Simulation and Automatic Design of
Nucleic Acid Reaction System

ABSTRACT

I propose an efficient simulation technique for analyzing biological reactions and an automatic design method for developing artificial reactions, both of which are targeted to systems mainly composed of nucleic acids.

DNA that is one type of nucleic acids is utilized as a genetic information carrier in biology, while RNA that is the other type of nucleic acids has a biological functionality because of its reactivity. For better understanding of biological reaction, it is inevitable to analyze such nucleic acid reaction systems in detail by using computer simulations. Analyzing nucleic acid reaction systems in detail is difficult by using conventional simulation models that deal with small number of variables from a phenomenological point of view. To simulate the systems in detail by enumerating all possible structures of nucleic acids, however, is also difficult because of the combinatorial explosion of the number of structures. One technique to prevent such combinatorial explosion is to use the model reduction by assuming the independence of reactions. Because concrete nucleic acid reaction systems do not satisfy the assumption in general, the model reduction cannot directly be applied to the systems.

As another background of this research, numbers of artificial reactions making use of the chemical property of nucleic acid have been developed, which is expected to have applications such as drug delivery. Because every interaction of molecules in a system has to be taken into account, designing such device driven by chemical reaction of nucleic acids and enzymatic reaction is not easy for human. One approach to overcome the difficulty is to combine a simple building block or assemble basic chemical reaction like computer programming. Conventionally proposed design methods of artificial nucleic acid reaction systems, however, do not take advantage of complex structures because they are based on models that restrict the structure of nucleic acids. Because human trial and error of programming is required, it is still not easy to design a system where human have to think of huge combination of nucleic acids.

The goal of this research is proposing a simulation model with small restriction of the structure of nucleic acids, efficient simulation technique to prevent combinatorial explosion, and automatic design method of artificial reaction systems. In particular, I define nucleic acid reaction system as a graph rewriting system that has small restriction and can numerically simulate the systems based on chemical kinetics. To prevent the combinatorial explosion of the number of structures, I propose a model reduction that focuses on a local structure, which is applied to simulate concrete nucleic acid reaction systems (hybridization chain reaction and RNA interference). By the abstraction focusing on a local structure, it is possible to ignore the combinatorial explosion caused by representing whole structure because the structure is represented by a set of local structures. To minimize the trial and error by human design, I propose an automatic design method using such simulation techniques and develop reaction systems such as logic gates and automaton, which can be implemented by actual nucleic acids. For the automatic design method, I define fitness functions of target systems and implement heuristic algorithm for the combinatorial optimization that can maximize the fitness value. Evaluation of this research is based on the discussion of the accuracy of the model reduction and the verification of the designed systems by actual chemical experiments. By this research, detail analysis of huge nucleic acid reaction systems and the development of artificial reaction systems that are nontrivial for human are expected.

,

.

DNA   ,   RNA

.   ,

.

,

,   .

,   .

,

.   ,

.

,

,   .

,

.

,

.

,   ,

.   ,

.

,   ,

,   .

,

.   ,

,   (

RNA   )   .   ,

,

.

,   ,

.

,

,   .

.

,   .

Acknowledgements

I would like to express my great appreciation to my supervisor Professor Masami Hagiya for constantly advising me to continue my research. Professor Satoshi Murata in Tohoku University welcomed me to his lab, and gives me an opportunity to conduct an experiment about spatiotemporal chemical reaction. I appreciate previous Assistant Professor Fumiaki Tanaka, now in National Institute of Advanced Industrial Science and Technology, for teaching the fundamentals of DNA computing and instructing the experimental procedures.

I thank Doctor Masahiro Hamano for the constructive discussion about the simulation and abstraction of RNAi in Chapter 6. Abstract model of nucleic acid reaction system and the discussion about the exactness in Chapter 4 was inspired by and discussed with Professor Satoshi Kobayashi in the University of Electro-Communications. I appreciate Doctor Richard Potter for correcting my English mistakes of the paper I submitted in 2012. Professor Hiroshi Iba and the member of his group suggested me to use a string of domains to represent the genotype for the evolutionary computation in Chapter 7. The knowledge of synthetic biology and its experimental operations are learned from Associate Professor Daisuke Kiga and the members of his group in Tokyo Institute of Technology.

Discussions with professors and members in molecular robotics and synthetic biology groups were significantly helpful for my research, especially Masahiro Takinoue, Ken Komiya, Shin-ichiro M. Nomura, Akihiko Konagaya, Akio Nishikawa, Miki Hirabayashi, and Nathanaël Aubert. I thank the anonymous reviewers of the conferences (International Meeting on DNA Computing and Molecular Programming, International Conference on Parallel and Distributed Processing Techniques and Applications, and International Conference on Computational Methods in Systems Biology) for sincerely reviewing and giving valuable comments. Work in this thesis is supported by Grant-in-Aid for JSPS Fellows (11J09247), Exploratory Research (11015189), and Scientific Research on Innovative Areas (24104005).

All members in Hagiya and Murata labs are important colleagues of my university life. I learned the significance of cooperation from the group members of The International Genetically Engineered Machine competition 2009 (Tokyo In-

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Organization of the Thesis

In this thesis, I propose computational techniques to efficiently simulate and automatically design chemical reaction systems composed of nucleic acids. Such techniques are required to analyze and construct elemental modules that can contribute in various fields such as biology, medicine, and engineering.

This thesis is composed of eight chapters, which are arranged in the follows order. In this Chapter 1, I first explain the potential applications of nucleic acid reaction systems. Then I point out the restriction of conventional researches to analyze and design reaction systems, and state the need of efficient simulation and design automation. Conventional researches highlighted in this chapter are closely related to the motivation of this thesis. In Chapter 2, I overview background of this thesis, including the basis, demonstration, simulation, and design of chemical and biological nucleic acid systems. By comparing this thesis and other researches, the target and the purpose of this thesis is clarified.

Chapter 3 explains a novel model to simulate a nucleic acid reaction system based on graph rewriting and chemical kinetics. In Chapter 4, I first point out that the model cannot address the combinatorial explosion of molecular species without any restrictions. Then I explain a technique to prevent the explosion and demonstrate efficient simulations of systems that utilize hybridization reactions. In Chapter 5, relation between simulation and abstraction is explained formally using a diagram. In the beginning of Chapter 6, another combinatorial explosion problem of a system based on enzymatic reactions is illustrated. A technique similar to that of the previous chapter is introduced to efficiently simulate the enzymatic reaction system. Chapter 7 explains an automatic design algorithm to search for a system with the function of interest. To evaluate the ability of the algorithm, *in silico* and *in vitro* experimental results are described. In Chapter 8, detailed restriction and trade-off of the techniques are discussed by comparing with related researches. Finally, future work and the contribution of this thesis

are summarized. Note that some of the methods and results are succeeded from the bachelor's and master's work of the author.

1.2 Nucleic Acid Reaction System

1.2.1 Potential Application and Importance

The programmability of deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) gives rise to the emerging filed that makes use of DNA and RNA to construct various functional systems. An arrangement of four types of building-blocks of nucleic acids allows the programmability because each building-block has different chemical properties. By exploring a deliberate arrangement and an appropriate combination of nucleic acids, functional systems are achieved. The functionality of nucleic acid systems can be further enhanced by harnessing the abilities of enzymes or chemically modifying nucleic acids.

Against those backgrounds, researchers have been developing functional modules such as combinatorial circuit [1, 2, 3], sensing switch [4, 5], and autonomous actuators [6, 7]. By combining such modules, researchers are recently demonstrating significant applications such as molecular sensors [8, 9, 10], drug delivery system [11], cell diagnosis system [12], and implementation of neural network [13]. Such systems are designed to contribute to several subjects which can be categorized as follows: engineering nanoscale machines, reconstructing living organism, and the treatment of diseases. For exploring further complex systems, techniques to analyze and strategies to design a new module play an important role in the field.

1.2.2 Nucleic Acids and Modeling

To model nucleic acid reaction systems, several simple representations of the chemical structure of nucleic acids (Figure 1.1) have been proposed. I explain one of them, which is a standard representation to simulate and design systems with dynamic behaviors (e.g. [1, 14]). The representation is further revised to a novel computational model in Chapter 3, while other conventional models are mentioned in the last section of this chapter.

Chemically, a strand of DNA is a sequence of nucleotides that is specified by a string of the four elemental bases 'A', 'T', 'G', and 'C', which represent adenine, thymine, guanine, and cytosine, respectively. On the other hand, RNA employs 'U', which represents uracil, instead of 'T'. Sequence of nucleotides is connected by a backbone that chemically consists of phosphate and sugar, which has a unique direction from 5' to 3'ends. Due to the chemical characteristic of

2

(a) Highlighting each atom

(b) Highlighting each nucleotide
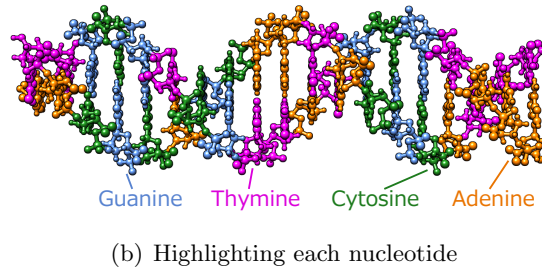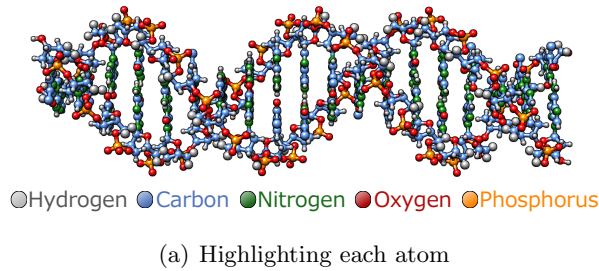
Figure 1.1: Chemical structure of double stranded DNA highlighting (a) each atom and (b) each nucleotide. Sequence of one DNA (ATGCGTCGTTATGCG-TAAA) is randomly chosen , while the other has a reverse complementary sequence (TTTACGCATAACGACGCAT). The atomic model was generated using Namot2 [16] and visualized by UCSF Chimera [17].

each base, nucleotides form hydrogen bonds between complementary pairs ('A'-'T' and 'G'-'C' for DNA and 'A'-'U' and 'G'-'C' for RNA). As a consequence, two complementary strands in anti-parallel directions hybridize into a well known double stranded structure [15].

To represent dynamical systems, a simple and customary representation have been exploited (Figure 1.2), which focuses on the topological information of DNA [18, 19]. In the representation, single stranded DNA (ssDNA) is drawn as a straight arrow representing the direction from 5' to 3' ends of DNA. A sequence of nucleotides is classified into subsequences according to a unit of reaction (called *domain* or *abstract base*), where each unit is denoted by a letter. In this thesis, uppercase and lowercase are complementary to each other. Hybridized complementary domains are connected by a line that corresponds to hydrogen bonds.

Reaction systems taking advantage of RNA can be also represented using the domain level representation. To distinguish RNA from DNA, zigzag arrow is assigned to represent RNA [20]. Because DNA is more stable and less reactive compared with RNA, it is possible to select optimal materials depending on

3
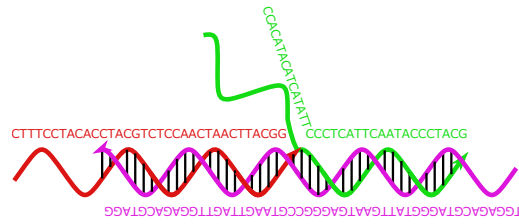
(a) Sequence and double helix
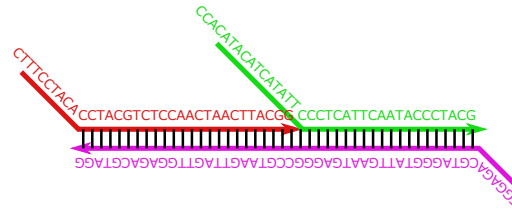
(b) Sequence and arrows
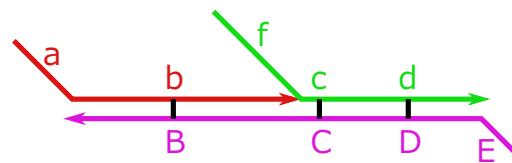
(c) Domain and arrow

Figure 1.2: Domain level representation of DNA of catalytic reaction gate [14], which is explained in a step-by-step manner. (a) All sequences of nucleotides and hydrogen bonds are shown using double helical arrows and lines. (b) Though the sequence is conserved, ssDNA is represented as a straight arrow. (c) Letters are allocated to reaction units (domains), where hybridized complementary domains are connected by lines. Letters 'a', 'b', 'c', 'd', 'e', and 'f' are allocated to the domains that consist of 10, 24, 4, 16, 6, and 16 long nucleotides, respectively.

purposes. For example, living cell takes advantage of RNA to moderate gene expressions by a mechanism called RNA interference (RNAi) [21, 22, 23].

1.2.3 Elemental Nucleic Acid Reactions

Hydrogen Bond Reactions

Dynamics of reaction systems arise as a consequence of repeating basic reactions of nucleic acids. Basic reactions of nucleic acids are related to hydrogen bonds, enzymes, and chemical modification. The reactions of hydrogen bonds are clas-

sified into three elemental reactions, which are hybridization, denaturation, and branch migration. In contrast, the variation of enzymes and modifications causes a wide variety of elemental reactions.

In the domain level representation, hybridization and denaturation are association and dissociation, respectively, between complementary domains. The kinetic speed of denaturation depends on the length of the sequence of complementary domains. The longer the sequence is, the more stable hybridized domains are, and the less likely the denaturation occurs. Branch migration exchanges hydrogen bonds using a neighboring hybridized domain as a toehold.

Only hydrogen bonds reactions (simply called hybridization reactions) take place in the example of catalytic reaction (Figure 1.3). Without catalyst, fuel and substrate do not interfere because any domains of the substrate are not accessible by the fuel. With the addition of catalytic input, they hybridize together and emit output and signal strands. Since the input serves as catalyst, it is repeatedly used for another round of reactions.

Enzymatic Reactions

Other types of reaction systems exploit enzymatic reactions to achieve complex functionalities. In general, enzymes are functional proteins discovered in the metabolic process of living organism and researched in field of molecular biology [24]. Modeling enzymatic reactions based on chemical species is indispensable for not merely the construction of artificial systems but the study of biology.

Enzymes that react with nucleic acids include, but not limited to, polymerase, nuclease, and ligase. Polymerase extends nucleic acids from free nucleotides and synthesizes a new sequence that is complementary to template nucleic acids. Depending on nucleic acids of the template, it is called DNA-directed or RNA-directed polymerase. Furthermore, depending on nucleic acids of the new strand, it is called DNA or RNA polymerase. For example, RNA-directed DNA polymerase synthesizes new DNA strand from RNA template.

