research programme just outlined will certainly be very difficult, and perhaps impossible, to achieve. Elsewhere (Cornish-Bowden and Cárdenas, 2007) we have hinted at an alternative way of looking at the model that may offer a way out of the impasse. Fig. 7 assumes a fundamental distinction between enzymes and metabolites, i.e. it is based on Fig. 6a. However, such a distinction may not correspond to reality, and Fig. 6b, in which all the catalytic processes are expanded into cycles of chemical reactions, may be considered more realistic. All enzymes are products of metabolism, and are therefore metabolites, and many metabolites (in the usual sense) participate in reaction cycles that regenerate them, and are thus catalysts. The distinction between enzymes and metabolites is thus more artificial than it is usually taken to be, and is simply a human interpretation of some very complicated chemistry. On the interpretation of Fig. 6b, therefore, the regeneration of all catalysts is automatically taken care of by the structure of the network and the laws of chemistry.

A similar point has been made by Kun et al. (2008), who analysed many well characterized metabolic networks and found that intermediary metabolism is always autocatalytic for ATP: although ATP is a product of metabolism it can never be produced without pre-existing ATP. Other metabolic cycles may be obligatorily autocatalytic in some networks, but not necessarily in all. They also comment, following Gánti (2003b), that “in addition to template and membrane growth, metabolism is also autocatalytic and, hence, results in replication”. We take this to refer to replication in Rosen’s sense, which, as we have mentioned, we prefer to call organizational invariance.

It is interesting to note that in biochemical systems theory (Savageau, 1976; Voit,



2000), the same sort of symbols are used for the concentrations of all components of a system, whether reactants or catalysts, in other words no necessary distinction between enzymes and metabolites is assumed. The distinction is maintained, however, in metabolic control analysis (Kacser and Burns, 1973; Heinrich and Rapoport, 1974; Fell, 1997), the other main approach to analysis of multienzyme systems.

In attempting here to relate the biochemical model to Rosen’s mathematical formalism we have oversimplified some points in the hope of remaining intelligible in chemical terms. It is hardly possible at the present state of understanding to resolve all the problems and arrive at a mathematically rigorous analysis, but we can note some points that will need to be clarified in the future. In Fig. 7 we took  $A$  to be the three-element set  $\{S, T, U\}$ , but more accurately this should be written as the three-element set  $A = \{(S, T), (ST, U), (S, U)\}$ , or better still the one-element set  $A = \{((S, T), (ST, U), (S, U))\}$ . A rigorous analysis will need to take account of this.

The underlying biological idea of  $\Phi$  is that it is a function that assigns molecular identities to the enzymes involved in

$$(S, T) \mapsto ST, (ST, U) \mapsto STU, (S, U) \mapsto SU$$

Understanding  $\beta$  ultimately amounts to understanding the unicity of  $\Phi$ , which requires an understanding of the constraints that  $\Phi$  must fulfil, which are not explicit in a purely combinatorial approach in which any molecule can (in principle) catalyse any reaction.

## 9.2 Other definitions of life

Fig. 6 is not only an  $(M, R)$ -system; it also satisfies the definition of an autocatalytic set (Kauffman, 1993) given in Section 3. In the case of Fig. 6 the autocatalytic set is  $\{STU, ST, SU\}$  and, as seen in Fig. 7a, the formation of every member of the set is catalysed by another member of the set (“at least one of the possible last steps in its formation [is] catalyzed by some member of the set”). The food set is  $\{S, T, U\}$ , and it is obvious from inspection that the second half of the definition (“connected sequences of catalyzed reactions lead from the maintained food set to all members of the autocatalytic set”) is also satisfied. We have been able to construct an autocatalytic set far smaller than those assumed by Kauffman because we have not incorporated his condition that closure is an inevitable consequence of the statistical properties<sup>15</sup> of the constituent molecules. It is worth recalling, however, that in a very large system we cannot assume that all of the individual molecules are nutrients: “instead of being ‘nutrients’ they are likely to be ‘poisons’ ... [and] a completely recycling system would have evolved chemically from a more complex system by losing those materials which would otherwise disrupt the recycling” (King, 1982). The model of Fig. 6 is, in fact, an autocatalytic system in King’s sense, and it is not by chance that Fig. 6 is somewhat similar in appearance to the subsystem illustrated in Fig. 1b of King (1982).

The minimal metabolic network that we have analysed can also be interpreted as an example of a minimal autopoietic system, because the three coupled transformations can be interpreted as “a network of processes that produces itself”, and if one

<sup>15</sup>The key word here is, of course, “inevitable”, not “statistical”: one cannot reasonably doubt that life started with a chance event, but it was not necessarily an event with high probability.

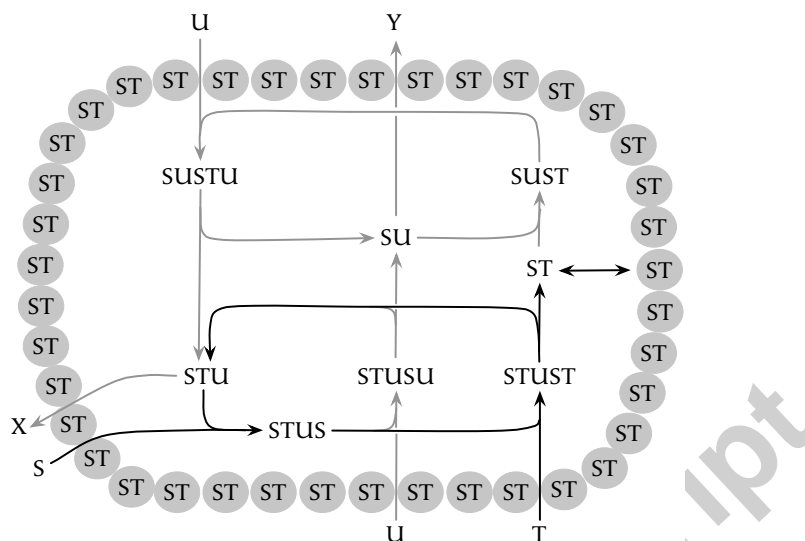


Fig. 8. A model of an  $(M,R)$ -system modified to incorporate enclosure of the system within a membrane, as in autopoietic models. The metabolic product  $ST$  is assumed to be capable of self-association to form a polymer that can form a membrane.

of the components ( $ST$ ) is further endowed with the property of self-assembling into a membrane the necessary encapsulation property of autopoietic systems is achieved (Fig. 8). Interestingly, this interpretation of a minimal  $(M,R)$ -system network as an autopoietic system overcomes an important shortcoming of previous models of autopoietic systems, which fail to explain (or sometimes even to mention) the origin of the necessary catalyst (see Section 5). In these previous models (Varela et al., 1974; Zeleny, 1981; McMullin and Varela, 1997; Breyer et al., 1998) the component responsible for catalysing the process  $S + S \rightarrow L$  was not produced by the system but rather imposed from the outside, making it a property of space rather than a consequence of biochemical interactions. In the language of efficient causation, therefore, these previous examples of autopoietic systems were not closed to efficient causation.

