

(M,R) Systems and RAF Sets: Common Ideas, Tools and Projections

S. Jaramillo¹, R. Honorato-Zimmer¹, U. Pereira¹, D. Contreras¹, B. Reynaert¹, V. Hernández¹,
J. Soto-Andrade¹, M.L. Cárdenas², A. Cornish-Bowden² and J.C. Letelier¹

¹ Facultad de Ciencias, Universidad de Chile, Casilla 653, Santiago, Chile.

² Bioénergétique et Ingénierie des Protéines, CNRS, Marseille, France.

letelier@uchile.cl

Abstract

There are deep underlying similarities between Rosen's (M,R) systems as a definition of life and the *RAF sets* (Reflexive Autocatalytic systems generated by a Food source) introduced by Hordijk and Steel as a way of analyzing autocatalytic sets of reactions. Using RAF concepts we have systematically explored the set of possible small idealized metabolic networks, searching for instances of (M,R) systems. This exhaustive search has shown that the central requirement of Rosen's framework, unicity of Φ , becomes harder and harder to obtain as the network grows in size. In addition, we give an expression for operators f , Φ and β in terms of RAF sets.

Introduction

Metabolic closure is easy to introduce *informally* but rather difficult to define. Although it is crucial for understanding living organization it was neglected until late in the 20th century. The rebirth of the scientific study of biological organization can be traced back to the 30-year period from 1958 to 1987, which saw the publication of several distinct perspectives on closure, including (M,R) systems (Rosen, 1958), the chemoton (Gánti, 1975), hypercycles (Eigen and Schuster, 1977), autopoiesis (Maturana and Varela, 1980), autocatalytic sets (Kauffman, 1986), and the first Artificial Life conference in Los Alamos in 1987 (organized by Christopher Langton). There was, however, an almost complete lack of cross-fertilization between the different schools of thought, with each theory developed with almost no reference to any of the others (Letelier et al., 2006; Cornish-Bowden et al., 2007; Cárdenas et al., 2010). The most extreme case of isolation is represented by Robert Rosen (1934-1998), who introduced the concept of (M,R) systems early in his career to represent biological metabolic networks. His isolation was aggravated by the intricate nature of his writings, in which biological ideas were mixed with abstract mathematics. Furthermore, he expressed his mathematical ideas in non-standard notations and without any effort to help the reader by giving examples or offering many needed clarifications.

In recent years, we have undertaken a systematic attempt to understand and explain the core notions of Rosen's the-

ory (Letelier et al., 2006). We have (a) clarified the relationship between (M,R) systems and autopoiesis (Letelier et al., 2003); (b) reframed Rosen's original formulation in terms of biochemical networks, with the introduction of the notion of "organizational invariance" for understanding Rosen's elusive mathematical operators (such as his β); (c) made a clear distinction between (M,R) systems in general and (M,R) systems with organizational invariance, a notion that is only implicit in Rosen's writing (he confusingly called these "replicative" (M,R) systems); (d) given mathematical and biological examples of simple idealized systems that can be understood within Rosen's intellectual framework; (e) clarified how these notions can be used to explore the origin of living systems and how they should be used in the context of what has come to be called "systems biology". Finally, we have also shown how our formulation of (M,R) systems can shed light on the problem of the computability of living systems (Cárdenas et al., 2010). This short summary is intended simply to underline how fruitful Rosen's view of metabolic closure has become, and to explain why we feel that the boundaries of our knowledge can be pushed to qualitatively new grounds by continuing the exploration of his ideas.

The systematic absence of examples (whether mathematical or biological) from Rosen's work has always been problematical, especially of simple examples that can serve as heuristic devices for enhancing theoretical research. In this paper we address the two points outlined above by pointing out the close relationship between (M,R) systems and a recent theory of living organization based on what have been called RAF sets. We show how many examples of simple (M,R) systems can be found by a computer algorithm constructed on the model of RAF sets. We discuss how the technical tools originating in RAF sets can be used to enhance the research of (M,R) systems, and specifically we address the problem of the nature and unicity of Rosen's Φ in the context of RAF sets.