DNA polymerase enforces primer-dependent polymerization, in which the enzyme always extends the 3' end of pre-existing nucleic acids (primer). This type of polymerase has been exploited to amplify the amount of DNA by polymerase chain reaction (PCR) for the detection of small amount of DNA. On the other hand, some RNA polymerases employ primer-independent polymerization. If new RNA strand is polymerized from double stranded DNA (dsDNA) template starting from a special promoter region, the reaction is called transcription.

Nucleases are capable of cleaving phosphate backbone of nucleic acids in var-
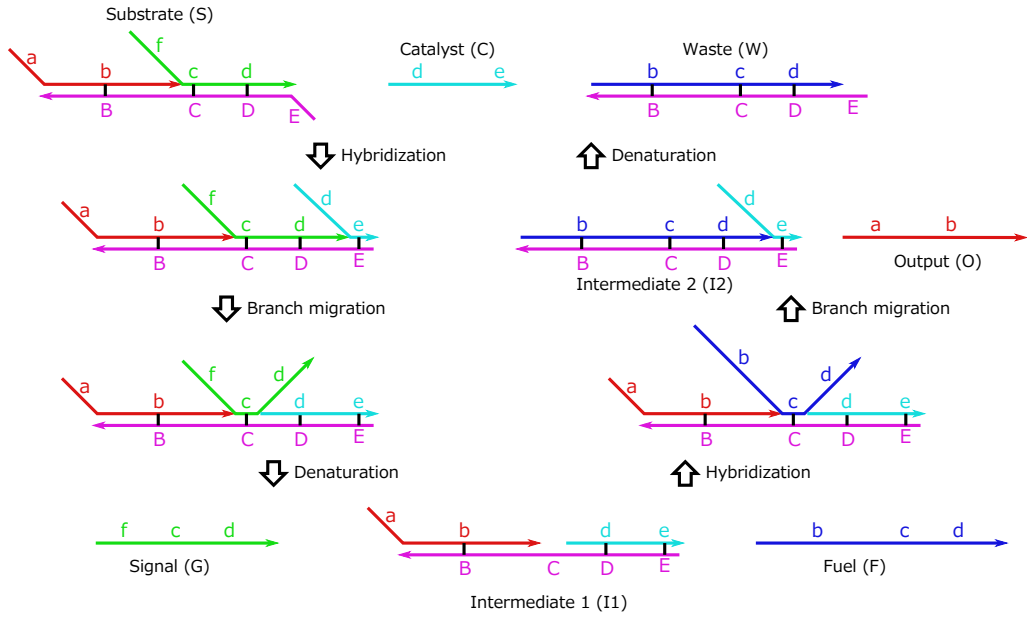
Figure 1.3: Hydrogen bonds reactions of catalytic system [14]. Each reaction is drawn one-by-one, while whole reactions start from the top and continue in anti-clockwise direction. The gate consists of the substrate (S) and the fuel (F), while the catalyst (C) serves as an input to the gate. To begin with, S and C hybridize together and result in an intermediate structure. As a result of subsequent branch migration and denaturation, the structure separates into the signal (G) and the intermediate 1 (I1). I1 and the F turn into the intermediate 2 (I2) and the output (O) by another sequential hybridization and branch migration. Finally, I2 releases C and becomes the waste (W) by denaturation.

ious manners. Deoxyribo- and ribo-nuclease repeat cleaving so that DNA and RNA are thoroughly decomposed. Nucleases known as nickase and restriction enzyme recognize specific sequence of dsDNA to cleave one and both sides of ds-DNA, respectively. On the other hand, nuclease known as Dicer non-specifically cleave long dsRNA into short dsRNAs. Ligase joint 3' and 5' end of two DNA strands, which has important function in replicating DNA [25].

Reverse-transcription-and-transcription-based autonomous computing system (RTRACS) [20, 26, 27] makes use of polymerization reactions for elaborate state transitions (Figure 1.4). Primer DNA is extended by a collaboration of two ssRNA inputs and polymerization to produce an intermediate DNA structure. With the structure, the converter form a full duplex DNA structure by hybridization and polymerization. The duplex structure contains double stranded promoter region where primer-independent polymerase can recognize to initiate the transcription. The system is an implementation of AND gate because the
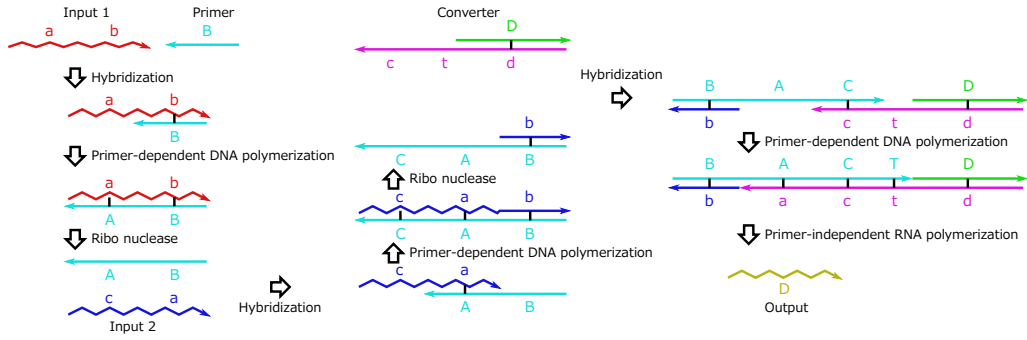
Figure 1.4: Enzymatic reactions of the AND gate of RTRACS [20]. The gate consists of the primer and the converter, while the inputs and output are ss-RNA. The primer that hybridizes to the input 1 is extended by an RNA-directed primer-dependent DNA polymerase. Ribo-nuclease decomposes only the Input 1 and the remained ssDNA hybridizes to the Input 2. Partially double stranded intermediate DNA structure is generated as a result of subsequent polymerization and ribo-nuclease decomposition. The structure hybridizes to the converter and form dsDNA by polymerization. Because the DNA promoter region (indicated by 'T' in the figure) is double stranded, downstream of the dsDNA is transcribed to the output by a primer-independent RNA polymerization.

output ssRNA is synthesized if and only if both inputs are added.

In Section 2.2, recent researches about systems utilizing hybridization and enzymatic reactions are exhibited. Because hybridization reactions sustain under diverse conditions, the robustness is one of the advantages of them. In contrast, enzymatic reactions are more efficient in specific functions compared with hydrogen bond reactions.

1.3 Simulation of Chemical Reaction System

1.3.1 Chemical Kinetics and Ordinary Differential Equation

Chemical kinetics is commonly applied to model the behavior of chemical reaction systems. A chemical reaction is represented by an equation where molecular species before and after the reaction are in the left-hand and right-hand sides, respectively. In the example of the catalytic reaction (Figure 1.3), the equations described in the paper [14] are

$$S + C \xrightarrow{k_1} I1 + B$$
$$I1 + F \xrightarrow{k_2} I2 + O$$
$$I2 \xrightarrow{k_3} C + W,$$
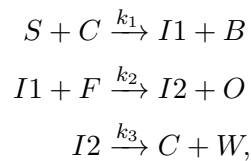
where

$$k_1 = 6.5 \times 10^5 [M^{-1}sec^{-1}]$$
$$k_2 = 4.2 \times 10^5 [M^{-1}sec^{-1}]$$
$$k_3 = 4.0 \times 10^{-3} [sec^{-1}].$$

Sequential reactions of hybridization, branch migration, and denaturation are collected into the first equation. Similarly, second equation describes a sequence of hybridization and branch migration. When left-hand side contains one and two molecules, the reactions are called unimolecular and bimolecular reactions, respectively.

Speed of each reaction is described by a rate constant over an arrow. The rate constants were experimentally measured in the paper [14] so that the results of computer simulation and chemical experiment coincide. If the purpose of the equations is to roughly predict the behavior, coarse values can be adopted to the rate constants. In that case, it is important to define the rate constants that are empirically correct on the order of magnitude.

To numerically predict the dynamic behavior of systems that are modeled by the chemical equations, two types of mathematical techniques can be considered. One is a stochastic simulation and the other is a continuous simulation, both of which are capable of tracking the time evolution of the concentrations of molecular species. For stochastic simulation, discrete numbers are assigned to molecular species to represent the amounts of molecules.

Gillespie algorithm [28] is a typical technique for the stochastic simulation based on statistical viewpoint. It is possible to accurately simulate each individual chemical reaction one-by-one according to the distribution of the possibilities reactions. Since random numbers are generated in the algorithm, the result of the simulation differs each time by an uncertainty of noise. Therefore, statistical treatment by simulating multiple traces is required to correctly understand the dynamics of systems by stochastic simulation.

In contrast, ordinary differential equations (ODE) are formulated from the chemical equations in order to simulate a system continuously. In the example,

the formulas are defined as

$$\frac{d}{dt}[S] = -k_1[S] \times [C]$$
$$\frac{d}{dt}[C] = -k_1[S] \times [C] + k_3[I2]$$
$$\frac{d}{dt}[I1] = k_1 S \times [C] - k_2[I1] \times [F]$$
$$\frac{d}{dt}[F] = -k_2[I1] \times [F]$$
$$\frac{d}{dt}[I2] = k_2[I1] \times [F] - k_3[I2]$$
$$\frac{d}{dt}[W] = k_3[I2],$$

where t is a time parameter and $[X]$ represents the concentration of chemical species X. The system of RTRACS (Figure 1.4) can also result in similar type of equations and formulas. Time evolution of each molecular species is uniquely tracked by numerically solving formulas with specific initial values. In this thesis, continuous simulation are applied for the simulation of nucleic acid reactions in most of the cases with the exception in Section 3.3.1.

In the simulation of nucleic acid reaction systems, formulas are carefully developed by human optimization to fit experimental results [1, 14, 2]. For the sake of simplicity, sequential reactions may be collected in one chemical equation as in the example of the catalytic gate. In systems biology, however, it is common to semi-automatically build such formulas and analyze the behavior of a system from an initial configuration of signaling pathway (metabolic modelings in [29]). Therefore, automatically developing formulas of nucleic acid reaction systems in similar way is important to save the labor of human optimization.

1.3.2 Simulation based on Programming Language

For such purpose, formal notations of DNA and the transition rules of the notation have been proposed [19, 30]. Only two types of domains are assumed in the papers, one is a long and stable domain, and the other is short and unstable domain. Long domain is too long to denature, while short domain (called toehold) is short enough to spontaneously denature.

The notation restricts DNA to a linear structure where single lower and multiple upper strands are connected in one direction. To represent such linear form structure, domains that compose lower strands are aligned from left to right. When a domain in lower strand is hybridized to a complementary upper domain, the domain is denoted in brackets [,]. Domains of upper strands are denoted in brackets <,>, which can also become overhangs from a hybridized upper strands.

Table 1.1: Notation of DNA structures. First column corresponds to the name of structures in Figure 1.3. Second column corresponds to the notation of each structure [30]. Other intermediate structures without names are also possible to represent by the notation. `A`, `B`, and `D` are long stable domains, while `C` and `E` are short toehold domains. Note that only one dimensional linear structures are represented by the notation.

C	`<D E>`
S	`<A> [B] : <F> [C D] : E`
F	`<B C D>`
I1	`<A> [B] : C : [D E]`
I3	`[B C D] : <D> [E]`
W	`[B C D] : E`

Each nick, where the upper strands are separated, is denoted by a colon :. For example, the notations of strands in the catalytic gate (Figure 1.3) are summarized in Table 1.1.

Transition rules of the notations are defined to represent hybridization, denaturation, and branch migration (called strand displacement in the paper). By applying the transition rule of denaturation, a transition from I2 to W and C is represented, for instance. Thanks to the mathematical definitions, most of the process to simulate DNA hybridization reaction systems can be automated. Concretely speaking, it is possible to automatically construct whole chemical reactions and further apply to continuous or stochastic simulations only with initial sets of strands and concentrations.

The ability of the formal definition is powerful so that formal analysis such as probabilistic model checking has been applied [31]. Nonetheless, restriction of the DNA structure is one of the limitation, because the structures of DNA are not limited to linear ones in the field of DNA computing. Other important structures comprise hairpin loop [32, 33, 34], bulge loop [35], and circle [36] structures.

Moreover, only hybridization reactions are considered as transition rules since the purpose of the paper was to simulate a system driven by strand displacement reactions. Other transition rules are necessary to simulate a system driven by enzymatic reactions such as restriction enzymes [37, 38, 39], polymerase [40, 41], and nickase[42]. A novel graph based representation of nucleic acid and reactions among them, which overcome such restrictions, are explained in Chapter 3.

1.3.3 Combinatorial Explosion Problem of Graph Based Model

Graph (equivalent to network) has been a powerful representation of biological phenomena in systems biology [43, 44]. In such graph based representations, nodes and edges correspond to biological components (RNA and proteins for example) and reactions (enzymatic reactions for example), respectively [45]. Graph and graph rewriting are also applied to represent molecules and chemical reactions, where nodes and edges correspond to atoms and chemical bonds in computational chemistry [46, 47, 48].

By modeling biological and chemical systems using graph, it is possible to simulate the dynamics of chemical and biological systems in a similar way of nucleic acid reactions. It is pointed out, however, that the simulation becomes intractable in some cases because of the explosion of the number of molecular species [49] and reactions [50]. I give concrete examples of nucleic acid reaction systems that show such combinatorial explosion in the beginning of Chapter 4 and 6.

The concept of combinatorial explosion can be easily understood by a simple example [49] that produces exponential number of molecular species (Figure 1.5(a)). In the figure, molecule H can bind to n molecules S_i $(1 \leq i \leq n)$ independently. Suppose H and S_i are unbound for all i $(1 \leq i \leq n)$ at the initial state, is it possible to simulate a system where the rate constants for binding and unbinding reactions are arbitrary? When n increases, the number of chemical equations exponentially increases because $2^n + n$ structures are produced from only $n + 1$ molecular species. With such straightforward model and simulation, it is impossible to simulate the system with large n.

1.3.4 Model Reduction by Subgraph

To prevent such explosion, model reduction techniques are commonly applied to reduce the number of parameters of simulation. I explain one conventional technique called κ-calculus that focuses on the locality of species [47, 51, 49, 52], while other techniques focus on the hierarchy of systems [53, 54, 55]. The problem is fundamentally caused by the exhaustive enumeration of structures even though the binding and unbinding reactions occur locally. The number of species is decreased by adopting new subgraphs that store the information of local connectivity between species. In the example [49], the new subgraphs are called *fragments*, where H is separated in n local fragments (Figure 1.5(b)).

