DNA as a Universal Substrate for Chemical Kinetics
(Extended Abstract)

David Soloveichik, Georg Seelig, and Erik Winfree

California Institute of Technology, Pasadena, CA, USA
dsolov@caltech.edu, seelig@dna.caltech.edu, winfree@caltech.edu

Abstract. We show that a DNA-based chemical system can be constructed such that it closely approximates the dynamic behavior of an arbitrary system of coupled chemical reactions. Using strand displacement reactions as a primitive we explicitly construct reaction cascades with effectively unimolecular and bimolecular kinetics. Our construction allows for individual reactions to be coupled in arbitrary ways such that reactants can participate in multiple reactions simultaneously, correctly reproducing the desired dynamical properties. Thus arbitrary systems of chemical equations can be compiled into chemistry. We illustrate our method on a chaotic Rössler attractor; simulations of the attractor and of our proposed DNA-based implementation show good agreement.

1 Introduction

Chemical reaction equations and mass action kinetics provide a powerful mathematical language for describing and analyzing chemical systems. For well over a century, mass action kinetics has been used to model chemical experiments, in order to predict and explain the evolution of the various species over time, and to elucidate the dynamical properties of the system under investigation. Chemistry exhibits complex behavior like oscillations, limit cycles, chaos or pattern formation, all of which can be explained by the corresponding systems of coupled chemical reactions [1,2,3]. While the use of mass action kinetics to describe existing chemical systems is well established, the inverse problem of experimentally implementing a given set of chemical reactions has not been widely considered. Many systems of coupled chemical equations appear to not have realizations in known chemistry.

Here we propose a method for implementing an arbitrary system of coupled chemical reactions using nucleic acids. We develop an explicit implementation of unimolecular and bimolecular reactions which can be combined into arbitrarily coupled reaction networks. In a formal system of chemical reactions such as

$$A \xrightarrow{k_1} B$$
$$A + B \xrightarrow{k_2} C + D$$
$$C \xrightarrow{k_3}$$

A. Goel, F.C. Simmel, and P. Sosík (Eds.): DNA 14, LNCS 5347, pp. 57–69, 2009.

a species may need to participate in multiple reactions, both as a reactant and/or as a product (species A, B or C) and these reactions need to progress at rates determined by the rate constants (k_1, k_2 and k_3). This imposes onerous constraints on the chemical properties of the species participating in these reactions. For example, it is likely hard to find a physical implementation of the chemical reaction equations using small molecules, since small molecules have a limited set of reactivities. Information-bearing biopolymers such as proteins or nucleic acids provide a more promising physical substrate for implementing arbitrary chemical reactions. Nucleic acids have the unique advantage that interactions between different single-stranded species can be programmed since sequence determines reactivity through Watson-Crick base pairing.

In our DNA implementation, we assign each formal species (e.g., A, B, C, D) to a set of DNA molecules. In some instances it may be possible to map a formal species to a single oligonucleotide but more generally a single formal species will correspond to several DNA species in order to reproduce the correct kinetics. Effective interactions between the species are mediated by an additional set of DNA complexes. Since the underlying chemistry involves aqueous-phase nucleic acid hybridization and strand exchange reactions, arbitrarily large rate constants and concentrations cannot be attained. However, any system of coupled chemical reactions can be scaled to use smaller rate constants and concentrations without affecting the kinetics except by a scaling factor (see Section 6). While our constructions are purely theoretical at this point, they are based on realistic assumptions and provide a roadmap for future experiments.

In the next section we describe strand displacement reactions that will serve as the basic building block for our construction. In the following section we show how to implement arbitrary unimolecular reactions, and then extend our construction to cover bimolecular reactions. In the final section we give a demonstration of our approach by describing the implementation of a system due to Willamowski and Rössler [4] with 3 species and 7 reactions exhibiting chaotic concentration fluctuations. Numerical simulations of the original formal system and our DNA-based chemical reactions using realistic rate constants and concentrations are in good agreement.

2 Cascades of Strand Displacement Reactions

We use strand displacement reactions as the basic primitive for our constructions (Fig. 1). Strand displacement has been found to be a flexible method for designing complex behaviors with nucleic acids including motors, logic gates, and catalysts [5,6,7,8]. Although a strand displacement reaction involves multiple elementary steps, including a random walk process, it is well modeled as a second-order process for a wide range of reaction conditions [9,10]. The effective rate constant of the second-order process is governed by the degree of sequence complementarity between the toeholds on the single-stranded species and on the partially double-stranded species [10].

