CHAPTER 11

CHEMICAL ORGANISATION THEORY

PETER DITTRICH AND PIETRO SPERONI DI FENIZIO*

Bio Systems Analysis Group, Jena Centre for Bioinformatics and Department of Mathematics and Computer Science, Friedrich Schiller University Jena, D-07743 Jena, Germany

Abstract: Complex dynamical reaction networks consisting of many molecular species are difficult to understand, especially, when new species may appear and present species may vanish completely. This chapter outlines a technique to deal with such systems. The first part introduces the concept of a chemical organisation as a closed and self-maintaining set of molecular species. This concept allows to map a complex (reaction) network to its set of organisations, providing a new view on the system's structure. The second part connects dynamics with the set of organisations, which allows to map a movement of the system in state space to a movement in the set of organisations. The relevancy of this approach is underlined by a theorem that says that given a differential equation describing the chemical dynamics of the network, then every stationary state is an instance of an organisation. Finally, the relation between pathways and chemical organisations is sketched

Keywords: reaction networks, constraint based network analysis, hierarchical decomposition, constructive dynamical systems

1. INTRODUCTION

The rapidly increasing size and complexity of reaction system models requires novel mathematical and computational techniques in order to cope with their complexity.[9] This chapter describes a technique that allows to identify for a given reaction network important sub-structures, called chemical organisations.[10,38] These

*Both authors contributed equally. May 10, 2006

M. Al-Rubeai and M. Fussenegger (eds.), Systems Biology, 361–393.

organisations allow to explain the (potential) behaviour of the reaction system from a global and more abstract perspective.

The theory aims at those systems where different combinations of molecular species (compounds) are present at different points in time. These systems are characterised by the fact that they are changing not only *quantitatively*, that is, by a change in the concentration of a molecular species, but also *qualitatively* when new molecular species appear or a present species completely vanish. Fontana and Buss[20] called systems that display the production of novelty *constructive (dynamical) systems*.

Classical approaches describe the dynamics of a reaction system as a "quantitative" movement in a fixed state space,[22] where a state is usually described by a concentration vector.[16] Here, we will operate on a higher level of abstraction and consider *qualitative* movements from a set of molecular species to another set of molecular species. We can interprete this qualitative change as a movement that goes from state space to state space, as new molecular species appear and old species disappear.

The lack of a theory for such constructive dynamical systems has been presented, identified, and discussed in detail by Fontana and Buss[20] in the context of a theory for biological organisation. As a partial solution, they suggest the important concept of a (biological) organisation as a set of molecules that are algebraically closed and dynamically self-maintaining.

Closure means that no new molecular species can be generated by reactions among molecules inside the organisation (Section 3.1). As such no novelty can spontaneously appear. Note that closure, as a property of a set of molecules, should not be confused with the thermodynamical closure of a system, which are two different and separated concepts.

Self-maintaining roughly means that every consumed molecule of the organisation has a way to be generated within the organisation such that it does not disappear from the system (Section 3.3).

Although closure and self-maintenance do not assure that a set of species will remain unchanged in time, the lack of them does imply that the system will eventually *qualitatively* move to a different set of molecules (Section 4.2). In a vast class of systems, which we call *consistent* (Section 3.5), it is possible to define a generator operator such that for any set of molecules an organisation is uniquely defined. The organisation generated by a set A represents the largest possible set of molecules that can stably exists when starting with A.

This implies that organisations partition the set of all possible sets of molecules, where a partition consists of all combinations of molecular species that generate the same organisation. Thus, as the system qualitatively progress from one set to another we can follow it on the more tractable set of all possible organisations (Section 4.3, Figure 4). The study of this movement together with a theorem relating fixed points to organisations will be the core concepts of the dynamical part of chemical organisation theory (Section 4).

Before we enter into the theory, some prerequisites are introduced in the following section.

2. REACTION SYSTEMS

The theory described herein aims at understanding *reaction systems*. A reaction system consists of molecules, and interaction rules among molecules that lead to the appearance or disappearance of other molecules.

Note that we have to distinguish between a reaction system as an abstract description of all possible molecular species (and their reactions), and an actual reaction vessel, which contains some concrete instances of molecules from the set of all possible molecules. In order to refer to an element of a reaction system, we will use a series of terms equivalently: *molecule*, *compound*, *molecular species*, or simply *species*; keeping in mind that the term "molecule" is somehow imprecise, since it can also refer to a concrete physical instance. Similarly, we call an interaction rule among molecules shortly a *reaction*.

In passing we note that reaction systems are not only used to model chemical phenomena. Their applications range from ecology,[35] protobiology,[36] systems biology[33] to computer science[2] and reach even the study of language and social systems.[11]

The description of a reaction system can be subdivided into three parts:[12] (1) the set of all possible molecules $\mathcal{M}$, (2) the set of all possible reactions among all the possible molecules $\mathcal{R}$, and (3) the dynamics (e.g. kinetic laws), which describes how the reactions are applied to a collection of molecules inside a reaction vessel.[16,23]

2.1 The Molecules $\mathcal{M}$

Step (1) requires that we identify all players, that is, the set of all molecular species that can appear in the model. The easiest way to specify this set is to enumerate explicitly all molecules. For example: $\mathcal{M} = \{H_2, O_2, H_2O, H_2O_2\}$. Alternatively, $\mathcal{M}$ can be defined implicitly, e.g., $\mathcal{M} =$ {all polymers that can be made from two monomers}. In this case, the set of all possible molecules can even become infinite, or at least quite large. For simplicity, but without loss of generality, we will consider here only small, explicit sets of molecular species. And we will refer to molecular species just by an index $i \in \mathcal{M}$, neglecting their structure.

2.2 The Reaction Rules $\mathcal{R}$

A reaction rule like $2H_2 + O_2 \rightarrow 2H_2O$ can be interpreted as a transformation of molecules, e.g., the transformation of hydrogen and oxygen molecules into water molecules. We will represent reaction rules by two matrixes $(l_{i,\rho})$ and $(r_{i,\rho})$, where $l_{i,\rho}$ and $r_{i,\rho}$ is the stoichiometric coefficient of molecule $i \in \mathcal{M}$ in reaction $\rho \in \mathcal{R}$ on

the lefthand side and on the righthand side, respectively. The set of all molecules $\mathcal{M}$ and a set of reaction rules $\mathcal{R}$ define a reaction network:

Definition 1 (reaction network)[1] *Given a set $\mathcal{M}$ of elements, called molecules or molecular species, and a set of reaction rules $\mathcal{R}$ given by the lefthand side and righthand side stoichiometric matrices $(l_{i,\rho})$ and $(r_{i,\rho})$, respectively, with $i \in \mathcal{M}$ and $\rho \in \mathcal{R}$. We call the pair $\langle \mathcal{M}, \mathcal{R} \rangle$ a reaction network (or algebraic chemistry, as in Ref.[10]).*

In the following $\mathrm{LHS}(\rho)$ denotes the set of molecules that appear on the left-hand side of reaction $\rho \in \mathcal{R}$. And $\mathrm{RHS}(\rho)$ the molecules on the righthand side. Furthermore we define $\mathcal{R}_A \subseteq \mathcal{R}$ as the subset of reactions that can "fire" when the molecules of the set A are present; formally $\mathcal{R}_A = \{\rho \in \mathcal{R} | \mathrm{LHS}(\rho) \subseteq A\}$.

The *stoichiometric matrix* $\mathbf{S}$ is defined as

$$\mathbf{S} = (s_{i,\rho}) = (r_{i,\rho} - l_{i,\rho}). \tag{1}$$

An entry $s_{i,\rho}$ of the stoichiometric matrix denotes the net amount of molecules of type i produced in reaction ρ.

Example 1 (reaction network, three species) *The reaction network consists of $m = 4$ molecular species $\mathcal{M} = \{H_2, O_2, H_2O, H_2O_2\}$ and $r = 2$ reaction rules $\mathcal{R} = \{\rho_1 : 2H_2 + O_2 \to 2H_2O, \rho_2 : 2H_2O_2 \to 2H_2O + O_2\}$.*

$$\mathrm{LHS}(\rho_1) = \{H_2, O_2\}, \quad \mathrm{RHS}(\rho_1) = \{H_2O\} \tag{2}$$

$$\mathrm{LHS}(\rho_2) = \{H_2O_2\}, \quad \mathrm{RHS}(\rho_2) = \{H_2O, O_2\}. \tag{3}$$

$$For\ A = \{H_2\}: \quad \mathcal{R}_A = \{\} \quad . \tag{4}$$

$$For\ A = \{H_2O_2\}: \quad \mathcal{R}_A = \{\rho_2 : 2H_2O_2 \to 2H_2O + O_2\}. \tag{5}$$

$$For\ A = \{H_2, O_2\}: \quad \mathcal{R}_A = \{\rho_1 : 2H_2 + O_2 \to 2H_2O\}. \tag{6}$$

$$(l_{i,\rho}) = \begin{pmatrix} 2\,0 \\ 1\,0 \\ 0\,0 \\ 0\,2 \end{pmatrix}, \quad (r_{i,\rho}) = \begin{pmatrix} 0\,0 \\ 0\,1 \\ 2\,2 \\ 0\,0 \end{pmatrix} \begin{matrix} H_2 \\ O_2 \\ H_2O \\ H_2O_2 \end{matrix}, \tag{7}$$

$$\mathbf{S} = \begin{pmatrix} -2 & 0 \\ -1 & 1 \\ 2 & 2 \\ 0 & -2 \end{pmatrix} \begin{matrix} H_2 \\ O_2 \\ H_2O \\ H_2O_2 \end{matrix}. \tag{8}$$

[1] From a theoretical point of view, a reaction network is a directed bipartite graph whose nodes represent molecules and reaction rules, respectively, and whose edges are weighted by stoichiometric coefficients. Further note that a reaction network as defined here with whole-numbered stoichiometric coefficients is equivalent to a Petri net.[31]

In the following, we will denote molecules with lowercase characters like *a, b, c, d* and sets of molecules by uppercase characters like *A, B, C, O, S*. A character like *a* stands for the name of a molecular species like H_2O. The following abstract example will be used throughout this chapter to illustrate the various concepts.

