Mapping multivalued onto Boolean dynamics

Gilles Didier, Elisabeth Remy, Claudine Chaouiya

HAL Id: hal-00656334
https://hal.science/hal-00656334

Submitted on 4 Jan 2012

Author's Accepted Manuscript

Mapping multivalued onto Boolean dynamics

Gilles Didier, Elisabeth Remy, Claudine Chaouiya

www.elsevier.com/locate/yjtbi

Cite this article as: Gilles Didier, Elisabeth Remy and Claudine Chaouiya, Mapping multivalued onto Boolean dynamics, *Journal of Theoretical Biology*, doi:10.1016/j.jtbi.2010.09.017

Mapping multivalued onto Boolean dynamics

Gilles Didier[1], Elisabeth Remy[1], Claudine Chaouiya[2,3*]

[1]Institut de Mathématiques de Luminy, Marseille, France
[2]Instituto Gulbenkian de Ciência, Oeiras, Portugal
[3] Inserm U928 - TAGC, Marseille, France

September 8, 2010

Abstract

This paper deals with the generalized logical framework defined by René Thomas in the 70's to qualitatively represent the dynamics of regulatory networks. In this formalism, a regulatory network is represented as a graph, where nodes denote regulatory components (basically genes) and edges denote regulations between these components. Discrete variables are associated to regulatory components accounting for their levels of expression. In most cases, Boolean variables are enough, but some situations may require further values. Despite this fact, the majority of tools dedicated to the analysis of logical models are restricted to the Boolean case. A formal Boolean mapping of multivalued logical models is a natural way of extending the applicability of these tools.

Three decades ago, a multivalued to Boolean variable mapping was proposed by P. Van Ham. Since then, all works related to multivalued logical models and using a Boolean representation rely on this particular mapping. We formally show in this paper that this mapping is actually the sole, up to cosmetic changes, that could preserve the regulatory structures of the underlying graphs as well as their dynamical behaviours.

Keywords: Gene Regulatory Networks; Boolean modelling; Multivalued modelling; Cell dynamics

1

1 Introduction

The generalized logical modelling of regulatory networks, introduced by R. Thomas and collaborators [26, 25], was successfully applied to study networks controlling various biological processes including the segmentation of the *Drosophila* embryo (*e.g.* [20, 21, 5, 19]), the cell cycle in eukaryotes (see [7] for a review), the activation and differentiation of T lymphocytes [11, 12, 13, 18].

In brief, in this formalism, a regulatory network is represented as a graph where the nodes represent the regulatory components and the edges represent the regulations. Each regulatory component is associated to a discrete variable accounting for its functional level (of expression, concentration or activity). Observing that regulatory effects generally occur around specific thresholds, for each component, its associated (Boolean) variable corresponds to a discretization of its activity range, accounting for situations where the actual level of the component is below or above the threshold (respectively 0 or 1). Finally, the evolution of each component level is defined according to the levels of its regulators and an updating rule (synchronous, asynchronous, ...).

In most cases, Boolean variables adequately convey the role of regulatory components but, in some situations, this all-or-none description might be too crude; it may be necessary to consider further levels when, for example, a component does not act at similar levels on the components that it regulates, or when its effect on a component differs depending on its functional level. To take into account such cases, multivalued variables were introduced [25]. However, it should be noted that increasing the range of values taken by the component levels increases the complexity of the model and hence its analysis.

In [28], a Boolean mapping was introduced that amounts to associate, for each regulatory component, one Boolean variable to each threshold. This mapping was used to set up the generalized logical framework and to relate it to the differential description [28, 26, 23]. More recently, a number of theoretical works on methods to analyse logical models have been published. Most of these works focus on Boolean models, for instance concerning the relationships between regulatory circuits and dynamical properties (*e.g.* [3, 17, 16]), the delineation of the structures of the dynamics and their attractors (*e.g.* [8, 29, 22]) or the identification of functional modules [27]. A great literature is also devoted to Boolean networks, in the same vein as [9] and [10]. For a recent review, see [4] and the references herein. The work presented here might help to transpose the results obtained about Boolean models onto the multivalued case. Anyhow, a proper Boolean mapping allows the use of implementations of Boolean methods to analyse multivalued models.

In this paper, we focus on Boolean mappings of multivalued logical models. Various Boolean mappings have been considered in computer science, especially in the community working on propositional reasoning [2]. Notably, the Van Ham Boolean mapping has been adopted as the one that should be used in the case of logical models of regulatory networks. Here, we formally prove that it is indeed the case since this is the sole Boolean mapping, up to cosmetic changes (by switching 0 and 1 or by adding constant Boolean components), that could

preserve both the structures of the regulatory networks and their dynamical behaviours.

This paper is organized as follows. In Section 2, we introduce basic definitions and notations, notably the Van Ham Boolean mapping [28], we present a formal statement of the problem and we prove our main result. This states that the Van Ham mapping is essentially the sole which satisfies basic requirements in keeping dynamical and regulatory features. Section 3 intends to provide the reader with the intuition of this mapping and to discuss key related issues. For this purpose, we present the graph-based representations of the dynamics and illustrate the main points of the paper on a published model. The paper ends with some conclusions and prospects.

2 Formal results

2.1 Dynamics and regulatory networks

We note $\mathcal{G}$ the (finite) set of regulatory components and consider a $\mathcal{G}$-dimensioned vector $\mathbf{m}$ where $\mathbf{m}_g$ is a positive integer representing the maximum level which can be reached by g, for all $g \in \mathcal{G}$. In other words, the level of the component g takes its value in $\{0, \ldots, \mathbf{m}_g\}$.

A configuration of the system is an element of $\mathcal{C}(\mathcal{G}, \mathbf{m}) = \prod_{g \in \mathcal{G}} \{0, \ldots, \mathbf{m}_g\}$. The g-coordinate of a configuration $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ is noted x_g. The distance $\mathrm{d}(x, x')$ between two configurations x and x' is defined as $\mathrm{d}(x, x') = \sum_{g \in \mathcal{G}} |x_g - x'_g|$. We say that two configurations x and x' are *direct neighbours* if $\mathrm{d}(x, x') = 1$.

Let $g \in \mathcal{G}$, $\delta \in \{-1, 1\}$ and $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$. The notation $x^{(g,\delta)}$ designates the direct neighbour of the configuration x where the g-coordinate moves in the direction δ:

- $x_g^{(g,\delta)} = x_g + \delta$ if $0 \leq (x_g + \delta) \leq \mathbf{m}_g$,

- $x_g^{(g,\delta)} = x_g$ if $(x_g + \delta) < 0$ or $(x_g + \delta) > \mathbf{m}_g$,

- $x_{g'}^{(g,\delta)} = x_{g'}$ if $g' \neq g$.

A *dynamics* over a set of regulatory components of $\mathcal{G}$ is represented by an evolution function f from $\mathcal{D}_\mathrm{f} \subseteq \mathcal{C}(\mathcal{G}, \mathbf{m})$ to $\mathcal{D}_\mathrm{f}$. In practical applications, we have generally $\mathcal{D}_\mathrm{f} = \mathcal{C}(\mathcal{G}, \mathbf{m})$, that is f is defined over the whole configuration space, but it will not be necessarily the case for the evolution functions considered in this paper.

A particular case of such evolution function arises when the maximum levels vector is of the form $\mathbf{m} = (1, \ldots, 1) = \mathbf{1}$. All components take values in $\{0, 1\}$ and we deal with a so-called Boolean dynamics [16].

The *regulatory network* associated to f is the directed graph $\mathcal{R}_\mathrm{f}$ where the set of vertices is $\mathcal{G}$ and the edges are signed, *i.e.* labelled by $\{+, -\}$ and defined in the following way. Let g, g' be two components of $\mathcal{G}$ and $\sigma \in \{+, -\}$. We have (g, g', σ) in $\mathcal{R}_\mathrm{f}$ if there exists a configuration $x \in \mathcal{D}_\mathrm{f}$ with $x^{(g,1)} \in \mathcal{D}_\mathrm{f}$ and

3

such that $f(x^{(g,1)})_{g'} \neq f(x)_{g'}$ and σ is the sign of $(f(x^{(g,1)})_{g'} - f(x)_{g'})$. In other words, there is an edge from g to g' if f defines a regulatory effect of g onto g', that is the evolution (next level) of component g' in a given configuration depends on the level of component g in this configuration (see Section 3 for an illustration). Notice that it is possible to have simultaneously $(g, g', +)$ and $(g, g', -)$ in a same regulatory graph.

2.2 Boolean mappings of dynamics

The object of the paper is to explore generic ways to recode a set of multivalued configurations of $\mathcal{C}(\mathcal{G}, \mathbf{m})$ into a set of Boolean configurations included in $\mathcal{C}(\mathcal{G}', \mathbf{1})$, where $\mathcal{G}'$ is a set of regulatory components, in such a way that they allow to transform multivalued dynamics over $\mathcal{C}(\mathcal{G}, \mathbf{m})$ into Boolean dynamics over $\mathcal{C}(\mathcal{G}', \mathbf{1})$. This can be done in a very natural way when the recoding is one-to-one.

