JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

Cite this article: Fellermann H, Cardelli L. 2014 Programming chemistry in DNA-addressable bioreactors. *J. R. Soc. Interface* **11**: 20130987.
http://dx.doi.org/10.1098/rsif.2013.0987

Received: 25 October 2013
Accepted: 3 March 2014

Subject Areas:
nanotechnology, biomimetics, bioinformatics

Keywords:
compartmentalization, programmable chemistry, biochemical engineering, theoretical computer science, membrane computing, DNA computing

Author for correspondence:
Harold Fellermann
e-mail: harold.fellermann@newcastle.ac.uk

Electronic supplementary material is available at http://dx.doi.org/10.1098/rsif.2013.0987 or via http://rsif.royalsocietypublishing.org.

Royal Society Publishing

Programming chemistry in DNA-addressable bioreactors

Harold Fellermann[1,2] and Luca Cardelli[3]

[1]School of Computing Science, Newcastle University, King's Gate, Newcastle upon Tyne NE1 7RU, UK
[2]Center for Fundamental Living Technology, University of Southern Denmark, Campusvej 55, 5230 Odense M, Denmark
[3]Microsoft Research Cambridge, 21 Station Road, Cambridge CB1 2FB, UK

We present a formal calculus, termed the *chemtainer calculus*, able to capture the complexity of compartmentalized reaction systems such as populations of possibly nested vesicular compartments. Compartments contain molecular cargo as well as surface markers in the form of DNA single strands. These markers serve as compartment addresses and allow for their targeted transport and fusion, thereby enabling reactions of previously separated chemicals. The overall system organization allows for the set-up of programmable chemistry in microfluidic or other automated environments. We introduce a simple sequential programming language whose instructions are motivated by state-of-the-art microfluidic technology. Our approach integrates electronic control, chemical computing and material production in a unified formal framework that is able to mimic the integrated computational and constructive capabilities of the subcellular matrix. We provide a non-deterministic semantics of our programming language that enables us to analytically derive the computational and constructive power of our machinery. This semantics is used to derive the sets of all constructable chemicals and supermolecular structures that emerge from different underlying instruction sets. Because our proofs are constructive, they can be used to automatically infer control programs for the construction of target structures from a limited set of resource molecules. Finally, we present an example of our framework from the area of oligosaccharide synthesis.

1. Introduction

Living systems are unique in that they integrate molecular recognition and information processing with material production on the molecular scale. The predominant locus of this integration is the cytoplasm, where a multitude of biochemical compounds are highly organized in vesicular compartments that co-locate reactants of desired reactions, whereas separating those of undesired reactions. Surface markers on these compartments are used for vesicular trafficking, as well as vesicle budding and fusion, thereby allowing for the fine-tuned control of biochemical reaction cascades [1–3].

The desire to employ this complex molecular organization in next-generation chemical synthesis has led to various studies on supermolecular compartments as nanoscale 'bioreactors' [4–7]. Several pathways for vesicle fusion [8–10] have been suggested. In particular, Hadorn *et al*. [11–13] employ short single-stranded DNA tags for the specific interaction of various reaction compartments. In the European Commission-funded project MATCHIT [14–17], we are working on creating an artificial cellular matrix that seamlessly integrates information processing and material production in much the same way as its biological counterpart: the MATCHIT framework employs DNA-addressable bioreactors (termed *chemtainers* in the following), to mimic the topological organization of the cytoplasm. DNA tags open up for DNA computing operations akin to the 'key–lock' information processing mechanism found in biological protein–protein interactions. This form of molecular information processing is governed by autonomous chemical reaction kinetics and allows for tight integration of

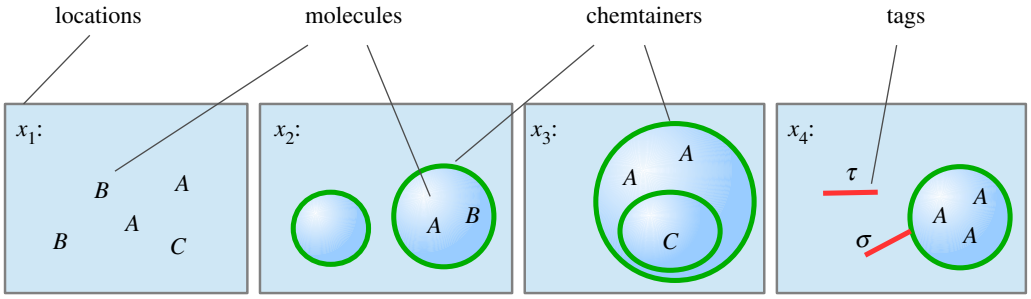

Figure 1. Example of a state of the artificial cellular matrix, where four locations (x_1, x_2, x_3 and x_4) hold content. All other locations are empty. Location x_1 contains a number of molecules in solution ($2A + 2B + C$), location x_2 contains two chemtainers (depicted as circles), one of them containing the molecules A and B, location x_3 holds a chemtainer encapsulated within another chemtainer, and location x_4 holds a chemtainer with three A's being decorated with a tag σ (straight line attached to the circle), with a separate tag τ in solution (separate straight line). Using the syntax defined in §2.1, this system state is described by the string $x_1 : 2A + 2B + C \circ x_2 : (\!|\,|\!) + (\!|\,A + B\,|\!) \circ x_3 : (\!|\,2A + (\!|\,C\,|\!)\,|\!) \circ x_4 : \tau + \sigma(\!|\,3A\,|\!)$. (Online version in colour.)

chemical production and information processing. In particular, we here employ the DNA *join-and-fork* gates to implement Boolean computation [18].

Whereas natural cells ultimately employ genomic information to regulate the resulting material production network, we envision programmable electronic control by embedding chemtainers into mechanoelectronic microsystems [19,20]. In such devices, dedicated microfluidic channels can be designed for vesicle generation [21], DNA tag insertion, tag and chemtainer trafficking [22], specific or unspecific fusion [23], vesicle rapture and encapsulation [24]. Possibly paired with real-time feedback, this allows for control of chemical reaction cascades at the molecular level. This interplay of autonomous molecular computation and external electronic chemtainer manipulation enables the programmatic set-up of complex reaction cascades that allow for automated chemical production of a desired target molecule from a limited set of chemical resources.

Here, we propose a formal calculus to describe system states and transitions in an abstract representation of the artificial cellular matrix. We call this the *chemtainer calculus*. In essence, the chemtainer calculus allows us to describe the organization and manipulation of chemical compounds down to the molecular level in possibly nested, addressable reaction compartments. Sets, or more precisely, multisets of such compartmentalized reagents are situated at locations, e.g. at positions of a microfluidic machine. Several calculi for compartmentalized reaction systems have been proposed previously [25–27]. Our syntax for nested and tagged compartments follows relatively closely the one from *brane calculi* [27]. Those calculi offer transitions for compartment transformations, such as fusion and splitting as well as molecular reactions. Transitions are integral components of the system state, and the calculus employs a process algebra to deduce the set of possible transitions from the current state. In the chemtainer calculus, we additionally define an explicit transition system that operates on states externally. This organization better reflects the difference between the chemical state transitions and external mechanoelectronic control.

The calculus was designed for the architecture described in reference [20]. The tool chain for compilation of high-level directives of the chemtainer calculus into eventual electrode configurations to control the mechanoelectronic microfluidic hardware is presented in reference [28]. Yet, the general framework discussed here is not tailored towards one specific technology. For example, instead of microfluidic channel segments, locations could equally denote test tube arrays, or wells in a high-throughput chip.

In this article, we first design a syntax that allows us to express the rather complex system states—or rather system *arrangements*—that we encounter in the artificial cellular matrix: located multisets of tagged and possibly nested chemtainers that carry cargo. Figure 1 schematically depicts an example of a possible state. We then introduce transitions between states that model possible changes of the physical state. Some of these transitions, e.g. DNA hybridization, capture autonomous chemistry, whereas other transitions are thought to be induced by control operations of the artificial cellular matrix. We proceed by defining a simple programming language that consists of sequences of parametric instructions that induce transitions on the system state. In §3, we discuss how the underlying instruction set of the chemtainer calculus affects the set of constructable objects, and we give our main result (theorem 3.3) that a set of eight instructions is sufficient to construct any well-formed target state. Because we use a non-deterministic semantics, our proofs demonstrate only the possibility, not the probability of creating a certain desired state. In §4, we apply our framework to the area of chemical manufacturing, where we present an algorithm for synthesis of branched oligomer structures by means of controlled chemtainer fusion and DNA computing.

2. The chemtainer calculus

2.1. System state

Objects of the calculus are molecules, address tags and chemtainers. Chemtainers represent compact microreactors, such as vesicles, oil droplets in water or water droplets in oil, DNA nanocages, etc. They can be decorated with address tags, and can hold chemical content and even other chemtainers within them. Here, we do not distinguish between different chemtainer types, but we could imagine a type system for chemtainers to specify their physical properties. All components of the system are situated at specified, discrete locations.

Our notion of space is rather simplistic: we assume a fixed set of locations; each component of the system state is situated at a certain location; we will introduce transitions that allow collocated components to interact (while objects at different locations may not interact), and we will introduce transitions that induce transport from one location to another by means of state transitions. Note that we currently do not introduce a specific topology on the set of locations (e.g. to move from one location to another, one might need to cross a third one), but such an extension is possible.

For countable index sets $J_{\mathcal{L}}$, $J_{\mathcal{M}}$ and $J_{\mathcal{T}}$, let $\mathcal{L} = \{x_i; i \in J_{\mathcal{L}}\}$ denote the set of locations, $\mathcal{M} = \{m_j; j \in J_{\mathcal{M}}\}$ some set of molecules, and $\mathcal{T} = \{s_k; k \in J_{\mathcal{T}}\}$ a set of DNA tags. We take $\mathcal{T} = \{\rho, \sigma, \tau, \ldots\}$ if tags are explicitly given. Note that $\mathcal{M}$ might intersect with $\mathcal{T}$ or not.