Finally, we should consider our model in the light of the *chemoton* of Gánti (2003a). In a sense this is more general than the  $(M,R)$ -system, and is therefore more than just a special case of an  $(M,R)$ -system, because it explicitly includes what Schrödinger (1944) called a *codescript*, i.e. a module for storing and using information, what in modern terminology we would call replication (though emphatically *not* what Rosen called replication).<sup>16</sup>

Gánti's chemoton has two fundamental aspects, cycle stoichiometry and the subdivision of metabolism into three subsystems. Like Rosen, Maturana and Varela, he aimed to define the special organization that a complex system of biochemical reactions must have in order to exhibit lifelike properties. He described the key in the

<sup>16</sup>Hofmeyr (2007) has argued that the chemoton is more limited than  $(M,R)$ -systems, because he considers that simple molecules implicit in Gánti's model cannot have the specificity needed to prevent the system from dissipating into unspecific side reactions. However, this is a problem with all current theories of life, none of which explains the origin of the necessary specificity.

following words: “Within a living system almost all events are connected with each other. These interactions happen in a special order, and the unique properties of living systems originate just from this special order of the interactions” (Gánti, 1975). Thus he identified the need for processes to occur in a defined order, but not the need for closure, a vital difference from  $(M,R)$ -systems and autopoiesis.

His three interlocking subsystems with a self-maintaining metabolism are the cyclic (later called metabolic), the genetic and the membrane subsystems. In adopting as necessary a genetic subsystem (essentially an informational system based on templates similar to the nucleic-acid based system found in current living organisms) he incorporated a restriction not found in  $(M,R)$ -systems or autopoiesis, which are not concerned with reproduction or evolution. To produce a system that can produce its own enzymes, and also to take into account that enzymes are catalysts that need to be continuously regenerated by the system itself, Gánti introduced the notion of cyclic stoichiometry, and organized the three subsystems as loops that produce their own enzymes, as illustrated in Fig. 1 of Gánti (1997).

Thus the chemoton is equivalent to a particular implementation of an autopoietic system that explicitly deals with the origin of enzymes via cycles and also makes explicit the steps to produce the encapsulating membrane. But important conceptual differences remain: in particular, organizational closure is the central concept in autopoiesis and  $(M,R)$ -systems, whereas it is absent from Gánti’s writings.

### 9.3 Comparison between theories of life

As we have noted briefly before (Cornish-Bowden and Cárdenas, 2009), a puzzling feature of the theories we have discussed about describing cellular organization from new principles (Rosen, 1958a; Gánti, 1966; Varela et al., 1974) has been the complete lack of dialogue between them or with autocatalytic sets (Kauffman, 1993), despite important overlap between some of their ideas. A survey of the core literature reveals that Rosen, in a long period between 1958 and 1971, only cited himself. Similar observations apply to the authors of autopoiesis; likewise the first English version of the chemoton theory (Gánti, 1975) also ignores the previous work on autopoiesis (understandable as it was only published some months before in 1974) and Rosen (less easy to explain, as the core papers were published in 1959 in a well known journal). By 1975, therefore, these three competing theories existed in complete isolation from one another. Full analysis of this autarky would be interesting, but it is beyond the scope of the current paper. A partial explanation may perhaps be found in the fact that Gánti’s original work (from 1966 to 1974) was published in Hungarian, and Maturana and Varela initially published in Spanish and were not in the mainstream of biochemistry (both were neurophysiologists more interested in artificial intelligence than in artificial life). In any case the fact that Gánti, Maturana and Varela did not cite Rosen’s work constituted a serious oversight due, perhaps, to the opaque style systematically maintained by Rosen.

By 1995 these three theories had evolved almost independently in the literature. Rosen had already published his seminal, but difficult, book *Life Itself* (Rosen, 1991), but continued with his entrenched tradition of ignoring competing viewpoints. A similar statement can be made for autopoiesis, which by then had already being found and expanded by many people in the “cybernetics” community, but the principal

authors did not cite Rosen or Gánti. A curious departure from this trend is provided by Gánti's only publication in this Journal (Gánti, 1997), in which he acknowledged that Varela was also examining the nature of life, but ignored Rosen's work completely, despite mentioning Varela in a section entitled "Life itself" in a paper whose title also evokes the title of Rosen's book.

Thus, for reasons that historians of science will need to explain in detail, the authors of three of the main theories concerning the inner organization of living metabolism chose, systematically, to ignore each other. In fact one of our first contributions to this field was an attempt to end this internal separation and to make explicit the relation between  $(M,R)$ -systems and autopoiesis (Letelier et al., 2003).

Part of the explanation for the isolation may be sought, no doubt, in the very different backgrounds of the principal players, Rosen as a theoretical biologist from the school of Nicolas Rashevsky, Gánti as a chemical engineer, and Maturana and Varela as neurophysiologists concerned primarily with the working of the nervous system, with no initial intention of developing a theory of life. They set out not to understand life but to understand how to avoid the infinite regress implicit in most models of brain function. That of course parallels the problem that Rosen sought to solve, and the solution of closure they proposed parallels Rosen's solution in  $(M,R)$ -systems.

## 10 Conclusions

Efforts to mathematically disprove Rosen's contention that an organism cannot have simulable models have not resolved the question. Louie (2007) has been highly critical of some of the arguments (Chu and Ho, 2006), and, as we have discussed in Section 3, there are problems also with some of the others. Other supposed contradictions can be attributed to the use of loose definitions in place of Rosen's very precise ones. As noted above, for example, the definition of computability used by Mossio et al. (2009) does not require termination of the program in a finite number of steps. Their definition of computability is widely accepted, but a more serious problem is their representation of Rosen's scheme with an incorrect set of equations. Similarly Wells (2006) replaced Rosen's precise definition of a *mechanism* by a vague one based on everyday ideas of what a machine is, and used it to claim that Rosen's conclusions were mistaken.