Let f be an evolution function defined over $\mathcal{D}_f \subseteq \mathcal{C}(\mathcal{G}, \mathbf{m})$. Under a one-to-one map b defined from $\mathcal{C}(\mathcal{G}, \mathbf{m})$ to $\mathrm{Im}_b(\mathcal{C}(\mathcal{G}, \mathbf{m})) \subseteq \mathcal{C}(\mathcal{G}_b, \mathbf{1})$, where $\mathcal{G}_b$ is a set of regulatory components, f induces its *conjugated evolution function* f_b over $\mathcal{D}_{f_b} = \mathrm{Im}_b(\mathcal{D}_f) \subseteq \mathcal{C}(\mathcal{G}_b, \mathbf{1})$ defined by:

$$f_b(y) = b \circ f \circ b^{-1}(y), \text{for all } y \in \mathcal{D}_{f_b}.$$

The definition of f_b can be represented by the following diagram:

$$
\begin{array}{ccc}
\mathcal{D}_f & \stackrel{f}{\longrightarrow} & \mathcal{D}_f \\
b \downarrow & & \downarrow b \\
\mathcal{D}_{f_b} & \stackrel{f_b}{\longrightarrow} & \mathcal{D}_{f_b}
\end{array}
$$

As for any evolution function, a regulatory network, noted $\mathcal{R}_{f_b}$, is associated to this conjugated evolution function f_b.

The following definition formalizes the Van Ham Boolean mapping, referred to in the text below as b_0.

Definition 1 *Let $\mathcal{G}$ be a set of components and $\mathbf{m}$ the vector of their maximum values. A set of $\mathbf{m}_g$ Boolean components $\{h_1^g, \ldots, h_{\mathbf{m}_g}^g\}$ is associated to each component g of $\mathcal{G}$. The set $\mathcal{G}_{b_0}$ is defined as $\bigcup_{g \in \mathcal{G}} \{h_1^g, \ldots, h_{\mathbf{m}_g}^g\}$. The map b_0 associates to each configuration x of $\mathcal{C}(\mathcal{G}, \mathbf{m})$, the configuration $b_0(x)$ of $\mathcal{C}(\mathcal{G}_{b_0}, \mathbf{1})$ which is such that, for all components g,*

- $b_0(x)_{h_w^g} = 1$ *for all* $1 \leq w \leq x_g$;

- $b_0(x)_{h_w^g} = 0$ *for all* $x_g < w \leq \mathbf{m}_g$.

Hence, we have $\sum_{w=1}^{\mathbf{m}_g} b_0(x)_{h_w^g} = x_g$. Figure 1 illustrates the mapping b_0. From its definition, b_0 associates each value v of the multivalued g-coordinate x_g of a configuration x to a unique configuration of the subset of Boolean components $(h_1^g, h_2^g, \ldots, h_{\mathbf{m}_g}^g)$ corresponding to g. The map b_0 is basically one-to-one. Note that, except when $\mathbf{m} = \mathbf{1}$, we have $\mathrm{Im}_{b_0}(\mathcal{D}_f) \neq \mathcal{C}(\mathcal{G}_{b_0}, \mathbf{1})$. Indeed some

4

configurations are not considered in the conjugated dynamics, that are such that $h_v^g < h_w^g$ with $1 \leq v < w \leq \mathbf{m}_g$. We will say that these configurations are not admissible.

x_g	$b_0(x)_{h_1^g}$	$b_0(x)_{h_2^g}$	$b_0(x)_{h_3^g}$	$b_0(x)_{h_4^g}$	$\ldots$	$b_0(x)_{h_{\mathbf{m}_g}^g}$
0	0	0	0	0	$\ldots$	0
1	1	0	0	0	$\ldots$	0
2	1	1	0	0	$\ldots$	0
3	1	1	1	0	$\ldots$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathbf{m}_g$	1	1	1	1	$\ldots$	1

Figure 1: Illustration of the Van Ham Boolean mapping b_0, showing the relationships between the multivalued variable x_g and its associated Boolean variables $h_1^g, h_2^g, \ldots, h_{\mathbf{m}_g}^g$. There are exactly $\mathbf{m}_g + 1$ configurations allowed and, for example, a configuration with $h_1^g = 0$ and $h_2^g = 1$ is not admissible (*i.e.* does not belong to $\mathcal{D}_{f_{b_0}}$).

2.3 Statement of the problem

In order to be relevant in a biological study, a conjugated Boolean dynamics has to be related to the initial one in such an amount that observations made over the Boolean system can be somehow interpreted in the initial one. We are thus interested in generic Boolean maps b for which the conjugated Boolean dynamics both keeps the basic topological properties of the initial multivalued dynamics and is such that the conjugated regulatory networks contain somehow the same information as that of the initial regulatory networks. After introducing some distinguished subsets of $\mathcal{G}_b$, we give the formal counterparts of keeping dynamics topological properties and regulatory features.

Let us consider a component $g \in \mathcal{G}$. We note $\Delta_{b,g}$ the set of components $h \in \mathcal{G}_b$ such that there exists a configuration $x \in \mathcal{D}_f$ with $b(x^{(g,1)})_h \neq b(x)_h$. Intuitively, the Boolean components of $\Delta_{b,g}$ are the ones which are used by b to follow the evolutions of the g-coordinate. For instance, with the notations of Definition 1, we have $\Delta_{b_0,g} = \{h_1^g, \ldots, h_{\mathbf{m}_g}^g\}$.

A map b is *neighbour-preserving* if for all configurations $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, all components $g \in \mathcal{G}$ and all $\delta \in \{-1, 1\}$, there exists a component $h \in \mathcal{G}_b$ and $\delta' \in \{-1, 1\}$ such that $b(x^{(g,\delta)}) = [b(x)]^{(h,\delta')}$. In other words, two direct neighbour configurations in $\mathcal{D}_f$ are mapped to two direct neighbour configurations in $\text{Im}_b(\mathcal{D}_f)$. For a neighbour-preserving map b and a component $g \in \mathcal{G}$, we have $h \in \Delta_{b,g}$ if and only if there exists a configuration $x \in \mathcal{D}_f$ and δ in $\{-1, 1\}$ with $b(x^{(g,1)}) = [b(x)]^{(h,\delta)}$.

We are interested in maps b for which it is possible to reconstruct $\mathcal{R}_f$, the initial regulatory network, from $\mathcal{R}_{f_b}$, the one determined by the conjugated Boolean dynamics. Formally, we say that a map b is *regulatory-preserving*, if

5

there exists a transformation H_b from the set of regulatory networks over $\mathcal{G}_b$ to the set of regulatory networks over $\mathcal{G}$ such that $H_b(\mathcal{R}_{f_b}) = \mathcal{R}_f$ for all evolution functions f over $\mathcal{C}(\mathcal{G}, \mathbf{m})$.

The main object of this work is to characterize the maps b from $\mathcal{C}(\mathcal{G}, \mathbf{m})$ to $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$ which are both one-to-one, neighbour- and regulatory-preserving. A direct consequence of the definitions is that the Van Ham mapping b_0 verifies all these properties. We will show that in fact all the Boolean maps sharing these properties are strongly related to the Van Ham mapping b_0.

2.4 Characterization

The following proposition shows that if a neighbour-preserving one-to-one map b is regulatory-preserving, then each component of $\mathcal{G}_b$ can be used to follow the evolution of at most one component g of $\mathcal{G}$ (in the sense it belongs to $\Delta_{b,g}$). Note that some of the components of $\mathcal{G}_b$ can be not used at all.

Proposition 1 *Let* b *be a neighbour-preserving one-to-one map from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$. *If* b *is regulatory-preserving then we have* $\Delta_{b,g} \cap \Delta_{b,g'} = \emptyset$ *for all* g *and* g' *in* $\mathcal{G}$ *such that* $g \neq g'$.