(M,R) systems

Rosen's original formulation of (M,R) systems (Rosen, 1958), relied on a view of metabolism as a graph, and on a very abstract view of enzymes as functions (in the mathematical sense). The metaphor of metabolism as a graph, new in 1958, has subsequently been adopted by many people, without attribution to Rosen. The view of enzymes as functions has not attracted a wide following as Rosen's formulation seems unnecessarily abstract, without bringing practical or theoretical benefits. He used this approach in order to be able to use *category theory* for framing his important intuition about metabolic closure. Although this demanding mathematical approach has some advantages, as described in our previous work, we shall not use it here as the fundamental ideas exposed by Rosen can be explained using set theory, and thereby become accessible to mainstream biologists.

Our analysis of (M,R) systems, together with our examples, shows that the crucial aspect to understand organizational invariance is to understand the nature of the equation

$$\Phi(b) = f$$

Here Φ represents the aspect of biological organization that relates how catalysts are produced by the system. This equation seems to imply that a living system is organized in such a way that knowing b (right-hand side of biochemical equations) should be enough to unambiguously assign the catalysts (represented by f) to the reactions in the network.

Rosen, moreover, requires that there be only way to carry out this assignment, i.e., that there is only one mapping Φ such that $\Phi(b) = f$, a demanding assumption indeed. In other words, that we can reverse the procedure that gives f back from Φ . The reverse procedure is Rosen's β , so that

$$\beta(f) = \Phi$$

Mathematically, β is just the inverse of the "evaluation at b " operator that evaluates every function at b . Biologically, β represents the mechanisms that specify how the process of creating catalysts is maintained over time, i.e., *organizational invariance*.

To clarify these notions, we created a small metabolic network where they can be embodied in actual molecules that implement the functions Φ and β (Letelier et al., 2006).

RAF sets

We now give a brief introduction to the work of Hordijk and Steel (2004), who constructed a formal framework to study autocatalytic systems. Their main aim appears to have been to expand Kauffman's formalism about autocatalytic sets (Kauffman, 1993), to respond the criticisms that arose out of Kauffman's assumptions. At the same time, their analysis developed interesting algorithms that handle this expanded

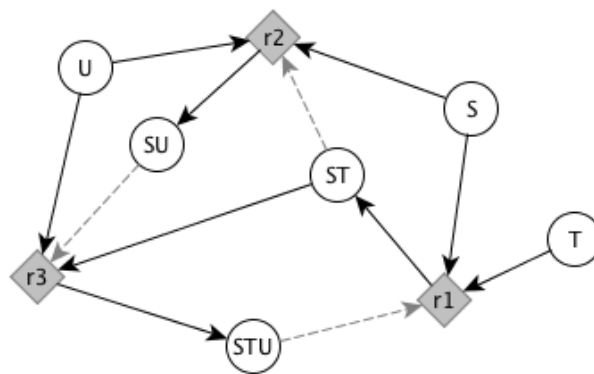


Figure 1: (M,R) system described by a *catalytic reaction graph*. Gray squares represent reactions and circles denote metabolites and enzymes. The black arrows represent chemical transformations while gray dashed arrows indicate catalyzations. This small network also contains a RAF set generated by the food set (S, T, U).

framework. As a result, they have produced a powerful approach that can be used to analyze a wide variety of systems, and here we shall describe how it applies to (M,R) systems. Their formalism depends on the following two sets: X , the set of molecules involved in metabolism as metabolites, catalysts or external input material (termed *food* in the formalism), and \mathcal{R} , the set of reactions that defines the metabolic network.

Each reaction r is represented as a tuple (A, B) , where $A, B \subset X$, $A \cap B = \emptyset$, A are the reactants and B the products of reaction r . This formalism is similar to Rosen's treatment of enzymes as transformations between two sets of molecules.

Further, to formalize the notion of catalysis, a specific set C (called the set of "catalyzations" by Hordijk and Steel), is introduced. Each catalyzation c is a tuple (x, r) , where $x \in X$ is the catalyst and $r \in \mathcal{R}$ is the reaction catalyzed by x . The similarity with Rosen (1958) is evident, as any given catalyzation $c = (x, r)$ can be rewritten as $c = (x, r) = (x, (A, B)) = (A, x, B)$, making transparent the fact that molecule x catalyzes the reaction $A \rightarrow B$.