We have recently used strand displacement cascades to construct DNA-based logic circuits [6,8]. Here we use some of the nomenclature and ideas from that

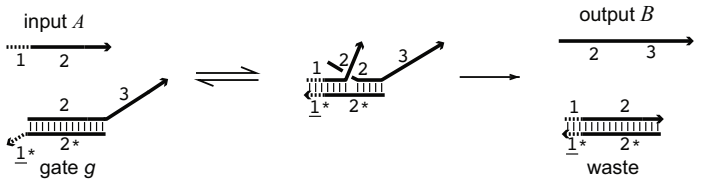

Fig. 1. Strand displacement reactions. The $3'$ end of each strand is indicated by an arrow. Functional domains are numbered and the star indicates complementarity. We use the underline notation $\underline{1}^*$ to indicate that this domain may not be completely complementary to domain 1. The reaction between input strand A and gate g is initiated at the toehold (dashed, domain 1^*). The reaction then proceeds through multiple short-lived intermediates and leads to the release of an output strand B and the formation of a chemically inert double-stranded waste product. Kinetically, the overall reaction is well approximated as being bimolecular, i.e., $A + g \xrightarrow{k} B$, where we omit the inert waste product. The value of the rate constant k depends on reaction conditions (salt, temperature), length and sequence composition of the toehold as well as the degree of complementarity between the toeholds on the strand and gate (domains 1 and $\underline{1}^*$). In practice, toehold domains are typically 2–8 nucleotides long, and the domains undergoing strand displacement are typically 20–30 nucleotides long.

work. Fig. 2 shows a two-stage strand displacement cascade where an input single-stranded nucleic acid species (strand) initiates a strand displacement cascade between two complexes (gates) leading to the release of an output strand. In strand displacement cascades, a strand is functionally inactive before its release from a gate and becomes active upon becoming completely single-stranded. For example, intermediate strand o cannot react with translator gate t before it is released from gate g because its toehold domain 3, which is required for initiating the reaction with t, is double-stranded. Similarly, output Bs cannot initiate a downstream strand displacement cascade until it is released from translator gate t because its toehold domain 4 is double-stranded. However, upon the addition of free As, intermediate strand o is released through strand displacement, which then causes the release of output Bs. The release of strand Bs makes it capable of initiating other strand displacement cascades in turn. Note that the binding of a toehold domain to its complement is transient unless a strand displacement reaction can be initiated because the toehold domains are short. Thus, for example, the 3 domain of input As does not block the 3^* domain of translator gate t.

An input or output strand has two regions: a recognition region which can participate in a strand displacement reaction, and a history region which cannot. The sequence of the history region (e.g., domain 7 on strand Bs) is determined by the translator gate from which the strand was released. All strands with the same recognition region react equivalently and we do not distinguish between them. For example, any strand with recognition region 1-2-3 is called As and any strand with recognition region 4-5-6 is called Bs. Since there are no sequence constraints (i.e., complementarity or equality) between the recognition region of

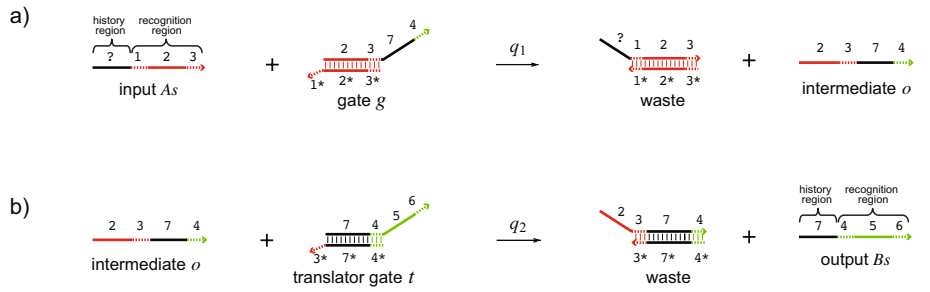

Fig. 2. Two-stage strand displacement cascade. Functional domains are numbered and all toehold domains are dashed. Different recognition regions are shown in different color. Input or output strands with identical recognition regions react equivalently and are therefore grouped into the same species. For example, As is any strand with recognition region 1-2-3, and Bs is any strand with recognition region 4-5-6, irrespective of their history regions. The two-stage cascade shown produces Bs with history region 7. Note that the sequences of the recognition regions of input and output strands As and Bs (1-2-3 and 4-5-6) may be completely unrelated to one another and therefore such a two-stage strand displacement cascade can link any input with any output species. **a)** Input strand As binds to gate g and by a strand displacement reaction releases the intermediate strand o. **b)** The intermediate strand o binds translator gate t and by a strand displacement reaction releases the output Bs.