Example 2 (reaction network, five species) *There are five molecular species* $\mathcal{M} = \{a, b, c, d, s\}$*, which react according to the following reaction rules*

$$\mathcal{R} = \begin{cases} \rho_1 : a + s \to 2a, & \rho_6 : \emptyset \to s, \\ \rho_2 : b + s \to 2b, & \rho_7 : a \to \emptyset, \\ \rho_3 : a + b \to 2a, & \rho_8 : b \to \emptyset, \\ \rho_4 : b + c \to 2b, & \rho_9 : c \to \emptyset, \\ \rho_5 : a + d \to c + d, & \rho_{10} : s \to \emptyset \end{cases}.$$

A graphical representation of this reaction network can be found in Figure 1. The lefthand side and righthand side stoichiometric matrixes read:

$$(l_{i,\rho}) = \begin{pmatrix} 1\,0\,1\,0\,1\,0\,1\,0\,0\,0 \\ 0\,1\,1\,1\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,1\,1\,0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 1\,1\,0\,0\,0\,0\,0\,0\,0\,1 \end{pmatrix} \begin{matrix} a \\ b \\ c \\ d \\ s \end{matrix} \tag{9}$$

and

$$(r_{i,\rho}) = \begin{pmatrix} 2\,0\,2\,0\,0\,0\,0\,0\,0\,0 \\ 0\,2\,0\,2\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1\,0\,0\,0\,0 \end{pmatrix} \begin{matrix} a \\ b \\ c. \\ d \\ s \end{matrix} \tag{10}$$

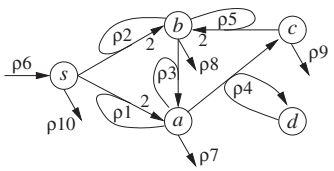

Figure 1. Reaction network of Example 2. A "2" refers to the stoichiometry, e.g. in the reaction rule $\rho_1 : a + s \longrightarrow 2a$

Subtracting $(l_{i,\rho})$ *from* $(r_{i,\rho})$ *leads to the stoichiometric matrix* $\mathbf{S} = (s_{i,\rho}) = (r_{i,\rho}) - (l_{i,\rho})$*:*

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{matrix} a \\ b \\ c. \\ d \\ s \end{matrix} \tag{11}$$

2.3 Dynamics

A reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ specifies the structure of a reaction system, but does not contain any notion of time. A common way to specify the dynamics of the reaction system is by using a system of ordinary differential equations of the following form:

$$\dot{\mathbf{x}}(t) = \mathbf{S}\mathbf{v}(\mathbf{x}(t)) \tag{12}$$

where $\mathbf{x} = (x_1, \ldots, x_m)^T \in \mathbb{R}^m$ is a concentration vector depending on time t, $\mathbf{S}$ a stoichiometric matrix, and $\mathbf{v} = (v_1, \ldots, v_r)^T \in \mathbb{R}^r$ a flux vector depending on the current concentration vector. A flux $v_\rho \geq 0$ describes the velocity or turnover rate of reaction $\rho \in \mathcal{R}$. The actual value of v_ρ depends usually on the concentration of the species participating in the reaction ρ (i.e., $\mathrm{LHS}(\rho)$). In order to avoid unwanted mathematical effects, we demand that $\mathbf{v}$ is differentiable, meaning intuitively that it depends "smoothly" on the concentration vector $\mathbf{x}$. Beside this mathematical assumption, there are (at least) two further assumptions that are due to the nature of reaction systems. These assumptions relate the function $\mathbf{v}$ to the reaction rules $\mathcal{R}$:

Assumptions 1: If a species i is necessary for a reaction ρ to take place, it must appear on the lefthand side of that reaction (i.e., $i \in \mathrm{LHS}(\rho)$). This implies that for all molecules $i \in \mathcal{M}$ and reactions $\rho \in \mathcal{R}$ with $i \in \mathrm{LHS}(\rho)$, if $x_i = 0$ then $v_\rho = 0$. The flux v_ρ must be zero, if the concentration x_i of a molecule appearing on the lefthand side of this reaction is zero. This assumption meets the obvious fact that a molecule has to be present to react.

Assumptions 2: If all species $\mathrm{LHS}(\rho)$ of a reaction $\rho \in \mathcal{R}$ are present in the reactor (e.g. for all $i \in \mathrm{LHS}(\rho), x_i > 0$) the flux of that reaction is positive, (i.e., $v_\rho > 0$). In other words, the flux v_ρ must be positive, if all molecules required for that reaction are present, even in small quantities (cf. Definition 14 (instance) and Definition 15 (abstraction)).

There is a large amount of kinetic laws fulfilling these assumptions, including all laws that are usually applied in practice. The most fundamental of such kinetic laws is mass-action kinetics, which is just the product of the concentrations of the interacting species:

$$v_\rho = \prod_{i \in \mathcal{M}} x^{l_{i,\rho}}. \tag{13}$$

It should be noted that more complicated laws like Michaelis-Menten kinetics are derived from mass-action kinetics. This is especially true for many laws describing inhibition.[23] So, we may interprete these more complicated laws as "syntactic sugar" that allows to describe large reaction systems in a more compact way, that is, with a smaller number of species.

2.4 Modifiers

When applying chemical organisation theory, we can neglect modifiers that inhibit a reaction, because they are not necessary for the reaction to occur (Assumption 2) and they usually do not switch off the reaction completely, which follows from the fact that laws of inhibition are derived from Michaelis-Menten kinetics. However, in case an inhibitory effect is so strong that it practically switches off a reaction completely, it has to be considered (not shown here).

For the theory, we can also neglect modifiers that enhance a reaction, as long as they are not necessary for that reaction to take place. Note that a modifier i that is necessary for a reaction ρ has to appear as a catalyst on the lefthand side of that reaction, that is, $i \in \mathrm{LHS}(\rho)$ (Assumption 2) and on its righthand side (i.e., $i \in \mathrm{RHS}(\rho)$).

2.5 Input and Output

There are many processes that give rise to an inflow and outflow, such as, incident sunlight, decaying molecules, or a general dilution flow. In this chapter we interpret the reaction rule $\emptyset \to a$ as an input of a, and $a \to \emptyset$ as an output of a ($\emptyset$ denoting the empty set, which is, from a mathematical point of view, assumed to be always present). In Example 2 (Figure 1) there is an inflow of s whereas a, b, c, and s decay spontaneously.

2.6 Unbalanced Reaction Systems

As sometimes otherwise stated, in chemical reaction system models the masses on the left hand side and right hand side can differ. That means that, formally, a reaction can produce or consume mass. This might appear unrealistic, since it is generally assumed that in a real chemical reaction mass is conserved.

However in a reaction system *model* it makes sense to consider unbalanced reaction rules, too, which can lead to more elegant and simpler models. We have already encountered two examples in the previous section, namely, inflow and outflow reaction rules. Further examples are models of exponential growth (e.g., $a \to 2a$) and other models assuming implicitly an unlimited substrate. This substrate is removed to obtain a simpler model, compare Example 9 (hypercycle without an explicit substrate, Section 3.5.1) with Example 10 (hypercycle with an explicit substrate, Section 3.5.2).

Chemical organisation theory can also deal with unbalanced reaction systems, including those where some molecular species cannot operate at steady state and can exhibit unlimited growth.

3. CHEMICAL ORGANISATION THEORY: STATIC PART

The first part of the theory deals with the static structure of a reaction system, that is, the molecules $\mathcal{M}$ and the reactions $\mathcal{R}$. Instead of considering a state, e.g. a concentration vector, we limit ourself to the analysis of the set of molecules present in that state.

In classical analysis, we study the movement of the system in state space. Instead, here, we consider the movement from one set of molecules to another. As in the classical analysis of the dynamics of the system, where fixed points and attractors are considered more important than other states, some sets of molecules are more important than others.

In order to find those sets, we introduce some properties that define them, namely: closure, semi-self-maintenance, semi-organisation, self-maintenance, and finally being an organisation. All definitions herein refer to a reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$.

3.1 Closed Sets

The first property of a set of molecules, called closure, assures that no new molecular species can be generated by the reactions inside the set or equivalently that all molecules that can be generated by reactions inside the set are already inside that set.

Definition 2 (closed set[20]) *A set $C \subseteq \mathcal{M}$ is closed, if for all reactions $\rho \in \mathcal{R}_C$,* $\mathrm{RHS}(\rho) \subseteq C$.

Given a set $A \subseteq \mathcal{M}$, we can always *generate* its closure $G_{CL}(A)$ according to the following definition:

Definition 3 (generate closed set[20]) *Given a set of molecules $A \subseteq \mathcal{M}$, we define $G_{CL}(A)$ as the smallest closed set C containing A. We say that A generates the closed set $C = G_{CL}(A)$ and we call C the closure of A.*

In passing we note that this definition is unambiguous: Let us suppose, ad absurdum, that we can find two smallest closed sets $C_1 \neq C_2$ both containing C. Against our assumption, their intersection would be an even smaller closed set containing C, because the intersection of closed sets is obviously closed, too.

We can generate the closed set for a given set A efficiently by the following algorithm: add all reaction products among molecules of A and insert them into A and repeat this procedure until no new molecule can be inserted anymore (for infinite systems, a limit has to be taken).

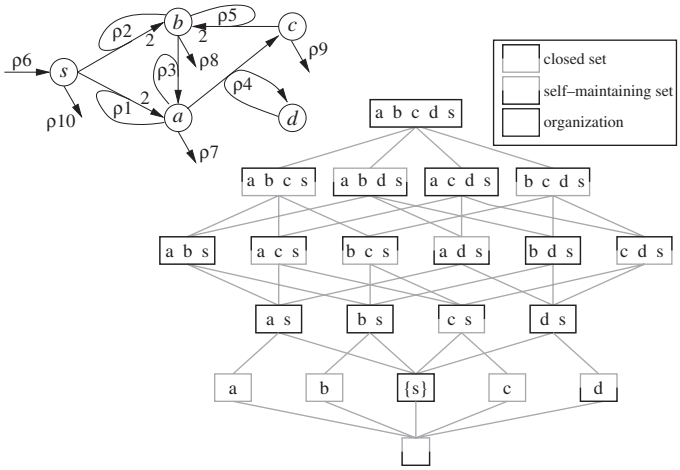

Figure 2. Fraction of the lattice of all sets of molecules from Example 2. Shown are all sets containing molecule s and small sets like $\{a\}$ and $\{\}$. Note that an organisation is a closed *and* self-maintaining set

Example 3 (closed sets) *In our Example 2, Figure 2, there are 14 closed sets:* $\mathcal{O}_{CL} = \{\{s\}, \{a, s\}, \{b, s\}, \{c, s\}, \{d, s\}, \{a, b, s\}, \{a, c, s\}, \{b, d, s\}, \{c, d, s\}, \{a, b, c, s\}, \{a, c, d, s\}, \{b, c, d, s\}, \{a, b, c, d, s\}\}$. *The empty set is not closed, because there is an inflow of molecule s. The set* $A = \{a, d, s\}$ *is not closed, because c can be produced by the reaction* $a + d \rightarrow d + c$. *The closed set generated by A is:* $C = G_{CL}(A) = \{a, c, d, s\}$.

A common algebraic concept that we shall use very often, from now on, is the lattice. A lattice, is a partially ordered set (poset) in which any two elements have a greatest lower bound (here, their intersection) and a least upper bound (here, their union).

Given the generate closed set operator we can define two basic operations, a union operation ($U \sqcup_{CL} V$) and an intersection operation ($U \sqcap_{CL} V$) on closed sets:

$$U \sqcup_{CL} V \equiv G_{CL}(U \cup V), \quad \text{and} \tag{14}$$

$$U \sqcap_{CL} V \equiv G_{CL}(U \cap V). \tag{15}$$

Trivially, closed sets, with the operations $\sqcup_{CL}$ and $\sqcap_{CL}$, form a lattice $\langle \mathcal{O}_{CL}, \sqcup_{CL}, \sqcap_{CL} \rangle$.

Closure is important because the closed set generated by a set (its closure) represents the largest possible set that can be reached from a given set of molecules. Furthermore a set that is closed cannot generate new molecules and is in that sense more stable with respect to novelty. As such the concept of closure alone

can already give valuable insight into the structure and organisation of complex chemical networks as shown by Ebenhöh et al..[13]

3.2 Semi-Self-Maintaining Sets

Before we introduce the second important property called "self-maintenance", an intermediate step will be taken by defining a property that is necessary for a set to be self-maintaining but not sufficient. This property, called *semi-self-maintenance*, is easier to check then the property self-maintenance. Furthermore, in reaction systems like *catalytic flow system* (Section 3.5.1), every semi-self-maintaining set is also self-maintaining, and thus it is sufficient to check in those systems just the semi-self-maintenance property.

