

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

David Soloveichik, Georg Seelig, and Erik Winfree

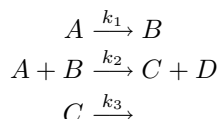
California Institute of Technology, Pasadena, CA, USA
dsolov@caltech.edu, seelig@dna.caltech.edu, winfree@caltech.edu

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 [1–3]. 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 is typically much harder, and has not been solved in general. 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 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 (eg 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.

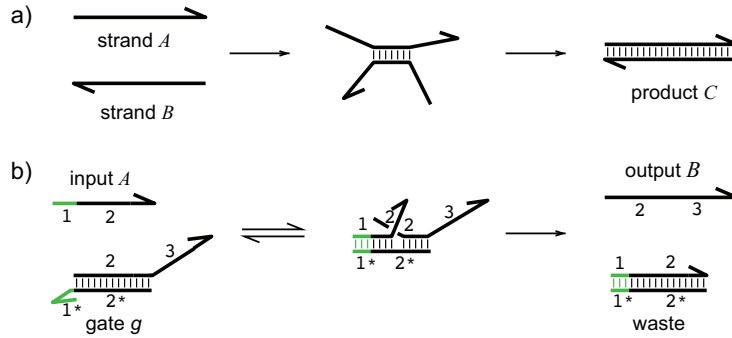


Fig. 1: Hybridization and strand displacement reactions. **a)** Hybridization reaction. Two complementary strands A and B react with each other to form a double helix C . The hybridization reaction proceeds through a set of partially hybridized intermediates. Nevertheless, the overall reaction kinetics is well approximated as a bimolecular reaction $A + B \xrightarrow{k} C$. The 3' end of each strand is indicated by an arrow. **b)** Strand displacement. Functional sub-domains are numbered and the star indicates complementarity. The reaction between input strand A and gate g is initiated at the toe-hold (green, sub-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 toe-hold as well as the degree of complementarity between the toe-holds on the strand and gate.

In the next section we describe strand displacement reactions which 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 on 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

Single-stranded nucleic acids with complementary sequences hybridize to form an inert double-helical molecule. Although hybridization reactions involve multiple elementary steps, for short oligonucleotides the kinetics is approximately a second order process [5, 6]. However, hybridization between two complementary strands is insufficient to implement systems of coupled bimolecular reactions. For instance, the double-stranded product is inert, and thus incapable of acting as a reactant in another reaction.

Strand displacement reactions provide for a more promising primitive. The basic principle is illustrated in Fig. 1(b). Although a strand displacement reaction involves multiple elementary steps, likely including a random walk process, it is well modeled as a second order process for a wide range of reaction conditions [7, 8]. 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.

We have recently used strand displacement cascades to construct DNA-based logic circuits [9]; here we use some of the nomenclature and ideas from that work. Fig. 2 shows how a single-stranded nucleic acid species (a strand) can initiate a strand displacement cascade between two complexes (gates) leading to the release of one or more strands. A strand is active for the initiation of a strand displacement cascade only if its toehold region is single-stranded. The release of strands B_s and C_s makes them active, and therefore capable of initiating other strand displacement cascades in turn. (We use notation B_s and C_s rather than B and C to be consistent with discussion to follow.) There are no sequence constraints (ie complementarity or equality) between the initiating strand A_s and the output strands B_s and C_s . This is possible in a two step cascade as shown; a one step cascade forces a region of the output strands to have sequence equality with the input strand (see “full translator” in Ref. [9]).

An input or output strand has two domains: a recognition region which can participate in downstream strand displacement reactions, and a history domain which cannot. The sequence of the history domain is determined by the gate from which the strand was released. All strands with the same recognition region