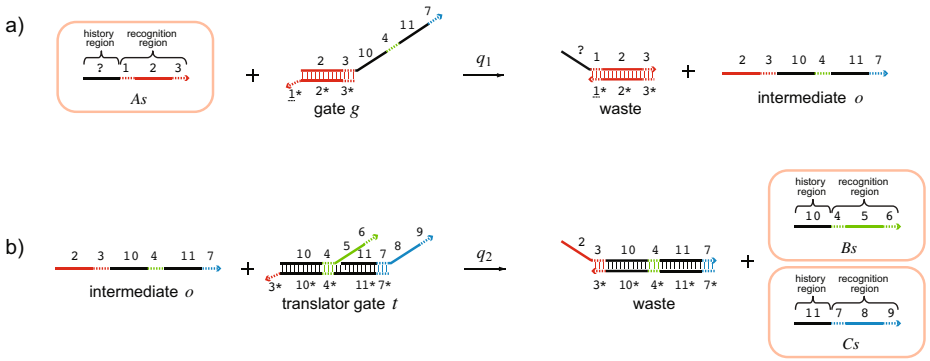

Fig. 3. Molecular implementation of the unimolecular reaction $A \rightarrow B + C$. Orange boxes highlight the DNA species As, Bs, and Cs that correspond to the formal species A, B, and C. Rate constant q_1 can be reduced by decreasing the complementarity between domains 1 and $\underline{1}^*$. The sequences of the recognition regions of input and output strands As, Bs, and Cs (regions 1-2-3, 4-5-6, and 7-8-9, respectively) may be completely unrelated to one another. The regime for desired unimolecular kinetics (concentrations of g, t and rate constants q_1, q_2) is described in the text. **a)** Input strand As binds to gate g and by a strand displacement reaction releases the intermediate strand o. **b)** The intermediate o binds translator gate t and by a strand displacement reaction releases the outputs Bs and Cs.

the input strand As and the output strand Bs, arbitrary chains of such two-step cascades can be linked together. This is possible for two-step cascades as shown (see "full translator" in Ref. [6]); however, a one-step cascade would force a part of the recognition region of the output strand to have sequence equality with the input strand, complicating the sequence design process. We call the second gate a translator gate to emphasize its role in translating the input to the appropriate output. A two-step strand displacement cascade may output multiple strands if we attach two outputs to translator gate t and extend the intermediate strand o using one more distinct history region (as is shown in Fig. 3). Again no sequence constraints exist between the input and the output strands.

In the design of systems of coupled two-step cascades, nucleic acid sequences need to be constructed to avoid unintended interactions. For instance, we can first design all recognition regions to have maximally independent sequences, and then for every translator gate, design maximally independent history regions of its output strands. Then a gate can react with only one recognition region (g-type gates) or intermediate strand (translator gates), ensuring the specificity of interactions. In addition, all sequences must have minimal secondary structure, such as hairpin loops, because such structure can inhibit the desired interactions.

3 Arbitrary Unimolecular Reactions

As a first step we will implement the basic monomolecular reaction $A \xrightarrow{k} B$, such that A and B are single-stranded nucleic acid species with completely independent recognition regions. As we will show, the appropriate monomolecular kinetics can be obtained as a limiting case of the reaction kinetics for a two-step strand displacement cascade:

$$A \xrightarrow{k} B \quad \Rightarrow \quad \begin{cases} As + g \xrightarrow{q_1} o \\ o + t \xrightarrow{q_2} Bs \end{cases}$$

We use the notation As and Bs to mean the implementation of formal species A and B by DNA strands with recognition regions unique for A and B, respectively. We do not include inert waste products when writing the chemical reaction equations. We now discuss the conditions required to make the implementation valid. First, we assume that all non-designed interactions are negligible. We will work in a regime where the concentrations $[g]$ and $[t]$ are in large excess of $[As]$ and $[o]$ so that they remain effectively constant at initial values $[g]_0$ and $[t]_0$ respectively. Then the two-step strand displacement cascade becomes equivalent to a pair of monomolecular reactions:

$$\begin{aligned} As &\xrightarrow{q_1[g]_0} o \\ o &\xrightarrow{q_2[t]_0} Bs \end{aligned}$$

By varying the toehold strength of gate g which determines rate constant q_1, or the initial concentration $[g]_0$, we set $q_1[g]_0$ equal to the formal rate constant k and

attain $d[As]/dt = -k[As]$ as desired. To also ensure that $d[Bs]/dt = k[As]$, we make $q_2[t]_0$ large enough that intermediate strand $[o]$ settles to its quasi-steady-state value $q_1[g]_0[As]/(q_2[t]_0)$ on a much faster time scale than that on which $[As]$ changes. Then $d[Bs]/dt = q_2[t]_0[o] \approx q_1[g]_0[As] = k[As]$ as desired. To make the quasi-steady-state approximation hold in this example, we can increase the relative toehold strength of gate t compared to gate g, or use a much larger initial concentration $[t]_0$ than $[g]_0$.

While experimentally, it may be useful to vary the degree of toehold complementarity affecting q_1 or concentration of gates $[g]_0$ to tune the effective rate constant, for simplicity throughout this paper we control reaction kinetics by tuning toehold strengths, while treating all gates as being present at the same high concentration ξ. Thus we set q_1 equal to k/ξ.