Rosen did not reject the possibility of artificial life, and in one of his less often cited papers (Rosen, 1973), he made an important distinction between the abstract theory of  $(M,R)$ -systems, in which "the replication map<sup>17</sup> is always perfect" and the dynamical formalism that he was discussing, in which it was legitimate "to seek specific realizations of abstract  $(M,R)$ -systems in dynamical terms, and for each such dynamical realization, to seek actual physical processes whose equations of motion are precisely those of the realization".

However, the argument about simulability will certainly continue: the work of many groups, including those attempting to develop life *in silico* depends on the as-

<sup>17</sup>It is important to remember that "replication" in Rosen's terminology has nothing to do with DNA replication, but corresponds to what we have called organizational invariance. His terminology appears, not surprisingly, to have misled some authors.

sumption that computer simulation of living systems is in principle possible, and any claims that it is not possible can expect to meet vehement opposition. It is important, therefore, to examine arguments on both sides with care and attention. In this paper we have tried to do that, and supply a more biological insight into Rosen's ideas.

This work was supported by Fondecyt 1030371 (JCL), Fondecyt 1040444 (JSA) and the CNRS (AC-B, MLC).

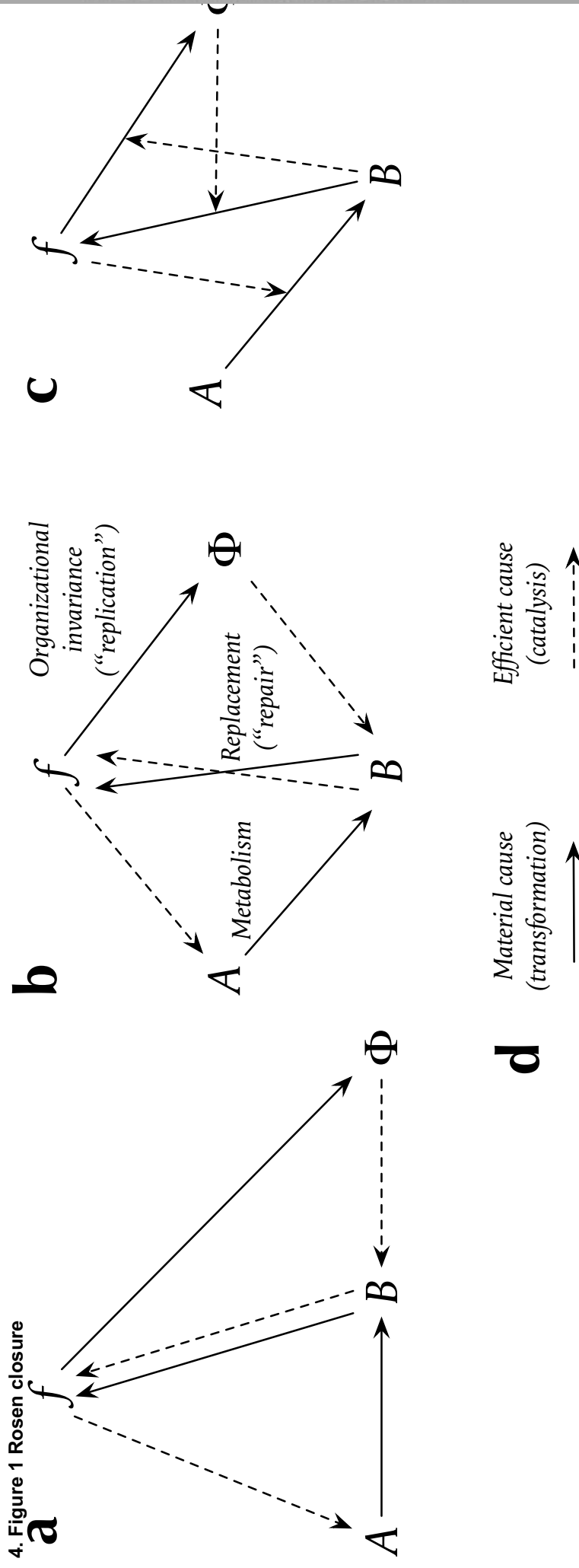
## References

- Breyer, J., Ackerman, J., McCaskill, J., 1998. Evolving reaction-diffusion ecosystems with self-assembling structures in thin films. *Artif. Life* 4, 25–40.
- Chemero, A., Turvey, M.T., 2006. Complexity and “closure to efficient cause”. In: Ruiz-Moreno, K., Barandiaran, R. (Eds.), *ALIFE X: Workshop on Artificial Autonomy*. MIT Press, Cambridge, MA, pp. 13–18.
- Chemero, A., Turvey, M.T., 2008. Autonomy and hypersets. *BioSystems* 91, 320–330.
- Chemero, A., Turvey, M.T., 2007. Complexity, hypersets, and the ecological perspective on perception-action. *Biol. Theory* 2, 23–36.
- Chu, D., Ho, W.K., 2006. A category theoretical argument against the possibility of artificial life: Robert Rosen's central proof revisited. *Artif. Life* 12, 117–134.
- Chu, D., Ho, W.K., 2007a. The localization hypothesis and machines. *Artif. Life* 13, 299–302.
- Chu, D., Ho, W.K., 2007b. Computational realizations of living systems. *Artif. Life* 13, 369–381.
- Cornish-Bowden, A., 1989. Metabolic control theory and biochemical systems theory: different objectives, different assumptions, different results. *J. Theor. Biol.* 136, 365–377.
- Cornish-Bowden, A., Cárdenas, M.L., 2007. Organizational invariance in  $(M, R)$  systems. *Chem. Biodivers.* 4, 2396–2406.
- Cornish-Bowden, A., Cárdenas, M.L., 2007. Self-organization at the origin of life. *J. Theor. Biol.* 252, 411–418.
- Cornish-Bowden, A., Cárdenas, M.L., 2009. Catalysis at the origin of life viewed in the light of the  $(M, R)$ -systems of Robert Rosen. In: Kettner, C. and Hicks, M.G. (Eds.), *Systems Chemistry*, Beilstein Institute, Frankfurt, pp. 21–33.
- Cornish-Bowden, A., Cárdenas, M.L., Letelier, J.C., Soto-Andrade, J., 2007. Beyond reductionism: metabolic circularity as a guiding vision for a real biology of systems. *Proteomics* 7, 839–845.
- Cottam, R., Ranson, W., Vounckx, R., 2007. Re-mapping Robert Rosen's  $(M, R)$ -systems. *Chem. Biodivers.* 4, 2352–2368.
- Dittrich, P., Ziegler, J., Banzhaf, W., 2001. Artificial chemistries: a review. *Artif. Life* 7, 225–275.
- Fell, D., 1997. *Understanding the Control of Metabolism*, Portland Press, London.
- Fernando, C., Rowe, J., 2007. Natural selection in chemical evolution. *J. Theor. Biol.* 247, 152–167.
- Fernando, C., Rowe, J., 2007. The origin of autonomous agents by natural selection. *BioSystems* 91, 355–373.
- Fontana, W. and Buss, L.W., 1996. The barrier of objects: from dynamical systems to bounded organizations. In: Casti, J., Karlqvist, A. (Eds.), *Boundaries and Barriers*, Addison-Wesley, Reading, MA, pp. 56–116.
- Gancedo, C., Flores, C.L., 2008. Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* 72, 197–210.
- Gánti, T., 1966. *Forradalom az Élet Kutatásában (Revolution in the Research of Life)*, Gondolat, Budapest.