The property *semi-self-maintaining* assures that every molecule that is consumed within a set, is produced within that set.

We say that a molecule $i \in \mathcal{M}$ is *produced* within a set $A \subseteq \mathcal{M}$, if there exists a reaction $\rho \in \mathcal{R}_A$ with $s_{i,\rho} > 0$. In the same way, we say that a molecule $i \in A$ is *consumed* within the set A, if there is a reaction $\rho \in \mathcal{R}_A$ with $s_{i,\rho} < 0$.

Definition 4 (semi-self-maintaining set[10]) *A set of molecules $S \subseteq \mathcal{M}$ is called semi-self-maintaining, if all molecules $i \in S$ that are consumed within S are also produced within that set S.*

Example 4 (semi-self-maintaining sets) *In our Example 2 there are 13 semi-self-maintaining sets:* $\mathcal{O}_{SSM} = \{\{\}, \{s\}, \{d\}, \{a, s\}, \{b, s\}, \{d, s\}, \{a, b, s\}, \{a, d, s\}, \{b, d, s\}, \{a, b, c, d\}, \{a, b, d, s\}, \{a, c, d, s\}, \{a, b, c, d, s\}\}$.

Note that the concept of (semi-) self-maintenance is closely related to the concept of an autocatalytic set.[14,27,34] An autocatalytic set is usually defined as a set of molecules such that each molecule is produced by at least one catalytic reaction within that set.[25]

3.3 Self-Maintaining Sets

In a semi-self-maintaining set, all molecules that are consumed are produced; yet this does not guarantee that the total amount of mass can be maintained.

Example 5 (reversible reaction) *A simple counterexample is the following reversible reaction in a flow reactor:* $\mathcal{M} = \{a, b\}$, $\mathcal{R} = \{a \to b, b \to a, a \to \emptyset, b \to \emptyset\}$. *Both molecules, a and b, decay.* $S = \{a, b\}$ *is a semi-self-maintaining set, because a is produced by the reaction* $b \to a$, *and b is produced by the reaction*

$a \to b$. But, obviously, the system $\{a, b\}$ is not stable, in the sense that there cannot be a stationary state in which the two molecules a and b have positive concentrations: both molecules decay and cannot be sufficiently reproduced, and thus they will finally vanish.

The solution to this problem is to consider the overall ability of a set to maintain its total mass. We call such sets simply *self-maintaining*.

Definition 5 (self-maintaining[10]) *Given an algebraic chemistry $\langle \mathcal{M}, \mathcal{R} \rangle$ with $m = |\mathcal{M}|$ molecules and $r = |\mathcal{R}|$ reactions, and let $\mathbf{S} = (s_{i,j})$ be the $(m \times r)$ stoichiometric matrix implied by the reaction rules $\mathcal{R}$, where $s_{i,\rho}$ denotes the number of molecules of type i produced in reaction ρ. A set of molecules $S \subseteq \mathcal{M}$ is called self-maintaining, if there exists a flux vector $\mathbf{v} \in \mathbb{R}^r$ such that the following three conditions apply: (1) for all reactions $\rho \in \mathcal{R}_S$ the flux $v_\rho > 0$; (2) for all remaining reactions $\rho \notin \mathcal{R}_S$, the flux $v_\rho = 0$; and (3) for all molecules $i \in S$, the production rate $(\mathbf{Sv})_i \geq 0$.*

v_ρ denotes the element of $\mathbf{v}$ describing the flux (i.e., velocity) of reaction ρ. $(\mathbf{Sv})_i$ is the production rate of molecule i given flux vector $\mathbf{v}$. It is practically the sum of fluxes producing i minus the fluxes consuming i.

Example 6 (reversible reaction (cont.)) *For the reversible reaction, Example 5, the stoichiometric matrix becomes $\mathbf{S} = ((-1, 1), (1, -1), (-1, 0), (0, -1))$, and we can see that there is no positive flux vector $\mathbf{v} \in \mathbb{R}^4$, such that $\mathbf{Sv} \geq \mathbf{0}$. In other words, no matter how we chose the velocity of the reactions, it is not possible to keep a and b in a reaction vessel. In fact, in that example, only the empty set $\{\}$ is self-maintaining. In case a and b would not decay, $\mathcal{R} = \{a \to b, b \to a\}$, the set $\{a, b\}$ would be (as desired) self-maintaining, because there is a flux vector, e.g., $\mathbf{v} = (1.0, 1.0)$, such that $\mathbf{Sv} = \mathbf{0} \geq \mathbf{0}$ with $\mathbf{S} = ((-1, 1), (1, -1))$.*

Example 7 (self-maintaining) *In our five species network (Example 2, Figure 2) all semi-self-maintaining sets except $\{a, b, c, d\}$ are also self-maintaining. The criterion for self-maintenance will be illustrated by looking at the self-maintaining set $S = \{s, a, d\}$ in more detail. First we have to find all the reactions active within S, this means to find all reactions $\rho \in \mathcal{R}_S \subseteq \mathcal{R}$ where the left hand side consists of molecules from S (i.e., $\mathrm{LHS}(\rho) \subseteq S$). There are five such reactions: $\mathcal{R}_S = \{\rho_1 : a + s \to 2a, \rho_5 : a + d \to d + c, \rho_6 : \emptyset \to s, \rho_7 : a \to \emptyset, \rho_{10} : s \to \emptyset\}$ These reactions correspond to column 1, 5, 6, 7, and 10 of the stoichiometric matrix $\mathbf{S}$, Eq. (11), and to the fluxes v_1, v_5, v_6, v_7, and v_{10}. According to the definition of self-maintenance, these five fluxes must be positive while the remaining fluxes must be zero. Now, in order to show that $S = \{s, a, d\}$ is self-maintaining, we have to find positive values for v_1, v_5, v_6, v_7, and v_{10} such that a, c, and s are produced at a non-negative rate.*

Here, this can be achieved by setting $v_1 = 3, v_5 = 1, v_6 = 10, v_7 = 1, v_{10} = 1$. *When multiplying this flux vector with the stoichiometric matrix,*

$$\mathbf{Sv} = \begin{pmatrix} 1 & 0 & 1 & 0 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} 3 \\ 0 \\ 0 \\ 0 \\ 1 \\ 10 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 6 \end{pmatrix} \begin{matrix} a \\ b \\ c, \\ d \\ s \end{matrix}$$

we can see that all molecules of $S = \{a, d, s\}$ *are produced at a non-negative rate. For example,* d *is produced at rate* 0 *and* s *is produced at rate 6. So, we can conclude that* $\{a, d, s\}$ *is self-maintaining. Note further that, in Example 2, molecule* d *will always be produced at zero rate* $((\mathbf{Sv})_4 = 0)$*, independently on how we chose* **v**.

In a self-maintaining set S every molecule that is consumed by a reaction $\rho \in \mathcal{R}_s$ must be also produced by a reaction within that set, in order to achieve a non-negative production of that molecule. Therfore we can conclude, as mentioned previously:

Lemma 1 [10] *Every self-maintaining set is semi-self-maintaining (proof in Ref.* [10]*).*

In other words, if a set is not semi-self-maintaining – a property easy to check – the set cannot be self-maintaining. The opposite is not generally true; there are sets that are semi-self-maintaining but not self-maintaining, such as $\{a, b, c, d\}$ in Example 2.

Note that the empty set is always self-maintaining (as it has nothing to maintain) and the set of all possible molecules is always closed (as there is nothing that can be added to it). On the other hand the empty set is not necessarily closed, and the set of all possible molecules is not necessarily self-maintaining.

3.4 Organisations

Together, closure and self-maintenance lead to the central definition of this approach[2]:

Definition 6 (organisation[10,20]) *A set of molecules* $O \subseteq \mathcal{M}$ *that is both closed and self-maintaining is called an organisation.*

[2] Note that our definition of an organisation intentionally reads like the definition by Fontana and Buss.[20] However, we are using a more general definition of self-maintenance, which includes the definition by Fontana and Buss and thus ensures "compatibility" with their approach.

An organisation represents an important combination of molecular species, which are likely to be observed in a large reaction vessel on the long run (cf. Theorem 1, Section 4.2). A set of molecules that is not closed would not exist for a long time, because new molecules will appear, changing that set. A set of molecules that is not self-maintaining will also inevitably change, since molecules not-maintained will vanish.

In the same way as we defined an organisation, we can define a *semi-organisation* as a closed and semi-self-maintaining set of molecules. From Lemma 1 trivially follows that every organisation is a semi-organisation.

Example 8 (organisations) *In our example (Example 2) there are 14 closed sets, 13 semi-self-maintaining sets, 12 self-maintaining sets, and 8 semi-organisations. All 8 semi-organisations are also organisations:* $\mathcal{O} = \{\{s\}, \{a, s\}, \{b, s\}, \{d, s\}, \{a, b, s\}, \{b, d, s\}, \{a, c, d, s\}, \{a, b, c, d, s\}\}$. *Although the reaction system is small, its organisational structure is already difficult to see when looking at the rules or their graphical representation. In Figure 2, a Hasse diagram with all 16 possible sets of molecules containing the substrate s together with some smaller sets is shown.*

Finding all organisations of a general reaction system appears to be computationally difficult (Section 6.2). One approach is to find the semi-organisations first, and then check, which of them are also self-maintaining.

3.5 Consistent Reaction Systems

The property of the set of organisations and semi-organisations depends strongly on the type of system studied. In this section we discuss a class of systems, called consistent reaction systems, where the set of organisations always forms an algebraic lattice, that is, there is a unique union and intersection of organisations and there is a unique largest organisation (if there is a finite number of organisations).

Definition 7 (consistent[10]) *A reaction network is called* consistent*, if (1) given any two (semi-)self-maintaining sets A and B then their set-union* $A \cup B$ *is also (semi-)self-maintaining; and (2) the closure* $G_{CL}(A)$ *of any (semi-)self-maintaining set A is (semi-)self-maintaining.*

Now, in Sections 3.5.1–3.5.3, we will present three types of consistent systems, which can be easily (i.e., in linear time) identified by looking at the reaction rules only. In practice it makes sense to check first, whether the system to be analysed falls into one of these classes.

3.5.1 Catalytic flow system

In a *catalytic flow system* all molecules $i \in \mathcal{M}$ are consumed by first-order reactions of the form $i \rightarrow \emptyset$ (dilution) and there is no molecule consumed by any other reaction. So, each molecule i decays spontaneously, or equivalently, is removed by a dilution flow. Apart from this, each molecule can appear only as a catalyst (without being consumed).

Definition 8 (catalytic flow system[10]) *A reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ is called a catalytic flow system, if for all molecules $i \in \mathcal{M}$: (1) there exists a first-order decay reaction $\rho \in \mathcal{R}$ with $\mathrm{LHS}(\rho) = \{i\}$ and $s_{i,\rho} = -1$; (2) for all reaction $\rho \in \mathcal{R}$ with $s_{i,\rho} < 0$, $s_{i,\rho} = -1$ and $\mathrm{LHS}(\rho) = \{i\}$ (reactions that consume i must be first-order decay reactions).*

Examples of *catalytic flow system* are the replicator equation,[35] the hypercycle,[14,15] the more general *catalytic network equation*,[40] various models of autocatalytic sets,[3,24] and AlChemy.[19] Furthermore some models of genetic regulatory networks and social system[11] are *catalytic flow systems*.