Proof: Assume there exist two components g and g' in $\mathcal{G}$ with $g \neq g'$ and a component $h \in \Delta_{b,g} \cap \Delta_{b,g'}$. The fact that b is neighbour-preserving altogether with $h \in \Delta_{b,g} \cap \Delta_{b,g'}$ imply there are two configurations x and x' in $\mathcal{C}(\mathcal{G}, \mathbf{m})$, δ and δ' in $\{-1, 1\}$ such that $x^{(g,1)} \neq x$, $b(x^{(g,1)}) = [b(x)]^{(h,\delta)}$, $x'^{(g',1)} \neq x'$ and $b(x'^{(g',1)}) = [b(x')]^{(h,\delta')}$. Let g'' be a component of $\mathcal{G}$. We consider the evolution function f defined for all configurations $y \in \mathcal{D}_f = \mathcal{C}(\mathcal{G}, \mathbf{m})$ by:

$$f(y) = \begin{cases} x & \text{if } y_{g''} = 0 , \\ x^{(g,1)} & \text{if } y_{g''} \geq 1 . \end{cases}$$

In the case where $\delta = \delta'$, we define the evolution function f' over $\mathcal{D}_{f'} = \mathcal{C}(\mathcal{G}, \mathbf{m})$ by :

$$f'(y) = \begin{cases} x' & \text{if } y_{g''} = 0 , \\ x'^{(g',1)} & \text{if } y_{g''} \geq 1 . \end{cases}$$

If $\delta \neq \delta'$, the evolution function f' is defined by:

$$f'(y) = \begin{cases} x'^{(g',1)} & \text{if } y_{g''} = 0 , \\ x' & \text{if } y_{g''} \geq 1 . \end{cases}$$

In short, the regulatory network $\mathcal{R}_f$ contains only one edge, which is $(g'', g, +)$ and $\mathcal{R}_{f'}$ contains only one edge, which is (g'', g', σ), where σ is the sign of the product $\delta\delta'$.

The conjugated evolution function f_b of f is then defined over $\text{Im}_b(\mathcal{C}(\mathcal{G}, \mathbf{m}))$ by:

$$f_b(b(y)) = \begin{cases} b(x) & \text{if } y_{g''} = 0 , \\ [b(x)]^{(h,\delta)} & \text{if } y_{g''} \geq 1 . \end{cases}$$

6

The definition of f′ ensures that the evolution function f'_b conjugated to f′ verifies for all $b(y) \in \mathrm{Im}_b(\mathcal{C}(\mathcal{G}, \mathbf{m}))$:

$$f'_b(b(y)) = \left\{ \begin{array}{ll} p & \text{if } y_{g''} = 0 \ , \\ p^{(h,\delta)} & \text{if } y_{g''} \geq 1 \ , \end{array} \right.$$

where p is equal to $b(x')$ or $b(x'^{(g',1)})$ depending on the sign of the product $\delta\delta'$.

From its expression, f_b (*resp.* f'_b) takes values in $\{b(x), [b(x)]^{(h,\delta)}\}$ (*resp.* in $\{p, p^{(h,\delta)}\}$). Moreover, we have, for all configurations $q \in \mathrm{Im}_b(\mathcal{C}(\mathcal{G}, \mathbf{m}))$, $f_b(q) = b(x)$ (*resp.* $f_b(q) = [b(x)]^{(h,\delta)}$) if and only if $f'_b(q) = p$ (*resp.* $f'_b(q) = p^{(h,\delta)}$). It follows that $\mathcal{R}_{f_b} = \mathcal{R}_{f'_b}$. Since $\mathcal{R}_f \neq \mathcal{R}_{f'}$, we cannot have simultaneously $H_b(\mathcal{R}_{f_b}) = \mathcal{R}_f$ and $H_b(\mathcal{R}_{f'_b}) = \mathcal{R}_{f'}$, whatever the transformation H_b from the set of regulatory networks over $\mathcal{G}_b$ to the set of regulatory networks over $\mathcal{G}$. As a consequence, the map b cannot be regulatory-preserving and we get our conclusion. □

The next proposition implies that if b is both neighbour- and regulatory-preserving then, for all components $h \in \Delta_{b,g}$, the h-coordinates $b(x)_h$ depend only on x_g.

Proposition 2 *Let* b *be a neighbour-preserving one-to-one map from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$. *If* b *is such that for all g and g′ in* $\mathcal{G}$ *with* $g \neq g'$, $\Delta_{b,g} \cap \Delta_{b,g'} = \emptyset$ *then for all components* $g \in \mathcal{G}$, *all levels* $0 \leq v \leq \mathbf{m}_g$, *there exists a partial Boolean configuration* $\beta^{g,v} \in \{0,1\}^{\Delta_{b,g}}$ *such that for all configurations* $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, *we have* $x_g = v$ *if and only if* $b(x)_h = \beta^{g,v}_h$ *for all* $h \in \Delta_{b,g}$.

Proof: Let g be a component of $\mathcal{G}$ and x a configuration in $\mathcal{C}(\mathcal{G}, \mathbf{m})$. We prove that for all $x' \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, $x_g = x'_g$ implies $b(x)_h = b(x')_h$ for all $h \in \Delta_{b,g}$ by induction over the distance $d(x, x')$. It is basically true when $d(x, x') = 0$. Assume that the property holds for all x'' such that $d(x, x'') \leq k$ and let x' be a configuration such that $x_g = x'_g$ and $d(x, x') = k+1$. There is a configuration y with $x_g = y_g$, $d(x, y) = k$ and $d(y, x') = 1$. In other words y and x' are direct neighbours and we can write $y = x'^{(g',\delta)}$ for a component $g' \neq g$ and $\delta \in \{-1, 1\}$. Since b is neighbour-preserving, there exists $h' \in \mathcal{G}_b$ such that $b(x')_{h'} \neq b(x'^{(g',\delta)})_{h'} = b(y)_{h'}$. By definition h' belongs to $\Delta_{b,g'}$ and, with $g' \neq g$ and our assumption, h' cannot be an element of $\Delta_{b,g}$. So we have $b(y)_h = b(x')_h$ for all $h \in \Delta_{b,g}$, and by induction hypothesis $b(x)_h = b(y)_h$ for all $h \in \Delta_{b,g}$.

Reciprocally, the proof of the fact that if z and z' are two configurations of $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$ such that $z_h = z'_h$ for all $h \in \Delta_{b,g}$, then $b^{-1}(z)_g = b^{-1}(z')_g$, can also be done by induction over the distance $d(z, z')$. Assume the property true for all z'' such that $d(z, z'') \leq k$ (it is basically true for $k = 0$) and let z' be a configuration such that $z_h = z'_h$ for all $h \in \Delta_{b,g}$ and $d(z, z') = k+1$. Let h' be such that $z_{h'} \neq z'_{h'}$, thus, with what precedes, $h' \notin \Delta_{b,g}$. Let us now consider the configuration $t = z'^{(h', z_{h'} - z'_{h'})}$. The definition of t ensures that $d(z, t) = k$ and, from the induction hypothesis, we have $b^{-1}(t)_g = b^{-1}(z)_g$. It remains to

prove that $b^{-1}(t)_g = b^{-1}(z')_g$. Let us note $x = b^{-1}(t)$ and $y = b^{-1}(z')$ and assume that $x_g \neq y_g$. We define the configuration x' by $x'_g = x_g$ and $x'_{g'} = y_{g'}$ for all components $g' \neq g$. There exists a path of components $x_{[0]}, x_{[1]}, \ldots x_{[n]}$ in $\mathcal{C}(\mathcal{G}, \mathbf{m})$ such that $x_{[0]} = x$, $x_{[n]} = x'$, $x_{[i]g} = x_g$ and $x_{[i]}$ is a direct neighbour of $x_{[i-1]}$, for all $0 < i \leq n$. The fact that the sets $\Delta_{b,g'}$ are disjoint ensures that, for all $0 < i \leq n$, $b(x_{[i]})_h = b(x_{[i-1]})_h$ for all $h \in \Delta_{b,g}$, since $x_{[i]g} = x_{[i-1]g}$ and a component $h \in \Delta_{b,g}$ not satisfying the preceding equality would belong to a set $\Delta_{b,g'}$ with $g' \neq g$. At this point, we have $b(x')_h = b(x)_h = b(y)_h$ for all $h \in \Delta_{b,g}$, $x'_g \neq y_g$ and $x'_{g'} = y_{g'}$ for all components $g' \neq g$. Since b is one-to-one, there is a component $h' \in \mathcal{G}_b$ such that $b(x')_{h'} \neq b(y)_{h'}$, thus $h' \notin \Delta_{b,g}$. Let us consider the path of components $y_{[0]}, y_{[1]}, \ldots y_{[y_g - x'_g]}$ (we assume $y_g > x'_g$, the other case being symmetrical), defined for all $0 \leq i \leq (y_g - x'_g)$ by $y_{[i]g} = x'_g + i$ and $y_{[i]g'} = x'_{g'} = y_{g'}$ for all components $g' \neq g$. Since $b(x')_{h'} \neq b(y)_{h'}$, there exists an integer $0 < i \leq (y_g - x'_g)$ such that $b(y_{[i-1]})_{h'} \neq b(y_{[i]})_{h'}$, which is in contradiction with the fact that $h' \notin \Delta_{b,g}$ because $y_{[i]} = y_{[i-1]}{}^{(g,1)}$. It follows that $x_g = b^{-1}(t)_g = b^{-1}(z')_g = y_g$, which, altogether with $b^{-1}(t)_g = b^{-1}(z)_g$, allow us to conclude that $b^{-1}(z)_g = b^{-1}(z')_g$ and to end the proof. $\square$

The next proposition establishes that each level of a component g in $\mathcal{G}$ uniquely relates to a component $h \in \Delta_{b,g}$ and reciprocally.