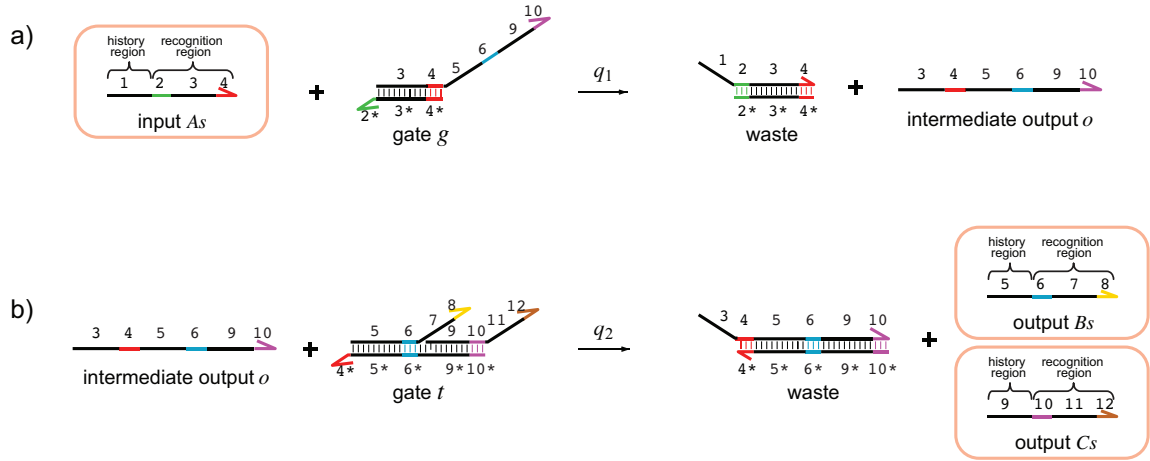


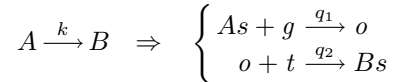
Fig. 2: Two-stage strand displacement cascade. Input As , and outputs Bs and Cs are single-stranded and can have arbitrary sequences. All toe-holds are indicated in color and regions of the same color bind to each other. **a)** Input strand As binds to gate g and by a strand displacement reaction releases the intermediate output strand o . **b)** The intermediate output o binds translator gate t and in turn releases the reaction products Bs and Cs . The sequences of output strands Bs and Cs in this two-stage cascade are completely unrelated to that of input strand As while the intermediate output o has some sequence overlap with As . This cascade implements the reaction $A \rightarrow B + C$ in the limit of large concentration of gate species as discussed in the text. Orange boxes highlight the DNA species that correspond to the formal species A , B and C of the bimolecular reaction $A \rightarrow B + C$. Other DNA species such as gates g and t have auxiliary roles and mediate effective interactions between the input and outputs. The recognition region of an input or output strand is the active part of the strand and is used to bind a down-stream target. All strands with the same recognition region behave identically and are grouped into the same species (for example, any strand with recognition domain 2-3-4 irrespective of its history domain is called As). Different history domains result if strands are released from different gates or positions within a gate.

react equivalently and therefore we do not distinguish between strands with different history domains if they have the same recognition domain.

In the design of such cascades, the sequences of all the strands are optimized through a genetic algorithm to have minimal unintended interactions. For example, we can first design the recognition domains of all input and output strands to minimize secondary structure and intermolecular interactions. Much of the sequences of the gates is then fixed. Finally, design the unassigned parts of the intermediate strands o_i to have minimal unintended interactions between themselves and the previously designed input and output strands.

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 sequences. 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:



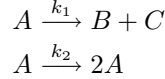
We use the notation As and Bs to mean the implementation of formal species A and B by DNA strands with recognition regions unique for A and B , respectively. We will now discuss the relationship between the formal rate constant k and actual molecular rate constants q_1, q_2 in this example. We ensure that $q_2[t]$ is much larger than $q_1[g]$, which can be done by changing the relative toehold strengths in gates g and t or by having a much larger initial concentration of gate t than of gate g . In this regime, we can make a steady state assumption for the reaction intermediate o : $d[o]/dt = 0$. Then the effective overall reaction rate is $q_1[g][As]$. If we further assume the concentration $[g]$ is in large excess of As so that it remains constant, the reaction becomes effectively first order in As with rate constant $k = q_1[g]_0$, where $[g]_0$ is the initial concentration of g , and we obtain:

$$-d[A]/dt = d[B]/dt = k[A], \quad k = q_1[g]_0.$$

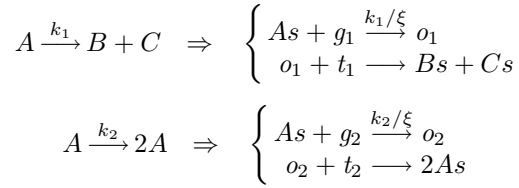
Note that the desired formal rate constant k can be experimentally varied by changing either the toehold of gate g which determines the rate constant q_1 , or the initial concentration $[g]_0$.

More complex monomolecular reactions with more than one product species (eg $A \rightarrow B + C$ or $A \rightarrow 2B$ or even $A \rightarrow 2A$) can be constructed in the same manner but using a translator gate t that releases multiple strands as in Fig. 2. Strands As, Bs, Cs corresponding to formal species A, B, C are identified by their distinct recognition regions. Removing the translator gate altogether allows for unimolecular decay reactions (eg $A \rightarrow$).