Example 9 (catalytic flow system) *The three-membered elementary hypercycle[14,15] under flow condition can be represented by three molecular species $\mathcal{M} = \{a, b, c\}$ and six reaction rules:*

$$\begin{aligned} \mathcal{R} = \{ & a + b \rightarrow a + 2b, \qquad a \rightarrow \emptyset, \\ & b + c \rightarrow b + 2c, \qquad c \rightarrow \emptyset \\ & c + a \rightarrow c + 2a, \qquad b \rightarrow \emptyset \}. \end{aligned}$$

We can see, that all three molecules decay by first-order reactions (representing the dilution flow), and that no molecule is consumed by any of the three remaining reactions. There are two organisations: $\{\}$ and $\{a, b, c\}$. The catalytic reaction rules are not balanced, since a substrate available at a constant concentration is implicitly assumed to be consumed.

In a *catalytic flow system* we can easily check a set for being self-maintaining, because:

Lemma 2 [10] *In a catalytic flow system, all semi-self-maintaining sets are self-maintaining (Proof in Ref.[10]).*

From which follows immediately:

Lemma 3 [10] *In a catalytic flow system, every semi-organisation is an organisation.*

So, in a *catalytic flow system* we can easily check, whether a set O is an organisation by just checking whether it is closed and whether each molecule in that set is produced by that set. Furthermore, given a set A, we can always generate

an organisation by adding all molecules produced by A until A is closed and then removing molecules that are not produced until A is (semi-) self-maintaining. With respect to the intersection and union of (semi-) organisations the set of all (semi-) organisations of a *catalytic flow system* forms an algebraic lattice (see below). A result which has already been noted by Fontana and Buss.[20]

3.5.2 Reactive flow system

In a *reactive flow system* all molecules are consumed by first-order reactions of the form $\{i\} \to \emptyset$ (dilution). But as opposed to the previous system, we allow arbitrary additional reactions in $\mathcal{R}$ and do not restrict these reactions to be catalytic. Thus note how a *catalytic flow system* is a particular kind of *reactive flow system*.

Definition 9 (reactive flow system[10]) *A reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ is called a reactive flow system, if for all molecules $i \in \mathcal{M}$, there exists a first-order decay reaction $\rho \in \mathcal{R}$ with* $\mathrm{LHS}(\rho) = \{i\}$ *and* $s_{i,\rho} = -1$.

This is a typical situation for chemical flow reactors or bacteria that grow and divide.[33] Note that in a cell that grows and divides, every molecule including the genome is subject to a dilution flow.

In a *reactive flow system*, semi-organisations are not necessarily organisations. Nevertheless, both the semi-organisations and the organisations form a lattice $\langle \mathcal{O}, \sqcup, \sqcap \rangle$. Moreover, the union ($\sqcup$) and intersection ($\sqcap$) of any two organisations is an organisation (see below).

Example 10 (reactive flow system) *As an example, we take the three-membered elementary hypercycle as before, but add an explicit substrate s. There are four molecules $\mathcal{M} = \{a, b, c, s\}$ and seven reaction rules:*

$$\begin{aligned} \mathcal{R} = \{ & a + b + s \to a + 2b, \qquad a \to \emptyset, \\ & b + c + s \to b + 2c, \qquad b \to \emptyset, \\ & c + a + s \to c + 2a, \qquad c \to \emptyset, \\ & \qquad\qquad\qquad\quad s \to \emptyset \}. \end{aligned}$$

We can see, that again all molecules decay by first-order reactions (representing the dilution flow), but now a molecule, s, is consumed by other reactions. Note that in this example there is only one organisation (the empty organisation), and even no semi-organisation that contains one or more molecules, because there is no inflow of the substrate. If we would add a reaction equation $\emptyset \to s$ representing an inflow of the substrate, we would obtain two organisations: $\{s\}$ and the "hypercycle" $\{a, b, c, s\}$.

3.5.3 Reactive flow system with persistent molecules

In a *reactive flow system with persistent molecules* there are two types of molecules: persistent molecules and non-persistent molecules. All non-persistent molecules are consumed (as in the two systems before) by first-order decay reactions, whereas a persistent molecule is not consumed by any reaction at all.

Definition 10 (reactive flow system with persistent molecules[10]) *A reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ is called a reactive flow system with persistent molecules, if we can partition the set of molecules in persistent P and non-persistent molecules $\bar{P}(\mathcal{M} = P \cup \bar{P}, P \cap \bar{P} = \emptyset)$ such that: (i) for all non-persistent molecules $i \in \bar{P}$: there exists a first-order decay reaction $\rho \in \mathcal{R}$ with $\mathrm{LHS}(\rho) = \{i\}$ and $s_{i,\rho} = -1$; and (ii) for all persistent molecules $i \in P$: there does not exist a reaction $\rho \in \mathcal{R}$ with $s_{i,\rho} < 0$.*

An example of a *reactive flow system with persistent molecules* is Example 2, where d is a persistent molecule. The *reactive flow system with persistent molecules* is the most general of the three systems where the semi-organisations and organisations always form a lattice, and where the generate organisation operator can properly be defined (see below). As in a *reactive flow system*, not all semi-organisations are organisations.

Lemma 4 [10] *A reactive flow system with persistent molecules is consistent (Proof in Ref.[10]).*

Remember that the intersection of two closed sets is again always closed. In a similar way, in consistent reaction systems, the union of self-maintaining sets is again self-maintaining. However, Section 3.7 will demonstrate that the latter is not necessarily true for general reaction systems.

3.6 Common Properties of Consistent Reaction Systems

Consistent reaction systems (including those discussed in Sections 3.5.1–3.5.3) possess some comfortable properties that allow us to present a series of useful definitions and lemmas.

In a consistent reaction system, given a set of molecules A, we can uniquely *generate* a semi-self-maintaining set, a semi-organisation, a self-maintaining set, and an organisation in a similar way as we have generated a closed set. And like for closed sets, we can define the union and intersection on semi-self-maintaining sets, semi-organisations, self-maintaining sets, and organisations, respectively. Furthermore, each, the semi-self-maintaining sets, semi-organisations, self-maintaining sets, and organisations form a lattice together with their respective union and intersection operators. This does not generalise to general reaction systems (see Section 3.7), because for a general reaction system we cannot uniquely generate

a semi self-maintaining set, a semi-organisation, a self-maintaining set, nor an organisation as is the case with consistent reaction systems.

3.6.1 Generate semi-self-maintaining set

Definition 11 (generate semi-self-maintaining set[10]) *Given a set of molecules $A \subseteq \mathcal{M}$ of a consistent reaction system, we define $G_{SSM}(C)$ as the biggest semi-self-maintaining set S contained in A. We say that A generates the semi-self-maintaining set $S = G_{SSM}(A)$.*

In order to calculate the semi-self-maintaining set generated by **A**, we remove those molecules that are consumed and not produced within **A**, until all molecules consumed are also produced, and thus reaching a semi-self-maintaining set. The operator G_{SSM} (generate semi-self-maintaining set) implies the union $\sqcup_{SSM}$ and intersection $\sqcap_{SSM}$ on semi-self-maintaining sets: Given two semi-self-maintaining sets S_1 and S_2, the semi-self-maintaining sets generated by their union $(S_1 \sqcup_{CL} S_2)$ and intersection $(S_1 \sqcap_{CL} S_2)$ are defined as: $S_1 \sqcup_{SSM} S_2 \equiv G_{SSM}(S_1 \cup S_2)$, and $S_1 \sqcap_{SSM} S_2 \equiv G_{SSM}(S_1 \cap S_2)$, respectively.

3.6.2 Generate self-maintaining set

Definition 12 (generate self-maintaining set[10]) *Given a set of molecules $C \subseteq \mathcal{M}$ of a consistent reaction system, we define $G_{SM}(C)$ as the biggest self-maintaining set S contained in C. We say that C generates the self-maintaining set $S = G_{SM}(C)$.*

For consistent reaction systems, $G_{SM}(C)$ is always defined, because the union $(\cup)$ of two self-maintaining sets is self-maintaining; and further, every set is either self-maintaining, or it contains a unique biggest self-maintaining set. Thus from every set we can generate a self-maintaining set. Note that self-maintaining sets are also semi-self-maintaining, $G_{SM}(G_{SSM}(S)) \equiv G_{SSM}(S)$, which is a useful property, because $G_{SSM}(S)$ is easier to compute. As usual, the union $\sqcup_{SM}$ and intersection $\sqcap_{SM}$ of self-maintaining sets S_1, S_2 are defined as $S_1 \sqcup_{SM} S_2 \equiv G_{SM}(S_1 \cup S_2)$, $S_1 \sqcap_{SM} S_2 \equiv G_{SM}(S_1 \cap S_2)$, respectively. Thus also the set of all self-maintaining sets $\mathcal{O}_{SM}$ forms a lattice $\langle \mathcal{O}_{SM}, \sqcup_{SM}, \sqcap_{SM} \rangle$. If S is self-maintaining, its closure $G_{CL}(S)$ is self-maintaining, too (again, not valid for *general reaction systems*).

3.6.3 Generate organisation

There are many ways in which we can generate an organisation from a set. We will present here the simplest one, which implicitly assumes that molecules are produced quickly and vanish slowly. This assumption leads to the largest possible organisation generated by a set:

Definition 13 (generate organisation[10,20]) *Given a set of molecules $A \subseteq \mathcal{M}$ of a consistent reaction system, we define*

$$G(A) = G_{SM}(G_{CL}(A)). \tag{16}$$

We say that A generates the organisation $O = G(A)$.

Equivalently we can also generate a semi-organisation $O = G_{SO}(A) = G_{SSM}(G_{CL}(A))$.

Following the same scheme as before, the union $\sqcup$ and intersection $\sqcap$ of two organisations O_1 and O_2 is defined as the organisation generated by their set-union and set-intersection:

$$O_1 \sqcup O_2 \equiv G(O_1 \cup O_2), \tag{17}$$

$$O_1 \sqcap O_2 \equiv G(O_1 \cap O_2). \tag{18}$$

Equivalently $G(A) = G_{SM}(G_{SSM}(G_{CL}(A)))$, which allows to compute the organisation generated by a set more easily in three steps.

Thus, for consistent reaction systems, also the set of all organisations $\mathcal{O}$ forms a lattice $\langle \mathcal{O}, \sqcup, \sqcap \rangle$. This important fact should be emphasised by the following lemma:

Lemma 5 [10] *Given a reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ of a consistent reaction system and all its organisations $\mathcal{O}$, then $\langle \mathcal{O}, \sqcup, \sqcap \rangle$ is a lattice.*

Knowing that the semi-organisations and organisations form a lattice, and that we can uniquely generate an organisation for every set, is a useful information. Then we know, for example, that there is a largest organisation. Furthermore we can map every state of a reaction vessel uniquely to an organisation (Section 4.1), which allows to characterise states and to partition the state space, uniquely.