The following context-free tree grammar G_{s} for system states formalizes the above verbal interpretation:

$$\text{global state} \quad \mathsf{S} := \emptyset \mid \mathsf{S} \circ \mathsf{S} \mid x_i : \mathsf{P}, \tag{2.1}$$

$$\text{local state} \quad \mathsf{P} := 0 \mid \mathsf{P} + \mathsf{P} \mid \mathsf{q}^* \langle\!(\mathsf{P})\!\rangle \mid \mathsf{q} \mid m_j \tag{2.2}$$

and

$$\text{tag} \quad \mathsf{q} := \mathsf{s} \mid \mathsf{s}^* \triangleright \mathsf{s}^*. \tag{2.3}$$

with the start symbol S. Here, the vertical broken bar is a meta-symbol that indicates syntactic choice. We denote the empty state by the symbol $\emptyset$. The binary operator $x_i : \mathsf{P}$ denotes localization, where x_i is a location identifier; the binary operator $\mathsf{S} \circ \mathsf{S}$ denotes composition of locations, whereas $\mathsf{P} + \mathsf{P}$ denotes composition within locations; 0 denotes the empty local state; * is the Kleene operator and signifies no or arbitrarily many repetitions of its argument. We write

$$\mathsf{q}^* = \diamond \mid \mathsf{q} + \mathsf{q}^* \tag{2.4}$$

and

$$\mathsf{s}^* = \diamond \mid \mathsf{s} + \mathsf{s}^*. \tag{2.5}$$

with the empty tag $\diamond$ to explicitly list tag and gate sets.

Following convention, we denote chemtainers by half-moon parentheses $\mathsf{q}^* \langle\!(\mathsf{P})\!\rangle$ that enclose the chemtainer content with address tags associated with the left parenthesis [27,29]. Finally, the relation $\mathsf{s}^* \triangleright \mathsf{s}^*$ denotes DNA gates, which will be explained later. We write G_{S}, G_{P}, G_{q^*} and G_{s^*} for the grammars with the start symbols S, P, q^* and s^*, respectively.

To generate a state with the grammar G_{X}, $\mathsf{X} \in \{\mathsf{S}, \mathsf{P}, \mathsf{q}^*, \mathsf{s}^*\}$ being a non-terminal, one recursively applies the above production rules starting from X until the resulting state tree contains no more non-terminal symbols. The language $L(G_{\mathsf{X}})$ is the set of all possible states that can be generated from the start symbol X. In what follows, $S, S', S'' \in L(G_{\mathsf{S}})$, $P, P', P'', P_i \in L(G_{\mathsf{P}})$, $q, q', q_k \in L(G_{\mathsf{q}^*})$ and $s, s', s'', s_k \in L(G_{\mathsf{s}^*})$ denote arbitrary states of the respective languages.

Informally, we interpret states of $L(G_{\mathsf{S}})$ to signify the following: the global system state is a composition of local states, where each local state has a location identifier x_i and an associated local state description. The latter are compositions of molecules m_j, gates q_k as well as chemtainers $q^* \langle\!(P)\!\rangle$ with content P and gate set q^*; gate sets, in turn, are compositions of gates q_k as well as individual tags s_k. See figure 1 for an example of a global state.

In order to conform with this interpretation, we introduce the following structural congruence relation (an equivalence relation with additional axioms that guarantee substitutivity of equals in context) over $L(G_{\mathsf{S}})$, $L(G_{\mathsf{P}})$, $L(G_{\mathsf{q}}^*)$ and $L(G_{\mathsf{s}}^*)$:

$$S \circ (S' \circ S'') \equiv (S \circ S') \circ S'', \tag{2.6}$$

$$S \circ S' \equiv S' \circ S, \tag{2.7}$$

$$S \circ \emptyset \equiv S, \tag{2.8}$$

$$P + (P' + P'') \equiv (P + P') + P'', \tag{2.9}$$

$$P + P' \equiv P' + P, \tag{2.10}$$

$$P + 0 \equiv P, \tag{2.11}$$

$$q_1 + (q_2 + q_3) \equiv (q_1 + q_2) + q_3, \tag{2.12}$$

$$q_1 + q_2 \equiv q_2 + q_1, \tag{2.13}$$

$$q + \diamond \equiv q, \tag{2.14}$$

$$s_1 + (s_2 + s_3) \equiv (s_1 + s_2) + s_3, \tag{2.15}$$

$$s_1 + s_2 \equiv s_2 + s_1, \tag{2.16}$$

$$s + \diamond \equiv s, \tag{2.17}$$

$$x_i : P \circ x_i : P' \equiv x_i : P + P', \tag{2.18}$$

$$\frac{S_1 \equiv S_2}{S \circ S_1 \equiv S \circ S_2}, \tag{2.19}$$

$$\frac{P_1 \equiv P_2}{x_i : P_1 \equiv x_i : P_2}, \tag{2.20}$$

$$\frac{P_1 \equiv P_2}{P + P_1 \equiv P + P_2}, \tag{2.21}$$

$$\frac{q_1 \equiv q_2}{q^* + q_1 \equiv q^* + q_2}, \tag{2.22}$$

$$\frac{s_1 \equiv s_2}{s^* + s_1 \equiv s^* + s_2}, \tag{2.23}$$

$$\frac{s_1^* \equiv s_2^*}{s_1^* \triangleright s^* \equiv s_2^* \triangleright s^*}, \tag{2.24}$$

$$\frac{s_1^* \equiv s_2^*}{s^* \triangleright s_1^* \equiv s^* \triangleright s_2^*}, \tag{2.25}$$

$$\frac{P_1 \equiv P_2}{q^* \langle\!(P_1)\!\rangle \equiv q^* \langle\!(P_2)\!\rangle} \tag{2.26}$$

and

$$\frac{q_1^* \equiv q_2^*}{q_1^* \langle\!(P)\!\rangle \equiv q_2^* \langle\!(P)\!\rangle}. \tag{2.27}$$

Thus, states that belong to the same equivalence class of $\equiv$ represent the same physical state, even though they might be syntactically different. Equations (2.6)–(2.18) induce monoidal structures on $(L(G_{\mathsf{S}}), \circ, \emptyset)$, $(L(G_{\mathsf{P}}), +, 0)$, $(L(G_{\mathsf{q}^*}), +, \diamond)$ and $(L(G_{\mathsf{s}^*}), +, \diamond)$, where ':' distributes over '$+$', and equations (2.19)–(2.27) guarantee that we can substitute equals in any context.

We introduce some notational shortcuts. We write $\langle\!(P)\!\rangle := \diamond \langle\!(P)\!\rangle$, $q \langle\!(\,)\!\rangle := q \langle\!(0)\!\rangle$, and we introduce the notation

$$nP := \underbrace{P + \cdots + P}_{n \text{ times}}, \tag{2.28}$$

which emphasizes the relation to mulitsets. We also allow ourselves to drop the explicit notation of empty locations by defining

$$x_i : 0 \equiv \emptyset. \tag{2.29}$$

With these definitions, the example state depicted in figure 1 can be written as

$$x_1 : 2A + 2B + C \circ x_2 : \langle\!(\,)\!\rangle + \langle\!(A + B)\!\rangle \circ x_3 : \langle\!(2A + \langle\!(C)\!\rangle)\!\rangle \circ x_4 : \tau + \sigma \langle\!(3A)\!\rangle. \tag{2.30}$$

2.2. Transitions

We are now ready to introduce a transition system that codifies the possible outcome of both autonomous chemical reactions as well as externally induced operations that manipulate the system state. Autonomous reactions, in turn, are either arbitrary chemical reactions among molecules, which we refer to as *application chemistry*, or the working of DNA computing operations.

2.2.1. Application chemistry

Reactions are just transitions of the form

$$P \to P'. \tag{2.31}$$

where $P = \sum_i \nu_i m_i$ and $P' = \sum_j \mu_j m_j$ are multisets of reactants and products with stoichiometric coefficients ν_i and μ_i.

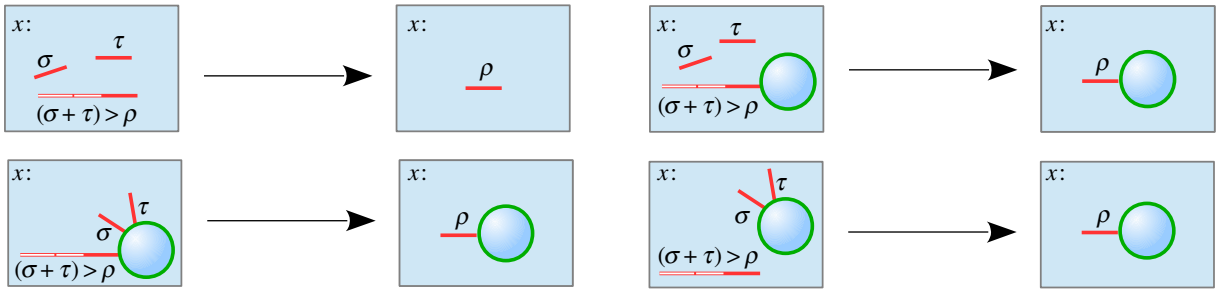

Figure 2. Transitions that encode gate reactions: a gate reacts with its respective input tags if the two are not separated by a chemtainer wall. Either the gate or its inputs might be attached to a chemtainer surface in that case the output tags of the gate will equally be bound to the surface. (Online version in colour.)