With the set of catalyzations defined, Mossel and Steel (2005) introduced a function γ that helps to simplify formulae in later sections:

$$\gamma_C(A, r) = \begin{cases} 1 & \text{if } \exists x \in A : (x, r) \in C, \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Additionally, a specific subset of X containing every molecule that is used but not produced by the metabolism is denoted F and it represents the food molecules.

Thus a *catalytic reaction system* over a food source F is composed by a triplet $\mathcal{L} = (X, \mathcal{R}, C)$ that defines the universe of molecules (X), the reactions occurring among these

molecules (\mathcal{R}) and the identity of the catalyst involved in each reaction (C) (see Figure 1). The following additional functions are defined: $\rho(r) = A$ and $\pi(r) = B$, which return the reactants and the products of any given reaction r , respectively. With the help of these elementary functions the same notion can be extended to a *set of reactions* \mathcal{R}' as $\rho(\mathcal{R}') = \bigcup_{r \in \mathcal{R}'} \rho(r)$, where $\mathcal{R}' \subseteq \mathcal{R}$. This definition captures the conglomerate of molecules that participate as reactants for a set of reactions. A similar definition holds for $\pi(\mathcal{R}')$, the products of a subset of reactions. With these ideas, we can define the *closure* of a subset $X' \subseteq X$ relative to $\mathcal{R}' \subseteq \mathcal{R}$ ($cl_{\mathcal{R}'}(X')$) as the set of reachable molecules that can be synthesized by starting from X' and applying all the reactions in \mathcal{R}' until no new molecule types appear. Then, a non-empty reaction subset \mathcal{R}' of \mathcal{R} is a *reflexively autocatalytic network* over F if $\rho(\mathcal{R}') \subseteq cl_{\mathcal{R}'}(F)$ and for each $r \in \mathcal{R}'$, $\gamma(\rho(\mathcal{R}') \cup \pi(\mathcal{R}'), r) = 1$. In other words every catalyst must be produced by a reaction in the same system or be part of the food set. This definition allows many reflexively autocatalytic networks in a *catalytic reaction system*. The network is F -generated if every reactant is either produced by the system or incorporated as a food item (i.e. formally $\rho(\mathcal{R}) \subseteq F \cup \pi(\mathcal{R})$). A network that is *reflexively autocatalytic* and F -generated is called a *RAF set* (see Figure 1).

RAF sets can be understood informally as an interdependent set of biochemical reactions where all of the metabolites are produced by the collection of reactions \mathcal{R}' . The advantage of this formalism is that it is precise enough to be coded in well defined algorithms that check whether a given reaction subset $\mathcal{R}' \subseteq \mathcal{R}$ is a RAF set over some food set F . We have implemented these algorithms, and we have created a simple framework in Lisp and Python, allowing us to carry out qualitative and quantitative analyses of (M,R) systems in terms of RAF formalism. Before discussing this, however, we need to show the extent to which RAF sets and (M,R) systems are equivalent.

RAF sets and (M,R) systems

Are (M,R) systems RAF sets? The original definition of an (M,R) system (Rosen, 1958) explicitly requires every catalyst (M in his original symbols) must be produced by the metabolism (R sub-systems are responsible for this task). This condition shows that (M,R) systems must be *reflexively autocatalytic* (RA) sets. Although, this does not necessarily imply that a RA set is an (M,R) system, because metabolic closure requires that no catalyst is given in the food set. In other words, a RA set is not in general an (M,R) system, but it may become one if all the catalysts in C are produced by the system and are not part of the food set F .

As (M,R) systems must be open to the flow of matter in order to satisfy thermodynamic requirements, their molecules derive ultimately from a food source, and they are, obviously, F -generated in the terminology of RAF sets. So (M,R) systems without organizational invariance are a sub-

set of RAF sets, as are (M,R) systems with organizational invariance. The latter must, however, have additional features (in the context of RAF) to explain the unusual properties of operators Φ and β .