In order to find the whole set of organisations, it is impractical just to check all the possible sets of molecules. Instead, we can start by computing the lattice of semi-organisations, and then test only those sets for self-maintenance. Furthermore, if the semi-organisations form a lattice, we can start with small sets of molecules and generate their semi-organisations, while the $\sqcup_{SO}$ operator can lead us to the more complex semi-organisations.

3.7 General Reaction Systems

When we consider general reaction systems, that is, reaction networks without any constraints, we cannot always generate a self-maintaining set uniquely. This implies that in a general reaction system neither the set of organisation nor the set of semi-organisations necessarily form a lattice. Examples of this can be found in planetary atmosphere chemistries.[8,42]

Example 11 (Reaction system without a lattice of organisation) *In this example we present a simple reaction network where the set of organisations does not form a lattice, and where we can not always generate an organisation for any given set of molecules:*

$$\mathcal{M} = \{a, b, c\}, \quad \mathcal{R} = \{a + b \rightarrow c\}. \tag{19}$$

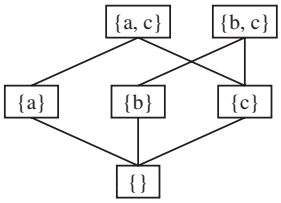

Figure 3. Example of a general reaction system that does not have a lattice of organisations. There are three molecules $\mathcal{M} = \{a, b, c\}$ and just one reaction rule $\mathcal{R} = \{a + b \rightarrow c\}$

This example can be interpreted as an isolated system, where there is no inflow nor outflow. a and b simply react to form c. Obviously, every set that does not contain a together with b is an organisation. So, there are 6 organisations: $\{\}$, $\{a\}$, $\{b\}$, $\{c\}$, $\{a, c\}$, *and* $\{b, c\}$. *As illustrated by Figure 3, there is no unique largest organisation and therefor the set of organisations does not form a lattice. Furthermore, given the set* $\{a, b, c\}$, *we can not generate an organisation uniquely, because there does not exist a unique largest self-maintaining set contained in* $\{a, b, c\}$. *There are two self-maintaining sets of equal size:* $\{a, c\}$ *and* $\{b, c\}$. *Why can this not happen in a reactive flow system with persistent molecules ? In a reactive flow system with persistent molecules, the set-union of two self-maintaining sets is again self-maintaining. Therefore there can not exist two largest self-maintaining sets within a set A, because their union would be a larger self-maintaining set within A (see Lemma 5). Note that each organisation makes sense, because each organisation represents a combination of molecules that can stably exists in a reaction vessel, which does not allow an outflow of any of the molecules according to the rules* $\mathcal{R}$.

4. CHEMICAL ORGANISATION THEORY: DYNAMICAL PART

The previous section deals with molecules $\mathcal{M}$ and their reaction rules $\mathcal{R}$, but not with the evolution of the system in time. However, the structures that we identified (i.e., the organisations) possess a strong relation to the potential dynamical behaviour of the reaction system. We will now get back the dynamics into our consideration.

To add dynamics to the theory, we have to formalise the dynamics of a system. In a very general approach, the *dynamics* is given by a *state space* X and a formal definition (mathematical or algorithmic) that describes all possible movements in X for any possible initial state $\mathbf{x}_0 \in X$. For simplicity, we assume a deterministic dynamical process described by a system of ordinary differential equations (Eq. (12), Section 2.3).

4.1 Connecting to the Static Theory

Let us assume now that $\mathbf{x} \in X$ represents the state of a reaction vessel, which contains molecules from $\mathcal{M}$. In the static part of the theory we consider just the

set of molecular species present in the reaction vessel, but not their concentrations, spatial distributions, velocities, and so on.

Now, given the state $\mathbf{x}$ of the reaction vessel, we need a function that maps uniquely this state to the set of molecules present. Vice versa, given a set of molecules $A \subseteq \mathcal{M}$, we need to know, which states from X correspond to this set of molecules. For this reason we introduce a mapping ϕ called *abstraction*, from X to $\mathcal{P}(\mathcal{M})$, which maps a state of the system to the set of molecules that are present in the system being in that state. The exact mapping can be defined precisely later, depending on the state space, on the dynamics, and on the actual application.

The concept of *instance* is the opposite of the concept of abstraction. While $\phi(\mathbf{x})$ denotes the molecules represented by the state $\mathbf{x}$, an instance $\mathbf{x}$ of a set A is a state where exactly the molecules from A are present according to the function ϕ.

Definition 14 (instance[10]) *Given a function $\phi : X \to \mathcal{P}(\mathcal{M})$ (called abstraction), which maps a state to a set of molecules, we say that a state $\mathbf{x} \in X$ is an instance of $A \subseteq \mathcal{M}$, if and only if $\phi(\mathbf{x}) = A$.*

In particular, we can define an instance of an organisation O (if $\phi(\mathbf{x}) = O$) and an instance of a generator of O (if $G(\phi(\mathbf{x})) = O$). Loosely speaking we can say that $\mathbf{x}$ *generates organisation* O.

Note that a state $\mathbf{x}$ of a consistent reaction system (Section 3.5) is *always* an instance of a generator of one and only one organisation O. This leads to the important observation that a lattice of organisations partitions the state space X, where a partition $X_O \subseteq X$ implied by organisation O is defined as the set of all instance of all generators of O:

$$X_O = \{\mathbf{x} \in X | G(\phi(\mathbf{x})) = O\} \quad \text{(states generating organisation } O\text{)}. \tag{20}$$

Note that as the system state evolves over time, the organisation $G(\phi(\mathbf{x}(t)))$ generated by $\mathbf{x}(t)$ might change (see below, Figure 4).

4.2 Fixed Points are Instances of Organisations.

Now we will describe a theorem that relates fixed points to organisations, and by doing so, underlines the relevancy of organisations. We will show that, given an ordinary differential equation (ODE) of a form that is commonly used to describe the dynamics of reaction systems, every fixed point of this ODE is an instance of an organisation. We therefore assume in this section that $\mathbf{x}$ is a concentration vector $\mathbf{x} = (x_1, x_2, \ldots, x_{|\mathcal{M}|})$, $X = \mathbb{R}^{|\mathcal{M}|}$, $x_i \geq 0$ where x_i denotes the concentration of molecular species i in the reaction vessel, and $\mathcal{M}$ is finite.

The dynamics is given by an ODE of the form $\dot{\mathbf{x}} = \mathbf{S}\mathbf{v}(\mathbf{x})$ where $\mathbf{S}$ is the stoichiometric matrix implied by the reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ (see Section 2.2). $\mathbf{v}(\mathbf{x}) = (v_1(\mathbf{x}), \ldots, v_r(\mathbf{x})) \in \mathbb{R}^r$ is a flux vector depending on the current concentration $\mathbf{x}$, where r denotes the number of reaction rules. A flux $v_\rho(\mathbf{x}) \geq 0$ describes the rate of a particular reaction ρ. For the function v_ρ we require only that $v_\rho(\mathbf{x})$ is positive,

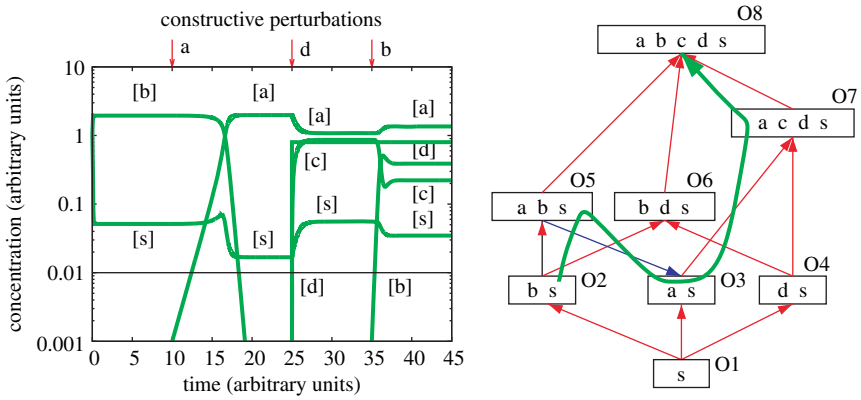

Figure 4. Example of a constructive dynamics in state space (left) and its visualisation in the lattice of organisations (right). Reaction network: Example 2 (five species) as shown in Figure 1. The following constructive perturbations are performed: At $t = 10$ we add a tiny amount (i.e., 0.001 units) of a; at $t = 25$ we add 0.8 units of d; and at $t = 35$ we add 0.001 units of b. Simulation parameters: $v_1 = 30[a]^2[s]$, $v_2 = 10[b]^2[s]$, $v_3 = [a][b]$, $v_4 = 10[b][c]$, $v_5 = [a][d]$, $v_6 = 2$, $v_7 = v_8 = v_9 = 1$. Initial state: $[a] = 0$, $[b] = 1$, $[c] = 0$, $[d] = 0$, $[s] = 1$. ODE: Eq. (12) with Eq. (11). Threshold $\Theta = 0.01$ (cf. Eq. (21))

if and only if the molecules on the left hand side of the reaction ρ are present in the state $\mathbf{x}$, and otherwise it must be zero (see Assumption 2, Section 2.3). Often it is also assumed that $v_\rho(\mathbf{x})$ increases monotonously, but this is not required here.

Given the dynamical system as $\dot{\mathbf{x}} = \mathbf{S}\mathbf{v}(\mathbf{x})$, we can define the abstraction of a state $\mathbf{x}$ formally by using a (small) threshold $\Theta \geq 0$ such that all fixed points have positive coordinates greater than Θ.

Definition 15 (abstraction[10]) *Given a dynamical system $\dot{\mathbf{x}} = f(\mathbf{x})$ and let $\mathbf{x}$ be a state in X, then the abstraction $\phi(\mathbf{x})$ is defined by*

$$\phi(\mathbf{x}) = \{i | x_i > \Theta, i \in \mathcal{M}\}, \quad \phi : X \to \mathcal{P}(\mathcal{M}), \quad \Theta \geq 0 \tag{21}$$

where x_i is the concentration of molecular species i in state $\mathbf{x}$, and Θ is a threshold chosen such that it is smaller than any positive coordinate of any fixed point of $\dot{\mathbf{x}} = f(\mathbf{x})$, $x_i \geq 0$.

Setting $\Theta = 0$ is a safe choice, because in this case ϕ always meets the definition above. But for practical reasons, it makes often sense to apply a positive threshold greater zero, e.g., when we take into consideration that the number of molecules in a reaction vessel is finite.

Theorem 1 *Let us consider a general reaction system whose reaction network is given by the reaction network $\langle \mathcal{M}, \mathcal{R} \rangle$ and whose dynamics is given by $\dot{\mathbf{x}} = \mathbf{S}\mathbf{v}(\mathbf{x}) = \boldsymbol{f}(\mathbf{x})$ as defined before. Let $\mathbf{x}' \in \boldsymbol{X}$ be a fixed point, that is, $\boldsymbol{f}(\mathbf{x}') = \mathbf{0}$, and let us consider an abstraction ϕ as given by Definition 15, which assigns a set of molecules to each state $\mathbf{x}$; then $\phi(\mathbf{x}')$ is an organisation (Proof in Ref.[10]).*

In other words, each fixed point $\mathbf{x}'$ is an instance of an organisation. Intuitively this has to be true, for the following reasons: Assume, ad absurdum, that there is a fixed point formed by a set of species that is not closed. These molecules would inevitably produce new species, which is in contradiction with the definition of a fixed point. Now assume, ad absurdum, that there is a fixed point formed by a set of species that is not self-maintaining. In this case, there is no flux vector such that all species can be produced at a non-negative rate, which again contradicts the definition of a fixed point.