We ensure that chemical reactions can occur among any co-located reactants that are not separated by chemtainer walls

$$\frac{P \to P'}{P + P'' \to P' + P''}, \tag{2.32}$$

$$\frac{P \to P'}{q^*(\!(P)\!) \to q^*(\!(P')\!)}, \tag{2.33}$$

$$\frac{P \to P'}{x_i : P \to x_i : P'} \tag{2.34}$$

and

$$\frac{S \to S'}{S \circ S'' \to S' \circ S''}. \tag{2.35}$$

We extend structural congruence to transitions by defining.

$$\frac{P \equiv P' \quad P' \to P'' \quad P'' \equiv P'''}{P \to P'''} \tag{2.36}$$

and

$$\frac{S \equiv S' \quad S' \to S'' \quad S'' \equiv S'''}{S \to S'''}. \tag{2.37}$$

In its ground form, the chemtainer calculus does not offer any transitions of the form (2.31). Instead, the user of the calculus is assumed to provide a set $\mathcal{R}$ of autonomous transitions as axioms.

Note that there is no particular need to restrict $\mathcal{M}$ to be finite. Our calculus can be applied without adaption to an infinite set of molecules, including polymers or branched structures—an example of which will be presented in §4. We also emphasize that we explicitly allow $\mathcal{M}$ and $\mathcal{T}$ to intersect. If they do, tags can occur in the reactant and product sides of chemical reactions, such that tags can be altered chemically.

2.2.2. DNA gate transitions

Here, DNA computation is implemented by *join-and-fork* gates [18] that irreversibly release a given set of output strands s_2^* once they have bound a set of inputs s_1^*, written $s_1^* \triangleright s_2^*$. Note that $s_1^* \triangleright s_2^*$ does not physically contain the strands s_1^*. Rather it means that the gate accepts those strands as input.

If a gate is co-located with all its input tags, it can release its output tag

$$s_1^* \triangleright s_2^* + s_1^* \to s_2^*. \tag{2.38}$$

We have to ensure that these transitions can occur between two co-located complementary tags in any context, unless they are separated by a chemtainer boundary. This is allowed by introducing the following inference rules

$$\frac{q + q' \to q''}{q + q'(\!(P)\!) \to q''(\!(P)\!)} \tag{2.39}$$

and

$$\frac{q + q' \to q''}{(q + q')(\!(P)\!) \to q''(\!(P)\!)}. \tag{2.40}$$

Examples of gate transitions are depicted in figure 2. Owing to the inferences (2.32)–(2.37), it is guaranteed that DNA computing operations perform in any context.

2.2.3. Induced transitions

We now introduce transitions that model the controlled manipulation of states through operations of the artificial cellular matrix. We introduce eight such operations, responsible for feeding and moving of chemtainers and tags, controlled fusion of chemtainers, encapsulation of material into chemtainers, chemtainer bursting and flushing of content. Our exact transitions are motivated by the functionalities of the underlying microfluidics architecture [20], but they also serve as a guideline as to how alternative transitions in other hardwares can be codified. The impact of the exact instruction set on the constructive capabilities of the resulting calculus will be discussed in §3.

Induced transitions are of the form

$$I : S \to S,$$

where I is a parametrized name. We first give the formal definition of these transitions (schematics are shown in figure 3) and will afterwards comment on their particular choice. In §2.3, we will introduce a small programming language based on these operations.

$$\textbf{feed}(x, m_i, \nu): \quad \emptyset \to x:(\!(\nu m_i)\!) \tag{2.41}$$

$$\textbf{feed\_tag}(x, s, \nu): \quad \emptyset \to x:\nu s \tag{2.42}$$

$$\textbf{move}(s, x, z): \quad x:(s + q^*)(\!(P)\!) \to z:(s + q^*)(\!(P)\!) \tag{2.43}$$

$$x:s \to z:s \tag{2.44}$$

$$\textbf{tag}(x): \quad x:s + q^*(\!(P)\!) \to x:(s + q^*)(\!(P)\!) \tag{2.45}$$

$$\textbf{fuse}(x): \quad x:q_1^*(\!(P)\!) + q_2^*(\!(P')\!) \to x:(q_1^* + q_2^*)(\!(P + P')\!) \tag{2.46}$$

$$\textbf{flush}(x): \quad x:P \to \emptyset \tag{2.47}$$

$$\textbf{wrap}(x, z): \quad x:P \to z:(\!(P)\!) \tag{2.48}$$

$$\textbf{burst}(x): \quad x:q^*(\!(P)\!) \to x:q^*(\!(\,)\!) + P \tag{2.49}$$

Transitions (2.41)–(2.49) operate strictly on the top level of the state, meaning that we do not allow, for example, tagging or fusion of chemtainers that reside within another chemtainer. To ensure this, we simply do not provide inference rules that would allow us to derive such transitions. However, the above transitions are allowed to operate in any context through the following inference

$$\frac{I : S \to S'}{I : S \circ \bar{S} \to S' \circ \bar{S}}. \tag{2.50}$$

We now comment on the individual choice of operations and their transitions. **feed** allows one to feed chemtainers that

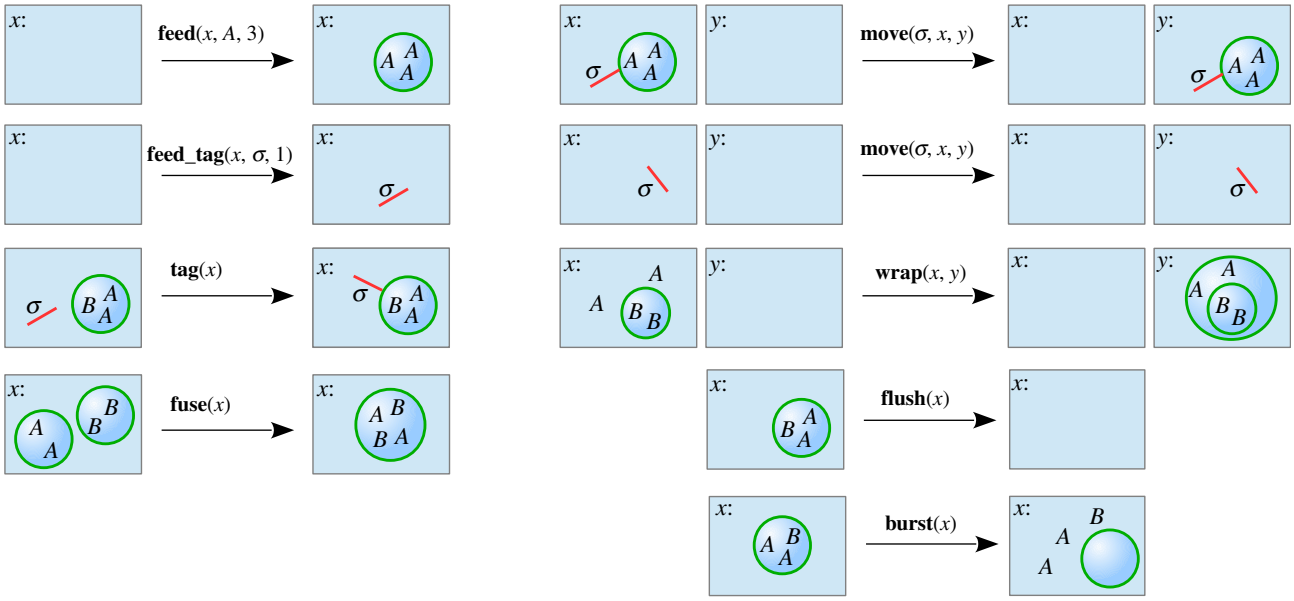

Figure 3. Schematics of programmable transitions and their instructions in the chemtainer calculus. (Online version in colour.)

are equipped with a certain number $\nu \in \mathbb{N}$ of molecules of a single molecular species into a location. This is the only means to provide elements of $\mathcal{M}$. Similarly, **feed_tag** allows one to feed a certain number $\nu \in \mathbb{N}$ of identical tags into a location. **tag** has been modelled to unspecifically attach some available tag to some container at a given location. No means are given to specify which tag or which chemtainer is transformed in this operation. The reason for this choice is that the actual linking of tags to a chemtainer involves linker molecules (e.g. biotin and streptavidin) that are common to all chemtainers and tags independent of the actual tag sequence or chemtainer type.

We provide a **flush** command to counteract the effect of feeding, where flushing simply removes content at a location. The operation is most easily implemented by literally flushing the material out of the system.

The **move** command allows one to move a specific tag, or a chemtainer decorated with a specific tag from one location to another. Notably, this allows for content separation: for example, the state $x:\sigma(\!(A)\!)+\tau(\!(B)\!)$ will transition into the state $x:\sigma(A)\circ y:\tau(B)$ in response to the operation **move**(τ, x, y). In a certain interpretation, this can be understood as chemtainer docking, where the locations correspond to volumes of bulk fluid that themselves move along a microfluidic channel decorated with the complementary tag (here τ). Assuming that both chemtainers are initially at location x which is decorated with τ, the complementarily tagged chemtainer is allowed to temporarily hybridize to the channel wall, whereas the bulk volume—and thus the locations—physically move along the channel. Non-matching content at the logical location x will remain at that location, although its physical position has changed. Matching content, however, will be at a different location, e.g. y and can be released into the corresponding bulk.

The operations **fuse** and **wrap** are more specific to the artificial cellular matrix and enable fusion and (vesicle) encapsulation. **fuse** induces unspecific fusion of co-located chemtainers, leading to mixing of both chemtainer surface and content. Besides unspecific fusion, tag-specific fusion of vesicles has been achieved experimentally [30,31]. A version of the chemtainer calculus featuring such tag-specific fusion is presented in [32]. **wrap** enables encapsulation of material into vesicles, e.g. using interface transfer: it is assumed that the material at location x resides in an aqueous phase which is first immersed into an oil phase where it forms surfactant-coated water droplets—providing the inner layer of the future vesicle. Next, the water droplet passes a surfactant covered oil–water interface which provides the outer layer of the vesicle membrane. If the original state at x contains chemtainers, these are equally transferred into the interior of the new chemtainer, leading to nested chemtainers [12]. Microfluidic implementations of this protocol are presented in [33]. We will later analyse the power of this transition system with and without the **wrap** operation.