The same scheme can be extended to more complex unimolecular reactions. Reactions with more than one product species (e.g., $A \to B + C$ or $A \to 2B$) including catalytic (e.g., $A \to A + B$) and autocatalytic reactions (e.g., $A \to 2A$) can be constructed using a translator gate t that releases multiple strands as in Fig. 3. Removing the translator gate altogether allows for unimolecular decay reactions (e.g., $A \to$). Fractional product stoichiometry (e.g., $A \to (1/3)B + C$) can be realized using a mixture of translator gates with some fraction having incomplete output strands. For example, reaction $A \to (1/3)B + C$ can be implemented if 2/3 of translator gates t in Fig. 3 are missing the 7-8 domains. Fractional product stoichiometries are equivalent to multiple reactions in which the same reactants produce different products, where the products are in integer stoichiometries. E.g. the two reactions $A \xrightarrow{2k/3} C$ and $A \xrightarrow{k/3} B + C$ are kinetically equivalent to a single reaction $A \xrightarrow{k} (1/3)B + C$. Conversely, all reactions with the same reactants but different products can always be combined into one reaction with possibly fractional product stoichiometries.

Arbitrary sets of unimolecular reactions can be coupled together by reusing the same recognition region in multiple reactions. Each reaction corresponds to a distinct two-step strand displacement cascade. For example, the system

$$\begin{aligned} A &\xrightarrow{k_1} B + C \\ B &\xrightarrow{k_2} 2B \end{aligned}$$

can be implemented with gate-mediated reactions

$$A \xrightarrow{k_1} B + C \quad \Rightarrow \quad \begin{cases} As + g_1 \xrightarrow{k_1/\xi} o_1 \\ o_1 + t_1 \longrightarrow Bs + Cs \end{cases}$$

$$B \xrightarrow{k_2} 2B \quad \Rightarrow \quad \begin{cases} Bs + g_2 \xrightarrow{k_2/\xi} o_2 \\ o_2 + t_2 \longrightarrow 2Bs \end{cases}$$

where unlabeled rate constants are much larger than k_1/ξ and k_2/ξ and initial concentrations $[t_i]_0$, $[g_i]_0 = \xi$ are high enough to remain effectively constant. The expressions for the DNA gate-mediated reactions in terms of formal rate

constants are obtained from the above analysis. As described in the previous section, the different two-step strand displacement cascades do not have significant undesired interactions. Thus each reaction should proceed without interference from the others except through the desired coupling of input and output strands.

4 Arbitrary Bimolecular Reactions

Consider the basic bimolecular reaction $A + B \xrightarrow{k} C$. Since a reaction between an input strand and a gate can be viewed as being bimolecular, it provides a possible implementation of this reaction. As before, A is mapped to strand As, but now B would have to be mapped to a gate. To emphasize that a gate is mapped to a formal species B we call the gate Bg. As in the case of unimolecular reactions, we can use the translator gate t to ensure sequence independence between recognition regions of As and Cs. The corresponding gate-mediated reactions therefore are:

$$A + B \xrightarrow{k} C \quad \Rightarrow \quad \begin{cases} As + Bg \xrightarrow{k} o \\ o + t \longrightarrow Cs \end{cases}$$

We set the unlabeled rate constant to be very large and the initial concentration of the translator gate $[t]_0 = \xi$ to be big enough to remain effectively constant. Then using the quasi-steady-state approximation for the intermediate strand o as in Section 3 we obtain the desired effective bimolecular reaction rate $k[As][Bg]$.

Having said that, this naive implementation has severe shortcomings. Since strand As must directly bind gate Bg, their sequences are not independent. Thus, gate Bg can react only with input As and cannot participate in reactions with other strand species. Further, it is not always possible to uniquely assign reactants to a gate or a strand. One such example is the following system:

$$\begin{aligned} A &\xrightarrow{k_1} B \\ A + B &\xrightarrow{k_2} C \end{aligned}$$

If we combine the implementation of monomolecular reactions developed in the previous section with the proposed bimolecular scheme, in the resulting system species B is mapped to two different forms, a strand Bs and a gate Bg_2:

$$A \xrightarrow{k_1} B \quad \Rightarrow \quad \begin{cases} As + g_1 \xrightarrow{k_1/\xi} o_1 & \text{(i)} \\ o_1 + t_1 \longrightarrow Bs & \text{(ii)} \end{cases}$$

$$A + B \xrightarrow{k_2} C \quad \Rightarrow \quad \begin{cases} As + Bg_2 \xrightarrow{k_2} o_2 & \text{(iii)} \\ o_2 + t_2 \longrightarrow Cs & \text{(iv)} \end{cases}$$

The concentrations of strand form Bs and gate form Bg_2 are entirely independent, and therefore the DNA reactions do not implement the desired formal chemical system.