Proposition 3 *Let* b *be a neighbour- and regulatory-preserving one-to-one map from $\mathcal{C}(\mathcal{G}, \mathbf{m})$ to $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$.*

For all $g \in \mathcal{G}$ and all levels $1 \leq v \leq \mathbf{m}_g$ there exists a unique component $h_v^g \in \Delta_{b,g}$ such that, for all configurations $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, we have $b(x^{(g,1)})_{h_v^g} \neq b(x)_{h_v^g}$ if and only if $x_g = v - 1$.

Reciprocally, for all components $h \in \Delta_{b,g}$, there exists a unique level $1 \leq v_h \leq \mathbf{m}_g$ such that, for all configurations $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, we have $b(x^{(g,1)})_h \neq b(x)_h$ if and only if $x_g = v_h - 1$.

Proof: From Proposition 2 and by using its notations, a configuration $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ is such that $x_g = v$ if and only if we have $b(x)_h = \beta_h^{g,v}$ for all $h \in \Delta_{b,g}$. For all $1 \leq v \leq \mathbf{m}_g$, there exists a unique $h_v^g \in \Delta_{b,g}$ such that $\beta_{h_v^g}^{g,v-1} \neq \beta_{h_v^g}^{g,v}$: existence comes from the fact that b is one-to-one and unicity from the neighbour-preserving property. This proves the first assertion.

Let us assume that, for $g \in \mathcal{G}$ and $h \in \Delta_{b,g}$, there exist two levels $1 \leq v, v' \leq \mathbf{m}_g$ with $v \neq v'$ and two configurations x and x' in $\mathcal{C}(\mathcal{G}, \mathbf{m})$ with $x_g = v - 1$ and $x'_g = v' - 1$ such that $b(x^{(g,1)})_h \neq b(x)_h$ and $b(x'^{(g,1)})_h \neq b(x')_h$. Without loss of generality, we assume that $v < v'$, and, possibly by replacing v' by a smaller level, that there exists no level v'' such that $v < v'' < v'$ and $b(x''^{(g,1)})_h \neq b(x'')_h$ for any configuration x'' with $x''_g = v''$. Considering Proposition 2 and the Boolean character of component h, it comes out that $b(x^{(g,1)}) = b(x)^{(h,\delta)}$ and $b(x'^{(g,1)}) = b(x')^{(h,\delta')}$ where δ and δ' are two elements of $\{-1, 1\}$ such that $\delta = -\delta'$.

Let g' be a component of $\mathcal{G}$ and let us consider the evolution function f

8

defined for all configurations y in $\mathcal{D}_{\mathrm{f}} = \{z \in \mathcal{C}(\mathcal{G}, \mathbf{m}) \mid z_{g'} \leq 1\}$ by:

$$\mathrm{f}(y) = \begin{cases} x & \text{if } y_{g'} = 0 \ , \\ x^{(g,1)} & \text{if } y_{g'} = 1 \ . \end{cases}$$

We define the evolution function f' over $\mathcal{D}_{\mathrm{f}'} = \mathcal{D}_{\mathrm{f}}$ by:

$$\mathrm{f}'(y) = \begin{cases} x'^{(g,1)} & \text{if } y_{g'} = 0 \ , \\ x' & \text{if } y_{g'} = 1 \ . \end{cases}$$

The regulatory network $\mathcal{R}_{\mathrm{f}}$ contains only one edge, which is $(g', g, +)$ and $\mathcal{R}_{\mathrm{f}'}$ contains only one edge, which is $(g', g, -)$.

Let $h_0^{g'}$ be the unique (see above) component of $\mathcal{G}_{\mathrm{b}}$ such that $\mathrm{b}(y^{(g',1)})_{h_0^{g'}} \neq \mathrm{b}(y)_{h_0^{g'}}$ for all configurations $y \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ with $y_{g'} = 0$. We are in the conditions of Proposition 2 and we note δ'', the difference $\beta_{h_0^{g'}}^{g',1} - \beta_{h_0^{g'}}^{g',0}$ (which belongs to $\{-1, 1\}$). For all $y \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ with $y_{g'} = 0$, we have $\mathrm{b}(y^{(g',1)})_{h_0^{g'}} - \mathrm{b}(y)_{h_0^{g'}} = \delta''$.

The evolution functions f_{b} and f_{b}', conjugated to f and f' respectively, are defined for all $\mathrm{b}(y) \in \mathrm{Im}_{\mathrm{b}}(\mathcal{D}_{\mathrm{f}})$, by:

$$\mathrm{f}_{\mathrm{b}}(\mathrm{b}(y)) = \begin{cases} \mathrm{b}(x) & \text{if } y_{g'} = 0 \ , \\ \mathrm{b}(x^{(g,1)}) & \text{if } y_{g'} = 1 \end{cases}$$

and

$$\mathrm{f}_{\mathrm{b}}'(\mathrm{b}(y)) = \begin{cases} \mathrm{b}(x'^{(g,1)}) & \text{if } y_{g'} = 0 \ , \\ \mathrm{b}(x') & \text{if } y_{g'} = 1 \ . \end{cases}$$

With the current assumptions, notations and by setting $a = \beta_{h_0^{g'}}^{g',0}$, these definitions can be rewritten for all $t \in \mathrm{Im}_{\mathrm{b}}(\mathcal{D}_{\mathrm{f}})$ as:

$$\mathrm{f}_{\mathrm{b}}(t) = \begin{cases} \mathrm{b}(x) & \text{if } t_{h_0^{g'}} = a \ , \\ [\mathrm{b}(x)]^{(h,\delta)} & \text{if } t_{h_0^{g'}} = a + \delta'' \end{cases}$$

and

$$\mathrm{f}_{\mathrm{b}}'(t) = \begin{cases} [\mathrm{b}(x')]^{(h,\delta')} & \text{if } t_{h_0^{g'}} = a \ , \\ \mathrm{b}(x') & \text{if } t_{h_0^{g'}} = a + \delta'' \ . \end{cases}$$

We recall that we have $\delta' = -\delta$. It follows that the regulatory networks $\mathcal{R}_{\mathrm{f}_{\mathrm{b}}}$ and $\mathcal{R}_{\mathrm{f}_{\mathrm{b}}'}$ are equal. They both contain only one edge which is $(h_0^{g'}, h, \sigma)$ where σ is the sign of $\delta\delta'' = (-\delta')\delta''$.

Since $\mathcal{R}_{\mathrm{f}} \neq \mathcal{R}_{\mathrm{f}'}$, there cannot exist a transformation H_{b} from the set of regulatory networks over $\mathcal{G}_{\mathrm{b}}$ to the set of regulatory networks over $\mathcal{G}$ such that both $H_{\mathrm{b}}(\mathcal{R}_{\mathrm{f}_{\mathrm{b}}}) = \mathcal{R}_{\mathrm{f}}$ and $H_{\mathrm{b}}(\mathcal{R}_{\mathrm{f}_{\mathrm{b}}'}) = \mathcal{R}_{\mathrm{f}'}$. This is in contradiction with the fact that the map b is regulatory-preserving and we get the proof of the second assertion of the proposition. $\qquad \square$

Before stating our main result, we need the following technical lemma.

9

Lemma 1 *Let* b *be a map from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_\mathrm{b}, \mathbf{1})$. *If a component* $h \in \mathcal{G}_\mathrm{b}$ *is such that there exist two configurations* x *and* y *in* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *with* $\mathrm{b}(x)_h \neq \mathrm{b}(y)_h$ *then there exists a component* $g \in \mathcal{G}$ *such that* $h \in \Delta_{\mathrm{b},g}$.

Proof: The fact that b is a map and $\mathrm{b}(x)_h \neq \mathrm{b}(y)_h$ ensures that $x \neq y$, thus $\mathrm{d}(x, y) > 0$. There exists a path of configurations $x_{[0]}, x_{[1]}, \ldots x_{[\mathrm{d}(x,y)]}$ in $\mathcal{C}(\mathcal{G}, \mathbf{m})$ such that $x_{[0]} = x$, $x_{[\mathrm{d}(x,y)]} = y$ and $x_{[i]}$ is a direct neighbour of $x_{[i-1]}$ (with $\mathrm{d}(x_{[i-1]}, x_{[i]}) = 1$), for all $0 < i \leq \mathrm{d}(x, y)$. Such a path can be built iteratively by setting $x_{[0]} = x$ and, for all $0 < i \leq \mathrm{d}(x, y)$, by defining $x_{[i]}$ as $\left[x_{[i-1]} \right]^{\left(g, \frac{y_g - x_g}{|y_g - x_g|} \right)}$ where g is a component of $\mathcal{G}$ such that $y_g \neq x_{[i]g}$ (since at each step we have $\mathrm{d}(x_{[i-1]}, y) = \mathrm{d}(x, y) - i + 1$, such a component g exists). Let now consider the image of such a path by b. Since $\mathrm{b}(x_{[0]})_h \neq \mathrm{b}(x_{[\mathrm{d}(x,y)]})_h$, there exists an integer $0 < j \leq \mathrm{d}(x, y)$ such that $\mathrm{b}(x_{[j-1]})_h \neq \mathrm{b}(x_{[j]})_h$. Moreover, the fact that $\mathrm{d}(x_{[i-1]}, x_{[i]}) = 1$ implies that we have either $x_{[i-1]}{}^{(g,1)} = x_{[i]}$ or $x_{[i]}{}^{(g,1)} = x_{[i-1]}$ for a component $g \in \mathcal{G}$. This matches the definition that $h \in \Delta_{\mathrm{b},g}$. $\square$

Our main theorem states that a one-to-one, neighbour- and regulatory-preserving map has a very special form.