Complementing encapsulation, we provide a **burst** operation that releases the content of a chemtainer. We envision that **burst** does not fully dissolve the chemtainer, but rather ruptures the chemtainer wall temporarily, for example by means of a heat or salt shock. After bursting, the chemtainer will reconstitute itself and remain in the system, available for further processing.

In order to capture obvious implementation constraints of the physical machine, some of the operations are restricted to subsets of $\mathcal{L}$, e.g. **feed**(x) might require $x \in \mathcal{L}_{\mathbf{feed}} \subset \mathcal{L}$. It can be shown that the constructive power of the calculus is not affected by this limitation, as long as **move**, **tag** and **burst** are allowed to operate on the entire set $\mathcal{L}$. Likewise, we could constrain the **move** command such that the target location has to be in a certain neighbourhood set of the source location in order to capture, e.g. the channel topology of microfluidic devices.

We wish to clarify one point that might be counterintuitive. Assume, for example, the state $x:10\sigma$. What will be the action to the operation **move**(σ, x, y)? Intuitively, we might expect that the system transitions into the state $y:10\sigma$, meaning that all instances of σ have been moved. However, $x:10\sigma$ is structurally equivalent to $x:9\sigma\circ x:\sigma$ owing to the distributive relation (2.18). To this state, we can apply the transition $S\circ x:\sigma \rightarrow S\circ y:\sigma$ equating $S = x:9\sigma$, which results in $x:9\sigma\circ y:\sigma$. Thus, **move**, and by similarity all other transitions, will operate only on *a single* instance. There is a simple way out, however: in order to move all 10 instances in the example above, we can simply execute the move operation 10 times in sequence. Our reason for this semantics is that it allows for

consistent assignment of stochastic rates in a potential stochastic semantics [34,35].

2.3. Programs

In §2.2, we have informally given parametric names to transitions, such as **move**(s, x, y), **tag**(x), a.s.o. We next interpret parametric names as elemental operations of a programming language that can be used to operate the artificial cellular matrix. We do this by first defining another formal language, the formal language of chemtainer programs, and then expressing the semantics of this language by means of state transitions, the latter employing the language $L(G_S)$ of chemtainer states.

As with states, we define programs by a recursive grammar, which we initially keep as simple as possible

$$\pi := \epsilon \mid I;\ \pi. \tag{2.51}$$

A program π can be either the empty program ϵ, or a (parametric) instruction followed by another program (the remainder), with instructions being separated by semicolons. Instructions I can be any of the formerly introduced parametric names **feed**(x_a, m_j, ν), **feed_tag**(x_a, qk, ν), **move**(sk, x_a, x_b), **tag**(x_a), **wrap**(x_a, x_b), **fuse**(x_a), **burst**(x_a) and **flush**(x_a) with $x_a, x_b \in \mathcal{L}$, $m_j \in \mathcal{M}$, $q_k \in L(G_{q^*})$ and $\nu \in \mathbb{N}$. Each instruction I has a set of associated transitions, and we write $I : S \to S'$ to denote any such associated transition (as was already done in the previous part).

Strictly, the concatenation operator ';' is only defined to concatenate single instructions with programs. In order to concatenate arbitrary programs π and π', we introduce an additional operator '**;**' for program concatenation through the following recursive definition

$$\epsilon;\ \pi = \pi \tag{2.52}$$

and

$$(I;\ \pi);\ \pi' = I;\ (\pi;\ \pi'). \tag{2.53}$$

Because both concatenation operators have different domains, it is clear from the context whether the concatenation refers to single instructions or programs. Therefore, we will drop the syntactic differentiation and use the symbol ';' for both operators.

A program π together with a system state S is expressed by the tuple $\langle \pi, S\rangle$ and we write $\langle \pi, S\rangle \to S'$ to denote that the program π can transform the state S into the state S', called a *result* of π. Results are defined by the following structural operational semantics [36]:

$$\langle \epsilon, S\rangle \to S, \tag{2.54}$$

$$\frac{\langle \pi, S''\rangle \to S \quad I : S' \to S''}{\langle I;\ \pi, S'\rangle \to S} \tag{2.55}$$

and

$$\frac{\langle \pi, S'\rangle \to S}{\langle I;\ \pi, S'\rangle \to S}. \tag{2.56}$$

In words, given a program that starts with instruction I with some associated transition from S' to S'', the second rule allows us to transform the system into S'', thereby reducing the original program to whatever remains after execution of I. If program execution encounters an instruction with no associated transition that is applicable with the current state S' (e.g. **tag** is asked to operate on an empty cell), the third rule allows us to skip the instruction without modifying the system state. This latter rule primarily addresses erroneous conditions and ensures that a program does not get stuck, simply because the respective molecules are not present at some point of the execution. Using non-deterministic semantics, rule (2.56) might even be used when rule (2.55) is applicable, thereby skipping instructions of the program sequence. This could be remedied with a stochastic semantics that assigns an arbitrarily small transition rate to rule (2.56).

We can now distinguish between autonomous and induced transitions, by extending the semantics with the rules

$$\frac{S \to S' \quad \langle \pi, S'\rangle \to S''}{\langle \pi, S\rangle \to S''} \tag{2.57}$$

and

$$\frac{\langle \pi, S\rangle \to S' \quad S' \to S''}{\langle \pi, S\rangle \to S''}, \tag{2.58}$$

where $S \to S'$ and $S' \to S''$ is an autonomous transition inferred through (2.35). These rules allow autonomous transitions to occur at any time during program execution without affecting the program.

We extend structural congruence to program application using the inference rule

$$\frac{S \equiv S'' \quad \langle \pi, S\rangle \to S' \quad S' \equiv S'''}{\langle \pi, S''\rangle \to S'''}. \tag{2.59}$$

Lemmas 2.1 and 2.2, proved in the electronic supplementary material, allow for uncomplicated composition of instructions.

Lemma 2.1. *Program application is not context sensitive: if the application of program π on the initial state S can lead to the result S', the application on the composition $S \circ \bar{S}$ can lead to the result $S' \circ \bar{S}$, containing S' as a substate. $\bar{S}$ is an invariant of π:*

$$\langle \pi, S\rangle \to S' \quad \Rightarrow \quad \langle \pi, S \circ \bar{S}\rangle \to S' \circ \bar{S}. \tag{2.60}$$

Lemma 2.2. *Programs can be concatenated, and the start configuration of the second program will be the result of the first program:*

$$\langle \pi', S\rangle \to S' \quad \wedge \quad \langle \pi'', S'\rangle \to S'' \quad \Rightarrow \quad \langle \pi';\ \pi'', S\rangle \to S''. \tag{2.61}$$

2.4. Extension: control flow directives and parallel execution

We could easily define control flow directives known from standard programming languages, such as branched execution and loops. This is particularly interesting in the context of feedback control. Let

$$p : S \to \{\text{true},\ \text{false}\} \tag{2.62}$$

be a conditional over the system state. For example, f could measure whether a fluorescent signal at a certain location exceeds a certain threshold. We could then extend our grammar to

$$\pi := \epsilon \mid I;\ \pi \mid \textbf{while}\ p(S)\ \textbf{do}\ \pi;\ \textbf{endwhile}, \tag{2.63}$$

to allow for conditional loops, where the semantics of the **while** statement are given by

$$\frac{\langle \pi, S\rangle \to S' \quad \langle \textbf{while}\ p\ \textbf{do}\ \pi;\ \textbf{endwhile}, S'\rangle \to S'' \quad p = \text{true}}{\langle \textbf{while}\ p\ \textbf{do}\ \pi;\ \textbf{endwhile}, S\rangle \to S''} \tag{2.64}$$

$$\frac{p = \text{false}}{\langle \textbf{while}\ p\ \textbf{do}\ \pi;\ \textbf{endwhile}, S\rangle \to S}. \tag{2.65}$$

As evidenced with these definitions, there is nothing particular about control flow directives in the chemtainer calculus

compared with other imperative languages. In order to introduce more elaborate control flows, the reader can simply apply standard text book procedures.

Similarly, it is straightforward to define the semantics for a language of communicating sequential processes [37,38] that parallelizes the production process

$$\frac{\langle \pi', S' \rangle \to \bar{S}' \quad \langle \pi'', S'' \rangle \to \bar{S}''}{\langle \pi' \mid \pi'', S' \circ S'' \rangle \to \bar{S}' \circ \bar{S}''}. \tag{2.66}$$

2.5. Example

To make the above definitions more familiar, we employ the chemtainer calculus in order to decorate chemtainers with addresses and use those to mix the contents of two chemtainers.

The left column shows which parametric instruction is executed, and the right column shows the effect of its associated transition. Read from top to bottom, the left column therefore simply gives a program that can construct some desired state denoted on the right.

$$\left.\begin{array}{ll} & \emptyset \\ \textbf{feed}(x, P) & x : (\!(P)\!) \\ \textbf{feed\_tag}(y, \sigma) & x : (\!(P)\!) \circ y : \sigma \\ \textbf{move}(\sigma, y, x) & x : \sigma + (\!(P)\!) \\ \textbf{tag}(x) & x : \sigma\, (\!(P)\!). \end{array}\right\} \tag{2.67}$$

To use this sequence later on in programs, we define it as the parametric macro **resource**(x, σ, P). Now, we use this macro to create two chemtainers and fuse them

$$\left.\begin{array}{ll} & \emptyset \\ \textbf{resource}(x, \sigma, P) & x : \sigma\, (\!(P)\!) \\ \textbf{resource}(y, \rho, Q) & x : \sigma\, (\!(P)\!) \circ y : \rho\, (\!(Q)\!) \\ \textbf{move}(\rho, y, x) & x : \sigma\, (\!(P)\!) + \rho\, (\!(Q)\!) \\ \textbf{fuse}(x) & x : (\sigma + \rho)(\!(P + Q)\!). \end{array}\right\} \tag{2.68}$$

Keep in mind that the right column only shows a possible response of the system state to the program on the left, not the necessary response. In general, because of the ambiguity of transitions, the system could have transitioned into alternative states in response to the same program.