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



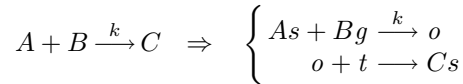
can be implemented with gate mediated reactions



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 steady state analysis as above. Our transformation from the formal chemical system to DNA reactions is not unique. For example, our choice that all gate species are initially present at the same concentration ξ is for convenience only.

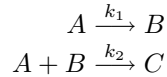
4 Arbitrary Bimolecular Reactions

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 As and Cs . The corresponding gate mediated reactions therefore are:

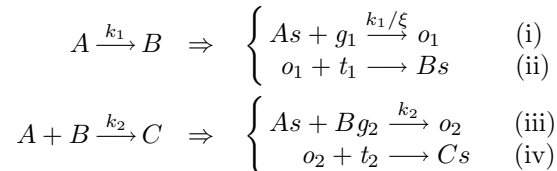


We set the unlabeled rate constant to be very large and the initial concentration of the translator gate $[t]_0 = \xi$ to be big enough to remain effectively constant. Then using the steady state approximation for the intermediate output o we obtain the effective reaction rate to be $k[As][Bg]$ as desired.

Having said that, this implementation has severe shortcomings. Since strand As must directly bind gate Bg , their sequences are not independent. 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:



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 :



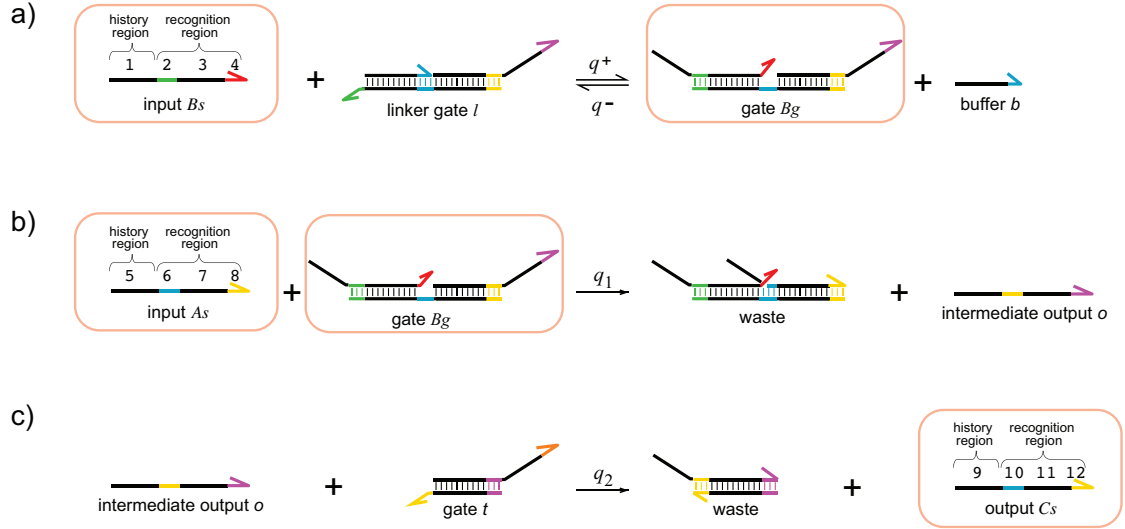
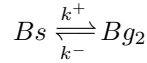


Fig. 3: Molecular implementation of the bimolecular reaction $A + B \rightarrow C$. As and Bs are single strands of DNA with unrelated sequences. **a)** Input strand Bs reversibly binds to the linker gate l forming the activated gate Bg , i.e. $B + l \rightleftharpoons Bg + b$. The equilibrium can be adjusted by varying the relative amounts of l and b . **b)** Input strand As binds to the activated gate complex Bg and irreversibly releases intermediate output strand o . **c)** The intermediate output o activates a translator gate in the same way as in Fig. 1(a), thereby eliminating any sequence constraints between input and product strands. As before, more than one output strand can be released as shown in Fig. 2(b). Orange boxes highlight the DNA species that correspond to the formal species A , B and C of the bimolecular reaction $A \rightarrow B + C$. Other DNA species such as gates g and t have auxiliary roles and mediate effective interactions between the input and outputs. Functional sub-domains of the inputs and outputs are numbered to highlight that there are no sequence constraints between these species.

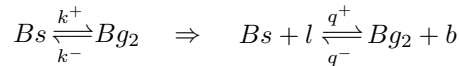
The concentrations of strand form Bs and gate form Bg_2 are entirely independent, and therefore the DNA reactions do not implement the desired formal chemical system.