Due to the fragments, it is possible to avoid redundant enumeration of the complexes between H and S_i. According to the paper, the chemical equations

(a) Combinatorial explosion

$H(s_i)$ $S_i(h)$

$S_i(h^0), H(s_i^0)$
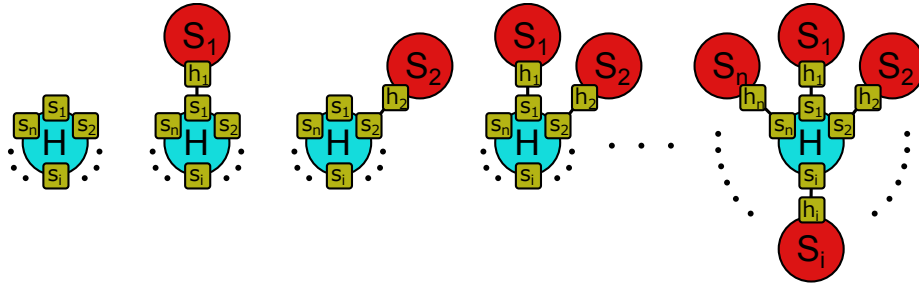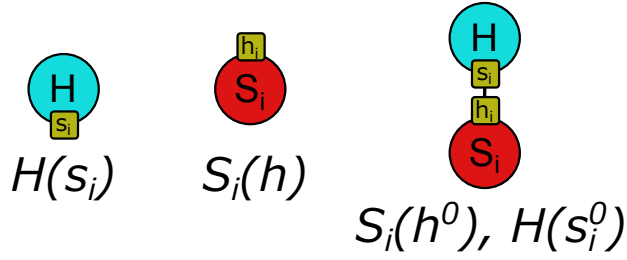
(b) Fragments

Figure 1.5: Simple example of the combinatorial explosion of graph based model and its abstraction [49]. Molecular species H has n sites where s_i binds or unbinds to the site h of S_i. Since all the binding and unbinding reactions are independent, combinatorial explosion in the number of molecular species is unavoidable. (a) H produces 2^n species by the combination of connections in a graph based representation. (b) Separating each site of H, only $3n$ number of species are taken into consideration. There are three types of fragments for each i ($1 \le i \le n$). Each fragment is named $H(s_i)$, $S_i(h)$, and $S_i(h^0), H(s_i^0)$ from the left to the right of the figure.

among fragments are

$$H(s_i) + S_i(h) \xrightarrow{k_i^+} S_i(h^0), H(s_i^0) \qquad\qquad (1 \le i \le n)$$

$$S_i(h^0), H(s_i^0) \xrightarrow{k_i^-} H(s_i) + S_i(h) \qquad\qquad (1 \le i \le n),$$
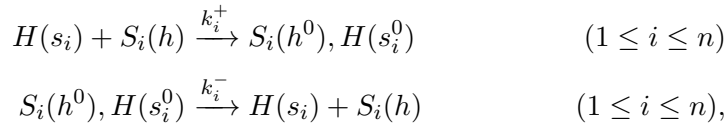
where k_i^+ and k_i^- are the rate constants of binding and unbinding between fragments. Consequently, the number of fragments successfully decreased to $3n$. Note that fragment $S_i(h^0), H(s_i^0)$ implicitly merges all global structures that binds s_i of H and h of S_i. This means that fragments are regarded as coverings and not as partitions of global structures.

A function to calculate the concentration of fragments from global structures is trivial. The concentration of unbound fragment $S_i(h)$ is equivalent to

that of corresponding structure. The concentration of the fragment $H(s_i)$ and $S_i(h^0), H(s_i^0)$ are computed by summing up all the concentration of structures where s_i of H and h of S_i are unbound and bound, respectively. By this function, simulations of global structures and fragments are related to each other.

Comparison between the time evolutions of fragments and global structures is one of the concerns to evaluate the ability of the reduction. From the aspect of bisimulation, it is possible to decide whether the time evolution of fragments exactly simulate the evolution of global structures or not. In the case of the example, independence of binding and unbinding reactions ensures the exactness [49]. However, the authors of the paper point out that the exactness cannot be achieved by slight modification to make the reactions dependent to each other.

An intriguing aspects of fragmentation include finding out an inverse function to compute the concentrations of global structures from those of fragments. In fact, it is possible to exactly recover the concentrations of global structures from the those of fragments in the example without any modifications [49]. It is recovered because the bound and unbound states are proportionally distributed. Concretely speaking, the ratio of bound state between the sites of index i is

$$\frac{[S_i(h^0), H(s_i^0)]}{[H()]},$$

while that of unbound state is

$$\frac{[H(s_i)]}{[H()]},$$

where $[H()]$ denotes the concentration of the structure H ignoring binding and unbinding states. It is calculated as $[H()] = [H(s_i)] + [S_i(h^0), H(s_i^0)]$, where $[H()]$ is identical value for any i $(1 \leq i \leq n)$.

Because of the independence, multiplying the ratios among all i $(1 \leq i \leq n)$ is possible. For example, the concentration of a global structure where all site s_i of H are bound to h of S_i $(1 \leq i \leq n)$ can be calculated as

$$[H()] \times \prod_{1 \leq i \leq n} \frac{[S_i(h^0), H(s_i^0)]}{[H()]}.$$

In most of the cases where the reactions are not independent, however, such inverse function to exactly recover the concentrations of global structures is not trivial. Because the model reduction by fragments was intended for analyzing signaling pathways such as phosphorylation cascades, direct application to nucleic acid reaction systems has not been shown. The reduction method cannot be simply applied since the definition of fragments depends on the purpose of simulation.

For calculating the equilibrium state of a hybridization reaction systems, where molecules hybridize to produce a large assembly, enumeration method using hypergraph has been proposed [56]. Similar to the fragmentation, the method focused on the locality of reactions to implicitly enumerate the combinatorially large number of assemblies. The author of the paper claims that the method can be applied to calculate the thermodynamic equilibrium state of DNA and RNA interactions. Though the thermodynamical equilibrium can be efficiently calculated by the method, the application to the time evolutionary simulation was not emphasized. By a theoretical constraint, the form of the assemblies was limited to a linear structures where elemental units repeat only in one dimensional direction.

Original reduction techniques to efficiently analyze nucleic acid reaction systems, which are independently developed from the fragmentation and equilibrium calculation, are explained in Chapter 4 and 6.

1.4   Design Strategy

1.4.1   Simplified and Restricted Modules

In addition to analyzing biological systems for understanding their dynamic behavior, such simulation techniques can be applied to design a system with a function of interest. In general, the designing process requires human trial and errors to construct an elaborate system. Simulation techniques play an important role for debugging a designed system by predicting its behavior.

Functional modules can be modeled by the programming language explained in Section 1.3.2, which is shown in Table 1.2. Simulation by the programming language delivers a schematic diagram that exhibits how reactions are related to each other, which I call *topology* of a module. By simulating the dynamics, it is possible to optimize the topology of a module and initial concentrations of structures to achieve a function of interest. Since such activity is similar to computer programming, human trial and error are inevitable to search for optimized conditions for a system.

Once a robust module is designed, scale-up systems can be designed by integrating the modules. For the purpose, a module called seesaw gate was developed [2], which was further integrated to construct combinatorial circuit [57] and neural network [13]. Although the gate can be represented by the programming language (Table 1.2), the authors of the paper developed their original representation and chemical equations. A module of see-saw gate is represented as a simplified node with multiple values and wires to describe its parameters. A

Table 1.2: Models of catalytic gate [14] and seesaw gate [2] represented in the paper of the programming language [30]. A module is put in parenthesis, where structures are separated by vertical lines. Initial concentration of each structure is specified by a value written in the front of it. For instance, PF*<B C D> means that the concentration of <B C D> is PF. Toehold domain <T^c> is not perfectly complementary to T because of a mismatch in nucleotides, which affects the kinetic speed of hybridization reactions.

Catalytic gate	(PF*<B C D> | PC*<D E> | PS*<A> [B] : <F> [C D] : E)
Seesaw gate	(PF*<S2 T S3> | PTh*[S3]:<T^c> | PI*<S3 T S4> | PGO*<S1>[T S3]:T)

topology of systems can be designed by connecting wires, which is similar to the design of electronic circuits.

Another similar method to integrate modules is the pathway programming [34]. A hairpin structure of DNA is simplified as a node and the reactions among nodes are represented by interactions between nodes (Figure 1.6). Using the representation, catalytic assembly, dendritic assembly, and autonomous locomotion of DNA structures have been demonstrated. In the paper, the simulation results by chemical equations that were manually developed had a good agreement with experimental results. Applying the programming language [30] to simulate the hairpin systems is impossible because hairpin structures is not supported by the notation.

1.4.2   Design Automation by Evolutionary Computation

Even with such programming techniques, however, the design of desired topology is still extremely difficult for humans. It is because the combination of nucleotides in a system increases rapidly as the number of molecular species increases, which makes it unintelligible to prevent undesirable reactions. To eliminate such complex process, evolutionary computation to search for a semi-optimal topology is effective.

In general, heuristic algorithms are exploited for the evolutionary computation. To execute such algorithms, computer models to describe a genotype of a system and a simulation method to predict the phenotype of the model are required. A candidate solution can develop diversity by a process of mutation or crossover. Fitness value is calculated from the result of the simulation, which affects the selection process of the evolution. By iterating those processes, a semi-optimal system that has relatively high fitness value can be selected.

Designing a topology of genetic network using such evolutionary computa-
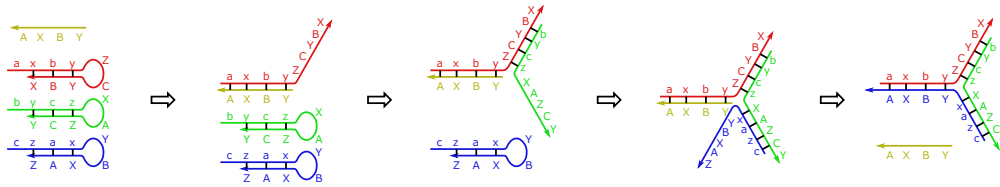
(a) Formation of three junction from hairpin structures
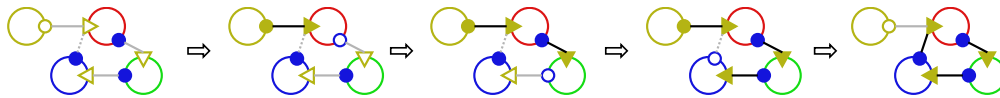
(b) Pathway programming and its execution

Figure 1.6: Pathway programming and its execution of three junction structure [34]. (a) Domain level representation of the reactions, and (b) pathway programming and its execution are shown. A hairpin structure of DNA is denoted as a node with triangle and circle sites. Triangle and circle are open states when they are accessible by other DNA strands, while circles may be solid by the formation of hairpin loop. Reactions propagate from circles to triangles by connected lines that are programmed in advance. When open triangle and circle are connected by a line, a reaction to open the hairpin occurs and both of them become solid. Once the hairpin is open, all the circles in the same node become open. When open triangle and solid circle are connected by dashed line and the circle is connected with another solid triangle by solid line, a reaction to displace ssDNA occurs and the first triangle become solid, while the other become open.

tion has been proposed. Genetic network is a system that controls the activity of genomes by a collaboration of various proteins. Demonstrations of genetic networks include bi-stable switch [58] and oscillatory behavior [59].

Applying evolutionary computation to the design of genetic network with various dynamics such as oscillatory behavior has been reported [60, 61]. The genotypes of them were based on models such as simplified gene [62], network representations [63, 64], and a combination of standard parts library [65]. Although the evolutionary computations had significant impact because of the design of nontrivial topologies, the designed genetic networks were not so easy to implement by an actual gene regulatory network. One of the reasons for the problem is that those models assumed unrealizable properties of genes and their regulation.

Although modeling and simulating techniques are reported for the nucleic acid reaction system so far, an application to the design of new topology by

evolutionary computation has not been addressed. For actual implementation, a system composed of nucleic acids is more suitable than genetic network because they are programmable materials and their reactions are realizable. I introduce an original approach to represent a genotype of nucleic acid reaction systems and apply the evolutionary computation to design a system with a function of interest in Chapter 7. The results of the evolutionary computation are verified not only by theoretical simulations but also by the chemical experiments.

Chapter 2

Background

2.1 DNA not Merely an Information Carrier

The idea of using molecules to build a machine was claimed since the Richard P. Feynman in around 1960s [66]. As the technology evolves, human has searched for several strategies to build artificial and complex molecular devices inspired by living organisms [67]. One of the directions is using DNA and RNA as building-blocks, because the programmability of nucleic acids can be applied to construct variety of functions and architectures.

In nature, DNA is utilized as an information carrier for biology and has been studied over decades. One of the famous researches is the suggestion of the structure of DNA [15]. As its physical and chemical properties are revealed, making use of DNA as a chemical material for other purpose has been searched.

The field of nucleic acid reaction systems can be roughly divided into four categories, which are constructing nanostructures, developing stimuli sensing devices, observing dynamic actuations, and conducting information processing. Among those categories, computer simulation and computer-aided design methods are widely exploited for various purposes. In this chapter, I overview noticeable researches of the field in order to make the position of this thesis clearer.

2.2 Nucleic Acid Module

2.2.1 Nanotechnology

Because of the programmability of DNA and RNA, precise positioning of atoms and building static structures by nucleic acids are one of the interests. Since the proposal and construction of simple structures of DNA in the early 1980s [68], fundamental structures have been reported, including a cube [69], a tetrahedron [70], an octahedron [71], Borromean rings [72], and junctions [73]. The breadth of the area can be understood by various recent review papers [74, 75, 76, 77, 78, 79, 80].

Combination of simple DNA tiles and motif was reported to build complex

structures such as geometric configuration [81, 82, 83], tubes [84], rings [85], and tensegrity triangle [86]. Using ssDNA as a tile also exhibits a powerful and beautiful constructions of 2D and 3D structures [87, 88]. Such assemblies were applied to various purposes such as positioning bio-molecules [89, 90, 91], computation [92], and photo-controllable capsule [93]. Those researches were pushed forward by not only an experimental observation of the behavior of tile assembly [94] but also a theoretical models of tile assembly [95, 96].

Artificial nanostructures were also constructed by employing RNA, which was inspired by natural architectures such as transfer RNA. RNA constructions include a square [97], a polyhedron [98], and two dimensional pattern [99]. Rich functionality of RNA is expected to expand the abilities of nucleic acid based architectures.