3. Instruction sets and their constructive power

The instructions given in equations (2.41)–(2.49) serve as examples of possible operations, rather than the final specification of an artificial cellular matrix. Potential real-world implementations might involve only a subset of the operations mentioned, or might enable other means of chemtainer manipulation. It then becomes interesting to determine whether different instruction sets are equivalent in what they can construct or whether the set of constructable states differs.

In order to approach the problem formally, we define the *constructive closure* $\omega^+(\Pi_x) \subset L(G_S)$ of the language Π_x as the set of states that can be constructed from the empty state through some program $\pi \in \Pi_x$. Formally

$$\omega^+(\Pi_x) := \{S \in L(G_S) : \exists \pi \in \Pi_x \wedge \langle \pi, \emptyset \rangle \to S\}. \tag{3.1}$$

Similarly, we define the *reset closure* $\omega^-(\Pi_x) \subset L(G_S)$ as the set of states that can be transformed back into the empty set. Formally

$$\omega^-(\Pi_x) := \{S \in \omega^+(\Pi_x) : \exists \pi \in \Pi_x \wedge \langle \pi, S \rangle \to \emptyset\}. \tag{3.2}$$

Of primary interest is of course the case $\omega^- = \omega^+$ where the machine can be reset from any state. In this case, one can formally transform any initial state into any target state, thereby programming sequences of states: if $\langle \pi, S \rangle \to \emptyset$ and $\langle \pi', \emptyset \rangle \to S'$, the concatenation $\pi; \pi'$ will transform S into S': $\langle \pi; \pi', S \rangle \to S'$. This result, however, is mostly of theoretical interest, as an interim reset of the machine might not be desired when programming such state sequences. In a practical application, one would define a distance measure between states and ensure that the transition path length of programs is minimized, thereby preferring for example a single **move** instruction over a sequence of **flush;feed_tag;move** instructions with equal outcome.

We consider the following three incremental instruction sets

$$I_{\min} = \{\textbf{feed}, \textbf{feed\_tag}, \textbf{move}, \textbf{tag}, \textbf{fuse}, \textbf{flush}\} \tag{3.3}$$

$$I_{\text{wrap}} = I_{\min} \cup \{\textbf{wrap}\} \tag{3.4}$$

$$I_{\text{burst}} = I_{\text{wrap}} \cup \{\textbf{burst}\}. \tag{3.5}$$

Let $\Pi_{\textbf{min}}$, $\Pi_{\textbf{wrap}}$ and $\Pi_{\textbf{burst}}$ be the sets of programs over the respective instruction set. We will show the following relations between the corresponding constructive and reset closures:

$$\begin{array}{ccccc} \omega^+(\Pi_{\textbf{min}}) & \subset & \omega^+(\Pi_{\textbf{wrap}}) & \subset & \omega^+(\Pi_{\textbf{burst}}) = L(G_S) \\ \| & & \| & & \cup| \qquad\qquad \vdots \vdots \\ \omega^-(\Pi_{\textbf{min}}) & \subset & \omega^-(\Pi_{\textbf{wrap}}) & \subset & \omega^-(\Pi_{\textbf{burst}}) \subseteq L(G_S). \end{array} \tag{3.6}$$

If and only if all locations are flushable, i.e. if $\mathcal{L}_{\text{flush}} = \mathcal{L}$, the right-most chain of relations simplifies to

$$\omega^+(\Pi_{\textbf{burst}}) = \omega^-(\Pi_{\textbf{burst}}) = L(G_S). \tag{3.7}$$

Diagram (3.6) can be decomposed into several theorems, for which we need to define the *nesting level* $d : L(G_S) \to \mathbb{N}$ of a state S or local state P as the following

$$d(\emptyset) = 0, \tag{3.8}$$

$$d(0) = 0, \tag{3.9}$$

$$d(m_j) = 0, \tag{3.10}$$

$$d(q^*) = 1, \tag{3.11}$$

$$d(x_i : P) = d(P), \tag{3.12}$$

$$d(S \circ S') = \max\{d(S), d(S')\}, \tag{3.13}$$

$$d(P + P') = \max\{d(P), d(P')\} \tag{3.14}$$

and

$$d(q\, (\!(P)\!)) = 1 + d(P). \tag{3.15}$$

With this definition, we can state the relations of diagram (3.6) more precisely:

Theorem 3.1. *The constructive closure of* $\Pi_{\textbf{min}}$ *is the set of states that do not contain nested chemtainers nor free floating molecules other than tags*:

$$\omega^+(\Pi_{\textbf{min}}) = \{S \in L(G_S) : d(S) \leq 1 \wedge S \not\equiv S' \circ x_i : m_j \text{ for } m_j \in \mathcal{M} \backslash \mathcal{T}\}. \tag{3.16}$$

Theorem 3.2. *The constructive closure of* $\Pi_{\textbf{wrap}}$ *is the set of states that do not contain free floating molecules other than tags:*

$$\omega^+(\Pi_{\textbf{wrap}}) = \{S \in L(G_S) : S \not\equiv S' \circ x_i : m_j \text{ for } m_j \in \mathcal{M} \backslash \mathcal{T}\}. \tag{3.17}$$

Theorem 3.3. *The constructive closure of* $\Pi_{\textbf{burst}}$ *is the entire set of states:*

$$\omega^+(\Pi_{\textbf{burst}}) = L(G_S). \tag{3.18}$$

Theorem 3.4. *The reset closure of all instruction sets is limited to states that do not contain free floating molecules other than tags in non-flushable locations:*

$$\omega^{-}(\Pi_{\mathrm{x}}) \subset \{S \in L(G_{\mathrm{S}}) : S \not\equiv S' \circ x_i : m_j \text{ for } x_i \notin \mathcal{L}_{\mathbf{flush}} \cup \mathcal{L}_{\mathbf{wrap\_in}},\ m_j \in \mathcal{M} \backslash \mathcal{T}\} \quad \forall \mathrm{x} \in \{\mathbf{min}, \mathbf{wrap}, \mathbf{burst}\} \tag{3.19}$$

Proofs for all theorems are given in the electronic supplementary material. Note that these proofs are constructive: for any given target structure $S \in \omega^{+}(\Pi_{\mathrm{x}})$, we can *automatically* generate a program that will construct S. This observation forms the core of a compiler that is able to generate a sequence of transitions for building a given target structure.

Theorems 3.1–3.4 assume that the entire set of molecules $\mathcal{M}$ can be fed as resources. We now address the case where only a subset $\mathcal{M}_{\mathrm{R}} \subset \mathcal{M}$ of compounds can be directly provided by **feed** and where chemical reactions are employed to produce compounds outside $\mathcal{M}_{\mathrm{R}}$.

For some reaction $\nu_1 m_1 + \nu_2 m_2 \rightarrow \nu_3 m_3 + \nu_4 m_4$ with $m_1, m_2 \in \mathcal{M}_{\mathrm{R}}$, we can employ the program

$$\left.\begin{array}{ll} & \emptyset \\ \mathbf{resource}(x, \psi, \nu_1 m_1) & x : \psi \langle\!(\nu_1 m_1)\!\rangle \\ \mathbf{resource}(y, \rho, \nu_2 m_2) & x : \psi \langle\!(\nu_1 m_1)\!\rangle \circ y : \rho \langle\!(\nu_2 m_2)\!\rangle \\ \mathbf{move}(\rho, y, x) & x : \psi \langle\!(\nu_1 m_1)\!\rangle + \rho \langle\!(\nu_2 m_2)\!\rangle \\ \mathbf{fuse}(x) & x : (\psi + \rho) \langle\!(\nu_1 m_1 + \nu_2 m_2)\!\rangle \\ & x : (\psi + \rho) \langle\!(\nu_3 m_3 + \nu_4 m_4)\!\rangle \end{array}\right\} \tag{3.20}$$

to initiate the chemical production of $P' = \nu_3 m_3 + \nu_4 m_4$, where m_3, m_4 are not necessarily elements of $\mathcal{M}_{\mathrm{R}}$. Applying this procedure repeatedly allows us to obtain successively bigger subsets of $\mathcal{M}$. However, without means for content separation, the set of producible states will be constrained by the stoichiometries of the reactions. If the artificial cellular matrix would offer content separation, e.g. by means of electrophoresis [39], we could encode this with the induced transition

$$\mathbf{separate}(x, y): \quad x : (q + q') \langle\!(P + P')\!\rangle \rightarrow y : q \langle\!(P)\!\rangle + q' \langle\!(P')\!\rangle. \tag{3.21}$$

where the exact partitioning of chemtainer content into P and P' would depend on its electrophoretic mobility. Continuing the above program, we could now derive

$$\left.\begin{array}{ll} & x : (\psi + \rho) \langle\!(\nu_3 m_3 + \nu_4 m_4)\!\rangle \\ \mathbf{separate}(x, y) & y : \psi \langle\!(\nu_3 m_3)\!\rangle + \rho \langle\!(\nu_4 m_4)\!\rangle. \end{array}\right. \tag{3.22}$$

The tagged chemtainers are ready to be used as input for subsequent reactions. This recovers the constructive power of the chemtainer calculus, provided that all elements of $\mathcal{M}$ can be constructed by some chain of reactions. However, in order to maintain a well-defined mapping from DNA tags to chemtainer content, the **separate** operation would need to assert the correct redistribution of surface tags along with the content separation such that products can be unambiguously identified by their tags.