Algorithmic search for simple metabolic (M,R) systems

In this section we explore the probability of occurrence of an (M,R) system with a unique assignment of catalysts. For this purpose we characterized all the possible graphs describing a system consisting of a number $\#F$ of initial molecules and $\#\mathcal{R}$ synthesis reactions between any two molecules in the system. More specifically, we analyzed systems that conformed with the requirement of being (M,R) systems, that is, we did not allow any catalyst to be food, nor a reactant nor a product in the reaction it catalyzed.

Attention must be paid to avoid having two apparently distinct reaction networks exhibiting the same topological structure. The mathematical term for this is *graph isomorphism* (see Figure 3). Two graphs are said to be isomorphic when they can be transformed into each other by a simple relabeling of their vertices. Isomorphic metabolisms can be grouped under an equivalence class.

Thus, for a given pair $(\#F, \#\mathcal{R})$ we enumerate the number of all possible different equivalence classes of reaction networks. Next, for each one of these reaction networks, we generated the set of all possible assignments for the catalysts complying with the restrictions stated previously. But again, by the argument of relabeling, the set of assignments can be

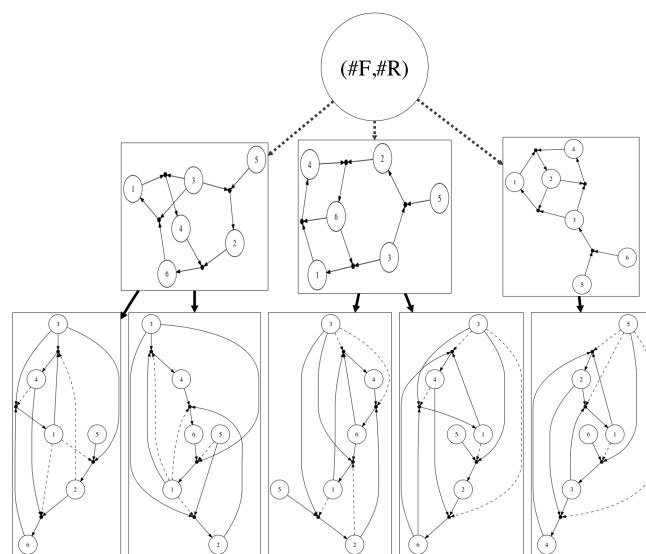


Figure 2: Diagram representing an example for the procedure to compute results from table 1. In the first step, the equivalence classes (3 in this example) are estimated for a given $(\#F, \#\mathcal{R})$; in the second step, all possible catalysts assignments for each equivalence class are calculated.