An innovative method to construct complex and large structures was DNA origami [100] (Figure 2.1). The method rapidly becomes standard in the filed of DNA nanotechnology, and researchers demonstrated a Möbius strip [101], a box with lid [102], stacked three dimensional shapes [103], curvature structures [104], rods [105, 106], assemblies of origami [107], and multilayer structures [108]. Applications of the method comprise channel construction [109], nano-pores [110, 111, 112], photo chemistry [113, 114, 115], positioning bio molecules [116, 117], positioning gold nano-particles [118, 119, 120], and the layout of carbon nanotubes [121]. To enhance the efficiency of the method, researches reported experimental optimization techniques such as controlling the conformation [122, 123, 124, 125], purifying the construction [126], and the analysis of its conformation [127, 128, 123, 129]. Similar methods derived from the original method, which include small scaffold [130], unidirectional scaffold [131], double stranded scaffold [132], chip-derived staples [133], repeated scaffold [134], and gridiron scaffold [135] origamis.

In the procedure to design origami architectures, arrangement of about 200 ssDNAs with total about 7000 nucleotide pairs is necessary. Obviously, the process cannot be accomplished without powerful computer modeling tools. The design process can be semi-automated by rapid prototyping tool [136] and further debugged by stability simulation [137]. Two dimensional origami can be automatically designed from an image that outlines the desired structure [138].

2.2.2 Sensor

Since the nucleotide bases have different chemical properties and steric constitutions, sensing bio-molecule or cell surface are possible by a special sequence
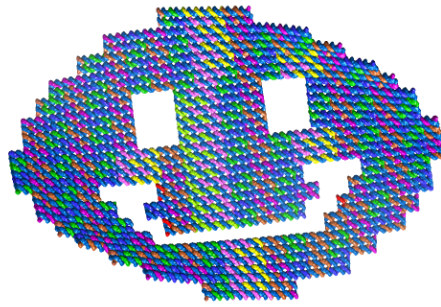
Figure 2.1: Computer graphics of the smiley face constructed by DNA origami [100]. The structure is made of one long scaffold strand and multiple short staple strands. The computer model was programmed on caDNAno [136] and visualized by a CanDo server [137].

of DNA or RNA. Such sequence that can detect specific target is called aptamer, which was invented in the begining of 1990s. As a result of *in vitro* selection, aptamers are chosen from a library of random sequences, whose technique is called systematic evolution of ligands by exponential enrichment (SELEX) [139, 140]. It is possible to recognize the state-of-the-art aptamers by some review papers [141, 8, 9, 10].

For example, DNA aptamers have been applied to detect thrombin [5, 142, 143, 144, 145, 146, 147], adenosine triphosphate (ATP) (Figure 2.2(a)) [4, 148, 149, 150, 151, 152], cocaine [153, 154, 155], platelet-derived growth factor (PDGF) [156, 157], hemin [158, 159], and leukemia cell [160, 161]. On the other hand, RNA aptamers and aptamer based sensors can detect protein kinase [162], Tat protein [163], malachite green [164, 165], theophylline [166], and cancer cell [167, 168]. Various strategies to combine aptamers with other technologies have been investigated for applications such as the chemical modification to enhance fluorophore [169], modular sensors [170, 171], protein conjugation [172], and gold nano-particle assembly [173]. Some of the aptamers were analyzed in more detail by experimental [174, 175, 176, 177] and computational [178] analyses. Computational approaches to search for aptamers by optimizing the process of SELEX have also been reported [179, 180].

Another type of nucleic acid sensor is called Dexyribozyme (DNAzyme) that has enzymatic activity in the presence of target metal ions. Some of the DNAzymes can cleave the backbone of ribo [181] or deoxyribo [182, 183] nu-

(a) ATP sensor	(b) UV sensor
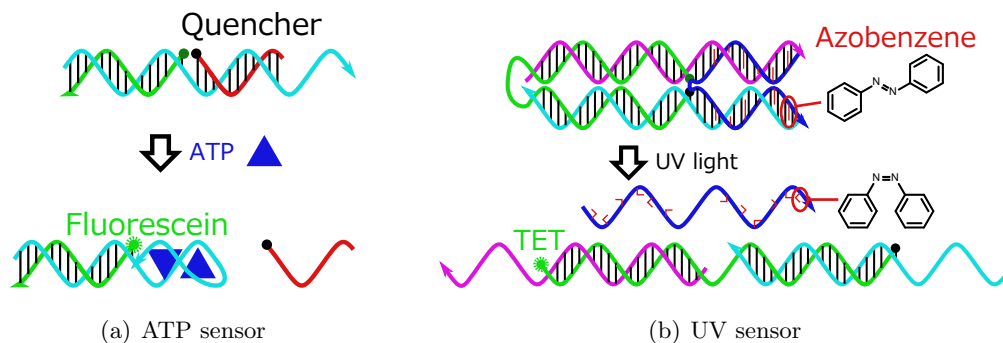
Figure 2.2: DNA sensors targeting ATP [149] and UV light [200], which emit fluorescence by the addition of inputs. (a) ATP aptamer changes its conformation by recognizing two ATPs (represented by triangles), by which a quencher labeled ssDNA is released from the sensor. (b) Azobenzenes (represented as straight and bent lines) inserted in the backbone of DNA change their conformations from trans to cis forms by irradiating UV light. Eventually, ssDNA with azobenzene can denature by a steric hindrance of the cis form of azobenezes. In both of the case, fluorophore molecules (fluorecein and TET) are quenched by a nearby quencher molecule in the absence of inputs, which is a phenomenon called Föster resonance energy transfer (FRET). Since the conformation changes by the inputs separate fluorophore and quencher molecules, output of the sensor can be measured as fluoresence intensity.

cleic acids, by which wide applications have been proposed [184]. The applications include sensors of DNA [185, 186], detection of ions [187], boolean gates [188, 189, 190, 3], control of gold nano-particles [191], and dynamical locomotion [192].

According to the condition of the environment, special sequence of DNA can change its conformation. For example, quadruplex structures called i-motif [193] and G-quadruplex [194] change their conformation when the pH and ionic conditions change, respectively [195]. Applications of i-motif include not only the sensors of pH [196] but also the mechanical motor [197], while that of G-quadruplex include the inhibition of protein activities [198].

Chemically attaching other molecules to DNA is another strategy to extend the functionality of DNA for the purpose of sensing. One example is to insert azobenzene molecules to DNA, by which nucleic acids become capable of responding to light irradiation (Figure 2.2(b)) [199, 200, 201]. Attaching thiole to conjugate nucleic acids with gold nano-particles is also possible for sensing heat [202, 203]. The aggregation of the gold nano-particles using DNA hybridization was adopted as a read-out of sensing [204, 205, 173, 206].

2.2.3 Actuator and Information Processing

A DNA fuelled machine was reported in 2000 [207] that made use of branch migration as a source of energy, which is one of the important inventions to build mechanical and dynamic nanoactuators. Combined with other techniques, numerous dynamic walking systems have been developed, including step-by-step walkers [208, 209, 210], autonomous walkers [34, 211, 6], DNAzyme driven walkers [212, 7, 213, 192]. Rotational movements were also demonstrated, which utilize multiple inputs [214] and conformation change of DNA from normal right-handed B-form to left-handed Z-form in ionic buffer [215]. The branch migration reaction was analyzed in more detail experimentally [216] and further extended to a remote toehold method [217].

The realization of a system using DNA as a computational resource was demonstrated in 1994 [218] by solving a small Hamiltonian path problem. After the demonstration, large number of researches appeared to use DNA as a material to carry mathematical computations [219]. In the early stage of the filed, massively parallel aspect of molecules in a solution was emphasized to solve classical combinatorial problems such as satisfiability problem [32].

Computations using other functions of DNA are also important, which comprise programmed nano-circles using i-motif [220]. State transition systems were developed using restriction enzymes [39, 37, 38], polymerase [41], and both i-motif and gold nano-particle [221]. Other examples include the combination of DNAzyme for a reversible computing gate [222], chemical reaction based dynamics [223], and dynamic oscillator using enzymatic reactions [224, 42]. One of the goal of this thesis is to automatically design such computational modules using evolutionary computation.

Combining such modules composed of nucleic acids, elaborate nanodevices or molecular robots can be constructed [225, 226, 227, 228, 229]. In addition to the statements in Section 1.2.1, I re-emphasize the importance of the research of nucleic acid reaction systems by noticeable applications. Robot-like behaviors have been demonstrated by an assembly line [230], navigation of DNA walker [231, 210], and Tug-of-War system [232]. Devices for observation detect target RNA [233], single molecule [234], and single nucleotide polymorphism [235]. For medical purpose, selective cell apoptosis [12, 11] and drug delivery [236, 237] have been demonstrated.

2.3 Computer Simulation and Automation

2.3.1 All Atom and Coarse-Grained Molecular Dynamics

Molecular dynamics method is a powerful technique to simulate the dynamic behavior of nucleic acid reactions. The method calculates interactions among particles for each time step, which can be formalized using energy functions. For instance, the transition of DNA from Z-form to B-form was simulated by computing the dynamics of all atoms [238]. All atom molecular dynamics method, however, is computationally expensive for the simulation of nucleic acids because of the enormous number of atoms in DNA and RNA.

To minimize the computational cost, coarse-grained molecular dynamics method optimized for the simulation of nucleic acids has been proposed [239, 240, 241]. In the coarse-grained models, single nucleotide that consists of around 35 atoms is often represented as three particles, each of which represents base, phosphate, and sugar. The coarse-grained models of nucleic acids have successfully been applied to study the thermodynamics [242], branch migration reaction [243], and a single step of DNA walker [244]. The method, however, is still computationally expensive to simulate nucleic acid reaction system such as the catalytic gate explained in Section 1.2.3 so that another simulation technique based on the domain level representation becomes essential.

2.3.2 Secondary Structure and Sequence Design

Predicting the secondary structure of nucleic acids is a problem to search for a set of base pairings with minimum free energy from a given nucleic acid sequence. Secondary structure is formed by intra-molecular hybridization as in the example of hairpin structure of the Figure 1.6. Solving the problem is important to design a desired secondary structure or to avoid any undesired structures in real experiments. The free energy of an entire secondary structure can be thermo-dynamically calculated as an ensemble of that of local structures such as base pairing, hairpin loop, and bulge loop. Based on such calculation, algorithms to efficiently solve the problem have been developed [245]. Moreover, advanced web services that are capable of solving the problem have recently appeared [246, 247].

On the other hand, the secondary structure of protein folding cannot be efficiently solved by computer algorithms. Secondary structure of protein is similar to that of nucleic acids folding, while protein has 20 basic building-blocks called amino acids. Solving the protein folding problem is important to understand and design a three-dimensional structure of proteins, which affects the functions of the

proteins [248]. To solve the problem, researchers have investigated the application of evolutionary computation [249, 250], though the range of the applications was limited to only a small part of the protein folding. Another interesting approach to solve the problem is extracting an algorithm from humans who are playing an online protein folding game [251, 252, 253].

An inverse problem which is to design a sequences of nucleotides that have desired secondary structure is called sequence design problem. Evolutionary computation is commonly applied to solve the problem because it is a multi objective optimization problem [254]. One of the objectives is the orthogonality among domains, by which sequences allocated to unrelated domains should not hybridize each other. Another objective is optimizing the contents of guanine and cytosine in each of domains in order to keep the melting temperature of nucleic acid constant. Obviously, the purpose of the optimizations is different from the design of the topology of a reaction system.

2.3.3 Phenomenological Simulation of RNAi

In contrast to the simulations that are based on chemical molecules, phenomenological simulations have been applied in biological systems. For example, the phenomenological model of RNAi [255] is shown in Figure 2.3. RNAi, also known as RNA silencing, is an internal cell mechanism commonly seen in both plants and animals to moderate gene expression [21, 22, 23]. In this mechanism, small RNAs produced from a dsRNA directly control gene expression. The system is known to involve specific structures named short interfering RNA (siRNA) and RNA-induced silencing complex (RISC), which are a 20-22 nucleotides long ssRNA and a complex of ssRNA with multiple proteins, respectively.

In the model, same chemical species appeared in different locations due to the pathway to produce the species. In such case, phenomenological model exhibits a redundancy of molecular species.

Moreover, it is difficult to understand how each molecular species behaves in RNAi by the phenomenological model, because the model is not based on chemical reactions. For instance, all siRNAs were treated equivalently in the model, which made it impossible to predict the distribution of the concentration of siRNAs. In fact, the concentrations of siRNAs are expected to have a gradient distribution, which is experimentally supported [256, 23], because polymerization extends RNA in only one direction (from 5' to 3'). Simulations of RNAi, where the target mRNA encodes Dicer gene [257] and aberrant RNA reproduces dsRNA [258], also have similar limitation.
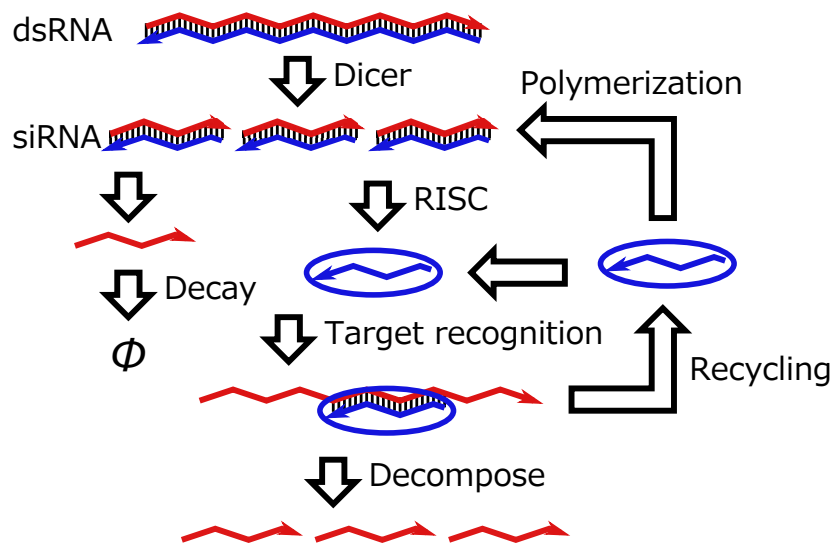
Figure 2.3: Phenomenological model of RNAi [255]. The representation of the Figure 1.2(b) are applied without the information of sequences. The target mRNA is decomposed when the dsRNA exists. DsRNA is first cleaved into siRNA by ribo-nuclease called Dicer. Then, the half of the siRNA is modified and turns into RISC which is denoted as a zigzag RNA surrounded by a circle. The RISC recognizes the target mRNA, by which mRNA is decomposed to short RNA strands. The RISC can be reused for another round of reaction. The model also includes a primer-independent polymerization so that RISC is possible to become siRNA again.

Such limitations can be avoided by simulating a model that is based on chemical reaction where all the molecular spices are treated independently. A simulation model of RNAi which distinguishes siRNAs depending on the their position was introduced [259]. In the model, however, the reactions were still phenomenological so that chemical structures of RNA and their transition were not taken into account.