4. Programmable reaction cascades

Here, we demonstrate how our calculus (even with the minimal instruction set $I_{\mathbf{min}}$) can be used to integrate chemical production with molecular computation in order to control programmable reaction cascades. This example is motivated by and closely mimics the synthesis of oligosaccharides in the Golgi apparatus [1,2]. Oligosaccharides are branched, heterogeneous polymers composed of typically 5–10 individual sugar monomers such as mannose, galactose and glucose. This diverse class of biochemicals is involved in various physiological processes pertaining, e.g. to cell–cell recognition, intra- and intercellular trafficking and metabolic modulation [40]. However, their combinatorial richness poses a challenge for chemical oligosaccharide synthesis based on conventional chemical manufacturing techniques [41].

Biological oligosaccharide synthesis proceeds in the Golgi apparatus by adding individual monomers one unit at a time to specific binding sites of the growing oligomer. Monomers are attached to enzymes that promote the specific binding reactions

$$P_i + E_j^* M_{\mathrm{i}} \rightarrow P_{i+1} + E_j. \tag{4.1}$$

Here, P_i denotes an intermediate oligomer, to which monomer M_i is added at the site j. If binding sites are unique, a given enzyme–monomer complex $E_j^* M_i$ contains all the information required to build the specific product P_{i+1} from P_i. Prior to these polymerization steps, monomers have to be attached to the respective enzymes

$$M_i + E_j^* \rightarrow E_j^* M_i. \tag{4.2}$$

Chemical one-pot synthesis of a given target structure is challenging, because repetition of bindings sites in the oligomer structure can lead to undesired side products. The number of potential side products can be controlled, however, by forcing some reaction steps to occur sequentially while others are allowed to proceed in parallel [42]. Weyland *et al.* [43] present an algorithm that identifies such optimal reaction cascades. For example, assume that the structure shown in figure 4 can be produced with the reaction cascade

$$\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Man} \rightarrow P_0 + 2\,E_{\mathrm{Gal}-4}, \tag{4.3}$$

$$E^*_{\mathrm{Man}-6}\mathrm{Man} + E^*_{\mathrm{Man}-6}\mathrm{Glc} + E^*_{\mathrm{Man}-3}\mathrm{Glc} \rightarrow P_1 E^*_{\mathrm{Man}-6} + E_{\mathrm{Man}-6} + E_{\mathrm{Man}-3}, \tag{4.4}$$

$$E^*_{\mathrm{Man}-3}\mathrm{Man} + E^*_{\mathrm{Man}-2}\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Glc} \rightarrow P_2 E^*_{\mathrm{Man}-3} + E_{\mathrm{Man}-2} + E_{\mathrm{Gal}-4} \tag{4.5}$$

and

$$P_0 + E^*_{\mathrm{Man}-6} P_1 + E^*_{\mathrm{Man}-3} P_2 \rightarrow P_3 + E_{\mathrm{Man}-6} + E_{\mathrm{Man}-3}, \tag{4.6}$$

where it has to be ensured that reactions (4.3)–(4.5) occur in isolation and prior to reaction (4.6).

Our strategy here is to employ the chemtainers in order to control the encounter of reactants and hence the order of reactions by encapsulating reactants within chemtainers. Fusion of chemtainers co-locates desired reactants and triggers their polymerization. Simultaneously, DNA gate computation on the chemtainer surface will reflect the change of chemtainer content after reaction. The same DNA computation can ensure that reactions are started if and only if other reactions have been performed beforehand. We use one location x_0 for processing and one location x_S for storing intermediate products.

We start by preparing chemtainers with enzyme complexes using the program given in equation (3.20). For

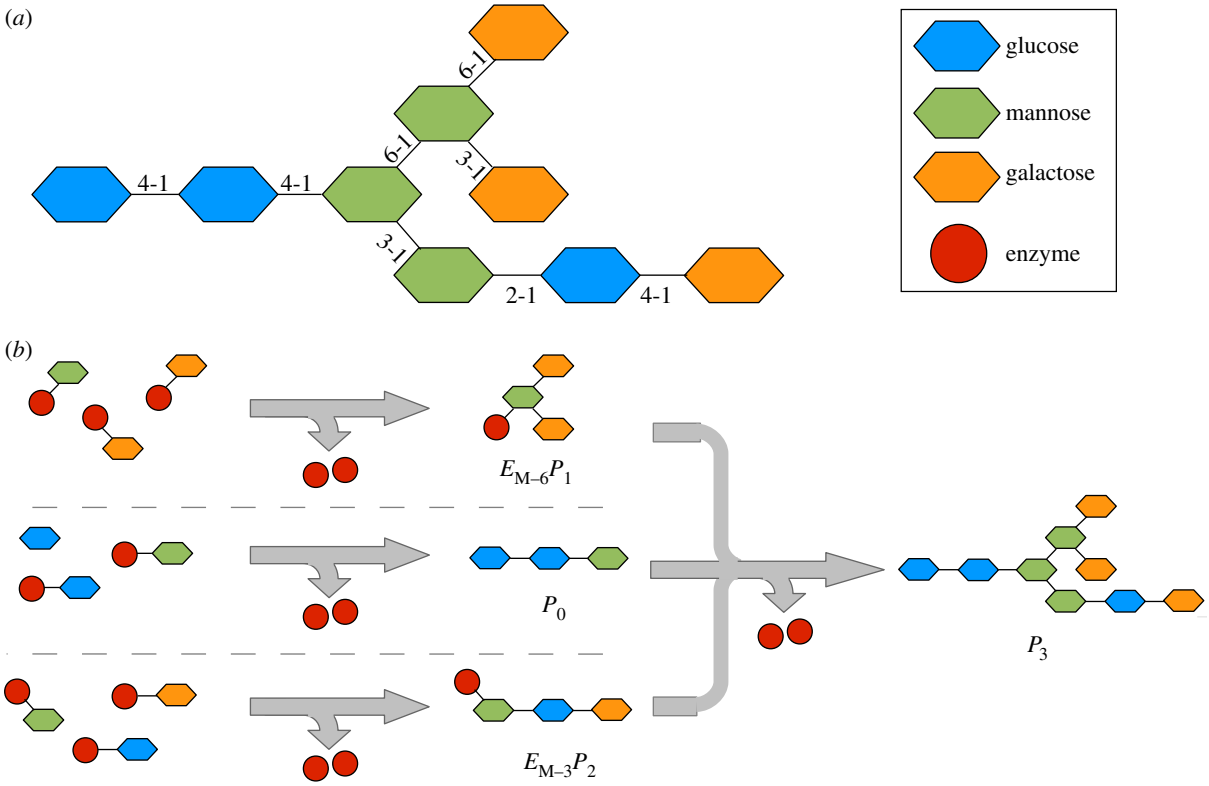

Figure 4. (*a*) Example of an oligosaccharide target structure, consisting of specifically connected monomers of different types (hexagons). (*b*) Reaction cascade to synthesize the target structure. Potential side reactions are avoided by encapsulating reactants into separate chemtainers (indicated by dashed lines). (Online version in colour.)

example

$$\begin{aligned}&\text{`}\mathbf{resource}(x_0, \sigma, \mathrm{Gal});\ \mathbf{resource}(x_0, \rho, E^*_{\mathrm{Gal}-4});\\ &\quad \mathbf{feed\_tag}(x_0, (\sigma+\rho)\triangleright\beta);\\ &\qquad\qquad \mathbf{fuse}(x_0);\ \mathbf{move}(\beta, x_0, x_S)\end{aligned} \tag{4.7}$$

creates the state $x_S : \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!)$. See figure 5 for a graphical representation of these steps. We use this algorithm repeatedly to set up the machine in the following state

$$\begin{aligned}x_S : {}&\alpha\,(\!(\mathrm{Gal})\!) + \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\ &+ \delta\,(\!(E^*_{\mathrm{Man}-6}\mathrm{Man})\!) + \epsilon\,(\!(E^*_{\mathrm{Man}-6}\mathrm{Glc})\!) + \zeta\,(\!(E^*_{\mathrm{Man}-3}\mathrm{Glc})\!)\\ &+ \eta\,(\!(E^*_{\mathrm{Man}-3}\mathrm{Man})\!) + \theta\,(\!(E^*_{\mathrm{Man}-2}\mathrm{Gal})\!) + \iota\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Glc})\!).\end{aligned} \tag{4.8}$$

With this mapping from chemical compounds to DNA tags, we translate the reaction cascade (4.3) and (4.6) into a set of DNA gates

$$(\alpha+\beta+\gamma)\triangleright\kappa, \tag{4.9}$$

$$(\delta+\epsilon+\zeta)\triangleright\lambda, \tag{4.10}$$

$$(\eta+\theta+\iota)\triangleright\mu \tag{4.11}$$

and

$$(\kappa+\lambda+\mu)\triangleright\omega. \tag{4.12}$$

To carry out reaction (4.3), we feed the gate (4.9) at location x_0. We then progressively move chemtainers from x_S to x_0, initiate their fusion and transport the intermediate product P_0 back to x_S (see figure 6 for a schematic representation):

$$\begin{aligned}
&& x_0 &: (\alpha+\beta+\gamma)\triangleright\kappa \circ x_S : \alpha\,(\!(\mathrm{Gal})\!) + \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{move}(\alpha, x_S, x_0) && x_0 &: (\alpha+\beta+\gamma)\triangleright\kappa + \alpha\,(\!(\mathrm{Gal})\!) \circ x_S : \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{move}(\beta, x_S, x_0) && x_0 &: (\alpha+\beta+\gamma)\triangleright\kappa + \alpha\,(\!(\mathrm{Gal})\!) + \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) \circ x_S : \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{move}(\gamma, x_S, x_0) && x_0 &: (\alpha+\beta+\gamma)\triangleright\kappa + \alpha\,(\!(\mathrm{Gal})\!) + \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{tag}(x_0) && x_0 &: (\alpha+(\alpha+\beta+\gamma)\triangleright\kappa)\,(\!(\mathrm{Gal} + \beta\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{fuse}(x_0) && x_0 &: (\alpha+\beta+(\alpha+\beta+\gamma)\triangleright\kappa)\,(\!(\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Gal})\!) + \gamma\,(\!(E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
\mathbf{fuse}(x_0) && x_0 &: (\alpha+\beta+\gamma+(\alpha+\beta+\gamma)\triangleright\kappa)\,(\!(\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Gal} + E^*_{\mathrm{Gal}-4}\mathrm{Man})\!)\\
&& x_0 &: \kappa\,(\!(P_0 + 2E_{\mathrm{Gal}-4})\!)\\
\mathbf{move}(\kappa, x_0, x_S) && x_S &: \kappa\,(\!(P_0 + 2\,E_{\mathrm{Gal}-4})\!)\\
\mathbf{flush}(x_0) && x_S &: \kappa\,(\!(P_0 + 2\,E_{\mathrm{Gal}-4})\!).
\end{aligned} \tag{4.13}$$

The **flush** instruction cleans the processing location in the case that the execution of an instruction was skipped because of rule (2.56). This prevents the subsequent fuse operations from operating on chemtainers that where intended to be kept separate.