Theorem 1 *A map* b *from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_\mathrm{b}, \mathbf{1})$ *is one-to-one, neighbour- and regulatory-preserving if and only if:*

1. *the set* $\{\Delta_{\mathrm{b},g} \mid g \in \mathcal{G}\}$ *partitions a subset of* $\mathcal{G}_\mathrm{b}$;

2. *for all components* $g \in \mathcal{G}$, *the set* $\Delta_{\mathrm{b},g}$ *contains exactly* $\mathbf{m}_g$ *components and there exists a numbering* $\{h_1, \ldots, h_{\mathbf{m}_g}\}$ *of the components of* $\Delta_{\mathrm{b},g}$ *which is such that, for all configurations* $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ *and all levels* $0 \leq v \leq \mathbf{m}_g$, *we have* $x_g \leq v$ *if and only if* $\mathrm{b}(x)_{h_w} = \mathrm{b}(\mathbf{0})_{h_w}$ *for all levels* $w > v$, *where* $\mathbf{0}$ *is the configuration* $(0, 0, \ldots, 0)$ *of* $\mathcal{C}(\mathcal{G}, \mathbf{m})$.

Proof: Let b be a one-to-one, neighbour- and regulatory-preserving map from $\mathcal{C}(\mathcal{G}, \mathbf{m})$ to $\mathcal{C}(\mathcal{G}_\mathrm{b}, \mathbf{1})$. Assertion 1 is nothing but a restatement of Proposition 1.

With the notations and results of Proposition 2, Proposition 3 implies that there exists a natural one-to-one and onto map φ_g between elements of $\Delta_{\mathrm{b},g}$ and $\{1, \ldots, \mathbf{m}_g\}$, where $\varphi_g(h)$ is the unique level of g such that $\beta_h^{g, \varphi_g(h)-1} \neq \beta_h^{g, \varphi_g(h)}$. It follows that $\Delta_{\mathrm{b},g}$ contains $\mathbf{m}_g$ components.

To finish to prove Assertion 2, let us take φ_g as numbering of $\Delta_{\mathrm{b},g}$. For a configuration $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ and $h \in \Delta_{\mathrm{b},g}$, we have $\mathrm{b}(x^{(g,1)})_h \neq \mathrm{b}(x)_h$ if and only if $x_g = \varphi_g(h) - 1$ (Proposition 3). For all configurations x with $0 \leq x_g < \varphi_g(h) - 1$, we have $\mathrm{b}(x^{(g,1)})_h = \mathrm{b}(x)_h$, thus, with Proposition 2, $\beta_h^{g, x_g+1} = \beta_h^{g, x_g} = \beta_h^{g, 0} = \mathrm{b}(\mathbf{0})_h$. Since h is a Boolean component, $\beta_h^{g, \varphi_g(h)-1} \neq \beta_h^{g, \varphi_g(h)}$ means that $\beta_h^{g, \varphi_g(h)} = 1 - \mathrm{b}(\mathbf{0})_h$. As previously, for all configurations x with $\varphi_g(h) \leq x_g < \mathbf{m}_g$, we have $\mathrm{b}(x^{(g,1)})_h = \mathrm{b}(x)_h$, thus, $\beta_h^{g, x_g} = \beta_h^{g, \varphi_g(h)} = 1 - \mathrm{b}(\mathbf{0})_h$. We conclude that, for all configurations $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, we have $\mathrm{b}(x)_h = \mathrm{b}(\mathbf{0})_h$ if and only if $0 \leq x_g < \varphi_g(h)$.

Reciprocally, if a Boolean mapping b satisfies Assertions 1 and 2 then b is basically one-to-one and neighbour-preserving. To show that it is also regulatory-preserving, we define the transformation $U_{\rm b}$ from the set of regulatory networks over $\mathcal{G}_{\rm b}$ to the set of regulatory networks over $\mathcal{G}$ by: an edge (g, g', σ) belongs to $U_{\rm b}(\mathcal{R}_{\rm f_b})$ if and only if there exists an edge (h, h', σ') in $\mathcal{R}_{\rm f_b}$ with $h \in \Delta_{{\rm b},g}$, $h' \in \Delta_{{\rm b},g'}$, $\sigma' = \sigma$ if ${\rm b}(\mathbf{0})_h = {\rm b}(\mathbf{0})_{h'}$ and $\sigma' = -\sigma$ otherwise. We have to show that if Assertions 1 and 2 are granted then $U_{\rm b}(\mathcal{R}_{\rm f_b}) = \mathcal{R}_{\rm f}$.

$\mathcal{R}_{\rm f} \subseteq U_{\rm b}(\mathcal{R}_{\rm f_b})$

By definition, there is an edge (g', g, σ) in $\mathcal{R}_{\rm f}$ if and only if there exists a configuration $x \in \mathcal{D}_{\rm f}$ such that $x^{(g',1)} \in \mathcal{D}_{\rm f}$, ${\rm f}(x^{(g',1)})_g \neq {\rm f}(x)_g$ and the sign of $({\rm f}(x^{(g',1)})_g - {\rm f}(x)_g)$ is σ. We detail only the case where $(g', g, +) \in \mathcal{R}_{\rm f}$, i.e. ${\rm f}(x^{(g',1)})_g > {\rm f}(x)_g$, the case where $(g', g, -) \in \mathcal{R}_{\rm f}$ being symmetrical. With the numbering $\{h_1, \ldots, h_{\mathbf{m}_g}\}$ of $\Delta_{{\rm b},g}$ considered in Assertion 2, we have ${\rm b}({\rm f}(x))_{h_v} = {\rm b}(\mathbf{0})_{h_v}$ and ${\rm b}({\rm f}(x^{(g',1)}))_{h_v} = 1 - {\rm b}(\mathbf{0})_{h_v}$, for all levels ${\rm f}(x)_g < v \leq {\rm f}(x^{(g',1)})_g$ (for all other levels v, ${\rm b}({\rm f}(x^{(g',1)}))_{h_v} = {\rm b}({\rm f}(x))_{h_v}$). Let us note v' the level $x_{g'}$. Again Assertion 2 implies that there exists a unique component $h'_{v'}$ of $\Delta_{{\rm b},g'}$ such that ${\rm b}(x)_{h'_{v'}} = {\rm b}(\mathbf{0})_{h'_{v'}}$ and ${\rm b}(x^{(g',1)})_{h'_{v'}} = 1 - {\rm b}(\mathbf{0})_{h'_{v'}}$. Putting all together, we get that, for all levels ${\rm f}(x)_g < v < {\rm f}(x^{(g',1)})_g$, we have $(h'_{v'}, h_v, +) \in \mathcal{R}_{\rm f_b}$ if ${\rm b}(\mathbf{0})_{h'_{v'}} = {\rm b}(\mathbf{0})_{h_v}$, and $(h'_{v'}, h_v, -) \in \mathcal{R}_{\rm f_b}$ otherwise. In other words, an edge $(g', g, +)$ in $\mathcal{R}_{\rm f}$ comes always with at least one edge (h, h', σ') in $\mathcal{R}_{\rm f_b}$ where $h \in \Delta_{{\rm b},g}$, $h' \in \Delta_{{\rm b},g'}$, $\sigma' = +$ if ${\rm b}(\mathbf{0})_h = {\rm b}(\mathbf{0})_{h'}$ and $\sigma' = -$ otherwise. The existence of an edge $(g', g, -)$ in $\mathcal{R}_{\rm f}$ leads to a symmetrical conclusion.

