# A NEW METHOD FOR THE BINARY ENCODING AND HARDWARE IMPLEMENTATION OF METABOLIC PAHTWAYS

## Carlos Recio Rincon, Paula Cordero, Juan Castellanos, Rafael Lahoz-Beltra

**Abstract**: In this paper we introduce a new method for the binary encoding of metabolic pathways. Our method assigns a 5-bit word to the functional groups of the molecules or metabolic intermediates, sorting the functional groups by its redox potential. We illustrate our approach modelling two very well known metabolic pathways, glycolysis and the Krebs cycle, showing how sugars and other glycolytic molecules could be modeled as binary matrices as well as LED dot matrices. The method enables the design of 'metabolic hardware' which may be useful in the study of the optimization of metabolic pathways as well as in the area of molecular and natural computing.

**Keywords**: molecular topology representation, computational chemistry, biosinpired architectures, molecular computing.

ACM Classification Keywords: F.1. Computation by abstract devices; F1.1. Models of computation

#### Introduction

Metabolic pathways are series of biochemical reactions occurring within a cell. In each pathway an enzyme catalyzes a reaction transforming a molecule or substrate to a new molecule or product. Modelling and simulation of self-organization in metabolic pathways (Fig. 1) has been addressed under different approaches. In fact the simulation of biochemical reactions is related to the history of computers and biomathematics [Mendes and Kell, 1996]. For instance, in the context of game theory [Melendez-Hevia, 1990; Melendez-Hevia et al., 1994] introduced 'the game of the pentose phosphate cycle', a mathematical game that gives a simple explanation of how the metabolic reactions of the pentose cycle find an optimal configuration. The authors found an optimal solution to the problem of transforming six pentoses in five hexoses applying the principle of Darwinian natural selection [Lahoz-Beltra and Perales-Gravan, 2010] and the simplicity theorem. The study of metabolic pathways is also required for the design of virtual cellular systems [Sipper, 1990; Takahashi et al., 2002] e.g. MCell, VCell and E-Cell, as well as in the field of molecular and natural computation [Stefanovic, 2008].

In all these studies, there are two main 'ingredients', metabolites or metabolic molecules and enzymes (Fig. 1). However, taken together these two components enzymes have received much attention since they are very important in the metabolic pathways by enabling the biochemical reactions take place to a reasonable speed. From a historical perspective, the application of the theory of finite automata has enabled the modelling of important biological molecules such as proteins (enzymes are a particular type of proteins). During the decade of the 90s several researchers in the area of molecular computing considered the possibility that future computers arise based upon an architecture composed of proteins [Hameroff et al., 1992]. However, and although there is currently a lack of studies on protein-inspired computers we believe that this is a promising field that will give interesting results in the future. To date most of the proposals has been based on studies with enzymes, proteins with catalytic function responsible for the thousands of chemical reactions that sustain life on Earth. In the scientific literature are described theoretical models [Birge, 1995; Bray, 1995] with no practical implementations as well as experimental devices using real enzymes [Hiratsuka et al., 1999].

21

Theoretical models of proteins and enzymes can be classified into two main groups. On the one hand, some models assume an analogy between an enzyme and a transistor, due to the computational and electronic characteristics of enzymatic processing. For instance, [Di Paola et al., 2004] proposed a model wherein the bacterial chemotaxis proteins are implemented as hardware using an operational amplifier (Fig. 2). On the other hand, there are a number of models that assume that an enzyme is a finite automaton with two or more states which correspond to conformational states of the enzyme. For example, [Marijuan, 1991] introduced a probabilistic model of an enzyme with its state table and transition probabilities. Within this group there are also models in which an enzyme is considered as a McCulloch-Pitts neuron [Okamoto et al., 1999; Di Paola et al., 2004]. In other instances it is possible to find models of protein assemblies. For instance, [Lahoz-Beltra et al., 1993] introduced a model bio-inspired in the microtubules of cellular cytoskeleton showing the possibility of molecular computation via Boolean operations in microtubules. In microtubules protein subunits are assembled and behaving according the theory of coherent excitations introduced by [Frohlich, 2012] and in consequence like automata which conformational changes occurring in an orchestrated fashion. Moreover, in 2008 [Lin and Chen, 2008] developed evolvable hardware bio-inspired in cytoskeleton. Therefore, in the models of this second group a network of proteins or enzymes is a network of finite automata capable of performing Boolean operations. Based on this approach [Lahoz-Beltra, 2001] introduced a model of electronic enzyme (Fig. 3) which is under Spanish patent [Lahoz-Beltra, 2003].