Repeating the above algorithm with gate (4.10) and tags δ, ϵ, ζ as well as gate (4.11) and tags η, θ, ι produces chemtainers with the intermediate products P_0, P_1 and P_2 tagged with κ, λ and μ, respectively. Eventually, the above algorithm with gate (4.12) and tags κ, λ, μ produces the chemtainer $\omega\,(\!(P_3 + \ldots)\!)$ which can readily be moved to some output location.

Note that the above program operates over a finite set of resources, tags and locations, in order to build a compound from a potentially unlimited universe of target molecules, and the mapping between tags and compounds is established

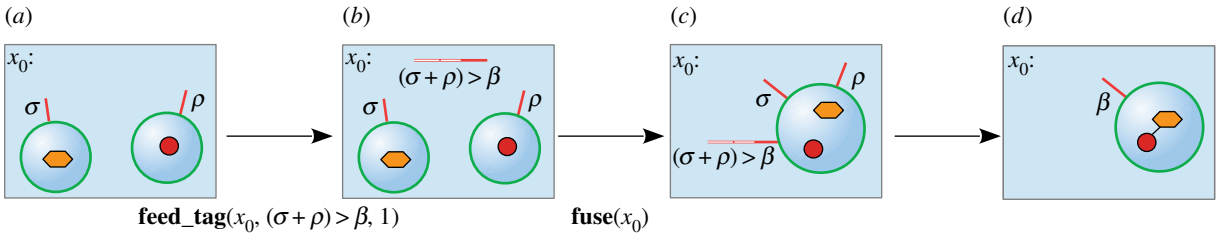

Figure 5. Steps for preparing enzymatically activated saccharide monomers. (*a*) Chemtainers with activated enzymes and monomers are fed to the same location. (*b*) A DNA gate is fed to the same location. (*c*) Fusion of chemtainers co-locates the reactants in a single chemtainer, to which the provided DNA gate attaches. (*d*) The reaction takes place inside the chemtainer, whereas the result of chemtainer fusion is reflected by the successful transition of the DNA gate. (Online version in colour.)

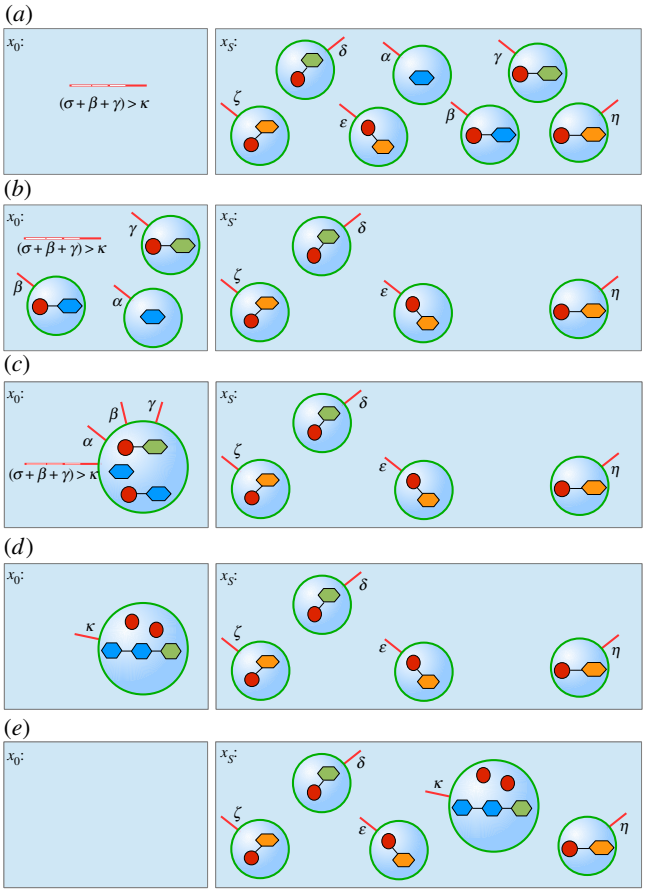

Figure 6. Steps in the programmed synthesis of compound P_0 from figure 4. (*a*) Initial system state, where resource chemtainers are provided at the storage location x_S and the DNA gate $(\alpha+\beta+\gamma)\triangleright\kappa$ is provided at the operation location x_0. (*b*) System state after chemtainers labelled α, β and γ are moved to x_0. (*c*) System state after fusion has been initiated in x_0, and the tag has been bound. (*d*) Chemical reactions and gate transitions occur autonomously at x_0. (*e*) The newly created chemtainer labelled κ with product P_0 is transferred back to x_S. (Online version in colour.)

only within the controlling program. This demonstrates again the universality of the programmatic synthesis approach.

5. Discussion

We have formally introduced the chemtainer calculus which is capable of capturing system states and transitions of an artificial cytoplasm. The chemtainer calculus allows us to describe the organization and manipulation of chemical compounds in possibly nested, addressable bioreactors. Elemental operations of a simple programming language are proposed that allow for the programmatic control of chemtainer transitions. A constructive proof is given that a programming language based on eight elemental instructions is capable of constructing any possible system configuration that can be expressed in the grammar of the chemtainer calculus. This proof can serve as a core component of a compiler for automated program inference. We have presented how our machinery can be used to compile and execute complex chemical synthesis protocols and have given an example from oligosaccharide synthesis which still poses a challenging task for conventional chemical manufacturing techniques. This framework for programmable chemical synthesis demonstrates how molecular computing and chemical production can be integrated in much the same way as in biological systems, for example in the Golgi apparatus.

One might object that chemical synthesis in chemtainers works the same as sequential mixings of reaction solutions in test tubes, which is a conventional methodology. What then is the added values of chemtainers? First, we point out that microfluidics in general (even without employing chemtainers), allows for the programmed set-up of reaction cascades in small volumes. This gives a general advantage over setting up reactions in test tubes either manually or even by liquid handling robots: as pointed out by Füchslin *et al.* [42], small spatial dimensions offer the possibility for rapid transport of intermediate compounds among different reaction environments. The accompanying reduction in processing times can be of crucial advantage in synthesis steps where intermediate compounds are only stable for short periods of time. Consequently, compounds not viable in present processing may become interesting candidates for, e.g. catalysis in miniaturized systems. Employing vesicles as reaction compartments allows for the use of even smaller reaction volumes on the femtolitre scale. Second, DNA-addressable reaction compartments can sometimes *avoid* the need for distinct reaction environments, in the sense that all reactants can be provided simultaneously in the same 'pot', but individual reactions are controlled by DNA-mediated vesicle fusion. This 'automated assembly' is programmed not in the microfluidic control, but in the DNA tagging of vesicles. Although this has not been exemplified in this paper, we can envision synthesis pathways where DNA tags participate in the reaction cascades, e.g. as aptamers. Folding the DNA tag set with the set of chemicals would allow chemical reactions to directly 'report' on their progress, such that, e.g. a DNA tag operation is triggered only after a chemical compound has been consumed. DNA-tagged reaction compartments further allow for the extraction and recovery of unreacted compounds by their unaltered DNA tag, and the extraction of reaction products by their altered DNA tag. Third,

encapsulation of compounds into vesicular reaction compartments offers additional advantages for chemical manufacturing in microfluidics. Compounds do not contaminate microfluidic channels as they are physically contained in vesicles. Vesicles can expose a unified physical 'interface' (in terms of friction, buoyancy, charge density, etc.) for microfluidic control independent of their content. By altering the composition of membrane molecules, these properties can be altered vastly independently of the vesicle content.

In all our derivations, we have made ample use of non-deterministic semantics: we have proved that there exists a program that is able to induce desired transitions, and is thus able to construct a desired state. We have not taken into account the *likelihood* of those transitions—especially with respect to possible but undesired side reactions. As this is an important issue in the area of programming chemistry, we have carefully designed our transition system with an extension towards stochastic semantics in mind [34]. Application of these known techniques to the chemtainer calculus would be the subject of future work.

Acknowledgements. Steen Rasmussen, Andrew Phillips, Rasmus Petersen, Rudolf Füchslin, Ehud Shapiro and members of the MatchIT consortium are acknowledged for helpful discussions.

Funding statement. The research leading to these results has received funding from the Danish National Research Foundation through the Center for Fundamental Living technologies (FLinT) and from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreements no. 249032 (MATCHIT) and no. 318671 (MICReagents).