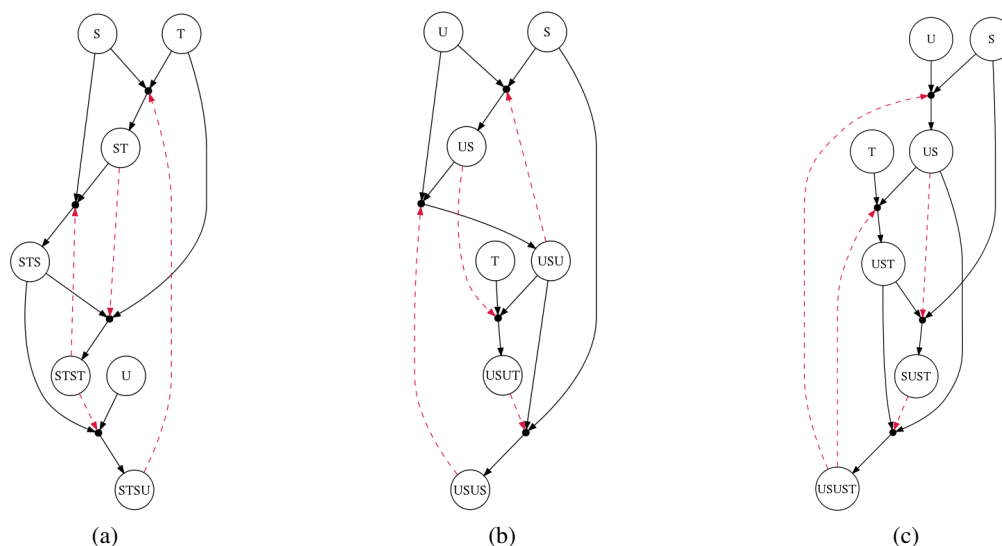


Figure 3: Three automatically generated RAF sets illustrating equivalence class and multiple catalyst assignments. Systems (a) and (b) have the same topological structure, i.e. there is an isomorphism from one to the other. Although this might not be obvious at first sight, a simple procedure of node relabeling transforms the reaction pathway in (a) to the one in (b). In spite of that, the systems differ in their catalyst assignments, i.e., even with the additional rules imposed by (M,R) systems, it is possible to make different choices when assigning the catalysts. System (c) has the same number of elements in the food set and the same number of reactions, but it belongs to another equivalence class.

also divided into equivalence classes (see Figure 2). Table 1 shows for $(\#F, \#\mathcal{R})$ the number of metabolic equivalence classes and the interquartile range¹ of the number of assignments. It can be seen that the number of possible assignments grows steeply with the number of reactions, so that it becomes more and more difficult to have a unique $\Phi(b) = f$ (Letelier et al., 2006).

There are some cases in which the range includes the critical value 1, which implies organizational invariance. Although, if we increase the number of food elements and leave the number of reactions unchanged, the generated reaction networks become shallower, and so we can consider the complexity of the network to be reduced and therefore the degrees of freedom of the assignment process are also reduced. In principle we could separate the trivial cases from those in which the unicity of the assignment reflects organizational invariance.

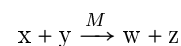
Rosen's triad in RAF formalism

The RAF formalism is not only useful for exploring the landscape of possible (M,R) systems, but it can also help to clarify some core concepts of (M,R) systems, namely Rosen's triad: f , Φ and β .

To explore the potential of the RAF formalism, we analyze the old problem in the theory of (M,R) systems of how

¹This refers to the range in which data falls after removing lower and upper 25%, thus giving a notion of the amplitude of the mean values

to treat molecules as functions. Consider the following biochemical reaction:



According to Rosen, this is the manifestation of the following function:

$$\begin{aligned} M &\in \text{Map}(X \times Y, W \times Z) \\ M &: X \times Y \rightarrow W \times Z \\ (x, y) &\rightarrow (w, z) \end{aligned}$$

The input elements are derived from the cartesian set $X \times Y$ that contains all the molecular types that, because of their structural similarities, can be used by the enzyme M as substrates. Our RAF-derived formalism extends the domain of function M to the whole set of molecules as follows: M is a function that, when given a set of molecules with the reactants, e.g. $(\dots, x, \dots, y, \dots)$, returns a set containing elements w and z . But if the original input set lacks elements x or y , we have $M(\text{input set}) = \emptyset$. Interestingly, with this formalism any molecule in the network ($x \in X$) can be treated as a function operating on any subset ($X' \subseteq X$) as follows:

$$x(X') = \pi(r_x) \text{ provided that } \rho(r_x) \subseteq X'$$

where r_x stands for the reaction that x catalyzes. If x catalyzes more than one reaction², then the above definition can

²This multifunctionality seems to be necessary for (M,R) systems (Letelier et al., 2006).

Number of food molecules	Number of reactions					
	3		4		5	
2	4	2-2	19	12-24	136	144-216
3	10	1-4	72	12-31	685	216-324
4	8	1-6	75	1-36	933	204-432
5	2	1-1	37	1-34	577	1-432
6	1	1-1	11	1-1	212	1-1

Table 1: Number of metabolic equivalence classes and the interquartile range of the number of their possible assignments. The number of equivalence classes increases dramatically with the number of reactions.

be generalized to:

$$x(X') = \{x_i : x_i \in \pi(r) \mid (x, r) \in C \wedge \rho(r) \subseteq X'\} \quad (2)$$

Note that defining x only requires the set of reactions each molecule catalyzes, not the whole reaction network. This means that every molecule-as-a-function definition depends only on local information.

In our earlier work, the following small metabolism was used as a testbed for exploring concepts related to (M, R) systems.