**Figure 1.** (Left) Metabolic pathways are series of biochemical reactions occurring within a cell. In each pathway an enzyme  $E_m$  (lines or edges) catalyzes a reaction transforming a molecule or substrate  $S_m$  (vertices or nodes) to a new molecule or product  $P_m$  (vertices or nodes). A set of metabolic pathways is called a metabolic network, e.g. glycolysis and the Krebs cycle [Source of the biochemical circuit diagram: Molecular Biology of the Cell. Alberts et al. Fourth edition]. (Right) Biochemical reaction:  $S_m + E_m \rightarrow P_m$ .



Figure 2. Modelling and simulation of bacterial taxis using operational amplifiers [Di Paola et al., 2004]

In this paper we did not study the computational role of enzymes in metabolic pathways, but we explored the possibility of using metabolic networks as hardware in the field of molecular and natural computing. We call to these biosinpired architectures as *metabolic hardware*. In particular, adopting as an example the intermediate molecules of two very well known metabolic pathways, glycolysis and the Krebs cycle, we introduce the methodology to translate the molecular structure or topology of their metabolic intermediates to a binary matrix.

#### Methods

From a historical perspective one of the first procedures to translate the molecular topology to a matrix was introduced by [Randic, 1974], taking an element  $a_{ij}$  the value 1 when the vertices are adjacent or 0 otherwise. Figure 4 illustrates an example of this method for vitamin A or retinol [Lahoz-Beltra, 2012].

Our method assigns a 5-bit word to the functional groups of the molecule. For that purpose we define a table or *Rosetta stone* (Table I) that includes the most frequent functional groups in metabolic intermediates, which were ordered by its redox potential (tendency of a functional group to acquire electrons).



**Figure 3.** An enzyme  $E_m(c_{1j}, c_{2j},..., c_{nm}, o_{1j}, o_{2j},..., o_{nm})$  has been defined as an automaton with a finite number of internal 'conformational' states represented by an *n*-bit word  $c_{1j}, c_{2j},..., c_{nm}$  and a set of operations or instructions modelling the 'active groups' of the active site and given by  $o_{1j}, o_{2j},..., o_{nm}$  Boolean operators (e.g. AND, XOR). We define an enzymatic reaction as  $S_m + E_m \rightarrow P_m$  where  $s_{1j}, s_{2j},..., s_{nm}$  and  $p_{1j}, p_{2j},..., p_{nm}$  are the *n*-bit words representing the substrate  $S_m$  and product  $P_m$  respectively of the enzymatic reaction performed by enzyme  $E_m$ . Based on above definitions the electronic enzyme 'catalyzes' a biochemical reaction conducting the Boolean operations given by:  $p_{1j} = s_{1j} o_{1j} c_{1j}, p_{2j} = s_{2j} o_{2j} c_{2j},..., p_{nm} = s_{nm} o_{nm} p_{nm}$  [Lahoz-Beltra, 2001].

steczki witaminy A:



to cząsteczka ta może być wyrażona za pomocą następującej macierzy:

	,													
Í	0	1	0	0	0	1	0	0	0	0	0	0	0	0
	1	0	1	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	1	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	1	0	1	0	0	0	0	0	0	0	0
	1	0	0	0	1	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	1	0	1	0	0	0	0	0	0
	0	0	0	0	0	0	1	0	1	0	0	0	0	0
	0	0	0	0	0	0	0	1	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	1	0	0	0
	0	0	0	0	0	0	0	0	0	1	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	1	0	1	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	1	0

Przyjmujemy dla  $x_{ij}$  wartość 1, jeżeli atomy *i*, *j* są między sobą związane, i wartość 0 w przypadku przeciwnym, to znaczy gdy wiązanie między nimi nie istnieje.