References

1. Rothman JE. 1981 The Golgi apparatus: two organelles in tandem. *Science* **213**, 1212–1219. (doi:10.1126/science.7268428)
2. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Watson P. 2002 *Molecular biology of the cell*. New York, NY: Garland Science Publishing.
3. Rothman JE. 1994 Mechanisms of intracellular protein transport. *Nature* **372**, 55–63. (doi:10.1038/372055a0)
4. Nardin C, Widmer J, Winterhalter M, Meier W. 2001 Amphiphilic block copolymer nanocontainers as bioreactors. *Euro. Phys. J. E* **4**, 403–410. (doi:10.1007/s101890170095)
5. Monnard P-A, Membr J. 2003 Liposome-entrapped polymerases as models for microscale/nanoscale bioreactors. *Biol.* **191**, 87.
6. Noireaux V, Libchaber A. 2004 A vesicle bioreactor as a step toward an artificial cell assembly. *Proc. Natl Acad. Sci. USA* **101**, 17 669–17 674. (doi:10.1073/pnas.0408236101)
7. Roodbeen R, van Hest JCM. 2009 Synthetic cells and organelles: compartmentalization strategies. *Bioessays* **31**, 1299–1308. (doi:10.1002/bies.200900106)
8. Richard A, Marchi-Artzner V, Lalloz M-N, Brienne M-J, Artzner F, Gulik-Krzywicki T, Guedeau-Boudeville M-A, Lehn J-M. 2004 Fusogenic supramolecular vesicle systems induced by metal ion binding to amphiphilic ligands. *Proc. Natl Acad. Sci. USA* **101**, 15 279–15 284. (doi:10.1073/pnas.0406625101)
9. Caschera F, Sunami T, Matsuura T, Suzuki H, Hanczyc M. 2011 Programmed vesicle fusion triggers gene expression. *Langmuir* **27**, 13 082–13 090. (doi:10.1021/la202648h)
10. Terasawa H, Nishimura K, Suzuki H, Matsuura T, Yomo T. 2012 Coupling of the fusion and budding of giant phospholipid vesicles containing macromolecules. *Proc. Natl Acad. Sci. USA* **109**, 5942–5947. (doi:10.1073/pnas.1120327109)
11. Hadorn M, Eggenberger Hotz P. 2010 DNA-mediated self-assembly of artificial vesicles. *PLoS ONE* **5**, e9886. (doi:10.1371/journal.pone.0009886)
12. Hadorn M, Bönzli E, Eggenberger Hotz P, Hanczyc MM. 2012 Hierarchical unilamellar vesicles of controlled compositional heterogeneity. *PLoS ONE* **7**, e50156. (doi:10.1371/journal.pone.0050156)
13. Hadorn M, Bönzli E, Fellermann H, Eggenberger Hotz P, Hanczyc MM. 2012 Specific and reversible DNA-directed self-assembly of oil-in-water emulsion droplets. *Proc. Natl Acad. Sci. USA* **109**, 20 320–20 325. (doi:10.1073/pnas.1214386109)
14. Amos M, Dittrich P, McCaskill J, Rasmussen S. 2011 In *Proc. from the 2nd European Future Technologies Conf. and Exhibition 2011* (*FET 11*), *4–6 May 2011, Budapest, Hungary*, pp. 56–60. Procedia Computer Science.
15. Rasmussen S, Albertsen AN, Fellermann H, Pedersen PL, Svaneborg C, Ziock H-J. 2011 In *Proc. 13th Annual Conf. Companion on Genetic and Evolutionary Computation* (*GECCO 11*), 12–16 July 2011, Dublin, Ireland, p. 1520. New York, NY: ACM.
16. Hadorn M *et al*. 2012 In *Proc. 13th Int. Conf. on the Simulation and Synthesis of Artificial Life 19–22 July, East Lansing, Michigan, USA* (eds C Adami, DM Bryson, C Ofria, RT Pennock). Cambridge, MA: MIT Press.
17. Fellermann H, Hadorn M, Bönzli E, Rasmussen S. 2012 In *Biomimetic and biohybrid systems*. Lecture Notes in Computer Science No. 7375 (eds TJ Prescott, NF Lepora, A Mura, PFMJ Verschure), pp. 343–344. Berlin, Germany: springer.
18. Cardelli L. 2011 Strand algebras for DNA computing. *Nat. Comput.* **10**, 407–428. (doi:10.1007/s11047-010-9236-7)
19. McCaskill JS, Wagler P. 2000 In *Field-programmable logic and applications: the roadmap to reconfigurable computing*, Lecture Notes in Computer Science No. 1896 (eds RW Hartenstein, H Grünbacher), pp. 286–299. Berlin, Germany: Springer.
20. Wagler PF, Tangen U, Maeke T, McCaskill JS. 2012 Field programmable chemistry: integrated chemical and electronic processing of informational molecules towards electronic chemical cells. *Biosystems* **109**, 2–17. (doi:10.1016/j.biosystems.2012.01.005)
21. Nishimura K, Suzuki H, Toyota T, Yomo T. 2012 Size control of giant unilamellar vesicles prepared from inverted emulsion droplets *J. Colloid Interface Sci.* **376**, 119–125. (doi:10.1016/j.jcis.2012.02.029)
22. Wagler PF, Tangen U, Maeke T, Mathis HP, McCaskill JS. 2003 Microfabrication of a BioModule composed of microfluidics and digitally controlled microelectrodes for processing biomolecules *Smart Mater. Struct.* **12**, 757–762. (doi:10.1088/0964-1726/12/5/012)
23. Tresset G, Takeuchi S. 2004 A microfluidic device for electrofusion of biological vesicles. *Biomed. Microdevices* **6**, 213–218. (doi:10.1023/B:BMMD.0000042050.95246.af)
24. Tanand Y, Hettiarachchi K, Siu M, Pan Y, Lee A, Amer J. 2006 Controlled microfluidic encapsulation of cells, proteins, and microbeads in lipid vesicles. *Chem. Soc.* **128**, 5656–5658. (doi:10.1021/ja056641h)
25. Păun G. 2000 Computing with membranes. *J. Comput. Syst. Sci.* **61**, 108–143. (doi:10.1006/jcss.1999.1693)
26. Regev A, Panina E, Silverman W, Cardelli L, Shapiro E. 2004 BioAmbients: an abstraction for biological compartments. *Theor. Comput. Sci.* **325**, 141–167. (doi:10.1016/j.tcs.2004.03.061)
27. Cardelli L. 2005 Brane calculi, interactions of biological membranes. In *Computational methods in systems biology* (eds V Danos, V Schachter), pp. 257–278. Berlin, Germany: Springer.
28. Fellermann H *et al*. 2013 Integrated production/computation IT architecture for MATCHIT. Tech. Rep. no. 249032-6.4. See http://ec.europa.eu/information_society/apps/projects/logos/2/249032/080/deliverables/001_MATCHITDeliverable64Final.pdf.
29. Wieczorek RM. 2011 Astronomical content in Rongorongo tablet Keiti. *Le Journal de la Socit des Ocanistes* **132**, 5–16. (doi:10.4000/jso.6272)
30. Stengel G, Zahn R, Höök F. 2007 DNA-induced programmable fusion of phospholipid vesicles. *J. Am. Chem. Soc.* **129**, 9584–9585. (doi:10.1021/ja073200k)

31. Chan YM, Lengerich BV, Boxer SG. 2009 Effects of linker sequences on vesicle fusion mediated by lipid-anchored DNA oligonucleotides. *Proc. Natl Acad. Sci. USA* **106**, 979–984. (doi:10.1073/pnas.0812356106)
32. Fellermann H, Krasnogor N. 2014 In *Language, life, limits*. Lecture Notes in Computer Science, vol. 8493 (eds A Beckmann, E Csuhaj-Varjú, K Meer), pp. 173-182. Berlin, Germany: Springer.
33. Shum HC, Thiele J, Kim S-H. 2014 In *Advances in Transport Phenomena 2011,* Advances in Transport Phenomena No. 3 (ed. L Wang*)*, pp. 1–28. Berlin, Germany: Springer.
34. Bacci G, Miculan M. 2012 Measurable stochastics for Brane Calculus. *Theor. Comput.* **431**, 117–136. (doi:10.1016/j.tcs.2011.12.055)
35. Miculan M, Sambarino I. 2012 In *Proc. the Sixth Workshop on Membrane Computing and Biologically Inspired Process Calculi MeCBIC 2012* (eds B Aman, G Ciobanu), p. 87101.
36. Plotkin GD. 2004 The origins of structural operational semantics. *J. Logic Algebr. Program.* **3**, 60–61.
37. Hoare CAR. 1978 Communicating sequential processes. *Commun. ACM* **21**, 666–677. (doi:10.1145/359576.359585)
38. Milner R. 1982 *A calculus of communicating systems*. New York, NY: Springer.
39. Tang T, Badal MY, Ocvirk G, Lee WE, Bader DE, Bekkaoui F, Harrison DJ. 2002 Integrated microfluidic electrophoresis system for analysis of genetic materials using signal amplification methods. *Anal. Chem.* **74**, 72533.
40. Varki A. 1993 Biological roles of oligosaccharides: all of the theories are correct. *Glycobiology* **3**, 97–130. (doi:10.1093/glycob/3.2.97)
41. Koeller KM, Wong C. 2000 Complex carbohydrate synthesis tools for glycobiologists: enzyme-based approach and programmable one-pot strategies. *Glycobiology* **10**, 1157–1169. (doi:10.1093/glycob/10.11.1157)
42. Füchslin RM, Dzyakanchuk A, Flumini D, Hauser H, Hunt KJ, Luchsinger RH, Reller B, Scheidegger S, Walker R. 2013 Morphological computation and morphological control: steps toward a formal theory and applications. *Artif. Life* **19**, 9–34. (doi:10.1162/ARTL_a_00079)
43. Weyland MS, Fellermann H, Hadorn M, Sorek D, Lancet D, Rasmussen S, Füchslin RM. 2013 The MATCHIT automaton: exploiting compartmentalization for the synthesis of branched polymers. *Comput. Math. Methods Med.* **2013**, 467428. (doi:10.1155/2013/467428)