$U_{\rm b}(\mathcal{R}_{\rm f_b}) \subseteq \mathcal{R}_{\rm f}$

Let us now assume that there is an edge $(h', h, +)$ in $\mathcal{R}_{\rm f_b}$. This means that we have a configuration $t \in {\rm Im}_{\rm b}(\mathcal{D}_{\rm f})$ with $t^{(h',1)} \in {\rm Im}_{\rm b}(\mathcal{D}_{\rm f})$ and such that ${\rm f}_{\rm b}(t^{(h',1)})_h > {\rm f}_{\rm b}(t)_h$. We first prove that there exists a unique component $g \in \mathcal{G}$ (resp. $g' \in \mathcal{G}$) such that $h \in \Delta_{{\rm b},g}$ (resp. $h' \in \Delta_{{\rm b},g'}$). Configurations t and $t^{(h',1)}$ are assumed to belong to ${\rm Im}_{\rm b}(\mathcal{D}_{\rm f})$, which, by construction, implies that ${\rm f}_{\rm b}(t^{(h',1)})_h \in {\rm Im}_{\rm b}(\mathcal{D}_{\rm f})$ and ${\rm f}_{\rm b}(t)_h \in {\rm Im}_{\rm b}(\mathcal{D}_{\rm f})$. Applying Lemma 1 to the reciprocal images by b of ${\rm f}_{\rm b}(t^{(h',1)})_h$ and ${\rm f}_{\rm b}(t)_h$ (resp. of t and $t^{(h',1)}$) gives us the existence of g (resp. of g') while unicity comes from Assertion 1. Assertion 2 then ensures that a numbering of $\Delta_{{\rm b},g}$ (resp. of $\Delta_{{\rm b},g'}$) has a strong property. These numberings are noted φ_g and $\varphi_{g'}$ respectively, i.e. $\varphi_g(h)$ denotes the number/level associated to $h \in \Delta_{{\rm b},g}$. From Assertion 2, the configurations ${\rm b}^{-1}(t)$ and ${\rm b}^{-1}(t^{(h',1)})$ are such that ${\rm b}^{-1}(t^{(h',1)}) = [{\rm b}^{-1}(t)]^{(g',\delta')}$, where $\delta' = 1$ if ${\rm b}(\mathbf{0})_{h'} = 0$ and $\delta' = -1$ otherwise. The fact that ${\rm f}_{\rm b}(t^{(h',1)})_h > {\rm f}_{\rm b}(t)_h$ means that we have ${\rm f}_{\rm b}(t^{(h',1)})_h = 1$ and ${\rm f}_{\rm b}(t)_h = 0$. Again with Assertion 2, this implies that:

- if ${\rm b}(\mathbf{0})_h = 0$ then ${\rm b}^{-1}({\rm f}_{\rm b}(t))_g < \varphi_g(h)$ and ${\rm b}^{-1}({\rm f}_{\rm b}(t^{(h',1)}))_g \geq \varphi_g(h)$,

- if ${\rm b}(\mathbf{0})_h = 1$ then ${\rm b}^{-1}({\rm f}_{\rm b}(t))_g \geq \varphi_g(h)$ and ${\rm b}^{-1}({\rm f}_{\rm b}(t^{(h',1)}))_g < \varphi_g(h)$.

11

We have $b^{-1}(f_b(t)) = f(b^{-1}(t))$ and $b^{-1}(f_b(t^{(h',1)})) = f(b^{-1}(t^{(h',1)}))$. With all that precedes we have an edge (g', g, σ) with $\sigma = +$ if $b(\mathbf{0})_{h'} = b(\mathbf{0})_h$ and $\sigma = -$ otherwise. The case of an edge $(h', h, -)$ in $\mathcal{R}_{f_b}$ is symmetrical.

Finally, there is an edge (g, g', σ) in $U_b(\mathcal{R}_{f_b})$ if and only if there is an edge (h, h', σ') in $\mathcal{R}_{f_b}$ with $h \in \Delta_{b,g}$, $h' \in \Delta_{b,g'}$, $\sigma = \sigma'$ if $b(\mathbf{0})_h = b(\mathbf{0})_{h'}$ and $\sigma = -\sigma'$ otherwise. With the definition of U_b, this implies that $U_b(\mathcal{R}_{f_b}) = \mathcal{R}_f$: the map b is regulatory-preserving. $\qquad\square$

The following corollary explicits the relation between a one-to-one, neighbour- and regulatory-preserving map and the Van Ham mapping b_0. It says that, if we require that all the Boolean components are used by a Boolean map b (*i.e.* their coordinates have to be different in at least two images by b), the map b_0 defined in Section 2.2 is, up to switching 0 and 1 in the Boolean components, the only one which is both one-to-one, neighbour- and regulatory-preserving.

Corollary 1 *Let* b *be a map from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$ *such that for all* $h \in \mathcal{G}_b$ *there exist two configurations* x *and* y *in* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *with* $b(x)_h \neq b(y)_h$. *The map* b *is one-to-one, neighbour- and regulatory-preserving if and only if for all configurations* $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ *and for all components* $h \in \mathcal{G}_b$, *we have:*

$$b(x)_h = \begin{cases} b_0(x)_h & \text{if } b(\mathbf{0})_h = 0, \\ 1 - b_0(x)_h & \text{if } b(\mathbf{0})_h = 1, \end{cases}$$

where $\mathbf{0}$ *is the configuration* $(0, 0, \ldots, 0)$ *of* $\mathcal{C}(\mathcal{G}, \mathbf{m})$.

Proof: Lemma 1 ensures that, under the assumptions of the corollary, for all components $h \in \mathcal{G}_b$, there exists a component $g \in \mathcal{G}$ such that $h \in \Delta_{b,g}$. By applying Theorem 1 to this case, Assertion 1 implies that $\{\Delta_{b,g} \mid g \in \mathcal{G}\}$ partitions the whole set $\mathcal{G}_b$. The relation between images by map b and images by map b_0 follows from Assertion 2 and Definition 1. $\qquad\square$

The proof of the following corollary comes directly with Theorem 1.

Corollary 2 *Let* b *be a neighbour- and regulatory-preserving one-to-one map from* $\mathcal{C}(\mathcal{G}, \mathbf{m})$ *to* $\mathcal{C}(\mathcal{G}_b, \mathbf{1})$. *For all* $x, y \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, *we have* $d(x, y) = d(b(x), b(y))$.

In other words, if a one-to-one map is both neighbour- and regulatory-preserving, it keeps the topological structure of the dynamics much beyond the direct-neighbours conservation.

3   Discussion and illustration

In this section, we first discuss the application of the Van Ham Boolean mapping b_0, in particular concerning the representation of the dynamics in terms of state transition graphs. A simple illustration is provided, based on a published model of the p53-Mdm2 network [1].

12

3.1 Graph representations of the dynamics

Temporal behaviours of logical regulatory networks can be represented as state transition graphs (called graphs of sequences of states in [26]), in which vertices represent logical states and edges account for transitions between states. This representation gathers paths that the systems can follow and conveniently supports the analysis of dynamical properties such as the existence of stable states, complex attractors, etc (see *e.g.* [15]). Hence, it is important to make sure of the impact of the Boolean mapping b_0 on this structure.

The asynchronous and synchronous state transition graphs (STG) of an evolution function f are two different ways of representing the dynamics induced by f. So far, in the multivalued case, STG have always been constructed under an additional constraint which is a stepwise change of the component values, *i.e.* there is a transition from a configuration x to a configuration y, both in $\mathcal{D}_f$, if:

- in an asynchronous updating: there exists $g \in \mathcal{G}$ such that $y = x^{(g,\delta)}$ with $\delta = \frac{f(x)_g - x_g}{|f(x)_g - x_g|}$,

- in a synchronous updating: for all $g \in \mathcal{G}$ such that $f(x)_g \neq x_g$, $y_g = x_g + \frac{f(x)_g - x_g}{|f(x)_g - x_g|}$, and for all $g' \in \mathcal{G}$ such that $f(x)_{g'} = x_{g'}$, $y_{g'} = x_{g'}$.

Note that, in the asynchronous case, edges (or transitions) in the STG connect only direct neighbours.

Proposition 4 *Let* f *be an evolution function over* $\mathcal{D}_f \subseteq \mathcal{C}(\mathcal{G}, \mathbf{m})$. *The asynchronous STG of* f_{b_0} *is isomorphic to the asynchronous STG of* f.

Proof: Since b_0 is one-to-one, it associates to each configuration of $\mathcal{D}_f$ a unique configuration in $\mathcal{D}_{f_{b_0}} = \mathrm{Im}_{b_0}(\mathcal{D}_f)$ and reciprocally. To finish to prove that b_0 is a graph isomorphism between the two STG, we have to show that there is an edge from a configuration x to a configuration y in the asynchronous STG of f if and only if there is an edge from $b_0(x)$ to $b_0(y)$ in the asynchronous STG of f_{b_0}. We recall that an edge of an asynchronous STG is either of the form $(x, x^{(g,1)})$ or of the form $(x, x^{(g,-1)})$, for a configuration x and a component g. We consider only the case where there is an edge of the first form, the other one being symmetrical.

By definition, there is an edge $(x, x^{(g,1)})$ in the asynchronous STG of f if and only if $x_g < f(x)_g$. From Definition 1 and with its notations, this inequality is equivalent to:

- $b_0(x)_{h_v^g} = b_0(f(x))_{h_v^g} = f_{b_0}(b_0(x))_{h_v^g}$ for all levels v with $1 \leq v \leq x_g$ or $f(x)_g < v \leq \mathbf{m}_g$,

- $b_0(x)_{h_v^g} < b_0(f(x))_{h_v^g} = f_{b_0}(b_0(x))_{h_v^g}$ for all levels v with $x_g < v \leq f(x)_g$.