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 link the two species Bs and Bg_2 through a fast reversible reaction



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 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 .

The above reaction establishing pseudoequilibrium between Bs and Bg_2 can be realized via a linker gate as shown in Fig. 3 (top). This corresponds to the following DNA reactions:



For the correct first order kinetics $Bs \xrightleftharpoons[k^-]{k^+} Bg$, 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.

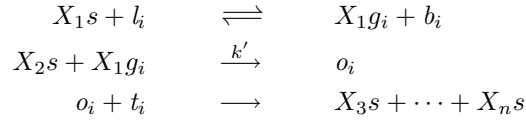
5 Systematic Construction

In this section we take the ideas we have developed in the previous sections 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.

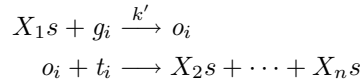
Let i index reactions and $X_j \in \{A, B, C, \dots\}$ index species. Let $f(X_j s)$ be the fraction of X_j in strand form $X_j s$. Similarly let $f(X_j g_i)$ be the fraction of X_j in gate form $X_j g_i$.

A bimolecular reaction $i : X_1 + X_2 \xrightarrow{k} X_3 + \dots + X_n$ is implemented by a gate system of Fig. 3 and is modeled by the DNA gate reactions below (where we omit inert waste products, and combine all strands with the same recognition domains into a single species).



Unlabeled rate constants are as high as possible, with the forward and reverse rates of the first reaction being equal. Rate constant $k' = \frac{k}{f(X_2 s)f(X_1 g_i)}$ is set 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.

The unimolecular reaction $i : X_1 \xrightarrow{k} X_2 + \dots + X_n$ is implemented by a gate system of Fig. 2, modeled by the DNA gate reactions below (where we again omit inert waste products, and combine all strands with the same recognition domains into a single species).

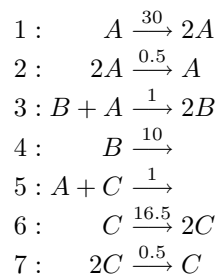


Again unlabeled rate constants as well as the initial concentrations $[g_i]_0 = [t_i]_0 = \xi$ are as high as possible. Rate constant $k' = \frac{k}{\xi f(X_1 s)}$ is set by varying the degree of complementarity of the toehold on gate g_i with the toehold on strand $X_1 s$.

Note that with the above construction 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.

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 to Willamowsky and Rössler [4]. We start with the following formal reactions, where the rate constants are from Ref. [10]:

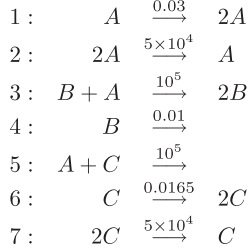


The strange attractor for the concentrations of A , B , and C is in the range of about 0 – 40.

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 [8]. 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,