From this theorem it follows immediately that a fixed point is an instance of a closed set, a semi-self-maintaining set, and of a semi-organisation. Let us finally mention that even if each fixed point is an instance of an organisation, an organisation does not necessarily possess a fixed point. A well known example is exponential growth: $\mathcal{M} = \{a\}$, $\mathcal{R} = \{a \to 2a\}$, $\dot{x} = \mathbf{S}\mathbf{v}(x)$ with $\mathbf{S} = 1$ and $\mathbf{v}(x) = x$. There are two organisation the empty organisation $\{\}$ and the organisation $\{a\}$, which represents an exponentially growing population. Obviously there is no fixed point with $x > 0$.

4.3 Movement from Organisation to Organisation

As opposed to an ODE, molecular systems are inherently discrete. The amount of molecules present in a reaction vessel is countable and can be represented by a natural number. In finite time, a molecular species can vanish completely, so that its concentration becomes exactly zero. Therefore, the composition of molecular species in a reaction vessel can change while new species enter or present species vanish.

This change of the composition of molecular species can be interpreted as a movement in the set of all possible sets of molecules $\mathcal{P}(\mathcal{M})$. In that case we can track the dynamics in the lattice of all possible sets of molecules. Note that $\mathcal{P}(\mathcal{M})$ is usually much smaller than X, however it can still be quite large, since it grows exponentially with the number of molecules.

As we can see in Figure 2, the lattice of all sets can already by quite complicated. A solution to this problem is to track a constructive dynamics in the lattice (or set) of organisations (Figure 4). This level of abstraction will filter out those changes that do not lead to a new organisation. Thus providing a quite high-level view.

4.3.1 ODEs and movement in the set of organisations

In an ODE like Eq. (12) molecules cannot vanish completely in finite time. They can only *tend to* zero as time *tends to* infinity. So, even if in reality a molecule disappears, in an ODE model it might still be present in a tiny quantity. A molecule whose concentration tends to zero in an ODE can be interpreted as a molecule that would vanish completely in reality, which would change the set of molecules present.

A common approach to overcome this problem is to introduce a concentration threshold Θ, below which a molecular species is considered not to be present. We

use this threshold in order to define the abstraction ϕ, which just returns the set of molecules present in a certain state. Additionally, we might use the threshold to manipulate the numerical integration of an ODE by setting a concentration to zero, when it falls below the threshold. In this case, a constructive perturbation (i.e., a perturbation that causes a new molecular species to appear) has to be greater than this threshold.

4.3.2 Downward movement

Not all organisations are stable. The fact that there exits a flux vector, such that no molecule of that organisation vanishes, does not imply that this flux vector can be realized when taking dynamics into account. As a result a molecular species can disappear. Each molecular species that disappears simplifies the system. Some molecules can be generated back. But eventually the system can move from a state that generates organisation O_1 into a state that generates organisation O_2, with O_2 always below $O_1 (O_2 \subset O_1)$. We call this spontaneous movement a *downward movement*.

Figure 4 illustrates this downward movement using the five-species example (Example 2). Starting with high concentration of the molecular species $\{a, b, s\}$ (at time $t = 15$ in Figure 4) the system moves spontaneously down to organisation $\{a, s\}$.

4.3.3 Upward movement

Moving up to an organisation above requires that a new molecular species appears in the system. This new molecular species cannot be produced by a reaction among present molecules (condition of closure). Thus moving to an organisation above is more complicated then the movement down and requires a couple of specifications that describe how new molecular species enter the system. Here we assume that new molecular species appear by some sort of random perturbations or purposeful interference, called *constructive perturbation*. We assume that a small quantity of molecules of that new molecular species (or a set of molecular species) suddenly appears.

Definition 16 (constructive perturbation) *A perturbation that moves a state* $\mathbf{x}$ *to a perturbed state* $\mathbf{x}'$ *where the molecular species present in* $\mathbf{x}$ *are different from those present in* $\mathbf{x}'$ *is called a constructive perturbation.*

Often, in practice, a constructive perturbation (appearance of new molecular species) has a much slower time scale than the internal dynamics (e.g., chemical reaction kinetics) of the system.

4.3.4 Visualising movements in the set of organisations

In order to display potential movements in the lattice or set of organisations, we can draw links between organisations. As exemplified in Figure 4, these links can indicate possible downward movements (down-link, blue) or upward movements

(up-link, red). A neutral link (black line) denotes that neither the system can move spontaneously down, nor can a constructive perturbation move the system up. Whether the latter is true depends on the definition of "constructive perturbation" applied. For the example in Figure 4 we defined a constructive perturbation as inserting a small quantity of *one* new molecular species.

The dynamics in between organisations is more complex than this intuitive presentation might suggest, for example in some cases it is possible to move from one organisation O_1 to an organisation O_2, with O_2 above (or below) O_1 without passing through the organisations in between O_1 and O_2.

5. ORGANISATIONS IN REAL SYSTEMS

Speroni et al.[39] have shown that artificial chemical reaction networks that are based on a structure-to-function mapping[3,20] possess a more complex lattice of organisation than networks created randomly. From this observation we can already expect that natural networks possess non-trivial organisation structures. Investigation of models of planetary photo-chemistries[8,42] and bacterial metabolism,[7,33] revealed lattices of organisations that vanish when the networks are randomised, indicating a non-trivial structure.

Here two brief examples from ongoing research[7,29] are presented. The first example, studied by Matsumaru et al.,[29] is a model of HIV-immune system dynamics comprising four species. The simplicity of that model allows to validate the approach analytically. The second example, taken from Ref.,[7] is based on a model of the central sugar metabolism of *E.coli* by Puchalka and Kierzek[33] comprising 92 species. It shows that non-trivial network complexity can be tackled and that sub-structures in the model can be identified not known before.

5.1 Example: HIV-Immune System Dynamics

Wodarz and Nowak[41] developed a model of immunological control of HIV in order to explain the effect of various drug treatment strategies. Especially the model shows, why a specific drug treatment strategy does not try to remove the virus, but aims at stimulating the immune defence, such that the immune system controls the virus at low but positive quantities.

In the model, there are four molecular species: $\mathcal{M} = \{\mathsf{x}, \mathsf{y}, \mathsf{w}, \mathsf{z}\}$: uninfected $CD4^+$ T cells x, infected $CD4^+$ T cells y, cytotoxic T Lymphocyte (CTL) precursors w, and CTL effectors z. The concentration of each species is specified by x, y, w, and z, respectively. The dynamics is given by an ordinary differential equation (ODE) with kinetic parameters a, b, c, d, h, p, q, β, and λ:

$$\begin{aligned} \dot{x} &= \lambda - dx - \beta xy, \\ \dot{y} &= \beta xy - ay - pyz, \\ \dot{w} &= cxyw - cqyw - bw, \\ \dot{z} &= cqyw - hz. \end{aligned} \tag{22}$$

From the given deterministic ODE model we derive chemical reaction rules, which form a reaction network (Fig. 5, right top):

$$\mathcal{R} = \{ \begin{array}{ll} \emptyset \to \mathsf{x}, & \\ \mathsf{x}+\mathsf{y} \to 2\mathsf{y}, & \mathsf{x} \to \emptyset, \\ \mathsf{y}+\mathsf{z} \to \mathsf{z}, & \mathsf{y} \to \emptyset, \\ \mathsf{x}+\mathsf{y}+\mathsf{w} \to \mathsf{x}+\mathsf{y}+2\mathsf{w}, & \mathsf{w} \to \emptyset, \\ \mathsf{y}+\mathsf{w} \to \mathsf{y}+\mathsf{z}, & \mathsf{z} \to \emptyset \}. \end{array}$$

The ODE model includes a decay term for each species. Therefore, for each species, we have a reaction rule transforming that molecular species into the empty set: $\mathsf{x} \to \emptyset$, $\mathsf{y} \to \emptyset$, $\mathsf{w} \to \emptyset$, and $\mathsf{z} \to \emptyset$. We observe in passing that in this particular case, since all species decay, the system is a reactive flow system (Definition 3.5.2), thus consistent (Lemma 4), and therefore the set of organisations must be a lattice (Lemma 5), with one well defined largest and one well defined smallest organisation.

A graphical representation of the network is shown in Fig. 5, upper right corner. The corresponding 4×9 stoichiometric matrix $\mathbf{S}$ reads:

$$\mathbf{S} = \begin{array}{c} \\ \mathsf{x} \\ \mathsf{y} \\ \mathsf{w} \\ \mathsf{z} \end{array} \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where each row corresponds to a molecular species x, y, w, z (from the top) and each column corresponds to a reaction. As mentioned previously, the stoichiometric matrix does not contain all information of the reaction network. For example, the reaction rule $\mathsf{y}+\mathsf{z} \to \mathsf{z}$ appears only as the column vector $(0, -1, 0, 0)^T$.

5.1.1 Lattice of organisations

For applying the theory we check every possible set of species (i.e., 16 sets) whether it is closed and self-maintaining. As a result three organisations are found. The Hasse diagram is depicted in Figure 5, middle right. The smallest organisation consists only of the "healthy cells" x (uninfected $CD4^+$ T cells). There can not be a smaller organisation (e.g., the empty set), because x is an input species and therefore the empty set is not closed. Since x is an input species, the set $\{\mathsf{x}\}$ is obviously self-maintaining. Looking at the reaction rules we can see that x alone can not produce anything else, thus the set $\{\mathsf{x}\}$ is closed, too. Formally, we can show that $\{\mathsf{x}\}$ is an organisation.

The second organisation, $\{\mathsf{x}, \mathsf{y}\}$ contains "healthy cells" x together with "ill cells" (infected $CD4^+$ T cells). Looking at the reaction network, we can see that $\{\mathsf{x}, \mathsf{y}\}$ is closed, because there is no reaction rule that allows to produce w or z just using x and y alone. With the flux vector $\mathbf{v} = (10, 1, 1, 0, 0, 1, 0, 0, 0)^T$ we can show that, according to Definition 5, $\{\mathsf{x}, \mathsf{y}\}$ is self-maintaining, e.g., $\mathbf{Sv} = (8, 0, 0, 0)^T$.