Then, treating every molecule as a function we have:

$$\begin{aligned} SU(S, T) &= \{ST\} \\ STU(S, U, T) &= \{SU\} \\ U(S, T, U, ST, STU) &= \emptyset \\ &\dots \end{aligned}$$

The last equation means that molecule U cannot transform the given mixture, because U is not a catalyst in the given metabolism. That said, we shall now analyze how concepts like f , Φ and β can be expressed with these ideas.

Metabolism: f

One of the basic equations in Rosen's model is $f(a) = b$, in which a represents the input materials (food set) needed by the organism to produce the complete set of metabolites and enzymes (b), i.e., every molecule reachable by the metabolism. Therefore, the function f is related to the notion of closure ($cl_{\mathcal{R}}(X')$). To be able to define f in our terms, let us define function $expand$.

$$expand_X(X') = X' \cup \bigcup_{x_i \in X} x_i(X') \quad (6)$$

Moreover, let us define how a molecule set (X') can be applied to another molecule set (Y').

$$\overrightarrow{X'}(Y') = \begin{cases} Y' & \text{if } expand_{X'}(Y') = Y', \\ \overrightarrow{X'}(expand_{X'}(Y')) & \text{otherwise} \end{cases} \quad (7)$$

Thus, we use a molecular set as a function (distinguished from regular molecular set by a "semi-arrow") by repeatedly applying $expand$ until no further additions occur. With these two last definitions, for any given catalytic reaction system $L = (X, \mathcal{R}, C)$, $f(a)$ can be defined as:

$$f(a) = \overrightarrow{catalysts(C)}(a) = b \quad (8)$$

where $catalysts$ is a function that returns every catalyst in the given catalyzation set C ($catalysts(C) = \{x : (x, r) \in C\}$). The function $catalysts$ is not required, as non-catalyst molecules do not modify the result. But it is used here as Rosen's formalism considers only catalysts as the core components of the metabolism.

Replacement: Φ

The formulation of Φ under RAF sets is more elaborate as we need to generate a function that using b as an input returns function f . The basic idea is to create mathematical objects that somehow keep track of which catalysts are produced and how these are created as a result of the metabolism. To begin we introduce operator Op . This operator returns the subset of molecules $X'' \subseteq X'$ that can act as catalysts upon the molecules in X' (the given molecule set).

$$Op(X') = \{x \in X' : x(X') \neq \emptyset\}$$

Then, for any given catalytic reaction system $L = (X, \mathcal{R}, C)$ over a food source F , $\Phi(b)$ will be defined as

$$\Phi(b) = \overrightarrow{Op}(cl_{\mathcal{R}}(b) \cup F) = f' \quad (9)$$

where $cl_{\mathcal{R}}(b)$ is the closure of b relative to the reaction set \mathcal{R} as defined above. Therefore, Φ returns the catalyst set that are reachable from b as a function (f'), because the "semi-arrow" over the expression transforms the resulting set into a function. Thus, f' is operationally equivalent to function f .

Organizational invariance: β

Finally, it remains to define β , which should take the metabolism f as input and give us the replacement system Φ . The function β receives a hypothetical metabolism f' in the form of a function, thus our first step will be to find which catalysts can be related to that function f' . For that purpose, let us define the function ν that given a molecular set b and a function f' , returns every reaction catalyzed by molecules in b , which produces part of the result of f' applied to F .

$$\nu(b, f', F) = \{r : \gamma(b \cup F, r) = 1\}$$

By using a new function μ , we filter out those reactions that cannot take place given the molecule set of interest ($b \cup F$).

$$\mu(b, f', F) = \{r \in \nu(b, f', F) : \rho(r) \subseteq b \cup F\} \quad (10)$$

This equation gives the reactions that are related to f' , therefore β can be defined. For simplicity we shall define it as applied to a molecular set b .

$$\beta(f')(b) = \overrightarrow{Op}_{\mu(b, f', F)}(b \cup F) \quad (11)$$

This formula is similar to that of Φ , the main difference being that it uses function μ to obtain \mathcal{R} instead of using \mathcal{R} directly. In this way β returns a function that, used in an (M, R) system, would relate unequivocally to Φ .

Conclusion

A formidable challenge for using (M, R) systems as a framework for modeling biological systems has been the lack of operational definitions for the important functions f , Φ and β . Here we have presented various definitions for those functions that can be used for any catalytic reaction system.