87

**Figure 4.** The molecule of vitamin A or retinol represented as a binary matrix [Lahoz-Beltra, 2012] (Transl: Polish).

Let S and P be two binary matrices which represent respectively the substrate  $S_m$  and product  $P_m$  of a biochemical reaction catalyzed by an enzyme  $E_m$ . Since that glycolysis and Krebs cycle all metabolites or metabolic intermediates are molecules of 3, 4, 5 or 6 carbon atoms, we will define the (1), (2), (3) and (4) matrices respectively:

$$C_{3} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \end{pmatrix}$$
(1) 
$$C_{4} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$
(3) 
$$C_{6} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{pmatrix}$$
(4)

Note that given a value *i*,  $(a_{i1} a_{i2} \dots a_{i5})$  is a row vector representing the functional group of the substrate  $s_{ij}$  or product  $p_{ij}$  molecules. Thus, each row in the matrices C<sub>3</sub>, C<sub>4</sub>, C<sub>5</sub> and C<sub>6</sub> represents a carbon atom in the molecule, having a total of 32 possible binary vectors from 00000 to 11111 (Table I). Using as a criterion the redox potential vectors were classified from its most reduced (addition of hydrogen or the removal of oxygen) form or alkyl group to the most oxidized (addition of oxygen or the removal of hydrogen) or CO<sub>2</sub>. However, since the metabolites of glycolysis and the Krebs cycle are the result of assembling functional groups among a total of 22 combinations of carbon, then 10 binary vectors are without chemical meaning. In order to perform future simulation experiments, molecules of CO<sub>2</sub> and acetyl-CoA were represented as a row vector (5) and 2x5 matrix (6) shown below:

$$CO_2 = \begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$$
 (5)  $acetyl-CoA = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$  (6)

#### Results

Applying the technique described above sugars and other glycolytic molecules were modeled as binary matrices as well as hardware (Fig. 5). The hardware representation was conducted implementing molecules as LED dot matrices using CEDAR Logic Simulator program [Sprague, 2007]. The route of glycolysis was modeled as shown below:

(1	1	0	0	0)	)	(1)	1	0	0	0)	)	(1	0	0	0	0)		(1	0	1	0	0)	
1	0	0	0	1		1	0	0	0	1		1	1	0	1	1		1	1	0	1	1	
1	0	0	1	0		1	0	0	1	0		1	0	0	1	0		1	0	0	1	0	
1	0	0	0	1		1	0	0	0	1		1	0	0	0	1		1	0	0	0	1	
1	0	0	0	1		1	0	0	0	1		1	0	0	0	1		1	0	0	0	1	
1	0	0	0	0	$\rightarrow$	1	0	1	0	0	$\rightarrow$	1	0	1	0	0	$\rightarrow$	1	0	1	0	0)	$\rightarrow$

 $\begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}$ 

representing each matrix the following metabolites of glycolysis:

Glucose  $\rightarrow$  Glucose-6-phosphate  $\rightarrow$  Fructose-6-phosphate  $\rightarrow$  Fructose-1,6-biphosphate  $\rightarrow$ 

Glyceraldehyde-3-phosphate  $\rightarrow$  1,3-biphosphateglycerate  $\rightarrow$  3-phosphoglycerate  $\rightarrow$  2-phosphoglycerate

 $\rightarrow$  phosphoenolpyruvate  $\rightarrow$  pyruvate

Using the same method the Krebs cycle was modeled as follows:

(	1	1	1	0	0)		(1	1	1	0	0)		(1)	1	1	0	0)		(1	1	1	0	0)	
	0	1	0	0	1		0	1	0	0	1		0	1	0	0	1		0	1	0	0	1	
	1	0	0	1	1_→		0	1	0	1	$0_{\rightarrow}$		0	1	0	0	1		0	1	0	0	1	
	0	1	0	0	1		1	0	0	1	0		1	1	0	1	1		1	1	1	0	1)	
	1	1	1	0	0)		(1	1	1	0	0 )		1	1	1	0	0							
													\ <del>•</del>		-		~ /							
←	(1	1	1	0	0)	$\rightarrow$	← (1	1	1	0	0)	$\rightarrow$	(-		-		-)	$\rightarrow$				$\downarrow$		
÷	(1	1 1	1 1	0 0	0) 0`	$\rightarrow$	(1)	1	1 1	0 0	0) 0`	$\rightarrow$	(1	1	1	0	0)	$\rightarrow$	(1	1	1	↓ 0	0	,
÷	(1)	1 1 1	1 1 0	0 0 1	0) 0` 1	$\rightarrow$	(1)	1 1 1	1 1 0	0 0 0	0) 0` 1	$\rightarrow$	$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$	1	1	0 0	0) 1	$\rightarrow$	$\begin{pmatrix} 1\\ 0 \end{pmatrix}$	1	1 0	↓ 0 0	0 <sup>°</sup> 1	,
ć	$\begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}$	1 1 1 1	1 1 0 0	0 0 1 0	0) 0 <sup>×</sup> 1 1	$\rightarrow$	(1)	1 1 1 0	1 1 0 0	0 0 0 1	0) 0` 1 0	$\rightarrow$	$\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$	1 1 1	1 1 1	0 0 0	0) 1 1	$\rightarrow$	$\begin{pmatrix} 1\\ 0\\ 0 \end{pmatrix}$	1 1 1	1 0 0	↓ 0 0	0 <sup>°</sup> 1 1	

where each matrix stands for one of the following metabolites:

Citrate  $\rightarrow$  Iso-citrate  $\rightarrow \alpha$  -Ketoglutarate  $\rightarrow$  Succinyl-CoA  $\rightarrow$ 

Oxalacetate ← Malate ← Fumarate ← Succinate

Note that we have used special notation for citrate and iso-citrate matrices, because the third carbon atom is bonded to three others (Fig. 6).

27



**Figure 5.** Glucose molecule (Left) and its hardware version as a matrix of LEDs (Right) simulated with CEDAR Logic Simulator.



Figure 6. Citrate molecule

## Conclusion

This paper presents a novel method to represent the topology of a molecule as a binary matrix. The method enables the design of 'metabolic hardware', developing an example with two well-known metabolic pathways, glycolysis and the Krebs cycle. In our opinion the binary representation of molecules or metabolites is a first step that will lead in a future to study the metabolic pathways in search of bioinspired architectures with special interest in the field of molecular and natural computing.

Decimal	Binary	"Func	Red-Ox scale				
0 1 2 3	000 00 000 01 000 10 000 11		Null VALUES				
4 5 6 7	001 00 001 01 001 10 001 11		NULL VALUES				
8 9 10 11	010 00 010 01 010 10 010 11	СН <sub>2</sub> - СН <sub>2</sub> - ÇН- Ç-	Alkyle	+ Red - Ox			
12 13 14 15	011 00 011 01 011 10 011 11	=CH <sub>2</sub> =CH- =¢- =C=	Alkene				
16 17 18 19	100 00 100 01 100 10 100 11	−Сн₂Он н-¢-Он © но-¢-н № -¢-Он	Alcohol				
20 21 22 23	101 00 101 01 101 10 101 11	-сн,о-® н-ҫ҅-о-® =с҅-о-® -ҫ҅-о-®	Ester				
24 25 26 27	11000 11001) 11010} 11010}	$-C_{H}^{O}$ $\longrightarrow Null values$ $-C-$	Carbonyle (aldehide, ketone)				
28 29 30 31	111 00 111 01 111 10 111 10 111 11	- C <sup>0</sup> → - C <sup>0</sup> OH 0 → - C <sup>0</sup> O S - C <sup>0</sup> O P CO2	Carboxile, Ether, CO <sub>2</sub>	-Red Ox			

TABLE I. Rosetta stone for the hardware implementation of metabolic pathways

### Bibliography

[Birge, 1995] R.R. Birge. 1995. Protein-based computers. Scientific American 272: 66-71.

[Bray, 1995]. D. Bary. 1995. Protein molecules as computational elements in living cells. Nature 376: 307-312.

- [Di Paola et al., 2004] V. Di Paola, P.C. Marijuan, R. Lahoz-Beltra. 2004. Learning and evolution in bacterial taxis: an operational amplifier circuit modelling the computational dynamics of the prokaryotic 'two component system' protein network. Biosystems 74: 29-49.
- [Frohlich, 2012] H. Frohlich. 2012. Coherent Excitations in Biological Systems. Proceedings in Life Sciences. Springer-Verlag Berlin and Heidelberg GmbH & Co. KG: 1-240.
- [Hameroff et al., 1992] S.R. Hameroff, J.E. Dayhoff, R. Lahoz-Beltra, A.V. Samsonovich, S. Rasmussen. 1992. Models for molecular computation: conformational automata in the cytoskeleton. IEEE Computer 25: 30-39.
- [Hiratsuka et al., 1999] M. Hiratsuka, T. Aoki, T. Higuchi. 1999. Enzyme transistor circuits for reaction-diffusion computing. IEEE Transactions on Circuits and Systems – I 46: 294-303.
- [Lahoz-Beltra and Perales-Gravan, 2010] R. Lahoz-Beltra, C. Perales-Gravan. 2010. A survey of nonparametric tests for the statistical analisis of evolutionary computational experiments. International Journal Information Theories and Applications 17: 49-61.
- [Lahoz-Beltra et al., 1993] R. Lahoz-Beltra, S.R. Hameroff, j.E. Dayhoff. 1993. Cytoskeletal logic: a model for molecular computation via Boolean operations in microtubules and microtubule-associated proteins. BioSystems 29: 1-23.