- Gánti, T., 1975. Organization of chemical reactions into dividing and metabolizing units: the chemotons. *BioSystems* 7, 15–21.
- Gánti, T., 1997. Biogenesis itself. *J. Theor. Biol.* 187, 583–593.
- Gánti, T., 2003. *The Principles of Life*, Oxford University Press, Oxford.
- Gánti, T., 2003. *Chemoton Theory*, Kluwer, New York.
- Goudsmit, A.L., 2007. Some reflections on Rosen's conceptions of semantics and finality. *Chem. Biodivers.* 4, 2427–2345.
- Heinrich, R., Rapoport, T.A., 1974. A linear steady-state theory of enzymatic chains: general properties, control and effector strength. *Eur. J. Biochem.* 42, 89–95.
- Hofmeyr, J.-H.S., 2007. The biochemical factory that autonomously fabricates itself: a systems biological view of the living cell. In: Boogerd, F.C., Bruggeman, F.J., Hofmeyr, J.H.-S., Westerhoff, H.V. (Eds.), *Systems Biology: Philosophical Foundations*, Elsevier, Amsterdam, pp. 219–242.
- Kacser, H., Burns, J.A., 1973. The control of flux. *Symp. Soc. Exp. Biol.* 27, 65–104.
- Kauffman, S.A., 1986. Autocatalytic sets of proteins. *J. Theor. Biol.* 119, 1–24.
- Kauffman, S.A., 1993. *The Origins of Order*, Oxford University Press, New York, p. 299.
- Kercel, S.W., 2007. Entailment of ambiguity. *Chem. Biodivers.* 4, 2369–2385.
- King, G.A.M., 1977a. Symbiosis and the origin of life. *Orig. Life* 8, 39–53.
- King, G.A.M., 1977b. Symbiosis and the evolution of prokaryotes. *BioSystems* 9, 35–42.
- King, G.A.M., 1982. Recycling, reproduction, and life's origins. *BioSystems* 15, 89–97.
- Kun, Á., Papp, B., Szathmáry, E., 2008. *Genome Biol.* 9, R51.
- Landauer, C., Bellman, K., 2002. Theoretical biology: organisms and mechanisms. In: Dubois, M. (Ed.), *Computing Anticipatory Systems: CASYS 2001 — Fifth International Conference*, American Institute of Physics. 59–70.
- Letelier, J.-C., Marín, J., Mpodozis, J., 2003. Autopoietic and  $(M, R)$ -systems. *J. Theor. Biol.* 222, 261–272.
- Letelier05 Letelier, J.-C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., Cárdenas, M.L., 2004. Metabolic closure in  $(M, R)$ -systems. In: Pollack, J., Bedau, M., Husbands, P., Ikegami, T., Watson, R.A. (Eds.), *Artificial Life*, vol. IX. MIT Press, Cambridge, MA, pp. 450–455. Academic Press, New York.
- Letelier, J.-C., Kuboyama, T., Yasuda, H., Cárdenas M.L., Cornish-Bowden, A., 2005. A self-referential equation,  $f(f) = f$ , obtained using the theory of  $(M, R)$ -systems: overview and applications. In: Anai, H., Horimoto, K. (Eds.), *Algebraic Biology 2005*, Universal Academy Press, Tokyo, pp. 115–126.
- Letelier, J.-C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., Cárdenas, M.L., 2006. Organizational invariance and metabolic closure: analysis in terms of  $(M, R)$ -systems systems, *J. Theor. Biol.* 238, 949–961.
- Louie, A.H., 2007. A living system must have noncomputable models. *Artif. Life* 13, 293–297.
- Louie, A.H., 2009. *More than Life Itself*, Ontos Verlag, Frankfurt.
- McMullin, B., 2004. Thirty years of computational autopoiesis: a review. *Artif. Life* 10, 277–295.
- McMullin, B., Varela, F.J., 1997. Rediscovering computational autopoiesis. In: Husbands, P., Harvey, I. (Eds.), *Proceedings of the Fourth European Conference on Artificial Life*, MIT Press, Cambridge, MA, pp. 38–47.
- Maturana, H.R., Varela, F.J., 1980. *Autopoiesis and Cognition: the Realisation of the Living*, D. Reidel Publishing Company, Dordrecht, The Netherlands.
- Morán, F., Moreno, A., Minch, E., Montero, F., 1996. Further steps towards a realistic description of the essence of life. In: Langton, C.G., Shimohara, K. (Eds.), *Artificial Life V*, MIT Press, Cambridge, MA, pp. 255–263.
- Mossio, M., Longo, G., Stewart, J., 2009. An expression of closure to efficient causation in terms of  $\lambda$ -calculus. *J. Theor. Biol.* 257, 498–498.
- Orgel, L.E., 2008. The implausibility of metabolic cycles on the prebiotic earth. *PLoS Biol.* 6,