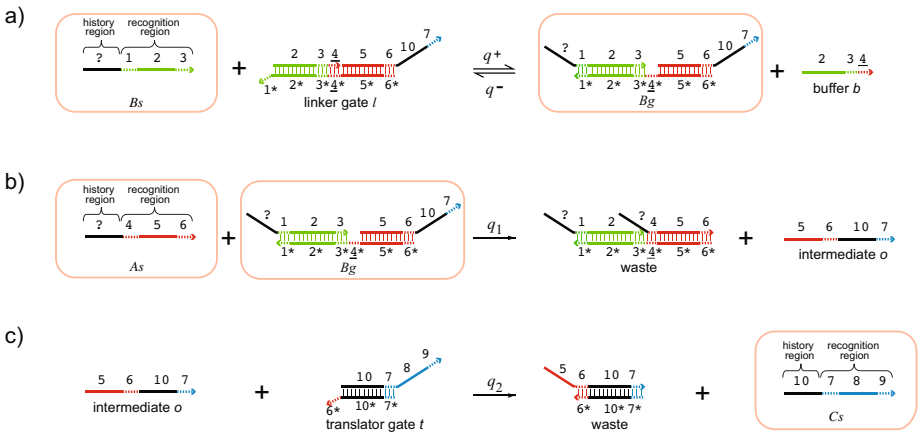

Fig. 4. Molecular implementation of the bimolecular reaction $A+B \rightarrow C$. Orange boxes highlight the DNA species As, Bs, and Cs that correspond to the formal species A, B, and C. Rate constant q_1 can be reduced by decreasing the complementarity between domains 4 and $\underline{4}^*$. The sequences of the recognition regions of input and output strands As, Bs, and Cs (regions 1-2-3, 4-5-6, and 7-8-9, respectively) are completely unrelated to one another. The regime for desired bimolecular kinetics (concentrations of l, b, t and rate constants q^+, q^-, q_1, q_2) is described in the text. **a)** Input strand Bs reversibly binds to the linker gate l forming the activated gate Bg, i.e., $B + l \rightleftharpoons Bg + b$. **b)** Input strand As binds to the activated gate complex Bg and irreversibly releases intermediate strand o through strand displacement. **c)** The intermediate strand o binds translator gate t and by a strand displacement reaction releases the output Cs.

However, if the two forms of B could be interchanged into one another on a time scale that is fast compared to the other reactions in the system, the correct behavior can be restored. We can link the two species Bs and Bg_2 through a fast reversible reaction

$$Bs \underset{k^-}{\overset{k^+}{\rightleftharpoons}} Bg_2$$

such that the two species achieve pseudoequilibrium. Then the formal species B exists in two different forms: $B = \{Bs, Bg_2\}$ and the total concentration of B is $[B] = [Bs] + [Bg_2]$. Let $f(Bg_2) = [Bg_2]/[B]$ be the fraction of B in gate form Bg_2. Under the pseudoequilibrium assumption, $f(Bg_2) = (k^+ + k^-)/k^+$ is a constant. Since the second formal reaction can only use the gate form Bg_2 as a reactant, and not all of B, we scale the rate constant of reaction (iii) by $1/f(Bg_2)$ so that the new rate constant is $k_2/f(Bg_2)$. Then the effective rate of the implementation of $A + B \xrightarrow{k_2} C$ is $(k_2/f(Bg_2))[As][Bg_2] = k_2[A][B]$ as desired. We can easily extend this idea to create a pseudoequilibrium between strand Bs and gates Bg_i for multiple reactions i.

We realize the above reaction establishing pseudoequlibrium between Bs and Bg_2 using a linker gate shown in Fig. 4(a). Strand Bs and buffer strand b

reversibly compete with each other via strand displacement reactions in a toehold exchange process [8]. Thus the reaction establishing pseudoequilibrium is implemented with gates as follows:

$$Bs \underset{k^-}{\overset{k^+}{\rightleftharpoons}} Bg_2 \quad \Rightarrow \quad Bs + l \underset{q^-}{\overset{q^+}{\rightleftharpoons}} Bg_2 + b$$

For the correct first-order kinetics $Bs \underset{k^-}{\overset{k^+}{\rightleftharpoons}} Bg$, the linker gate l and the buffer strand b must be in excess, such that their concentrations remain effectively constant. Then $k^+ = q^+[b]_0$ and $k^- = q^-[l]_0$ where $[b]_0$ and $[l]_0$ are the initial concentrations of the buffer and linker strands respectively. For simplicity we will use $[b_0] = [l]_0 = \xi$ and $q^+ = q^-$.

Lastly, we need to confirm the absence of unintended cross-reactions when implementing multiple coupled bimolecular reactions. As in the simple strand displacement cascades described in Section 2, gates can only react with specific recognition regions or intermediate strands. The exception to this rule is the reaction of gate Bg with the buffer strand b. Gate form Bg can react with any strand with accessible domains ...3-4. Because without loss of generality we can assume that there is only one formal reaction $A + B \rightarrow$ (see discussion of fractional product coefficients in Section 3), and domains 3 and 4 are unique to Bs and As respectively, nothing other than the correct buffer strand can react here.