Chapter 3

Analysis of Nucleic Acid Reaction System

3.1 Graph Rewriting Model of Nucleic Acids and Reaction

To predict the dynamic behavior of a nucleic acid reaction system, a flexible model to represent various structures of DNA and RNA is required. Here, I propose a novel model based on a graph data structure, which derives from the domain level representation explained in Section 1.2.2. The advantage of the model is the small restrictions on the secondary structures of nucleic acid, while conventional models were strictly restricted as described in Chapter 1. That kind of flexible representation is useful for not only the semi-automated simulation of nucleic acid reaction system but also the design of modules that take advantage of complex structure of nucleic acids.

The graph based model of DNA catalytic gate (Figure 1.2.2) is shown in Figure 3.1. I represent nucleic acids by a graph data structure, whose nodes, undirected edges, and directed edges represent domains, hydrogen bonds, and phosphate backbones, respectively. The proposed method is capable of representing various secondary structures, all of which cannot be represented by any one of the conventional design methods at the same time.

Basic secondary structures of nucleic acids include single strand, duplex, loop, and pseudoknot. A duplex is formed when two complementary domains are hybridized. A loop structure is a single-stranded region of nucleic acids extending from a hybridized structure, which can be categorized into hairpin loop, bulge loop, internal loop, and multi-branched loop. Pseudoknot is a geometrically entangled structure, which can be found in natural RNA structures.

The difference among domains is introduced by a parameter allocated to each domain. The parameter corresponds to the length of nucleotides sequence in a domain. For instance, 10, 24, 4, 16, 6, and 16 are allocated to the domains 'a', 'b', 'c', 'd', 'e', and 'f' in Figure 3.1(a), respectively, which is the same as the domain level representation of Figure 1.2. Those parameters are referred to during the simulation process explained in the next section.

(a) Catalytic gate   (b) Hairpin loop   (c) Bulge loop and pseudknot
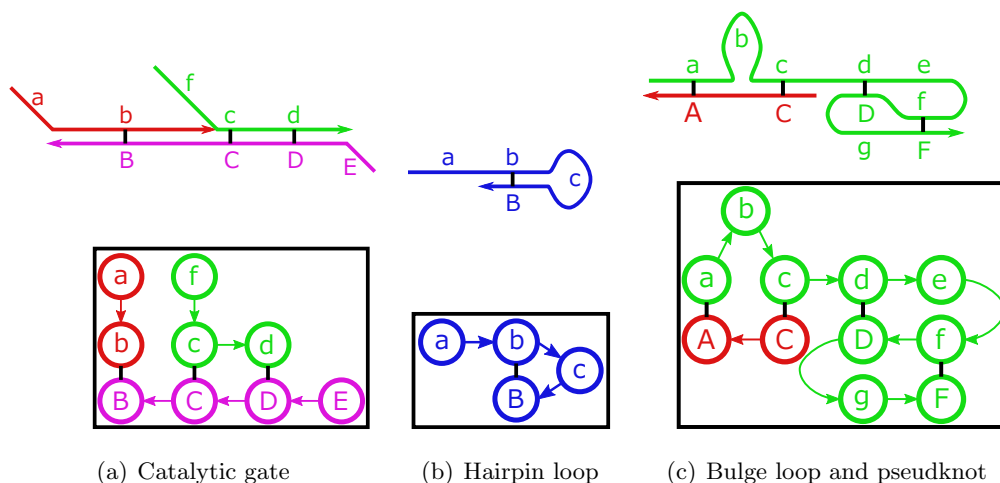
Figure 3.1: Graph based models of the (a) catalytic gate [14], (b) hairpin loop, (c) bulge loop, (c) and pseudoknot. Domain level and corresponding graph based representations are shown in the top and the bottom of the figure, respectively. Domain 'c' of (b), 'b' of (c), and from 'd' to 'F' of (c) are hairpin loop, bulge loop, and pseudoknot, respectively. Each connected graph is surrounded with a rectangle shape, since a connected graph corresponds to one structure.

A definition of chemical reaction is necessary to constructively predict the behavior of a nucleic acid reaction system represented by the graph based model. In particular, a chemical reaction is defined as a transformation of graph, or graph rewriting. A graph rewriting rule is in a form $L \to R$ that is similar to chemical equation, which has left-hand and right-hand sides. L and R are patterns of a graph corresponding to before and after the reaction of the rewriting rule, respectively. When the rewriting rule is applied to a graph, the subgraph that matches L is replaced by R.

In the example of catalytic gate explained in Section 1.2.3, three fundamental hybridization reactions are involved, which are hybridization, denaturation, branch migration (Figure 3.2). To apply the graph based model to RTRACS and RNAi, it is also necessary to define rewriting rules of enzymatic reactions. All elemental rewriting rules are summarized in the following Section 3.2.

3.2 Simulation based on Chemical Kinetics

3.2.1 Enumeration of Structure

The above graph rewriting system allows straight forward simulation by applying chemical kinetics and solving ODE. If the user determines initial configuration as a set of structures and their concentrations, the dynamic behavior of a system can

27
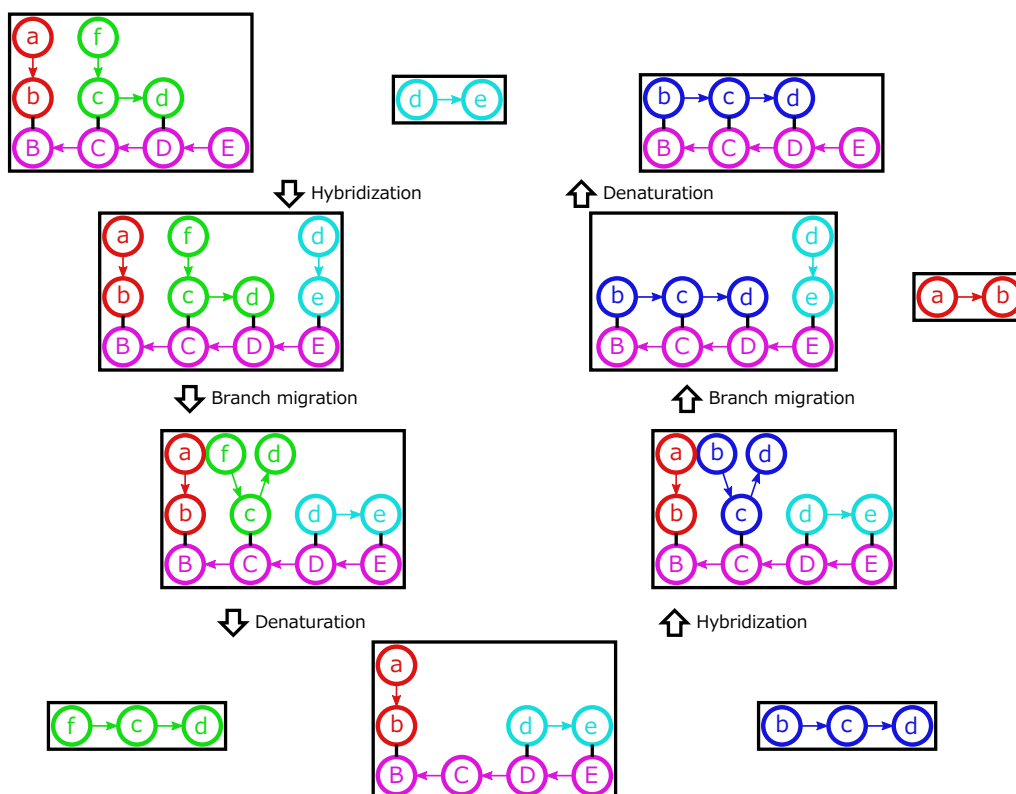
Figure 3.2: Transition of the catalytic gate [14] by graph rewriting. The figure is arranged in a similar manner to the Figure 1.3. Each reaction corresponds to a transition of graphs. The transition of branch migration reaction is applied twice for the reaction on the right to displace both 'b' and 'd'. As a result of graph transition by branch migration and denaturation, a graph can be separated into multiple connected graphs.

be automatically predicted. The simulator has two parts of calculations, which are to enumerate structures that are producible from initial structures, and to analyze the concentration changes by a numerical method.

At the beginning of a simulation, the simulator enumerates whole structures in a system to decide the number of variables. By iteratively applying rewriting rules from the initial set of structures, total number of structures is determined as shown in the example of Figure 3.3. During the enumeration, the simulator does not allow a structure that contains multiple identical strands such as dimer. This is because unbounded number of structures can be produced even from a small number of initial structures. As explained in Section 1.3.3, this kind of problem is a fundamental one in a simulation of bio-molecular systems, which I try to solve in Chapter 4 and 6.
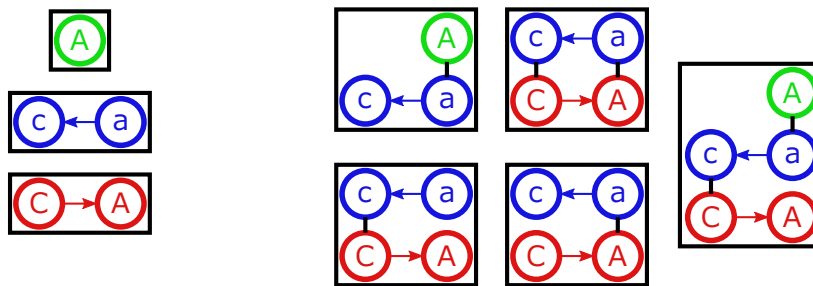
Figure 3.3: Enumeration of structures starting from three ssDNA. Three initial structures are shown in the left, while five possible structures by applying the rewriting rules are shown in the right.

3.2.2 Rewriting Rule
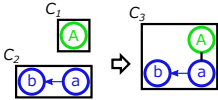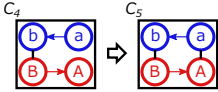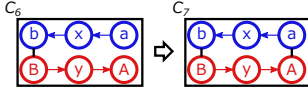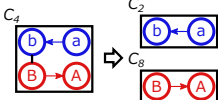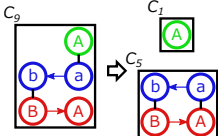
After the enumeration process, the simulator assigns variables to all molecular species in the system in order to represent their concentrations. By formalizing ODE of all reaction rules by applying chemical kinetics, they are numerically solved to obtain time evolution of the concentrations. Rate constants of reactions are defined on the order of magnitude, and the units of concentration and time are arbitrary. I actually employ a common numerical analysis called Runge-Kutta-Fehlberg-4,5 method to solve ODE [260].

Transitions by elemental rewriting rules of hybridization reactions are summarized in Table 3.1. I distinguish unimolecular hybridization reactions that extend hybridization region from those do not. Due to the directed edge of graph based model, exchange of hydrogen bond can occur in both directions by branch migration rule. In the table, only one of them is described. Although the rate constants of hybridization and branch migration are fixed, that of denaturation depends on the length of domains that are separating. If the length parameter of the domain is l, the kinetics k_d is calculated as $k_d = 8.0 \times (-2)^l$.

In the example of the catalytic gate (Figure 3.2), all possible reactions are not shown. For instance, denaturation reactions coincide with hybridization reactions in opposite direction. Branch migration reaction in the left of the figure is possible to happen in both directions. After listing ODE by applying rewriting rules, time-derivation term of each structure becomes a summation of all expressions. If both bimolecular hybridization and denaturation reactions in the Table 3.1 are applied, the term of C_2 is formalized as $\frac{d}{dt}C_2 = k_dC_4 - k_{h1}C_1C_2$.

Similarly, elemental rewriting rules of enzymatic reaction are summarized in Table 3.2. Nodes of RNA are represented by zigzag shapes, which is consistent with the domain level representation. Unlike the hybridization reactions, enzy-
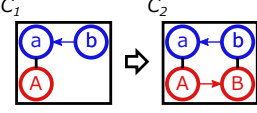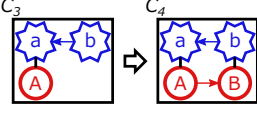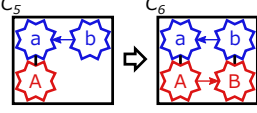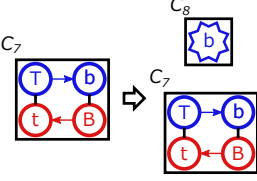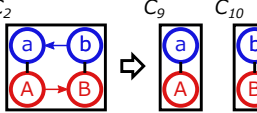
Table 3.1: Transition by elemental hybridization reactions. Elemental rewriting rules of hybridization reactions are summarized with typical examples. First column denotes the name of reaction. Second column corresponds to a typical example by applying the rewriting rule. Concentration of a structure is denoted by a variable written in the upper left of the rectangle surrounding the structure. Formulas of ODE are listed in the third column and default rate constants are listed in the fourth column.

Reaction name	Schematic example	Equations	Rate constant
Bimolecular hybridization		$\frac{d}{dt}C_1 = -k_{h1}C_1C_2$ $\frac{d}{dt}C_2 = -k_{h1}C_1C_2$ $\frac{d}{dt}C_3 = k_{h1}C_1C_2$	$k_{h1} = 0.1$
Unimolecular hybridization by extending		$\frac{d}{dt}C_4 = -k_{h2}C_4$ $\frac{d}{dt}C_5 = k_{h2}C_4$	$k_{h2} = 1.0$
Unimolecular hybridization with interval		$\frac{d}{dt}C_6 = -k_{h3}C_6$ $\frac{d}{dt}C_7 = k_{h3}C_6$	$k_{h3} = 0.1$
Denaturation		$\frac{d}{dt}C_4 = -k_dC_4$ $\frac{d}{dt}C_2 = k_dC_4$ $\frac{d}{dt}C_8 = k_dC_4$	k_d is defined as a function
Branch migration		$\frac{d}{dt}C_9 = -k_bC_9$ $\frac{d}{dt}C_1 = k_bC_9$ $\frac{d}{dt}C_5 = k_bC_9$	$k_b = 0.01$

matic reactions can operate the phosphate backbone of nucleic acids. Because of the restriction of the structures, any enzymatic reactions that produce a structure with more than two identical strands are neglected. As all the rate constants defined in this section are default values, they may be different in other sections.

Even with the restriction, combinatorial explosion of the number structures is inevitable. To ignore unimportant structures which may not be the main products of a system, I introduce a concentration threshold. Structures with the concentration lower than 10^{-5} are disregarded. Thanks to the threshold, the process of enumeration can be distributed among the duration of a simulation instead of computing all possible structures in the beginning. Similarly, stochastic simulation [28] can employ such strategy, since structures without concentration

Table 3.2: Transition by elemental enzymatic reactions. Meaning of columns are the same as in the previous Table 3.1. When the promoter domain 'T' becomes dsDNA, downstream of it is transcribed to an mRNA. By the decay reaction, all RNA nodes disappear from a graph. When the product by the decay reaction is an empty graph, it is denoted as Φ.