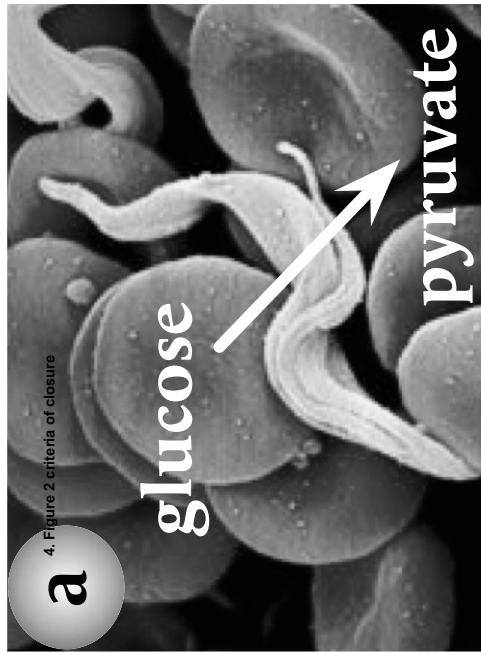
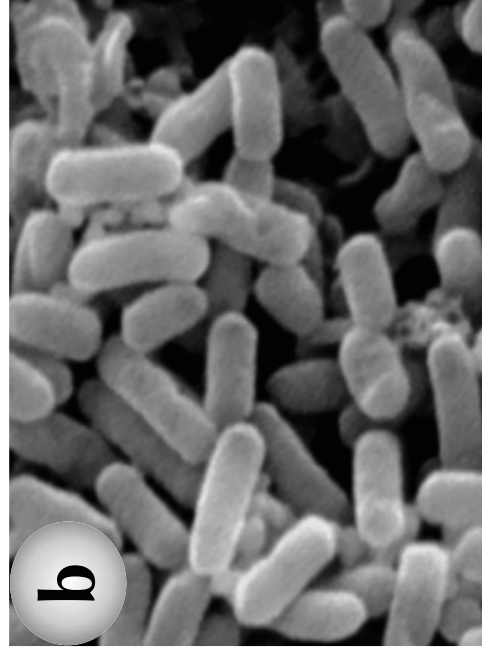
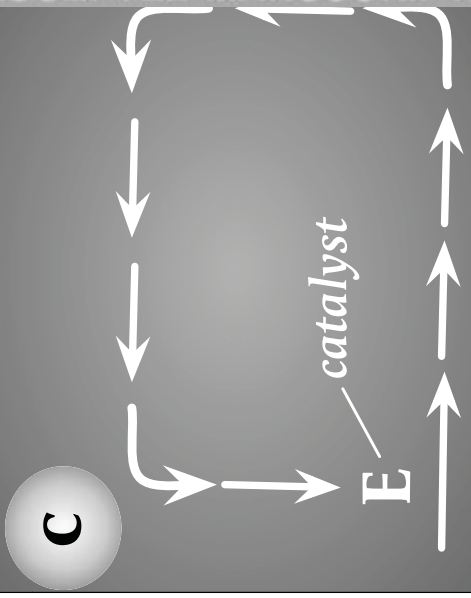
e18.

- Rosen, R., 1958a. A relational theory of biological systems. *Bull. Math. Biophys.* 20, 245–260.
- Rosen, R., 1958b. The representation of biological systems from the standpoint of the theory of categories. *Bull. Math. Biophys.* 20, 245–260.
- Rosen, R., 1959. A relational theory of biological systems. II. *Bull. Math. Biophys.* 21, 109–128.
- Rosen, R., 1966. A note on replication in  $(M, R)$ -systems. *Bull. Math. Biophys.* 28, 149–151.
- Rosen, R., 1971. Some realizations of  $(M, R)$ -systems and their interpretation. *Bull. Math. Biophys.* 33, 303–319.
- Rosen, R., 1973. On the dynamical realization of  $(M, R)$ -systems. *Bull. Math. Biol.* 35, 1–9.
- Rosen, R., 1991. *Life Itself: a Comprehensive Inquiry into the Nature, Origin and Fabrication of Life*. New York, Columbia University Press.
- Savageau, M.A., 1976. *Biochemical Systems Analysis: a Study of Function and Design in Molecular Biology*, Addison–Wesley, Reading, MA.
- Schrödinger, E., 1944. *What is Life?* Cambridge University Press, Cambridge.
- Sriram, G., Martinez, J.A., McCabe, E.R.B., Liao, J.C., Dipple, K.M., 2005. Single-gene disorders: what role could moonlighting enzymes play? *Am. J. Hum. Genet.* 76, 911–924.
- Tipton, K.F., O’Sullivan, M.I., Davey, G.P., O’Sullivan, J., 2003. It can be a complicated life being an enzyme, *Biochem. Soc. Trans.* 31, 711–715.
- Turing, A.M., 1936. On computable numbers, with an application to the Entscheidungsproblem. *Proc. Lond. Math. Soc.* (2) 42, 230–265.
- Varela, F.J., Maturana, H.R., Uribe, R., 1974. Autopoiesis: the organization of living systems, its characterization and a model. *BioSystems*, 5, 187–196.
- Voit, E.H., 2000. *Computational Analysis of Biochemical Systems: a Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge.
- Wells, A.J., 2006. In defense of mechanism. *Ecol. Psych.* 18, 39–65.
- Wolkenhauer, O., 2007. Interpreting Rosen, *Artif. Life* 13, 291–292.
- Wolkenhauer, O., Hofmeyr, J.-H.S., 2007. An abstract cell model that describes the self-organization of cell function in living systems. *J. Theor. Biol.* 246, 461–476.
- Zeleny, M. (Ed.), 1981. *Autopoiesis: a Theory of Living Organization*, North-Holland, New York.
- Zeleny, M., 1981. Ecosocieties: societal aspects of biological self-production. *Soz. Syst.* 1, 179–202.