Among all the Boolean components h_v^g with $x_g < v \leq f(x)_g$, there is only $h_{x_g+1}^g$ which is such that $[b_0(x)]^{(h_{x_g+1}^g, 1)}$ belongs to $\mathcal{D}_{f_{b_0}}$. Moreover $[b_0(x)]^{(h_{x_g+1}^g, 1)} =$

13

$b_0(x^{(g,1)})$, hence all the preceding is equivalent to have an edge $(b_0(x), b_0(x^{(g,1)}))$ in the asynchronous STG of f_{b_0}.

Reciprocally, let us assume that there is an edge $(b_0(x), b_0(y))$ in the asynchronous STG of f_{b_0}. From the definition of the asynchronous STG, this implies the existence of a component $h \in \mathcal{G}_{b_0}$ such that either $b_0(y) = b_0(x)^{(h,1)}$ or $b_0(y) = b_0(x)^{(h,-1)}$. Let us assume that we are in the first case (the other one being symmetrical). From Definition 1 and with its notation, if both $b_0(x)$ and $b_0(x)^{(h,1)}$ belong to $\mathcal{D}_{f_{b_0}}$ then there exists a component $g \in \mathcal{G}$ such that $h = h_{x_g+1}^g$. It comes that $y = x^{(g,1)}$. Moreover, an edge $(b_0(x), b_0(y))$ belongs to the asynchronous STG of f_{b_0} only if $f_{b_0}(b_0(x))_{h_v^g} > b_0(x)_{h_v^g}$, which is itself equivalent to $f(x)_g > x_g$ (Definition 1) and thus there is an edge (x, y) in the asynchronous STG of f. $\qquad\square$

Note that when the evolution function f, defined on $\mathcal{C}(\mathcal{G}, \mathbf{m})$, is such that $|f(x)_g - x_g| \leq 1$, $\forall g \in \mathcal{G}$ and $\forall x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, the STG of f (synchronous or asynchronous) is isomorphic to the STG of f_{b_0}. However, if there exists $g \in \mathcal{G}$ and $x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$ with $|f(x)_g - x_g| > 1$, the synchronous STG of f_{b_0} is no longer isomorphic to that of f (see Figure 2). In this case, it is always possible to consider a modified function f' defined from f as follows: $\forall g \in \mathcal{G}, \forall x \in \mathcal{C}(\mathcal{G}, \mathbf{m})$, if $f(x)_g = x_g$ then $f'(x)_g = x_g$, otherwise $f'(x)_g = x_g + \frac{f(x)_g - x_g}{|f(x)_g - x_g|}$. Then the synchronous STG (*resp.* the asynchronous STG) of f' and of f'_{b_0} are isomorphic. It must be noticed that this procedure has a drawback: although the asynchronous STG of f' obviously equals that of f, the regulatory networks $\mathcal{R}_f$ and $\mathcal{R}_{f'}$ are not the same. For a general dynamics f, $f(x)_g$, the g-level of the image of a configuration x, only depends on the levels of the regulators of g and not on its own level x_g, except in the case where g is auto-regulated. In contrast with the corresponding modified dynamics f', level $f'(x)_g$ is, whenever $f(x)_g \neq x_g$, one step up or down from the current g-level x_g. As a consequence, in the multivalued case, level $f'(x)_g$ may change with x_g, even if the component g is not auto-regulated in the initial dynamics f and all its regulators levels remain the same. Finally, the regulatory network $\mathcal{R}_{f'}$ of the modified dynamics may contain extra auto-regulations with regard to the initial regulatory network $\mathcal{R}_f$.

3.2 Application to the p-53-Mdm2 network

To illustrate and further discuss the Van Ham Boolean mapping of multivalued logical regulatory graphs, we chose the model defined in [1]. It describes the interactions between the tumour suppressor protein p53 with its main negative regulator, the ubiquitin ligase Mdm2. This is represented in its nuclear and cytoplasmic forms. The nuclear form of Mdm2 down-regulates the level of active p53, which in turn up-regulates the level of cytoplasmic Mdm2 and inhibits its nuclear translocation (see [1, 6] for further details and references).

As in [1], P, Mn and Mc stand for proteins p53, nuclear Mdm2 and cytoplasmic Mdm2, respectively. This regulatory network is shown in Figure 2.

In the course of the modelling process, the evolution function is defined by specifying the behaviour of each component, *i.e.* the level it tends to, depending

14

Table 1: Evolution function for the p53 network. On the left hand side, the configurations x and the evolution function $f(x)$ are considered in the multi-valued case (only p53 takes three values, the two other variables are Boolean). The right part of the table displays the Boolean mapping $b_0(x)$ and the associated evolution function $f_{b_0}(b_0(x))$. Notation P1 (resp. P2) denotes the Boolean component associated to value 1 (resp. 2) of P.

	x			$f(x)$			$b_0(x)$				$f_{b_0}(b_0(x))$			
P	Mc	Mn	P	Mc	Mn	P1	P2	Mc	Mn	P1	P2	Mc	Mn
0	0	0	2	0	1	0	0	0	0	1	1	0	1
1	0	0	2	0	0	1	0	0	0	1	1	0	0
2	0	0	2	1	0	1	1	0	0	1	1	1	0
0	1	0	2	0	1	0	0	1	0	1	1	0	1
1	1	0	2	0	1	1	0	1	0	1	1	0	1
2	1	0	2	1	1	1	1	1	0	1	1	1	1
0	0	1	0	0	1	0	0	0	1	0	0	0	1
1	0	1	0	0	0	1	0	0	1	0	0	0	0
2	0	1	0	1	0	1	1	0	1	0	0	1	0
0	1	1	0	0	1	0	0	1	1	0	0	0	1
1	1	1	0	0	1	1	0	1	1	0	0	0	1
2	1	1	0	1	1	1	1	1	1	0	0	1	1

on the levels of its regulators. In the p53/Mdm2 network, since P regulates two components (Mn and Mc), there is no reason to assume that the level of P beyond which it influences Mn is equal to that beyond which it influences Mc. Abou-Jaoudé *et al.* thus proposed that this variable could take values 0, 1 and 2 (whereas Mn and Mc are Boolean). They further considered different evolution functions, depending on whether P acts on Mn above its first threshold and on Mc above its second threshold (or the other way around) and depending on the combined effects of P and Mc on Mn. Table 1 explicitly defines f one of these evolution functions, together with its conjugated Boolean function f_{b_0} obtained with the mapping b_0.

Conversely, as explained in Section 2.1, it is possible to derive the regulatory network of an evolution function. From f given in Table 1, it is possible to recover the regulatory network of Figure 2. For example, we can infer that P inhibits Mn above its first threshold since $f([100])_{Mn} - f([000])_{Mn} = -1$, whereas P activates Mc only above its second threshold since $f([200])_{Mc} - f([100])_{Mc} = 1$ while $f([100])_{Mc} - f([000])_{Mc} = 0$. This corresponds to the situation depicted in Figure 2 (a) in [1].

As for the dynamics, it is worth to note the impact of the stepwise assumption. For example, in the multivalued case, the configuration [000], which image by f is [201] has two successors ([100] and [001]) in the asynchronous updating and a unique successor ([101]) in the synchronous updating (see Figure 2, top). In the Boolean counterpart, the configuration [0000], which image by f_{b_0} is [1101] has also two successors ([1000] and [0001]) in the asynchronous updating (recalling that the configuration [0100] is not admissible), implementing the same behaviour as in the multivalued model; whereas the unique successor [1101] of [0000] in the synchronous updating corresponds to the configuration [201], opposing the stepwise assumption (as P's level goes from 0 directly to 2).
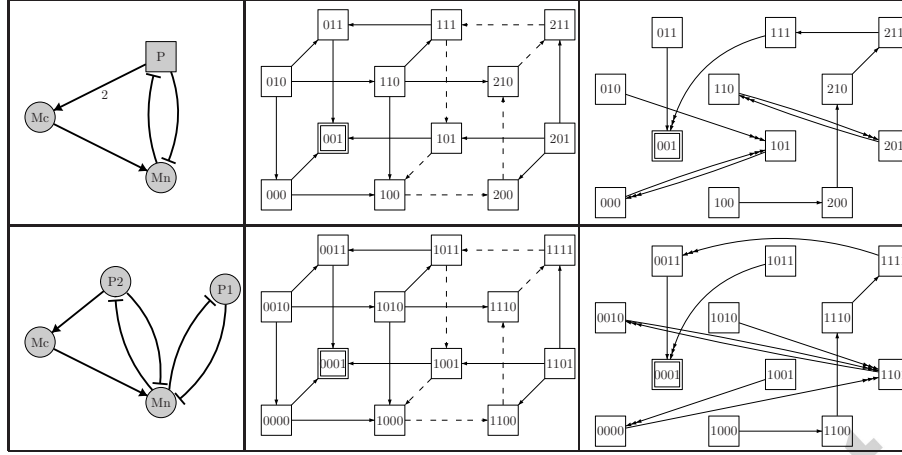
15

Figure 2: The p53/Mdm2 network. Top, the multivalued regulatory network $\mathcal{R}_f$ (f is defined in Table 1), together with the asynchronous and synchronous STG of f. In the regulatory network on the left, normal arrows denote activatory interactions, whereas blunt arrows denote inhibitory interactions. Rectangular node P is multivalued while circular nodes Mc and Mn are Boolean. Finally, arc labels indicate the source level for which the interaction is active (by default, this label is 1 and is omitted). Bottom, the Boolean version of the p53/Mdm2 network $\mathcal{R}_{f_{b_0}}$ inferred from f_{b_0} and the corresponding STG encompassing all the admissible states. In the asynchronous STG, dotted edges emphasize a transient cycle (which is lost in the synchronous updating mode). In the synchronous STG, multiple arrows denote multiple changes. The stable state (conserved for all updating modes) is displayed in a double frame.