Reaction name	Schematic example	Equations	Rate constant
Primer-dependent DNA-directed DNA polymerization	$C_1 \Rightarrow C_2$	$\frac{d}{dt}C_1 = -k_{ddd}C_1$ $\frac{d}{dt}C_2 = k_{ddd}C_1$	$k_{ddd} = 0.005$
Primer-dependent RNA-directed DNA polymerization	$C_3 \Rightarrow C_4$	$\frac{d}{dt}C_3 = -k_{drd}C_3$ $\frac{d}{dt}C_4 = k_{drd}C_3$	$k_{drd} = 0.005$
Primer-dependent RNA-directed RNA polymerization	$C_5 \Rightarrow C_6$	$\frac{d}{dt}C_5 = -k_{drr}C_5$ $\frac{d}{dt}C_6 = k_{drr}C_5$	$k_{drr} = 0.005$
Primer-independent DNA-directed RNA polymerization	$C_7 \Rightarrow C_7, C_8$	$\frac{d}{dt}C_8 = k_{idr}C_7$	$k_{idr} = 0.005$
Dicer	$C_2 \Rightarrow C_9, C_{10}$	$\frac{d}{dt}C_2 = -k_{dicer}C_2$ $\frac{d}{dt}C_9 = k_{dicer}C_2$ $\frac{d}{dt}C_{10} = k_{dicer}C_2$	$k_{dicer} = 0.01$
Decay	$C_8 \Rightarrow \Phi$	$\frac{d}{dt}C_8 = -k_{decay}C_8$	$k_{decay} = 0.001$

never react to produce new structures.

I employed the strategy so that the rewriting rules are applied to produce new structures only when the enumeration process becomes necessary. To implement this feature, the simulator produces structures dynamically and checks whether the concentration of each structure exceeds the threshold. More concretely, the duration of a simulation is divided into intervals, and the simulator checks the concentration at the beginning of each interval.

3.3 Analysis of System

3.3.1 Simulation of Catalytic Gate

To verify that graph based model is capable of predicting the behavior of nucleic acid reaction systems correctly, the catalytic gate [14] and the AND gate of RTRACS [20] are chosen as benchmarks. The results of kinetic and stochastic simulations of the catalytic gate are shown in Figure 3.4. Most of the results exhibited catalytic activities by which the concentration of output increased higher than that of input. The results also had a good agreement with the experimental results described in the paper. It is possible to assure that both simulations were predicting the behavior of catalytic gate correctly. Note that the concentrations of continuous simulation was from 0.0 to 1.0, while amount of stochastic simulation was from 0 to 100.

The result of continuous simulation, in which 0.01 concentration of input was added, did not increase the concentration of output because the concentration of input was too small to exceed the threshold. In contrast, the amount of output did increase for stochastic simulation that had only 1 copy of input strand. As described in Section 1.3.1, I focus on continuous simulation in the rest of this thesis.

3.3.2 Simulation of RTRACS

To examine whether the graph based model of RNA and enzymatic reactions were appropriate, I simulated the AND gate of RTRACS [20]. The graph rewriting model was developed by applying the transition rules in a similar manner (Figure 3.5(a)). Note that Dicer and RNA polymerization reactions were not involved in RTRACS. As in the result (Figure 3.7), the continuous simulator worked as expected because the output was synthesized if and only if both RNA inputs were added.

Because the output has to be synthesized by transcription, the outputs were never observed for the false combinations of inputs where promoter region cannot form dsDNA. This result shows that the simulator can correctly predict the behavior of a nucleic acid reaction system driven by enzymatic reactions, and it may be possible to adapt the simulator for further applications.

3.3.3 Simulation of RNAi

By extending the expression of the graph based model, simulation of RNAi based on molecular species is possible. I develop an original RNAi model (Figure 3.6)
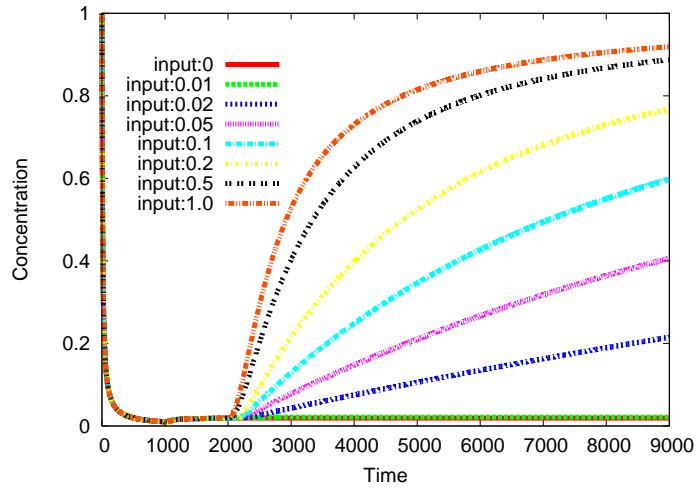
(a) Continuous simulation
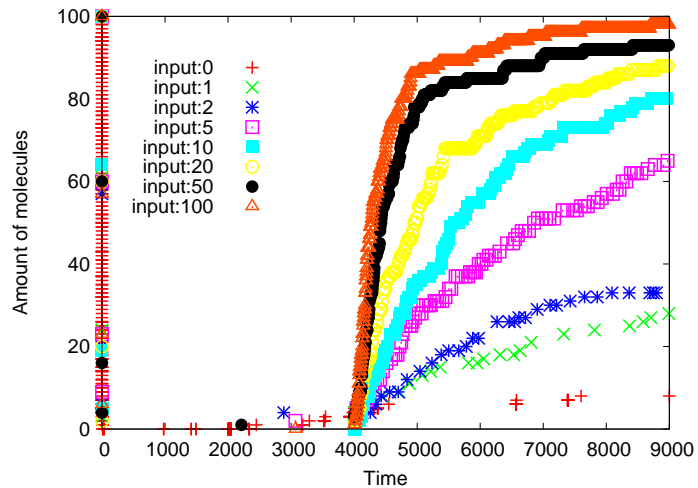
(b) Stochastic simulation

Figure 3.4: Continuous and stochastic simulations of the catalytic gate [14]. X and y axes correspond to time and concentration or amount of the output strand, respectively. First changes in first 2000 and 4000 time units were caused by making the gate in continuous and stochastic simulations, respectively. The graph legends indicate the concentration and amount of inputs.

by reference to the phenomenological models [259, 257, 258, 255]. Elemental reactions of hybridization, primer-dependent RNA polymerization, and Dicer are in common with the previous model, while denaturation, decay, and transcription (primer-independent RNA polymerization) reactions are modified. I further add RISC, decompose, aberrant, and primer-independent RNA-directed RNA polymerization to follow the reactions of RNAi.

(a) Graph rewriting model

(b) Continuous simulation
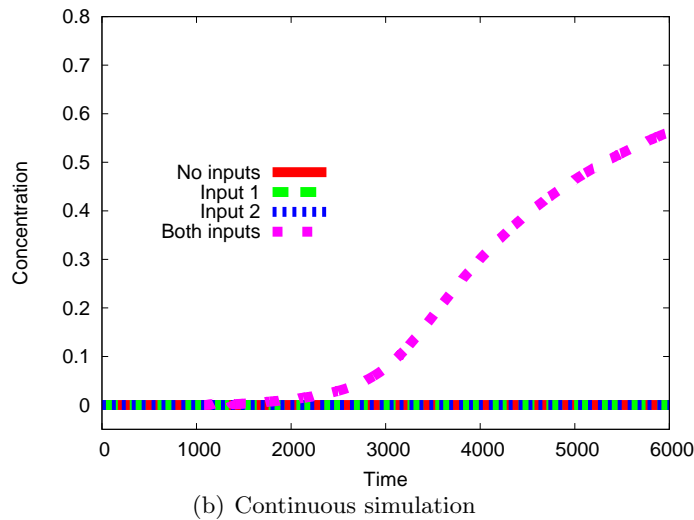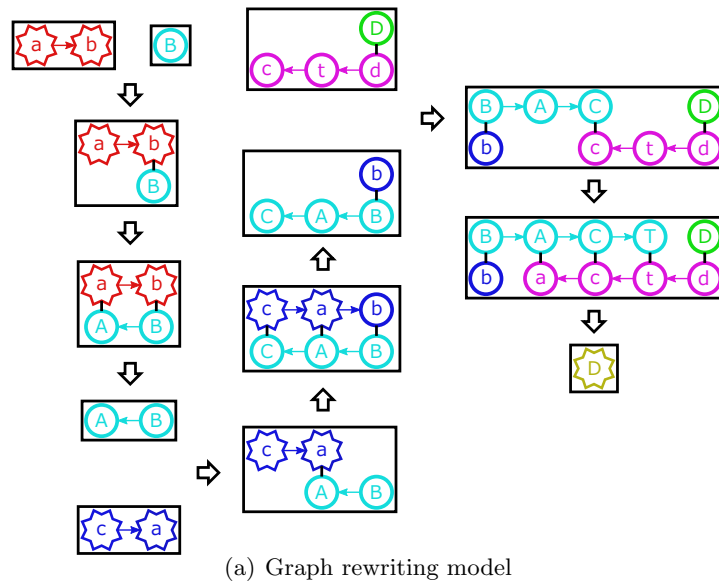
Figure 3.5: Graph rewriting model and continuous simulation of the AND gate of RTRACS [20]. (a) The graph rewriting model is shown, which is consistetn with the domain level representation of Figure 1.4. (b) X and y axes correspond to time and concentration of the output strand, respectively. The inputs were added in 1000 time units and the concentration of the output began to increase from about 2000 time units when both of the inputs were added. The graph legends indicate the combination of inputs.

Input of dsRNA initiates the cascades of reactions from the top of the figure. The dsRNA is first cleaved into siRNAs by a Dicer reaction, where the number of domains that are cleaved by the reaction is a variable of the model. Because siRNAs have a smaller number of hydrogen bonds, denaturation reactions spon-
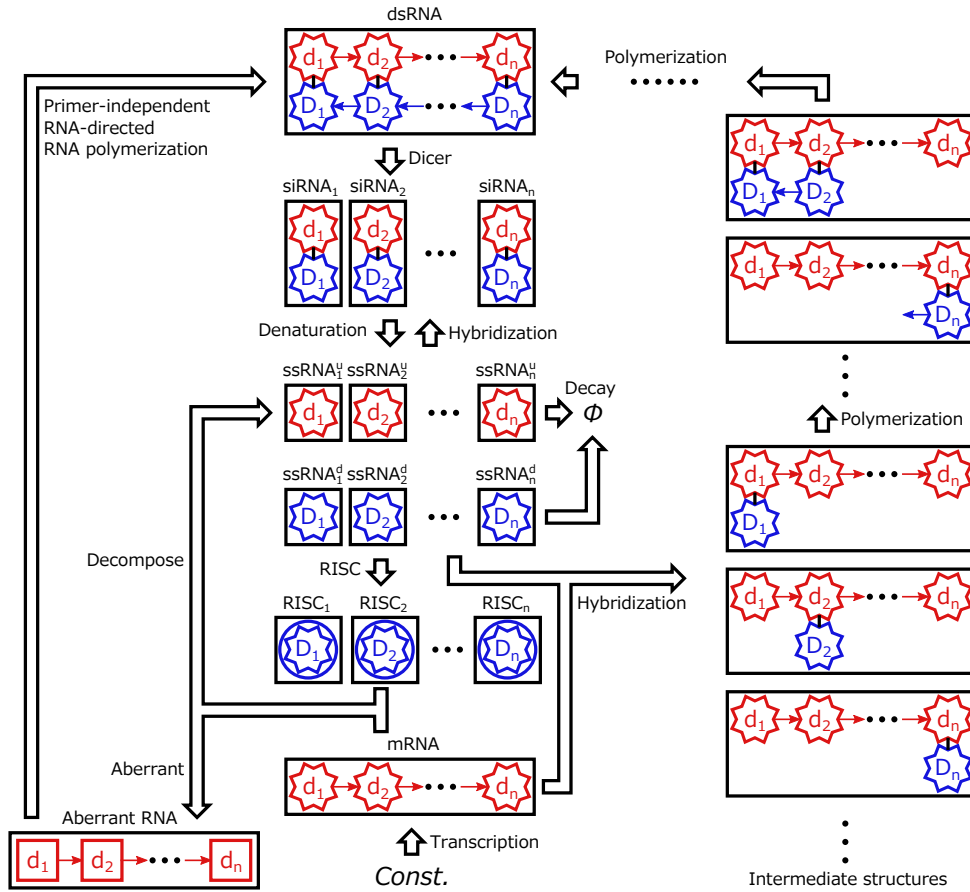
Figure 3.6: Transition of RNAi by graph rewriting. Name of each structure is indicated on the top of a rectangle that is surrounding the corresponding structure. n is a parameter of the system, which indicates the size of input dsRNA. In addition to normal RNA, the graph model has two special forms of nucleic acids, which are RISC and aberrant RNA. A node of RISC is represented by both circle and zigzag shape, while that of aberrant RNA is represented by a square. $Const.$ is a constant value that is introduced to constantly synthesize the target mRNA by transcription reaction.

taneously occur and split siRNAs to upper and lower ssRNAs. Inversely, upper and lower ssRNAs can hybridize together to compose siRNAs. I assumed that the denaturation is only applied to one domain RNA. In other word, denaturation is not applied to other hybridized structures such as dsRNA.

Unlike the previous definition of decay reaction where any appearances of RNA nodes disappear, only ssRNAs decay and become waste in the model. Transcription is activated continuously and generates the target mRNA constantly as in the bottom of the figure. Some of the remaining lower ssRNAs are recruited to form RISC, which in turn decompose mRNA by identifying the complementary

sequences.

Because lower ssRNAs and mRNA have complementary sequences, they hybridize together and produce partially double-helical structures (intermediate structures listed in the right bottom part of the figure). This structures can also denature because of the assumption that denaturation can occur only when the length of hydrogen bonds is one domain. After hybridization, primer-dependent polymerization reaction extends the lower part of the double-helical structures and synthesize longer (but still partial) double-helical structures. If the rightmost lower ssRNA is used as a primer, polymerization can reproduce the complete dsRNA which can initiate the whole reactions again.

There is also another pathway to reproduce complete dsRNA via aberrant RNA, which is observed in plant cells [21]. For the pathway, a target mRNA is aberranted by RISC, and the aberrant RNA is eventually duplicated to reproduce complete dsRNA by primer-independent RNA-directed RNA polymerization. Note that I distinguish all ssRNAs, siRNAs and RISCs based on the original position in the dsRNA. Thanks to the graph based model, all combination of partially double-helical structures are also distinguished.