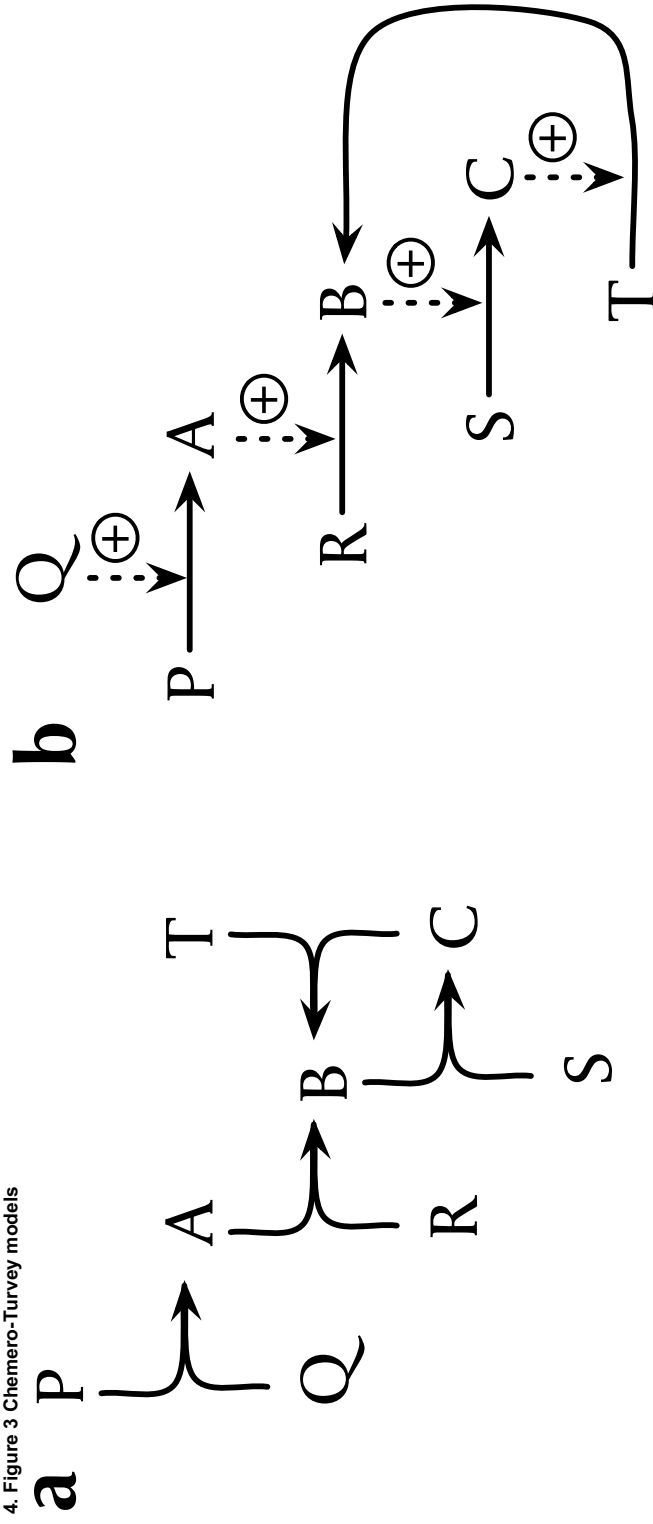
4. Figure 1 Rosen closure

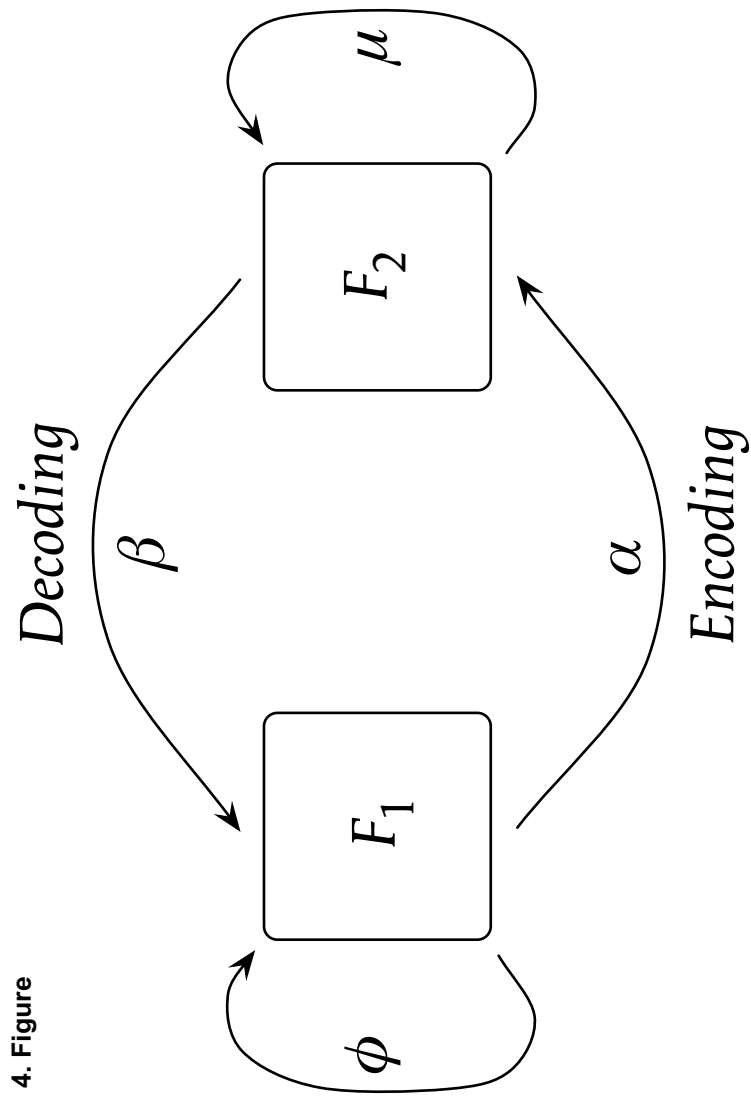




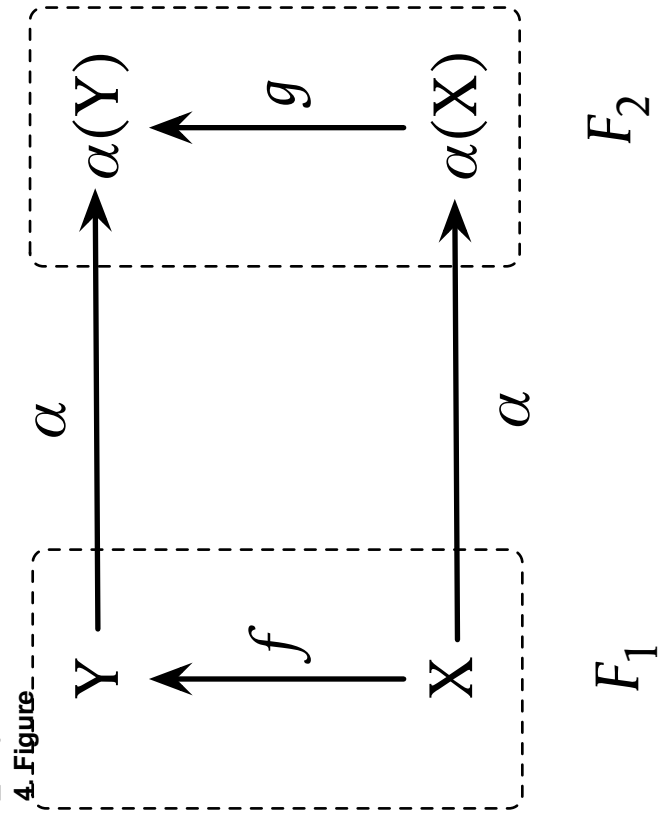
4. Figure 2 criteria of closure