5 Systematic Construction

In this section we take the ideas developed above and describe a systematic algorithm for compiling arbitrary unimolecular and bimolecular reactions into DNA gate-mediated chemistry. This algorithm is used in the next section to implement a Rössler attractor chaotic chemical system.

Without loss of generality we assume that every reaction has a unique combination of reactants. For example, the pair of reactions $A+B \overset{k_1}{\rightarrow} C$ and $A+B \overset{k_2}{\rightarrow} D$ are combined into a single reaction $A+B \overset{k_1+k_2}{\longrightarrow} (k_1/(k_1+k_2))C+(k_1/(k_1+k_2))D$ (see the discussion of fractional product coefficients in Section 3). Let i index reactions and $X_j \in \{A, B, C, \dots\}$ index species. Let $f(X_j s)$ be the fraction of X_j in strand form $X_j s$. Similarly let $f(X_j g_i)$ be the fraction of X_j in gate form $X_j g_i$.

Consider any unimolecular formal reaction i. Write the reaction as $X_1 \overset{k}{\rightarrow} \alpha_2 \cdot X_2 + \cdots + \alpha_n \cdot X_n$, where $0 < \alpha \leq 1$. We implement this reaction by a two-step strand displacement cascade (Fig. 3), modeled by the DNA gate reactions below (where we omit inert waste products, and combine all strands with the same recognition regions into a single species).

$$\begin{aligned} X_1 s + g_i &\xrightarrow{k'} o_i \\ o_i + t_i &\longrightarrow \alpha_2 \cdot X_2 s + \cdots + \alpha_n \cdot X_n s. \end{aligned}$$

Product fractions α_j are set by preparing translator gate t_i with α_j fraction of complete and $1-\alpha_j$ incomplete output strands for $X_j s$ as discussed in Section 3. Unlabeled rate constants as well as the initial concentrations $[g_i]_0 = [t_i]_0 = \xi$ are as high as possible. Rate constant k' is set to $\frac{k}{\xi f(X_1 s)}$ by varying the degree of complementarity of the toehold on gate g_i with the toehold on strand $X_1 s$. Note that by following the argument of Section 3, and using the fact that $[X_1] = [X_1 s]/f(X_1 s)$, the effective rate of this reaction is $k'[X_1 s]\xi = k[X_1]$ as desired.

Consider any bimolecular formal reaction i. Write the reaction as $X_1 + X_2 \xrightarrow{k} \alpha_3 \cdot X_3 + \cdots + \alpha_n \cdot X_n$, where $0 < \alpha \leq 1$. We implement this reaction by a linker gate mechanism combined with the two-step strand displacement cascade (Fig. 4) and is modeled by the DNA gate reactions below (where we again omit inert waste products, and combine all strands with the same recognition regions into a single species).

$$\begin{aligned} X_1 s + l_i &\rightleftharpoons X_1 g_i + b_i \\ X_2 s + X_1 g_i &\xrightarrow{k'} o_i \\ o_i + t_i &\longrightarrow \alpha_3 \cdot X_3 s + \cdots + \alpha_n \cdot X_n s \end{aligned}$$

Product fractions α_j are set by preparing translator gate t_i with α_j fraction of complete and $1-\alpha_j$ incomplete output strands for $X_j s$ as before. Unlabeled rate constants are as high as possible, with the forward and reverse rates of the first reaction being equal. Rate constant k' is set to $\frac{k}{f(X_2 s) f(X_1 g_i)}$ by varying the degree of complementarity of the toehold on $X_1 g_i$ with the toehold on strand $X_2 s$. The initial concentrations $[l_i]_0 = [b_i]_0 = [t_i]_0 = \xi$ are as high as possible. Following the argument of Section 4, and using the facts that $[X_2] = [X_2 s]/f(X_2 s)$ and $[X_1] = [X_1 g_i]/f(X_1 g_i)$, we see that the effective rate of this reaction is $k'[X_2 s][X_1 g_i] = k[X_1][X_2]$ as desired.

With the above construction, determining $f(X_j s)$ and $f(X_j g_i)$ is easy: for every i, j, $f(X_j s) = f(X_j g_i) = 1/(m+1)$ where m is the number of bimolecular reactions in which X_j appears as the first reactant.

The sequences of the DNA components can be designed as follows. First, for all formal species design maximally independent recognition regions with minimum secondary structure. Then, for each formal reaction, design the history regions for all products of that reaction to be maximally independent and have minimum secondary structure. At this point all auxiliary DNA species are fully specified. Significant unintended interactions between auxiliary species participating in different formal reactions cannot occur by the arguments in Sections 2 and 4. The system is initiated by adding appropriate starting amounts of the formal species in single-stranded form with arbitrary history regions.