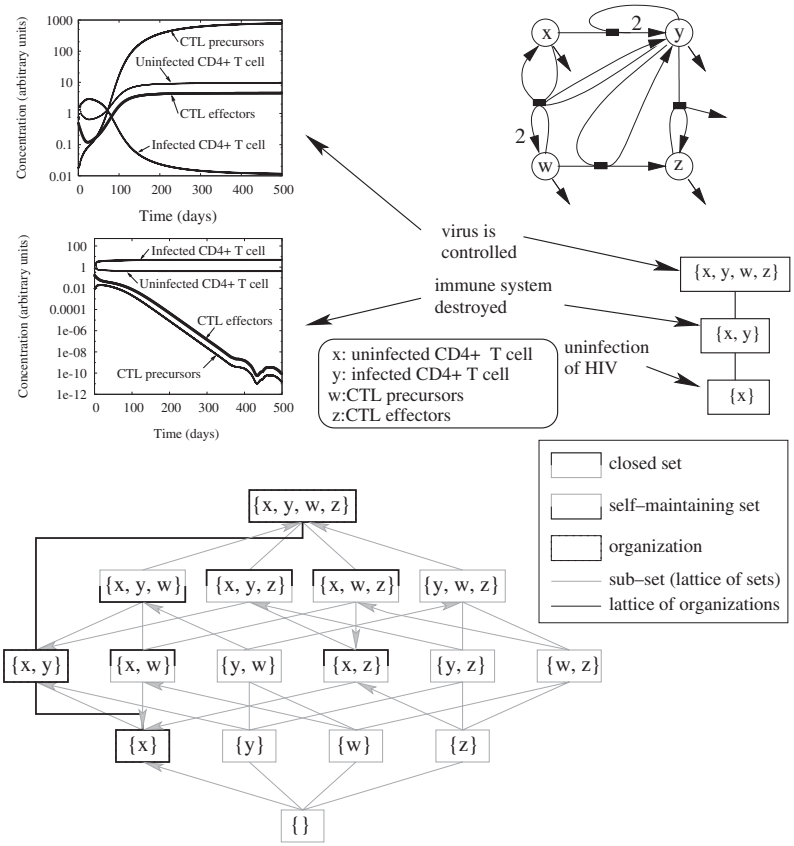

Figure 5. Illustration of the analysis of the HIV immunological response model by Wodarz and Nowak.[41] The ODE model given in Eq. 22 is transformed to a chemical reaction network (right top). The resulting hierarchy of organisations is shown as a Hasse diagram (right middle). Two of the organisations represent the attractors: virus under control (top organisation) and immune system destruction (middle organisation). Dynamic simulations[28] leading to both attractors are shown on the left. Parameters were taken from Ref.[41] as follows: $\lambda = 1$; $d = 0.1$; $\beta = 0.5$; $a = 0.2$; $p = 1$; $c = 0.1$; $b = 0.01$; $q = 0.5$; $h = 0.1$. Initial concentrations for left, top plot: $x = 0.74$; $y = 0.75$; $w = 0.018$; $z = 0.49$. Initial concentrations for left, bottom plot: $x = 0.75$; $y = 0.14$; $w = 0.0095$; $z = 0.17$. At the bottom, the full lattice of sets is shown including closed and self-maintaining sets. See Matsumaru et al.[29] for details. Figure reproduced from Ref.[10]

The largest organisation contains all species and is thus obviously closed. Looking at the reaction rules, we can see that since x can be produced at an arbitrarily high rate, we can also produce y, z, and w at arbitrarily high rates, because we can freely choose the flux vector $\mathbf{v}$ according to the definition of self-maintenence (Definition 5). Actually, the production rate $\mathbf{Sv}$ of all four species $\{x, y, w, z\}$ can be positive, when we chose for example a flux vector like $\mathbf{v} = (100, 1, 1, 1, 50, 1, 50, 10)^T$.

No further organisation exists, which implies according to Theorem 1 (Section 4.2) that there is no other combination of species that can form a stationary state.

5.1.2 Connecting with dynamics and explaining a drug treatment strategy

From a mathematical analysis[41] and simulation studies[28] it is known that the model has two modes of behaviour belonging to two asymptotically stable fixed points: One of the attractors is characterised by high virus load and no CTL precursors and effectors present. This state is interpreted as the complete destruction of the immune defence. The organisation $\{x, y\}$ represents this attractor. When the HIV virus is controlled by the immune defence, all four molecular species are present in the system, constituting the other attractor. This state is reflected in the largest organisation $\{x, y, w, z\}$. The smallest organisation $\{x\}$ can be interpreted as the condition where no $CD4^+$ T cell is infected by the HIV virus.

After identifying the lattice of organisations, we can use it to explain the strategy of a drug therapy: Looking at the lattice of organisations, we can describe two strategies for a drug therapy: The first one tries to move the system into the smallest organisations $\{x\}$, where no virus is present at all. An alternative strategy may move the system into the largest organisation, where the virus is present, but also an immune system response controlling the virus.

There are drugs available that can bring down the virus load by several orders of magnitude. If by this procedure the virus could be completely removed, the system would move into the smallest organisation, because the set $\{x, w, z\}$ generates[3] organisation $\{x\}$. However, it has been observed that although the virus load can be decreased below detection limit, the virus can not be fully removed so that the virus appears again after stopping the treatment. Therefore, the actual strategy of a drug therapy is not to move the system into the lowest organisation, but into the highest organisation. In practice, this is achieved by applying the drug periodically allowing the immune defence to increase.[41]

We can see that the strategy of a drug treatment can be explained on a relatively high (i.e., less detailed) level of abstraction using the lattice of organisations, namely as a movement from an organisation representing an ill state to an organisation representing a healthy state. It is important to note that choosing the right level of abstraction depends on what should be explained. The lattice of organisations is a suitable level of abstraction for describing the overall strategy, i.e., the quality of a drug treatment. However, *how* an actual drug treat should look like quantitatively in order to move the system into the largest organisation can not be answered by our theory. For this we have to chose a more detailed level of abstraction, e.g., the ODE model, which provides information on how the system can move from one organisation to another.

[3] Note that we use the word "generate" as a precisely defined technical term (Eq. (16), Section 3.6.3).

5.2 Example: Central Sugar Metabolism of E. Coli

In order to demonstrate that chemical organisation theory can reveal structures in networks of non-trivial size, Centler et al.[7] applied the theory to a large stochastic network model of the central sugar metabolism of E. Coli as introduced by Puchalka and Kierzek.[33]

The model consists of 92 species and 197 reactions, including gene expression, signal transduction, transport, and enzymatic activities. The model was simplified by ignoring inhibitory links. By doing so we assume that an inhibition has only a quantitative effect and does not shut off a reaction pathways completely (cf. Section 2.4).

Two slightly different reaction networks can be studied depending on whether we consider activators as necessary or not in the production of proteins.

When we consider activators to be necessary, a relatively complex lattice of organisations appears as depicted in Figure 6. The smallest organisation, O_1, contains 76 molecules including the glucose metabolism and all input molecules. The input molecules, chosen according to Ref.,[33] include the external food set (Glcex, Glyex, Lacex) and all promoters. Two other organisations, O_3 and O_4, contain the Lactose and Glycerol metabolism, respectively. Their union results in the largest organisation O_5 that contains all molecules.

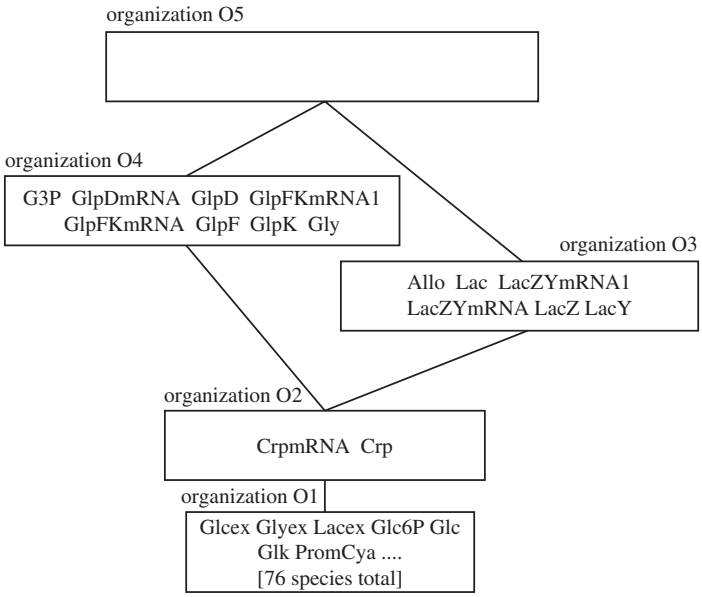

Figure 6. Lattice of organisations of a model of the central sugar metabolism of *E.coli.*.[33] In an organisation, only names of new molecular species are printed that are not present in an organisation below. The vertical position of an organisation correlates with the number of chemical species it contains. Organisation O_5 (top) is the largest one, containing all species from O_4 and O_3. Figure from Refs.[7] and[10]

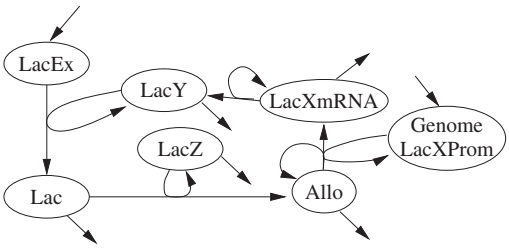

Figure 7. Illustration of the Mechanism leading to Organisation O_3 in Figure 6. *LacEx* represents external lactose, whereas *Lac* represents lactose inside the cell

In order illustrate the mechanism that leads to higher organisations, we take a closer look at organisation O_3, which includes the lactose uptake. Following Puchalka and Kierzek,[33] we assume that external lactose (*LacEx*) is alway present, as well as the genetic information, including the promoter (*LacXProm*) to make messenger RNA (*LacXmRNA*). This is considered by defining a fixed inflow for those molecular species: $\emptyset \rightarrow LacEx$ and $\emptyset \rightarrow LacXProm$. Figure 7 shows the fraction of the reaction network, which explains, why, in the model, the lactose uptake can not be generated from the input species: *LacY* is required for lactose uptake, which is expressed by the reaction $LacEx + LacY \rightarrow LacY + Lac$. However, *LacY* cannot be generated given only the input species, such as *LacEx* and *LacXProm*. Therefore, the closure of the input species does not include molecules like *LacY*. If we add *LacY* to the input species, organisation O_3 can be generated.

If instead we consider the more precise version of the model, which does not assume the necessary presence of activators for gene expression, we have a more realistic model, yet a model that focuses at a longer timescale (as the base level expression of a gene without the activator appears to be relatively slow). In this case the resulting lattice collapses into a single huge organisation. In such case what the theory suggests is a general stability of the system against various perturbations, since no matter what kind of initial combination of molecular species we start off, all other molecular species can be generated, given the input flux mentioned above and sufficient time such that a gene depending on an activator can be expressed in the absence of its activators.

6. DISCUSSION AND OUTLOOK

The theory of chemical organisation, as sketched here, creates a first, rough map of the structure and potential dynamical behaviour of a reaction system. The obtained scaffold (i.e., the set of organisations) can guide further more detailed analysis, which may study the dynamics within or in-between organisations using classical tools from dynamical systems theory. The results of more detailed studies can

in turn be explained and summarised with respect to the lattice of organisations resulting in a global picture.

6.1 Related Work

There are a number of other approaches that operate purely on the reaction network's topology in order to infer potential dynamical properties.

Classical reaction network theory provides powerful theorems, which can predict for a specific class of reaction networks whether a network possesses positive stationary states or whether positive periodic solutions are possible. For example, the deficiency-zero theorem states roughly that a weakly reversible mass-action reaction system with deficiency zero contains one unique equilibrium point in each positive reaction simplex.[18] This line of research provides probably the strongest mathematical results that link network structure to potential dynamics. However, the research focuses on *positive* solutions, i.e., solutions where all molecular species are present. Here, we are interested in states of the reaction system where only a subset of species are present. Furthermore, our theory is not restricted to systems with deficiency zero or one. We do not require for our theorem that the reaction system is governed by mass-action kinetics; in turn of course, chemical organisation theory can not predict the stability of an organisation. Furthermore, we do not focus on stationary behaviour, but we aim at understanding complex transitive dynamics, such as the movement between organisations.