An important unresolved matter is to make explicit how Rosen's equations can be fulfilled using concepts and definitions imported from RAF sets. Suppose that a given molecule set X and reaction set R compose an (M, R) system, how can that be proved using RAF-derived functions? First, let us distinguish a particular subset a of X , which contains every molecule that is not a product or a catalyst for any reaction. Then, we can write:

$$f(a) = b$$

This signifies "let the molecular system evolve until no further novelty can be produced". Now, we should expect that using the produced molecules as function will have the same effect as using f . In our terms, that means:

$$\Phi(b)(a) = b$$

This has the important consequence that f becomes equivalent (operationally) to $\Phi(b)$ in this molecular system.

β , as introduced here, does not explain Rosen's basic result ($\beta(f) = \Phi$, which means that Φ is uniquely determined by f). The definition of β and all associated formulae cannot explain Rosen's result, they merely serve as formal language that could help us to operate on modern metabolic data using Rosen's viewpoint.

Since the beginning of the 21st century there has been a resurgence of interest in the work of Robert Rosen, but it is not easy to understand and it is not apparent how to advance in a theory full of powerful but often obscure ideas (Letelier et al., 2006). Many attempts have been made to find the route to be followed in developing the theory (Wolkenhauer and Hofmeyr, 2007). Here we apply another formalism (RAF sets) that could be useful for clarifying the nature and properties of the operators f , Φ and β .

Finally, we have the caveat that living systems are not mere "soups of letters", and their complex properties are due to more than some combinatorics among molecules. It is apparent that to advance in our understanding of living organisms, it will be necessary to include further considerations into our current theory. These could be geometrical, thermodynamical, topological, or even merely historical, that is, relative to how life has come into existence, and later evolved here on Earth.

The RAF formalism may usher in an era in which the theory of (M, R) systems will demand reasoning tools that begin to resemble category theory more and more... Rosen would be amused!

References

- Cárdenas, M. L., Letelier, J. C., Gutierrez, C., Cornish-Bowden, A., and Soto-Andrade, J. (2010). Closure to efficient causation, computability and artificial life. *J. Theor. Biol.*, 263:79–92.
- Cornish-Bowden, A., Cárdenas, M. L., Letelier, J. C., and Soto-Andrade, J. (2007). Beyond reductionism: metabolic circularity as a guiding vision for a real biology of systems. *Proteomics*, 7:839–845.
- Eigen, M. and Schuster, P. (1977). A principle of natural self-organization. *Naturwissenschaften*, 64(11):541–565.
- Gánti, T. (1975). Organization of chemical reactions into dividing and metabolizing units: the chemotons. *Bio Systems*, 7(1):15–21.
- Hordijk, W. and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *J. Theor. Biol.*, 227:451–461.
- Kauffman, S. (1993). *The Origins of Order*. Oxford University Press, New York.
- Kauffman, S. A. (1986). Autocatalytic sets of proteins. *J. Theor. Biol.*, 119(1):1–24.
- Letelier, J. C., Marín, G., and Mpodozis, J. (2003). Autopoietic and (M, R) systems. *J. Theor. Biol.*, 222:261–272.

- Letelier, J. C., Soto-Andrade, J., Guíñez Abarzúa, F., Cornish-Bowden, A., and Cárdenas, M. L. (2006). Organizational invariance and metabolic closure: analysis in terms of (M,R) systems. *J. Theor. Biol.*, 238:949–961.
- Maturana, H. and Varela, F. (1980). *Autopoiesis and Cognition: the Realisation of the Living*. D. Reidel Publishing Company, Dordrecht, The Netherlands.
- Mossel, E. and Steel, M. (2005). Random biochemical networks: the probability of self-sustaining autocatalysis. *J. Theor. Biol.*, 233(3):327–336.
- Rosen, R. (1958). A relational theory of biological systems. *Bull. Math. Biophys.*, 20:245–260.
- Wolkenhauer, O. and Hofmeyr, J.-H. (2007). An abstract cell model that describes the self-organization of cell function in living systems. *J. Theor. Biol.*, 246:461–476.