After assigning initial concentrations of mRNA and dsRNA to be 1.0, continuous simulation was carried out. I carefully decided default rate constants as shown in Table 3.3 by checking the parameters of conventional studies [259, 257, 258, 255]. The rate constant of denaturation, decay, and primer-dependent polymerization are modified to 0.0078, 0.01, and 1.0, respectively.

To show the advantages of the model, I focus on the concentration changes of the complete dsRNA and the distribution of ssRNAs among their positions. By changing the rate constants of the Dicer and primer-independent polymerization, I successfully obtained three types of dynamic behaviors of dsRNA (Figure 3.7), where n was 3. The concentration of dsRNA increased or became saturated when the rate constant of primer-independent polymerization was non-zero, while it decreased when it was zero. The result indicates that the aberrant pathway is important to keep the concentration of dsRNA, which is hopefully explaining the mechanism of RNAi in plant cell [21].

The concentrations of ssRNAs are expected to have a distribution among position in the dsRNA where they gradually decreased from left to right because the polymerization reaction extends lower RNA only toward the left direction. The distribution of ssRNA was unclear by the simulation when n was 3. To predict the distribution, it is necessary to increase the parameter n. With large n, however, simulation became intractable due to the combinatorial explosion of intermediate structures, which is solved later in Chapter 6.

Table 3.3: Transition by elemental RNAi reactions. Meaning of columns are the same as in the previous Table 3.1

Reaction name	Schematic example	Equations	Rate constant
RISC		$\frac{d}{dt}C_1 = -k_{risc}C_1$ $\frac{d}{dt}C_2 = k_{risc}C_1$	$k_{risc} = 0.01$
Decompose		$\frac{d}{dt}C_3 = -k_{dec}C_2C_3$ $\frac{d}{dt}C_i^u = k_{dec}C_2C_3$ $(1 \le i \le n)$	$k_{dec} = 0.005$
Transcription		$\frac{d}{dt}C_3 = k_{irr}$	$k_{idr} = 0.05$
Aberrant		$\frac{d}{dt}C_4 = k_a C_2 C_3$	$k_a = 0.01$
Primer-independent RNA-directed RNA polymerization		$\frac{d}{dt}C_4 = -k_{irr}C_4$ $\frac{d}{dt}C_5 = k_{irr}C_4$	$k_{irr} = 0.01$
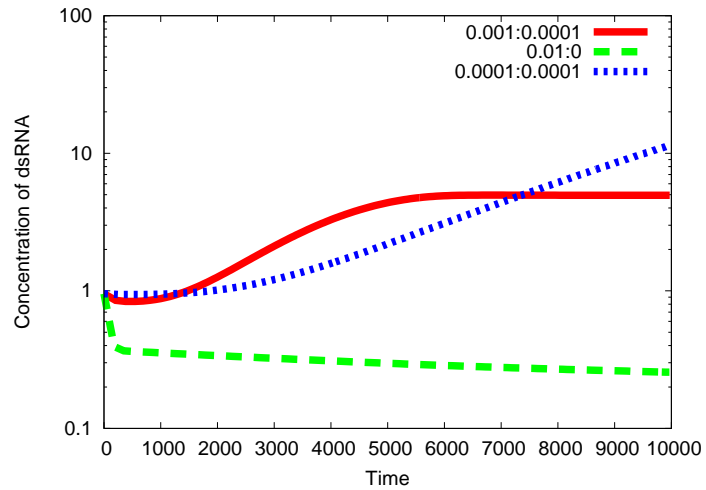
Figure 3.7: Continuous simulation of RNAi. X and y axes correspond to time and the concentration of dsRNA, respectively. First and second values in the legend are rate constants of Dicer and primer-independent polymerization, respectively.

Chapter 4

Abstraction of Hybridization Reaction

4.1 Simulation of HCR

4.1.1 Combinatorial Explosion of HCR

Combinatorially large number of structures can be produced in the straight-forward simulation employing the graph based model. I call the model in the previous Chapter 3 as *naïve* model, and a single graph of nucleic acids as *global* structure. In contrast, I propose a novel *abstract* model that focuses on the *local* structure of nucleic acids in this chapter, which appears in my paper [261].

Hybridization chain reaction (HCR) [33] is a typical example that generates an unbounded number of structures in the naïve model. HCR consists of two hairpin DNA strands and one initiator strand that starts the cascade of hybridization reactions (Figure 4.1(a)). The initiator first hybridizes to hairpin 1 and opens the hairpin by branch migration. After that, hybridization and branch migration occur alternately between hairpins 1 and 2. Because these reactions continue until the hairpins are thoroughly consumed, the number of global structures is essentially unbounded (Figure 4.1(b)).

4.1.2 Exact Abstraction of HCR

To avoid the unbounded number of structures, I introduce an abstraction technique. By the abstraction, variables are allocated to local structures (Figure 4.2), where the number of local structures is finite. A local structure contains information about the strand it belongs to and the information about how each domain is connected. If a domain is not connected to any other domains, its connectivity is defined as null. Otherwise its connectivity must identify both the strand and the position of complementary domain although this information is omitted in the figure.

I can develop an exact abstraction for the HCR under some appropriate assumptions as follows. In the *naïve* model, variables are allocated to all the struc-
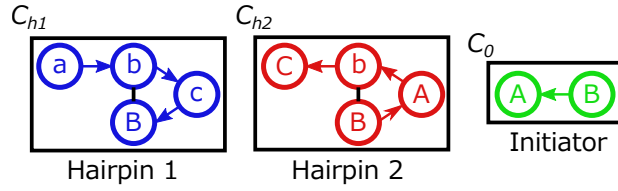
(a) Components of HCR
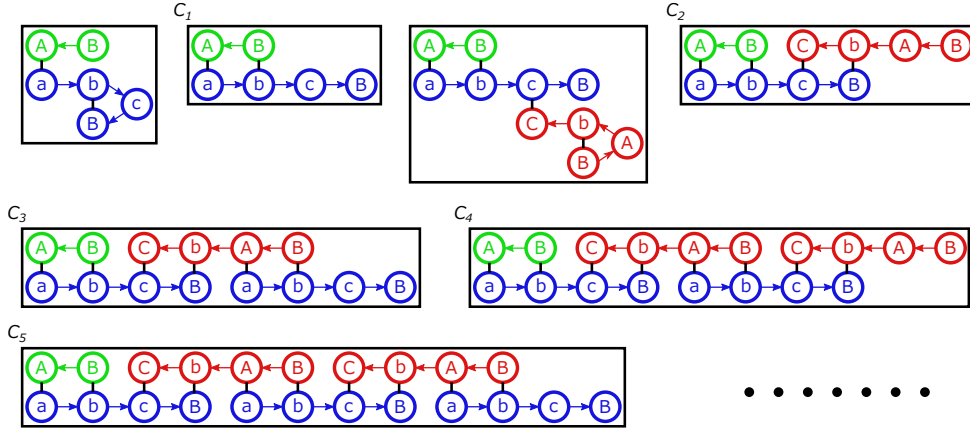
(b) Products of HCR

Figure 4.1: Components and products of HCR [33] represented by a naïve model. (a) HCR consists of two hairpin DNAs and one initiator DNA. Hairpin 1 and hairpin 2 cannot react each other without the addition of the initiator. (b) The figure lists the possible structures of the early stage of HCR. Alternative opening of hairpins continues as long as the components exist.

tures depending on the number of opened hairpins (Figure 4.1(b)). I allocate C_0, C_{h1}, C_{h2}, and C_i to the initiator strand, Hairpin 1, Hairpin 2, and a structure that has i opened hairpins, respectively. I ignore the denaturation reactions and also assume that hybridization and branch migration occur together.

ODE for the naïve model are formed as

$$\frac{d}{dt}C_0 = -k_b \times C_0 \times C_{h1}$$

$$\frac{d}{dt}C_{2n+1} = k_b \times C_{2n} \times C_{h1} - k_b \times C_{2n+1} \times C_{h2} \ \ (n \in \mathbb{N})$$

$$\frac{d}{dt}C_{2n+2} = k_b \times C_{2n+1} \times C_{h2} - k_b \times C_{2n+2} \times C_{h1} \ \ (n \in \mathbb{N})$$

$$\frac{d}{dt}C_{h1} = \sum_{i \in \mathbb{N}} -k_b \times C_{2i} \times C_{h1}$$

$$\frac{d}{dt}C_{h2} = \sum_{i \in \mathbb{N}} -k_b \times C_{2i+1} \times C_{h2},$$
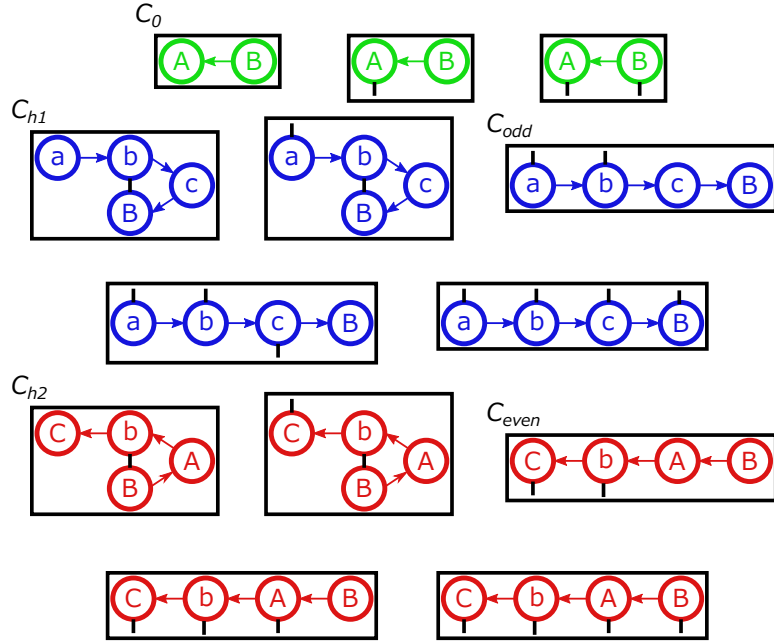
Figure 4.2: Local structures of HCR. Local structures are single strands with the annotation of their connectivity by hydrogen bonds. Though 13 local structures are shown in the figure to illustrate the concept of local structure, the actual number can be calculated in Section 4.2.2.

where $\mathbb{N}$ denotes the set of non-negative integers. Obviously, the number of variables is unbounded. By ignoring C_i for $i > 5$ to limit the number of global structures in the naïve model, I obtained the concentration changes from C_1 to C_5 shown in Figure 4.3(a). The initial values for C_0, C_{h1}, and C_{h2} were 0.01, 1.0, and 1.0, respectively (0 for other structures).

In contrast, exact abstract simulation using local structures does not require such cutoff. I regard all the opened hairpins that can bind to the next hairpin as one local structure, and allocate C_{odd} and C_{even} for hairpin 1 and hairpin 2, respectively (Figure 4.2). C_0, C_{h1}, and C_{h2} are allocated equally to the naïve model. ODE for these variables in the abstract model are formed as

$$\frac{d}{dt}C_0 = -k_b \times C_0 \times C_{h1}$$

$$\frac{d}{dt}C_{odd} = k_b \times C_0 \times C_{h1} + k_b \times C_{even} \times C_{h1} - k_b \times C_{odd} \times C_{h2}$$

$$\frac{d}{dt}C_{even} = k_b \times C_{odd} \times C_{h2} - k_b \times C_{even} \times C_{h1}$$

$$\frac{d}{dt}C_{h1} = -k_b \times C_0 \times C_{h1} - k_b \times C_{even} \times C_{h1}$$

$$\frac{d}{dt}C_{h2} = -k_b \times C_{odd} \times C_{h2}.$$
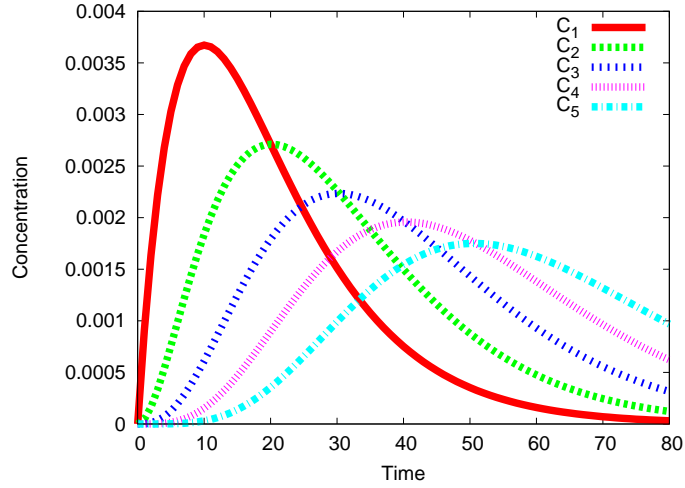
(a) Naïve simulation of HCR
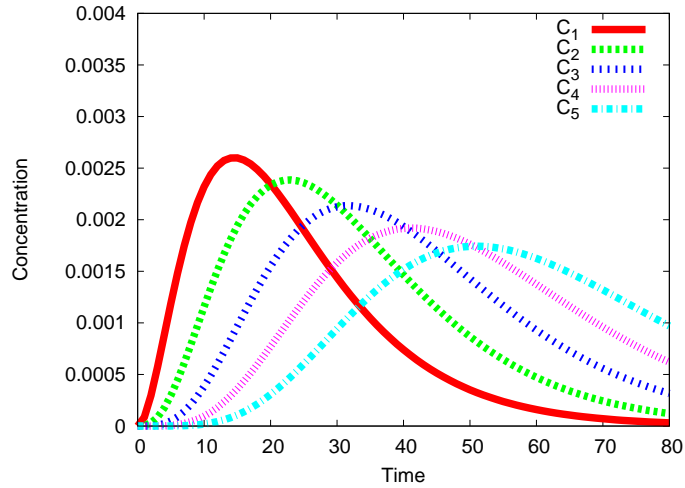
(b) Abstract simulation of HCR

Figure 4.3: Naïve and abstract simulations of HCR. X and y axes correspond to time and concentration of the global structures indicated in the graph legend, respectively. (a) The time evolution by the naïve simulation is shown. (b) Concentration of global structures recovered from abstract simulation is shown. In both simulation results, the peak shift of the concentrations is successfully predicted.

4.1.3 Recovering the Concentration of Global Structure

The inverse function to recover the concentration of global structures from local structures is not trivial. I try to recover each C_i by assuming Possion distribution

of the global structures:

$$C_i = (C_{\text{odd}} + C_{\text{even}}) \times \frac{\lambda^i \exp^{-\lambda}}{i!},$$

where λ is a function of time. After simulation of T time units, λ is calculated by

$$k_b \times \int_0^T (C_{\text{h1}} + C_{\text{h2}})dt,$$

which denotes the average number of the occurrences of the branch migration reaction. The simulation result by recovering the concentrations of global structures is shown in Figure 4.3(b). The recovered concentration was not exactly simulated because of the approximation under the assumption of Possion distribution.