6 Example

We illustrate our method of using DNA-based chemistry to implement arbitrary formal systems of coupled chemical equations on the chaotic system due

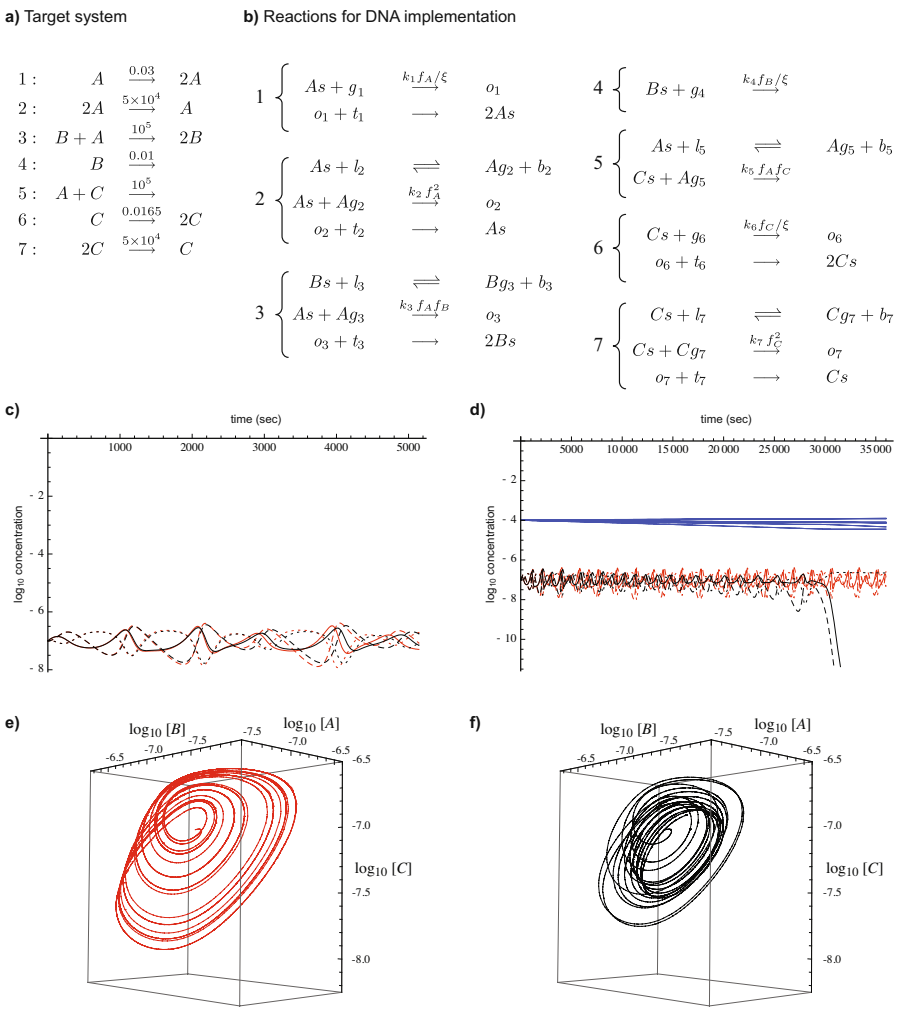

Fig. 5. Rössler attractor example. **(a)** The formal chemical reaction system to be implemented. **(b)** Reactions modeling our DNA implementation. Each bracket implements the formal reaction with the number indicated. Here k_1 through k_7 are the original rate constants for reactions 1 through 7 as in (a). Multiplicative factors $f_A = 1/f(As) = 1/f(Ag_2) = 1/f(Ag_5) = 3$, $f_B = 1/f(Bs) = 1/f(Bg_3) = 2$, $f_C = 1/f(Cs) = 1/f(Cg_7) = 2$. We use initial concentration of the gates and buffer strands $\xi = 10^{-4}$. Unlabeled rate constants are 10^5. **(c)** Plot of the log-concentrations of A (solid), B (dashed), C (dotted) for the original system (red), as well as their modeled concentrations (black). **(d)** Longer time plot showing also the log-concentrations of g_i (blue, decreasing) and b_i (blue, increasing). **(e,f)** Trajectories of the original system and DNA implementation in the 3-dimensional phase-space (first 5 hours).

to Willamowsky and Rössler [4]. We start with the following formal reactions, where the rate constants are from Ref. [11]:

$$\begin{aligned}
1: &\quad A \xrightarrow{30} 2A \\
2: &\quad 2A \xrightarrow{0.5} A \\
3: &\quad B + A \xrightarrow{1} 2B \\
4: &\quad B \xrightarrow{10} \\
5: &\quad A + C \xrightarrow{1} \\
6: &\quad C \xrightarrow{16.5} 2C \\
7: &\quad 2C \xrightarrow{0.5} C
\end{aligned}$$

The strange attractor for the concentrations of A, B, and C is in the range of about 0–40.