[Lahoz-Beltra, 2001] R. Lahoz-Beltra. 2001. Evolving hardware as model of enzyme evolution. BioSystems 61: 15-25.

- [Lahoz-Beltra, 2003] R. Lahoz-Beltra. 2003. Enzima electronico, con sitio activo reconfigurable. Patente No. P200000174 (27.01.2000). Oficina Española de Patentes y Marcas. http://invenes.oepm.es/InvenesWeb/ detalle?referencia=P200000174.
- [Lahoz-Beltra, 2012] R. Lahoz-Beltra. 2012. Matematica zycia. Modele matematyczne w biologii i ekologii. RBA (Transl: Polish): 1-159.
- [Lin and Chen, 2008] Y-H. Lin, J-C. Chen. 2008. Neuromolecularware A bio-inspired evolvable hardware and its application to medical diagnosis. In: ARC 2008 (R. Woods et al., Eds.), LNCS 4943: 324-329.
- [Marijuan, 1991] P.C. Marijuan. Enzymes and theoretical biology: sketch of an informational perspective of the cell. BioSystems 25: 259-273.
- [Melendez-Hevia et al., 1994] E. Melendez-Hevia, T.G. Waddell, F. Montero. 1994. Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle. Journal of Theoretical Biology 166: 201-220.
- [Melendez-Hevia, 1990] E. Melendez-Hevia. The game of the pentose phosphate cycle: a mathematical approach to study the optimization in design of metabolic pathways during evolution. Biomed. Biochem. Acta 49: 903-916.
- [Mendes and Kell, 1996] P. Mendes, D.B. Kell. 1996. Computer simulation of biochemical kinetics. In: BioThermoKinetics of the living cell (H.V. Westerhoff et al., Eds.), BioThermoKinetic Press: 254-257.
- [Okamoto et al., 1999] M. Okamoto, T. Sekiguchi, K. Tanaka, Y. Maki, S. Yoshida. 1999. Biochemical neuron: harwdare implementation of functional devices by mimicking the natural intelligence such as metabolic control systems. Computers and Electrical Engineering 25: 421-438.
- [Randic, 1974] M. Randic. 1974. On the recognition of identical graphs representing molecular topology. The Journal of Chemical Physics 60: 3920-3927.
- [Sipper, 1990] M. Sipper. 1990. The emergence of cellular computing. IEEE Computer 32:18-26.
- [Sprague, 2007] B. Sprague, M. Lewellyn, D. Knierim, J. Lansford, N. Harro. 2007. CEDAR Logic Simulator. Cedarville University. http://sourceforge.net/projects/cedarlogic/
- [Stefanovic, 2008] D. Stefanovic. 2008. Emerging models of computation: directions in molecular computing. In: Software-Intensive Systems (M. Wirsing et al., Eds.) LNCS 5380: 255-265.

[Takahashi et al., 2002] K. Takahashi, K. Yugi, K. Hashimoto, Y. Yamada, C.J.F. Pickett, M. Tomita. 2002. Computational challenges in cell simulation: a software engineering approach. IEEE Intelligent Systems 17: 64-71.

## Authors' Information

30



**Carlos Recio Rincon** - Biologist, co-author of the course on CD-ROM entitled 'Histologia animal y vegetal' (2007, Editorial Complutense). His main interest is agriculture and ecology, *currently* works in '*rural* tourism'; e-mail: <u>carlosrecio@gmail.com</u>

**Paula Cordero** – Natural Computing Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail: paula.cormo@gmail.com

Major Fields of Scientific Research: complex systems, natural computing, evolutionary computation and the design of bioinspired algorithms.



Juan Castellanos – Head of Natural Computing Group, Facultad de Informática, Universidad Politécnica de Madrid, Campus de Montegancedo s.n., 28660 Boadilla del Monte, Madrid, Spain; e-mail: jcastellanos@fi.upm.es

Major Fields of Scientific Research: natural computing, formal language and automata theory.



**Rafael Lahoz-Beltra** – Department of Applied Mathematics, Faculty of Biological Sciences, Complutense University of Madrid, 28040 Madrid, Spain; e-mail: <u>lahozraf@ucm.es</u> Major Fields of Scientific Research: evolutionary computation, embryo development modeling and the design of bioinspired algorithms.