4 Conclusions and prospects

A number of Boolean mappings have been studied by the computer science community (see *e.g.* [2]), some of them considering a logarithmic rather than linear number of Boolean variables. Hence, it was debatable whether some mapping more compact than the one proposed by Van Ham in [28] could be considered to map multivalued to Boolean variables representing levels of regulatory components. In this paper, we have proved that the Van Ham mapping, b_0, which uses a number of Boolean variables linear with the total number of levels in the initial dynamics, constitutes the sole mapping, up to switching 0 an 1 in some of the Boolean components or to add constant Boolean components, satisfying a number of requirements: one-to-one, neighbour preservation and regulatory structure conservation.

Given an evolution function f and its conjugate f_{b_0}, their associated state transition graphs (STG) are isomorphic if defined under an asynchronous updating. In the synchronous case, both STG are still isomorphic if f relates any configuration x to a configuration $f(x)$ which components differ from those of

16

x by at most one (that is to say $\forall g \in \mathcal{G}$, $|f(x)_g - x_g| \leq 1$). When it is not the case, it is always possible to consider a modified function f' such that the STG of f' and of f'_{b_0} are isomorphic (cf. the end of Subsection 3.1). But the regulatory networks $\mathcal{R}_f$ and $\mathcal{R}_{f'}$ are not the same. This problem relates to the reverse-engineering of regulatory networks from dynamics described as (partial) STG, for which there is still work to do.

It is worth noting that the Van Ham's Boolean mapping could be easily implemented in a software like GINsim, which allows the definition and analysis of logical models [14].

Finally, we would like to stress the potential interest in studying Boolean counterparts of multivalued logical models. In the introduction, we have already suggested the fact that methods dedicated to Boolean networks could then be used. Here we would like to remark that the Boolean mapping, by itself, could provide some insights in the functioning of regulatory circuits by, in a way, disentangling them. The role of regulatory circuits in the emergence of essential dynamical properties is well recognized [24]. These are known to be responsible for the generation of multi-stationarity (for positive circuits) or oscillations (for negative circuits). In Figure 2, the negative circuit in $\mathcal{R}_f$, between P, Mc and Mn, is defined for higher levels of P. This indicates that the Boolean mapping could be a way to disentangle intertwined regulatory circuits, facilitating the understanding of their dynamical roles. Indeed, the analysis of circuits is more difficult in the multivalued case and formal results were mainly established for Boolean networks.

Acknowledgements

References

[1] W. Abou-Jaoude, D. A Ouattara, and M. Kaufman. From structure to dynamics: frequency tuning in the p53-Mdm2 network I. Logical approach. *J. Theor. Biol.*, 258(4):561–577, 2009.

[2] C. Ansótegui and F. Manyà. Mapping problems with finite-domain variables to problems with boolean variables. In H. H. Hoos and D. G. Mitchell, editors, *SAT (Selected Papers)*, volume 3542 of *LNCS*, pages 1–15, 2004.

[3] J. Aracena, J. Demongeot and E. Goles. Positive and negative circuits in discrete neural networks. *IEEE Transaction on neural networks*, 15(1):77–83, 2004.

[4] S. Bornholdt. Boolean network models of cellular regulation: prospects and limitations. *J. Roy. Soc. Interface* **5**:S85–S94, 2008.

[5] M. Chaves, R. Albert, and E. D. Sontag. Robustness and fragility of Boolean models for genetic regulatory networks. *J. Theor. Biol.*, 235(3):431–49, 2005.

[6] A. Ciliberto, B. Novak, JJ. Tyson. Steady states and oscillations in the p53/Mdm2 network. *Cell Cycle*, 4(3):488–93, 2005.

[7] A Fauré and D Thieffry. Logical modelling of cell cycle control in eukaryotes: a comparative study. *Mol. Biosyst.*, 5:1569–81, 2009.

[8] A. Garg, I. Xenarios, L. Mendoza, and G. De Micheli. An efficient method for dynamic analysis of gene regulatory networks and *in silico* gene perturbation experiments. In *Recomb*, volume 4453 of *LNCS*, pages 62–76, 2007.

[9] L. Glass, and S.A. Kauffman. The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.*,**39**:103–129, 1973.

[10] S.A. Kauffman. *The Origins of Order. Self-Organization and Selection in Evolution.* Oxford University Press, 1993.

[11] M. Kaufman, J. Urbain, and R. Thomas. Towards a logical analysis of the immune response. *J. Theor. Biol.*,**114**(4):527–561, 1985.

[12] M. Kaufman, and R. Thomas. Model analysis of the bases of multistationarity in the humoral immune response. *J. Theor. Biol.*, **129**(2):141–162, 1987.

[13] L. Mendoza. A network model for the control of the differentiation process in Th cells. *Biosystems*, 84(2):101–14, 2006.

[14] A. Naldi, D. Berenguier, A. Fauré, F. Lopez, D. Thieffry, and C. Chaouiya. Logical modelling of regulatory networks with GINsim 2.3. *Biosystems* **97**(2):134–9, 2009.

[15] A. Naldi, E. Remy, D. Thieffry, and C. Chaouiya. A reduction of logical regulatory graphs preserving essential dynamical properties. in *CMSB'09*, volume 5688 of *LNCS*, pages 266–80, 2009.

[16] E. Remy and P. Ruet. From minimal signed circuits to the dynamics of Boolean regulatory networks. *Bioinformatics*, 24(16):i220–6, 2008.

[17] A. Richard and J.-P. Comet. Necessary conditions for multistationarity in discrete dynamical systems. *Discrete Applied Mathematics*, 155(18):2403-2413, 2007.

[18] J. Saez-Rodriguez, L. Simeoni, JA. Lindquist, R. Hemenway, U. Bommhardt, B. Arndt, U-U. Haus, R. Weismantel, ED. Gilles, S. Klamt, and B. Schraven. A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol*, 3(8):e163, 2007.

[19] L. Sánchez, C. Chaouiya, and D. Thieffry. Segmenting the fly embryo: a logical analysis of the segment polarity cross-regulatory module. *Int. J. Dev. Biol.*, 52(8):1059–75, 2008.

[20] L. Sánchez and D. Thieffry. A logical analysis of the drosophila gap-gene system. *J. Theor. Biol.*, 211(2):115–41, 2001.

[21] L. Sánchez and D. Thieffry. Segmenting the fly embryo: a logical analysis of the pair-rule cross-regulatory module. *J. Theor. Biol.*, 224(4):517–37, 2003.

[22] H. Siebert. Deriving behavior of Boolean bioregulatory networks from sub-network dynamics. *Math. Comput. Sci.*, 2(3):421-442, 2009.

[23] E.H. Snoussi and R. Thomas. Logical identification of all steady states: the concept of feedback loop characteristic states. *Bull. Math. Biol.*, 55(5):973–91, 1993.

[24] D. Thieffry. Dynamical roles of biological regulatory circuits. *Brief Bioinform.*, 8(4):220–5, 2007.

[25] R. Thomas. Regulatory networks seen as asynchronous automata: A logical description. *J. Theor. Biol.*, 153:1–23, 1991.

[26] R. Thomas and R. D'Ari. *Biological Feedback*. CRC Press, 1990.

[27] L. Tournier and M. Chaves. Uncovering operational interactions in genetic networks using asynchronous Boolean dynamics. *J. Theor. Biol.*, 260(2):196–209, 2009.

[28] P. Van Ham. How to deal with variables with more than two levels. In R. Thomas, editor, *Kinetic logic: a Boolean approach to the analysis of complex regulatory systems.*, volume 29. Lecture notes in Biomathematics, 1979.

[29] K. Willadsen and J. Wiles. Robustness and state-space structure of Boolean gene regulatory models. *J. Theor. Biol.*, 249(4):749–65, 2007.