a) Original system



b) Reactions for DNA implementation

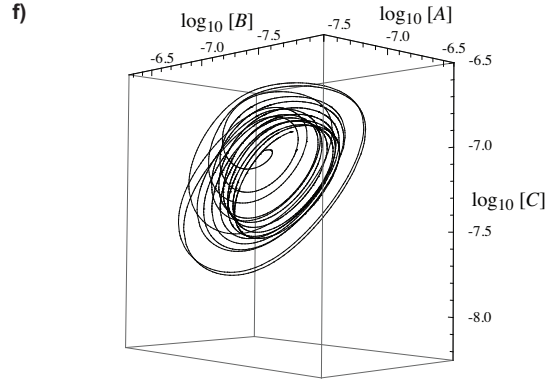
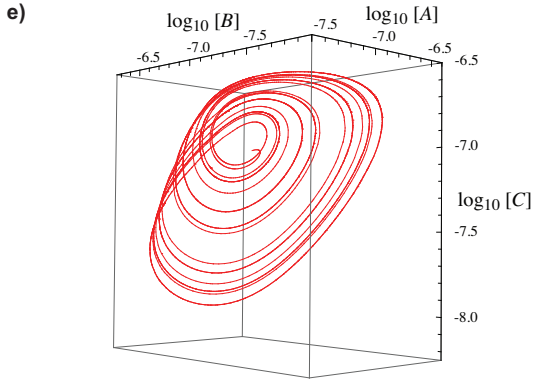
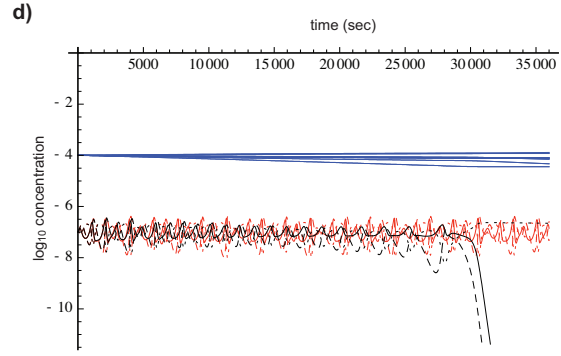
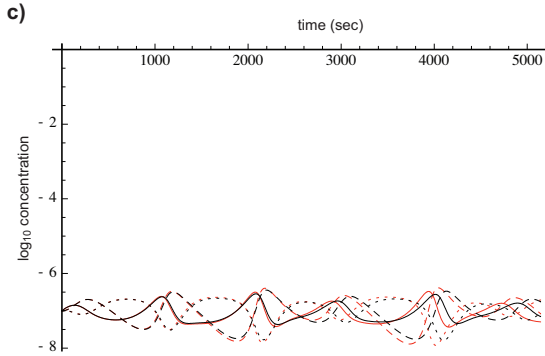
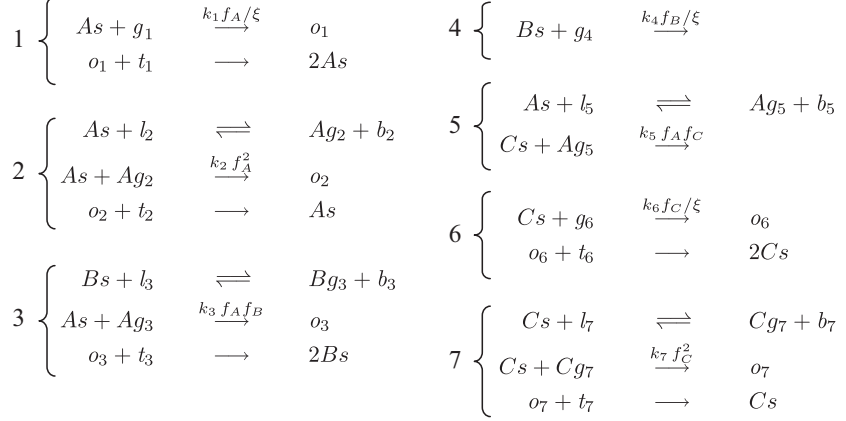


Fig. 4: Rössler attractor example. (a) The formal chemical reaction system to be implemented. (b) Reactions modeling our DNA implementation. Each bracket implements the formal reaction with the number indicated. Here k_1 through k_7 are the original rate constants for reactions 1 through 7 as in (a). 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).

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. 4(a).

Applying our construction yields a DNA implementation governed by equations Fig. 4(b). Simulations confirm (Fig. 4(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 to destroy the chaotic behavior at around 8 hours (see (d)).

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 [9], 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.

Acknowledgments. This work was supported by NSF Grant 0728703. We thank D. Zhang, J. Schaeffer and M. Magnasco for useful discussions.

References

1. G.R. Gavalas. *Nonlinear Differential Equations of Chemically Reacting Systems*. Springer-Verlag, 1968.
2. S.K. Scott. *Chemical Chaos*. Oxford University Press, 1991.
3. I.R. Epstein and J.A. Pojman. *An Introduction to Nonlinear Chemical Dynamics: Oscillations, Waves, Patterns, and Chaos*. Oxford University Press, 1998.
4. K. D. Willamowski and O. E. Rössler. Irregular oscillations in a realistic abstract quadratic mass action system. *Z. Naturforsch A*, 35:317–318, 1980.
5. C.R. Cantor and P.R. Schimmel. *Biophysical Chemistry, Part III: The behavior of biological macromolecules*. W. H. Freeman and Company, 1998.
6. L. E. Morrison and L. Stols. Sensitive fluorescence-based thermodynamic and kinetic measurement of dna hybridization in solution. *Biochemistry*, 32:3095–3104, 1993.
7. C. Green and C. Tibbetts. Reassociation rate limited displacement of dna strands by branch migration. *Nucleic Acids Research*, 9:1905–1918, 1981.
8. B. Yurke and A.P. Mills. Using DNA to Power Nanostructures. *Genetic Programming and Evolvable Machines*, 4(2):111–122, 2003.
9. G. Seelig, D. Soloveichik, D.Y. Zhang, and E. Winfree. Enzyme-Free Nucleic Acid Logic Circuits. *Science*, 314(5805):1585–1588, 2006.
10. P. Gaspard. *Encyclopedia of Nonlinear Science*, chapter “Rössler Systems”, pages 808–811. Routledge, 2005.