First we scale this system into a regime realistic for DNA-based chemistry which constrains reaction rates and concentrations. Second order rate constants for strand displacement reactions can be approximately in the range 0–$10^6/M/s$, with their value determined by the degree of toehold complementarity [10]. Typical experimental concentrations are on the order of 0–$10^{-3}M$. Similar to experimental implementations of other dynamical chemical systems, a flow reactor may be used to replenish the stock of unreacted gates and remove waste to maintain the appropriate reaction conditions [3]. This may make it possible to use lower gate concentrations.

Clearly, by scaling all rate constants by the same factor we simply speed up or slow down the system without affecting the dynamical behavior. We can scale the concentrations at which the chaotic system operates by scaling the bimolecular rate constants differently from the unimolecular ones. In general if $[X_j](t)$ are solutions to differential equations arising from a set of unimolecular and bimolecular reactions, then $\alpha[X_j](t)$ are solutions to the differential equations arising from the same set of reactions but in which we divide all bimolecular rate constants by α. We first slow down the system by multiplying all rate constants by 10^{-3}, and then use concentration scaling factor $\alpha = 10^{-8}$, obtaining the rate constants in Fig. 5(a).

Applying our construction yields a DNA implementation governed by the equations in Fig. 5(b). Simulations confirm (Fig. 5(c, d)) that the DNA implementation behaves very close to the formal system (a) until the depletion of linker gates l_i and the buildup of buffer strands b_i sufficiently alters the effective rate constants, gradually decoupling the gate implementation from the target system.

7 Conclusion

We have proposed a method for approximating an arbitrary system of coupled unimolecular and bimolecular chemical reactions using DNA-based chemistry. Our construction takes advantage of cascades of strand displacement reactions [6], and elementary techniques of approximation in chemical kinetics. Each formal species

occurring in the system of chemical reactions is represented as a set of strands and gates. The multiform representation is necessary because it is not always possible to find a single DNA species that is capable of participating in all reactions involving a given formal species. However, the different forms are constructed to be in equilibrium with each other and thus participate in kinetics as if they were a single species, up to a scaling of rate constants.

While we have taken care to provide a systematic algorithm for compiling a set of chemical reactions into DNA, in practice it may often be possible and preferable to reduce the complexity by optimizing the construction for the particular system of interest. For example, in many cases complete sequence independence between strands may not be necessary, possibly allowing one to eliminate some translator gates. Similarly, pseudoequilibrium linkage is unnecessary if mapping a species directly to a strand or gate does not cause problems.

For simplicity in our systematic construction rate constants are set by the degree of sequence complementarity between toehold domains. However, there are many other degrees of freedom available such as the relative concentrations of linker gate and buffer strand for bimolecular reactions. Probably in practice, toehold domains provide a rough order of magnitude control over formal rate constants, while adjusting concentrations of auxiliary species allows fine-tuning them.

Acknowledgments. This work was supported by NSF Grant 0728703. GS acknowledges support from the Swiss National Science Foundation and the Burroughs Wellcome Fund. We thank D. Zhang, J. Schaeffer, and M. Magnasco for useful discussions.

References

1. Gavalas, G.R.: Nonlinear Differential Equations of Chemically Reacting Systems. Springer, Heidelberg (1968)
2. Scott, S.K.: Chemical Chaos. Oxford University Press, Oxford (1991)
3. Epstein, I.R., Pojman, J.A.: An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos. Oxford University Press, Oxford (1998)
4. Willamowski, K.D., Rössler, O.E.: Irregular oscillations in a realistic abstract quadratic mass action system. Z. Naturforsch A 35, 317–318 (1980)
5. Yurke, B., Turberfield, A.J., Mills Jr., A.P., Simmel, F.C., Neumann, J.L.: A DNA-fuelled molecular machine made of DNA. Nature 406(6796), 605–608 (2000)
6. Seelig, G., Soloveichik, D., Zhang, D.Y., Winfree, E.: Enzyme-free nucleic acid logic circuits. Science 314(5805), 1585–1588 (2006)
7. Seelig, G., Yurke, B., Winfree, E.: Catalyzed relaxation of a metastable DNA fuel. Phys. Rev. Lett. 90, 118102–118111 (2006)
8. Zhang, D.Y., Turberfield, A.J., Yurke, B., Winfree, E.: Engineering entropy-driven reactions and networks catalyzed by DNA. Science 318(5853), 1121 (2007)
9. Green, C., Tibbetts, C.: Reassociation rate limited displacement of DNA strands by branch migration. Nucleic Acids Research 9, 1905–1918 (1981)
10. Yurke, B., Mills, A.P.: Using DNA to Power Nanostructures. Genetic Programming and Evolvable Machines 4(2), 111–122 (2003)
11. Gaspard, P.: Encyclopedia of Nonlinear Science. In: "Rössler Systems", pp. 808–811. Routledge (2005)