Another class of methods that operate on the network structure identifies so called *flux modes*.[23,30] A flux mode is a set of reaction rules that can operate at a steady state. Flux modes are similar to T-invariants, a concept from Petri net theory.[31] Obviously, flux modes can be linearly combined and thus form a complete lattice (when we also consider the empty flux mode). A flux mode implies a set of molecules, namely the set of molecules participating in the reactions of that flux mode. Therefore, a flux mode is similar to the concept of self-maintenence. However, the set of participating molecules is not necessarily self-maintaining nor closed.[26] And not all self-maintaining sets are represented by flux modes.[26] For example, a self-replicating molecule whose concentration growths for ever is self-maintaining. But, since it does not reach a steady state, it is not captured by a flux mode.

Boolean networks or logical networks are another approach to handle networks without requiring detailed kinetics. A boolean network consists of a set of boolean functions that take their own result of a previous time step as input. It is not yet clear, how stationary states of a boolean network are related to organisations, which is an interesting aspect for future investigations. Logical networks are a useful tool to model signalling and gene regulatory networks.[17] However, a boolean network does not consider stoichiometry, which is essential in general reaction networks as considered by our theory.

6.2 Computational Complexity

This chapter describes the mathematical base of our theory and does not focus on algorithmic issues. So far some preliminary algorithms are available. One of them computes the set of organisations from the bottom up by starting from the smallest organisation and then, recursively, adding molecules in order to generated all the organisations above. With this algorithm we can currently analyse networks containing up to 200 molecular species. For less than 30 species, a brute force algorithm appears faster. The brute force algorithm simply tests every possible set of molecules whether it is closed and self-maintaining.

6.3 Structure-to-Function Mapping

The aim of the theoretical framework introduced in this chapter is to deal with constructive dynamical systems. However, the examples we presented for illustrating the new concepts where relatively simple: A set of molecules has always been defined as a list of symbols. And also the reaction rules were given as an explicit list. In other words, in these examples, all molecules that could appear were already listed explicitly in the definition of the set of molecules.

When designing the presented theoretical framework we had already more complex reaction systems in mind, namely those where the set of molecules and reaction rules are defined *implicitly*. In these systems, molecules possess a structure, that is, there is a grammar specifying their syntax, and reaction rules are defined implicitly by referring to that structure. A simple example is the prime number chemistry,[2,4] where the set of molecules are all natural numbers and the reaction rules are defined by the numerical devision operator. More complex examples from the field of artificial chemistry are AlChemy,[19,20] the combinator chemistry,[37] or the more realistic toy chemistry by Benkö et al..[5] But also in biochemistry and systems biology we observe a growing number of models where the reaction network is defined implicitly, which usually leads to a combinatorial explosion in size. Examples are models of DNA assembly,[21] DNA computing,[1] or combinatorial signalling networks.[6,32]

Note that in our approach the set of molecules $\mathcal{M}$ and the set of reactions $\mathcal{R}$ of a reaction network can be defined implicitly. Furthermore, the dynamics that we assumed in Section 4 is quite general, so that we can theoretically apply our framework also to the systems mentioned above. However, computational tools for an automatic analysis of implicitly defined reaction systems have yet to be developed, which is a significant challenge for future research.

7. ACKNOWLEDGEMENTS

We grateful to Florian Centler, Naoki Matsumaru, and Christoph Kaleta contributing examples and valuable discussions. Financial support by the *Federal Ministry of Education and Research* (BMBF) Grant 0312704A and the *German Research Foundation* (DFG) Grant Di852/4-1 is greatly acknowledged.

8. REFERENCES

1. L. M. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266(5187):1021–1024, 1994.
2. J.-P. Banâtre and D. L. Métayer. The GAMMA model and its discipline of programming. *Sci. Comput. Program.*, 15(1):55–77, 1990.
3. W. Banzhaf. Self-replicating sequences of binary numbers – foundations I and II: General and strings of length n = 4. *Biol. Cybern.*, 69:269–281, 1993.
4. W. Banzhaf, P. Dittrich, and H. Rauhe. Emergent computation by catalytic reactions. *Nanotechnology*, 7(1):307–314, 1996.
5. G. Benkö, C. Flamm, and P. F. Stadler. A graph-based toy model of chemistry. *J. Chem. Inf. Comput. Sci.*, 43(4):1085–1093, 2002.
6. N. Borisov, N. Markevich, H. J.B., and K. B.N. Signaling through receptors and scaffolds: independent interactions reduce combinatorial complexity. *Biophys. J*, 89(2):951–966, 2005.
7. F. Centler, P. S. di Fenizio, N. Matsumaru, and P. Dittrich. Chemical organizations in the central sugar metabolism of escherichia coli. In *Modeling and Simulation in Science Engineering and Technology, Post-proceedings of ECMTB 2005*, pages 1–8. Birkhäuser, 2006. (in print).
8. F. Centler and P. Dittrich. Chemical organizations in atmospheric photochemistries: a new method to analyze chemical reaction networks. Submitted.
9. M. E. Csete and J. C. Doyle. Reverse engineering of biological complexity. *Science*, 295:1664–1669, 2002.
10. P. Dittrich and P. S. di Fenizio. Chemical organization theory. *Bull. Math. Biol.*, 2006. In print.
11. P. Dittrich, T. Kron, and W. Banzhaf. On the formation of social order – modeling the problem of double and multi contingency following luhmann. *Journal of Artifical Societies and Social Simulation*, 6(1), 2003.
12. P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries – a review. *Artificial Life*, 7(3):225–275, 2001.
13. O. Ebenhöh, T. Handorf, and R. Heinrich. Structural analysis of expanding metabolic network. *Genome Informatics*, 15(1):35–45, 2004.
14. M. Eigen. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, 1971.
15. M. Eigen and P. Schuster. The hypercycle: a principle of natural self-organisation, part A. *Naturwissenschaften*, 64(11):541–565, 1977.
16. P. Érdi and J. Tóth. *Mathematical Models of Chemical Reactions: Theory and Applications of Deterministic and Stochastic Models*. Pinceton University Press, Princeton, NJ, 1989.
17. C. Espinosa-Soto, P. Padilla-Longoria, and E. R. Alvarez-Buylla. A gene regulatory network model for cell-fate determination during Arabidopsis thaliana flower development that is robust and recovers experimental gene expression profiles. *Plant Cell*, 16(11):2923–2939, 2004.
18. M. Feinberg and F. J. M. Horn. Dynamics of open chemical systems and the algebraic structure of the underlying reaction network. *Chem. Eng. Sci.*, 29(3):775–787, 1973.
19. W. Fontana. Algorithmic chemistry. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, pages 159–210, Redwood City, CA, 1992. Addison-Wesley.
20. W. Fontana and L. W. Buss 'The arrival of the fittest': Toward a theory of biological organization. *Bull. Math. Biol.*, 56(1):1–64, 1994.
21. K. V. Gothelf and R. S. Brown. A modular approach to DNA-programmed self-assembly of macromolecular nanostructures. *Chemistry*, 11(4):1062–1069, Feb 2005.
22. C. Heij, A. e. C. M. Ran, and F. v. Schagen. *Introduction to Mathematical Systems Theory Linear Systems, Identification and Control*. Birkhäuser, 2006.
23. R. Heinrich and S. Schuster. *The Regulation of Cellular Systems*. Chapman and Hall, New York, NY, 1996.
24. S. Jain and S. Krishna. Autocatalytic sets and the growth of complexity in an evolutionary model. *Phys. Rev. Lett.*, 81(25):5684–5687, 1998.
25. S. Jain and S. Krishna. A model for the emergence of cooperation, interdependence, and structure in evolving networks. *Proc. Natl. Acad. Sci. U. S. A.*, 98(2):543–547, 2001.

26. C. Kaleta, F. Centler, and P. Dittrich. Analyzing molecular reaction:networks: from pathways to chemical organizations. *Molecular Biotechnology*, 2006. In print.
27. S. A. Kauffman. Cellular homeostasis, epigenesis and replication in randomly aggregated macromolecular systems. *J. Cybernetics*, 1:71–96, 1971.
28. N. Matsumaru, F. Centler, and P. D. Klaus-Peter Zauner. Self-adaptive scouting – autonomous experimentation for systems biology. In G. R. Raidl, S. Cagnoni, J. Branke, D. Corne, R. Drechsler, Y. Jin, C. G. Johnson, P. Machado, E. Marchiori, F. Rothlauf, G. D. Smith, and G. Squillero, editors, *Applications of Evolutionary Computing, EvoWorkshops 2004*, volume 3005 of *LNAI*, pages 52–62. Springer, Berlin, 2004.
29. N. Matsumaru, F. Centler, P. Speroni di Fenizio, and P. Dittrich. Chemical organization theory applied to virus dynamics. *it – Information Technology*, 48(3), 2006. In print.
30. J. A. Papin, J. Stelling, N. D. Price, S. Klamt, S. Schuster, and B. O. Palsson. Comparison of network-based pathway analysis methods. *Trends Biotechnol*, 22(8):400–405, Aug 2004.
31. C. A. Petri. *Kommunikation mit Automaten*. PhD thesis, University of Bonn, Bonn, 1962.
32. R. Pinkas-Kramarski, I. Alroy, and Y. Yarden. Erbb receptors and egf-like ligands: cell lineage determination and oncogenesis through combinatorial signaling. *J. Mammary. Gland. Biol. Neoplasia*, 2(2):97–107, 1997.
33. J. Puchalka and A. Kierzek. Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys. J.*, 86(3):1357–1372, 2004.
34. O. E. Rössler. A system theoretic model for biogenesis (in German). *Z. Naturforsch. B*, 26(8):741–746, 1971.
35. P. Schuster and K. Sigmund. Replicator dynamics. *J. Theor. Biol.*, 100:533–8, 1983.
36. D. Segré, D. Lancet, O. Kedem, and Y. Pilpel Graded autocatalysis replication domain (GARD): Kinetic analysis of self-replication in mutually catalytic sets. *Orig. Life Evol. Biosph.*, 28(4–6):501–514, 1998.
37. P. Speroni di Fenizio. A less abstract artficial chemistry. In M. A. Bedau, J. S. McCaskill, N. H. Packard, and S. Rasmussen, editors, *Artificial Life VII*, pages 49–53, Cambridge, MA, 2000. MIT Press.
38. P. Speroni Di Fenizio and P. Dittrich. Artificial chemistry's global dynamics. movement in the lattice of organisation. *The Journal of Three Dimensional Images*, 16(4):160–163, 2002.
39. P. Speroni di Fenizio, P. Dittrich, J. Ziegler, and W. Banzhaf. Towards a theory of organizations. In *German Workshop on Artificial Life*, Bayreuth, 5.-7. April, 2000. In print, available online: di.ttri.ch/p/SDZB2001gwal.html.
40. P. F. Stadler, W. Fontana, and J. H. Miller. Random catalytic reaction networks. *Physica D*, 63:378–392, 1993.
41. D. Wodarz and M. A. Nowak. Specific therapy regimes could lead to long-term immunological control of hiv. *Proc. Nat. Acad. Sci. USA*, 96(25):14464–9, 1999.
42. Y. L. Yung and W. B. DeMore. *Photochemistry of Planetary Athmospheres*. Oxford University Press, New York, 1999.