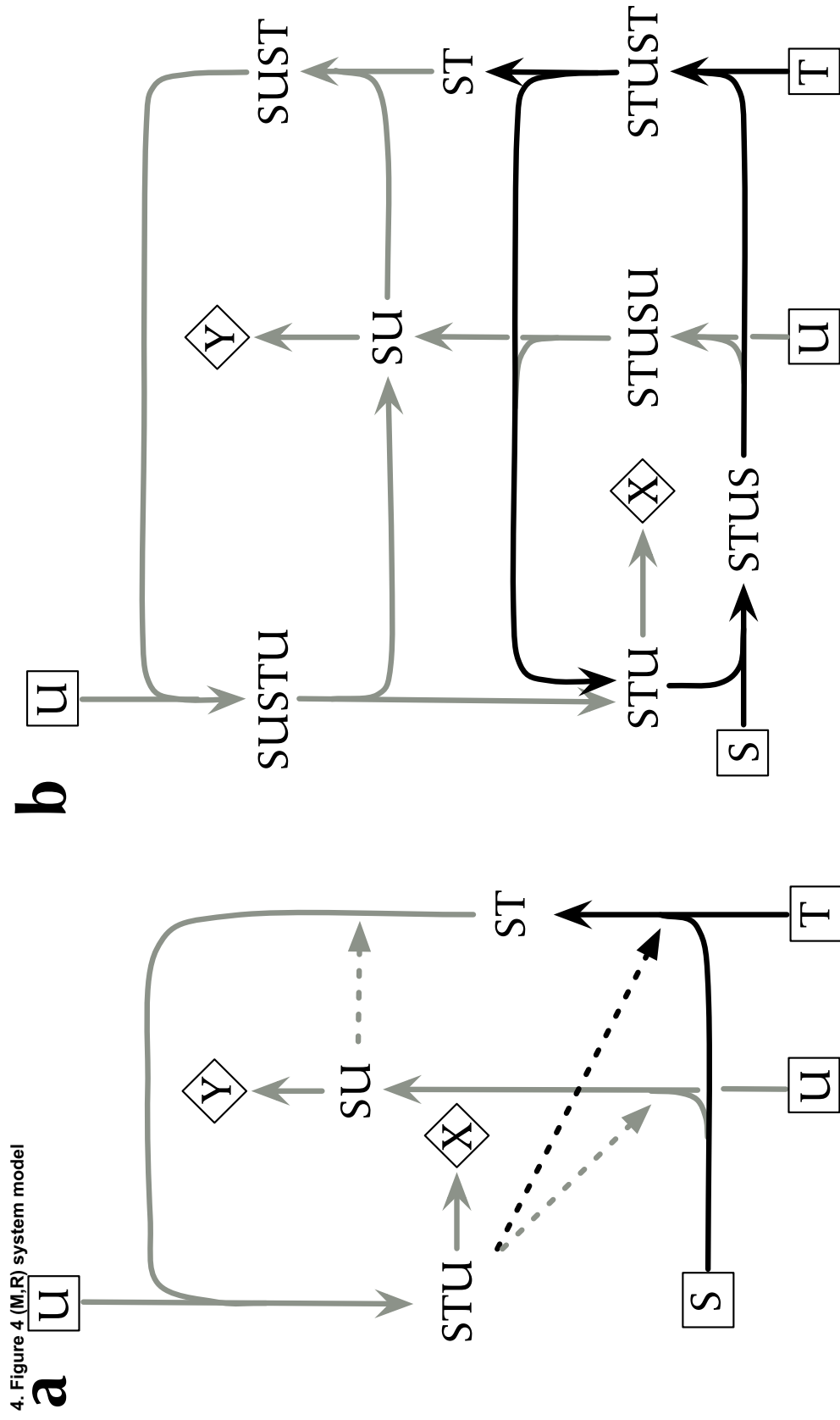
4. Figure 3 Chemero-Turvey models



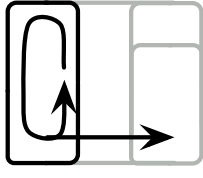
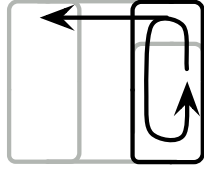
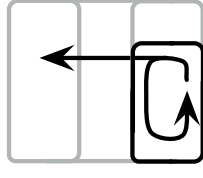


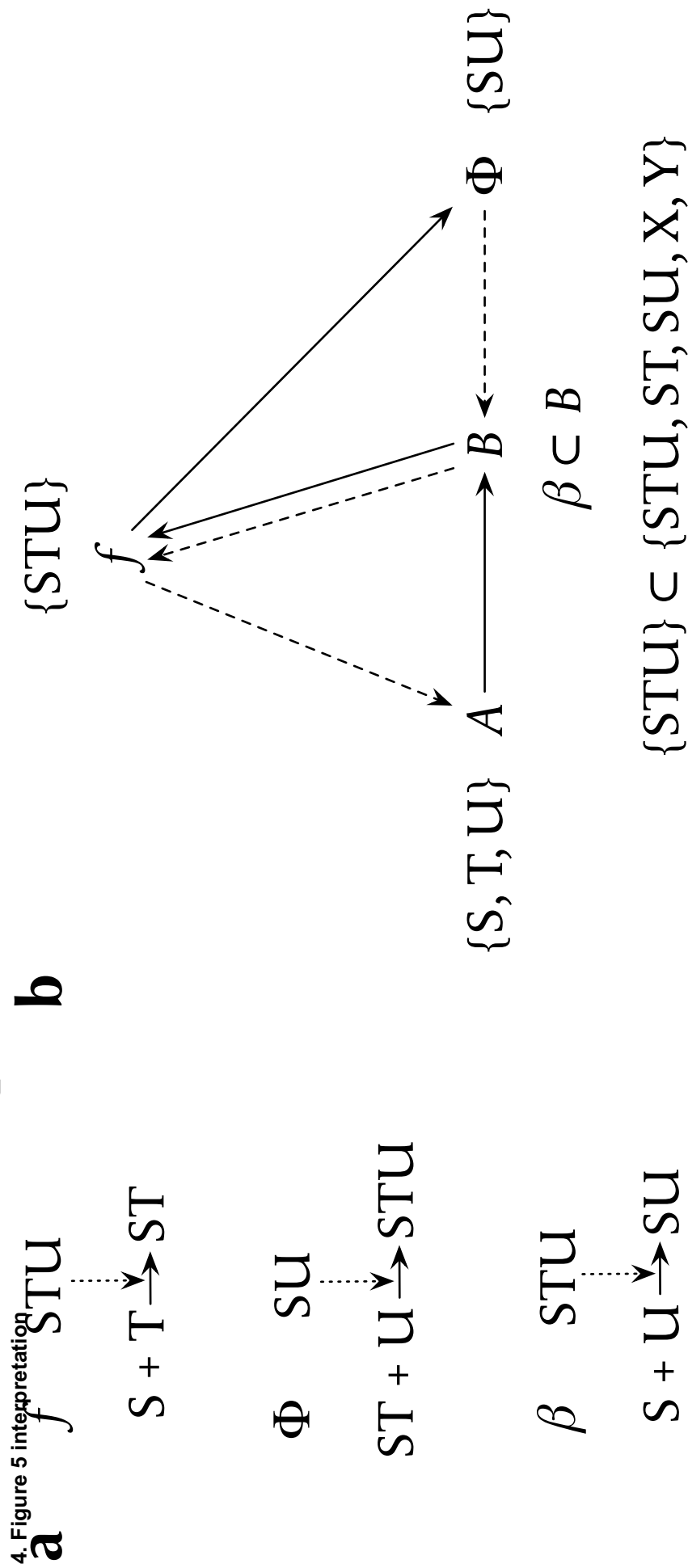
4. Figure

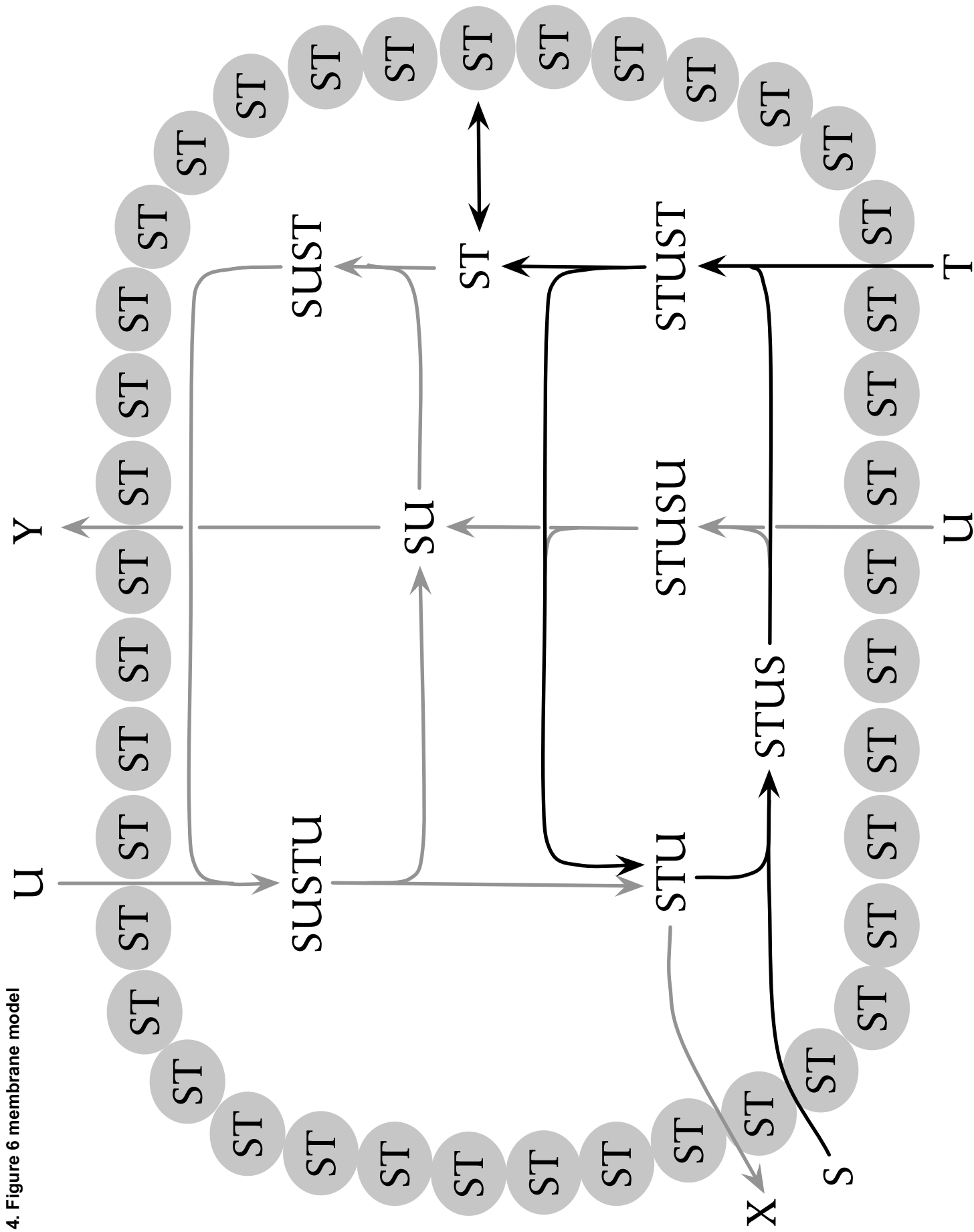




4. Figure 4 (M,R) system model

**C** *Replacemen**Metabolism**Organization invariance*





4. Figure 6 membrane model