4.2 General Abstraction of Hybridization Reaction System

4.2.1 Abstract Model

Though the exact abstraction of HCR was achieved by reasonable assumptions, the method cannot be applied to other reactions system such as the catalytic gate in Section 3.3.1. In this section, another abstract model for general hybridization reaction systems is defined formally. Note that the model utilizes only hybridization reactions so that none of the nucleic acids changes its phosphate backbone. Assume that an alphabet Σ and a set of single strands $S \subseteq \Sigma^*$ are given in advance, and the binary relation $X \subseteq \Sigma \otimes \Sigma$ is also defined to represent the complementary relationships of domains, where $\otimes$ represents a direct product of sets.

By distinguishing all domains of strands, I define the set of local domain $D \subseteq S \otimes \mathbb{N}$ as

$$D = \{(s, i) \mid s \in S, i \in \mathbb{N}, 1 \leq i \leq |s|\},$$

where $|s|$ denotes the length of s. I define a function $LETTER$, which is a map from D to Σ such that for any $d = (s, i) \in D$ and $s = a_1 \, a_2 \, a_3 \, \cdots$, $LETTER(d) = a_i$ holds. This function gives the corresponding letter of a given local domain.

As a consequence, the set of local structures $L \subseteq S \otimes (D \cup \epsilon)^*$ is defined as

$$L = \{(s, d_1 d_2 \cdots d_n) \mid s = a_1 \, a_2 \, \cdots \, a_n \in S, \text{ either } d_i = \epsilon \text{ or}$$
$$d_i \in D \text{ and } (a_i, LETTER(d_i)) \in X \text{ holds for all } 1 \leq i \leq n\}.$$

Note that I use ϵ as a symbol to represent unconnected domains, and sequence of ϵ is allowed in $d_1 d_2 \cdots$. Thus, $(s, d_1 d_2 \cdots d_n)$ corresponds to single stranded nucleic acid if $d_i = \epsilon$ holds for all $1 \leq i \leq n$.

For example, modeling an HCR by the graph based model (Figure 4.1) gives sets

$$\Sigma = \{\text{ 'a', 'A', 'b', 'B', 'c', 'C'}\}$$
$$S = \{\text{"abcB", "BAbC", "BA"}\}$$

and the relation

$$(\text{'a', 'A'}) \in X, (\text{'A', 'a'}) \in X, (\text{'b', 'B'}) \in X, \cdots.$$

Local domains and local structures are defined as

$$
\begin{aligned}
D = \{ & (\text{"abcB"}, 1), (\text{"abcB"}, 2), (\text{"abcB"}, 3), \\
& (\text{"abcB"}, 4), (\text{"BAbC"}, 1), (\text{"BAbC"}, 2), \cdots \} \\
L = \{ & (\text{"abcB"}, \epsilon\,\epsilon\,\epsilon\,\epsilon), (\text{"abcB"}, \epsilon\,\epsilon\,\epsilon\,(\text{"abcB"}, 2)), (\text{"abcB"}, \epsilon\,\epsilon\,\epsilon\,(\text{"BAbC"}, 3)), \\
& (\text{"abcB"}, \epsilon\,\epsilon\,(\text{"BAbC"}, 4)\,(\text{"abcB"}, 2)), \cdots \}.
\end{aligned}
$$

4.2.2 Enumeration of Local Structure

To evaluate the performance of the abstract model, the number of local structures is chosen as a measure to compare the efficiency of simulation. Enumeration of local structures for such purpose is performed by finding all possible $l \in L$. To enumerate the total number of local structures from a given alphabet and strands, I define functions $DOMAINS$, $COMPLEMENTS$, and $CONNECTIONS$.

First, $DOMAINS$ is a map from S to 2^D defined as

$$DOMAINS(s) = \{(s, 1), (s, 2), \cdots, (s, |s|)\},$$

which expresses all local domains in a given strand. Next, $COMPLEMENTS$ is a map from D to 2^D defined as

$$COMPLEMENTS(d) = \left\{ d' \mid (LETTER(d), LETTER(d')) \in X \right\}.$$

This finds all domains that are complementary to the given domain.

Then, $CONNECTIONS$ is a map from S to $\mathbb{N}$ defined as

$$CONNECTIONS(s) = \prod_{d \in DOMAINS(s)} (|COMPLEMENTS(d)| + 1),$$

where $|COMPLEMENTS(d)|$ denotes the cardinality of set $COMPLEMENTS(d)$. This calculates the number of all combinations of connections from a given strand. Finally, the total number of local structures is calculated by the following expression

$$\sum_{s \in S} CONNECTIONS(s).$$

43

By adopting the function, the numbers of local structures of HCR (Figure 4.1(a)) and catalytic gate reaction (Figure 3.2) are calculated to be 126 and 94, respectively.

4.2.3 Enumeration Efficiency

Using the enumeration function, the efficiencies of the naïve and abstract simulations were compared in terms of the number of structures. As a benchmark, I generated a random system as a random sequence of letters, which determines the set S. Note that the size of set Σ was fixed to 14 (letters from 'a' to 'g' and letters from 'A' to 'G'). I first fixed a maximum size of S, and then I generated and simulated 200 random systems to obtain the average and maximum numbers of structures. After that, I took another maximum size in turn and repeated the calculation for each maximum size. I tried 21 different sizes (whose values were suitable for simulations), which range over the x axis of Figure 4.4.

In the naïve simulations with or without threshold, the restriction of structure and the just-in-time enumeration were imposed as mentioned in Section 3.2.2. Without such restriction, the simulation of a system with large number of domains immediately becomes intractable.

As expected, the naïve simulation without threshold exhibited faster combinatorial explosion than the others because entire combinations of structures were tested by the execution. The increase in the number of structures in the abstract model seemed to be slower than that of the naïve simulation without threshold because of the limit on the number of local structures. These results indicate that an appropriate model and simulation are necessary for the efficient enumeration and simulation of hybridization reaction systems. Note that the results of naïve simulation with threshold and stochastic simulation are only shown as references in the figures. Direct comparison of the results is not fair because the values for the naïve simulation with threshold and the stochastic simulation depend on parameters such as threshold and copy number.

4.2.4 Reactions among Local Structure

Continuous simulation based on the abstract model is possible for the hybridization reaction system. This simulator analyzes the concentration changes of local structures by predicting three hybridization reactions, which is similar to the method explained in the previous Chapter 3.3. Although the same strategy is employed for formalizing ODE by chemical kinetics, the reaction rules of abstract model are different. Both left-hand and right-hand sides of rewriting rules consist
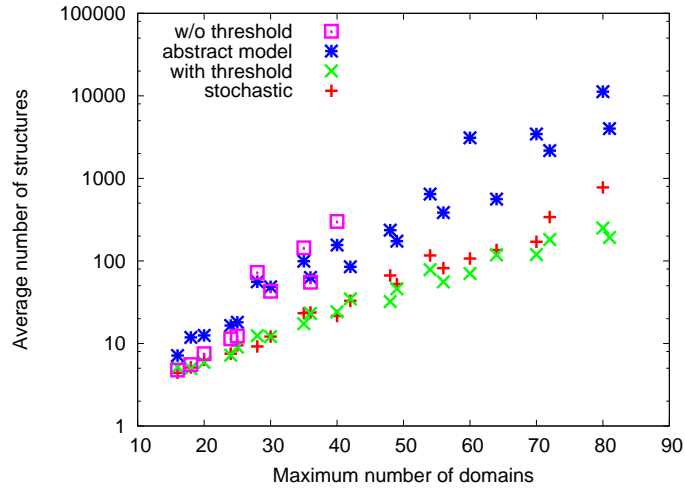
(a) Average number of structures
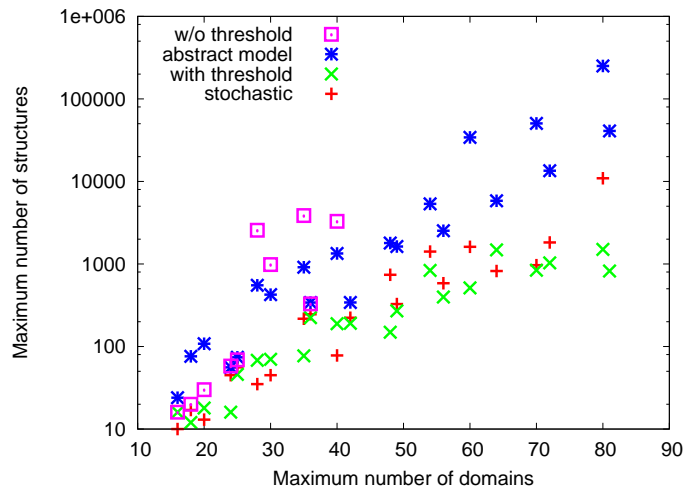
(b) Maximum number of structures

Figure 4.4: (a) Average and (b) maximum numbers of structures. The results were calculated from naïve simulation with and without threshold, enumeration of abstract model, and stochastic simulation. X axis corresponds to the maximum number of local domains, which is the number of letters in S to generate a random system. Y axis corresponds to the number of structures of a method indicated in the legend. Completion of the naïve simulation with a size greater than 40 was impossible due to an out-of-memory error.

of local structures. Since all the local structures are enumerated in advance of simulation without rewriting rules, it is correct to interpret the reaction rules as patterns to form ODE rather than rewriting rules to produce new structures.

More formally, matching to the pattern of hybridization reaction can be ex-

plained by local structures $l_1, l_2 \in L$, which are denoted as

$$l_1 = (s^1, d_1^1 d_2^1 \cdots d_m^1 \cdots d_{|s^1|}^1)$$
$$l_2 = (s^2, d_1^2 d_2^2 \cdots d_n^2 \cdots d_{|s^2|}^2),$$

where $s^1, s^2 \in S$, $d_i^1, d_j^2 \in D$ $(1 \le i \le |s^1|)$ $(1 \le j \le |s^2|)$ hold. When $d_m^1 = \epsilon$, $d_n^2 = \epsilon$, and $(LETTER((s^1, m)), LETTER(s^2, n)) \in X$ are satisfied, it is possible to apply hybridization. Note that the satisfied condition means the m-th and n-th local domains of l_1 and l_2 are unconnected and complementary to each other. In the local structures after the reaction, d_m^1 and d_n^2, both of which are ϵ, are replaced by (s^2, n) and (s^1, m), respectively.

Denaturation is interpreted as an inverse reaction of the hybridization. The pattern of branch migration is also interpreted in a similar manner, while the reaction involves three local structures. When l_1 is equal to l_2 in the hybridization reaction, it is naturally interpreted as unimolecular reaction. Even in such case, however, ODE are formed as bimolecular reaction. This means that hybridization to itself and hybridization to other molecular species with the same structure are not distinguished. The same situation can happen to denaturation and branch migration reactions.

Table 4.1 shows typical transitions by the reactions using schematic examples. Since the information of global structure is not included in local structure, hybridization is only defined as bimolecular reaction. In contrast, denaturation and branch migration are no more defined as unimolecular reaction, because only local structures are considered. Though the same rate constants are utilized for the reactions, function R is introduced to emulate unimolecular reaction. Function R calculates the ratio of concentration among all possible connections from reacting domains.

To calculate R_2 and R_3, I define a function $CONNECTED$ that is a map from D to 2^L. Suppose that C_l denotes the concentration of local structure l, $CONNECTED$ is

$$CONNECTED(d) = \{l \mid d^* \in D^*, l = (s, d^*), d \text{ appears in } d^*\}.$$

This $CONNECTED$ finds entire local structures that are connected to the given local domain. Function R_2 for denaturation, where domains d_1 of local structure l_1 and d_2 of l_2 separate, is defined as

$$R_2(C_{l_1} C_{l_2}) = \frac{C_{l_1} C_{l_2}}{\sum\limits_{l \in CONNECTED(d_1)} C_l},$$

Table 4.1: Transition by elemental hybridization reactions of abstract model. The column is the same as in Table 3.1. Rate constants and the function for k_d are the same to those of naïve simulation although unimolecular hybridizations are neglected in the abstract model.

Reaction name	Schematic example	Equations	Rate constant
Hybridization		$\frac{d}{dt}C_1 = -k_{h1}C_1C_2$ $\frac{d}{dt}C_2 = -k_{h1}C_1C_2$ $\frac{d}{dt}C_3 = k_{h1}C_1C_2$ $\frac{d}{dt}C_4 = k_{h1}C_1C_2$	$k_{h1} = 0.1$
Denaturation		$\frac{d}{dt}C_5 = -k_dR_2(C_5C_6)$ $\frac{d}{dt}C_6 = -k_dR_2(C_5C_6)$ $\frac{d}{dt}C_2 = k_dR_2(C_5C_6)$ $\frac{d}{dt}C_7 = k_dR_2(C_5C_6)$	k_d is defined as a function
Branch migration		$\frac{d}{dt}C_8 = -k_bR_3(C_6C_8C_9)$ $\frac{d}{dt}C_9 = -k_bR_3(C_6C_8C_9)$ $\frac{d}{dt}C_6 = -k_bR_3(C_6C_8C_9)$ $\frac{d}{dt}C_1 = k_bR_3(C_6C_8C_9)$ $\frac{d}{dt}C_{10} = k_bR_3(C_6C_8C_9)$ $\frac{d}{dt}C_{11} = k_bR_3(C_6C_8C_9)$	$k_b = 0.01$

which is equivalent to

$$R_2(C_{l_1}C_{l_2}) = \frac{C_{l_1}C_{l_2}}{\sum\limits_{l\in CONNECTED(d_2)} C_l}.$$

The equivalence is due to the fact that reactions of specific complementary pairing always occur in both ends of the connection. Formally speaking, equation $\sum\limits_{l\in CONNECTED(d_1)} C_l = \sum\limits_{l\in CONNECTED(d_2)} C_l$ is always satisfied.

Function R_3 for branch migration was defined similarly. Suppose a reaction where connection between d_1 of l_1 and d_2 of l_2 is displaced with connection between d_3 of l_3 and d_2 of l_2 while connection between d_4 of l_2 and d_5 of l_3 remain. In this reaction, the ratio was defined as

$$R_3(C_{l_1}C_{l_2}C_{l_3}) = \frac{C_{l_1}C_{l_2}C_{l_3}}{\sum\limits_{l\in CONNECTED(d_1)} C_l \times \sum\limits_{l\in CONNECTED(d_5)} C_l}.$$

Similar to denaturation, there are three other equivalent expressions.

47
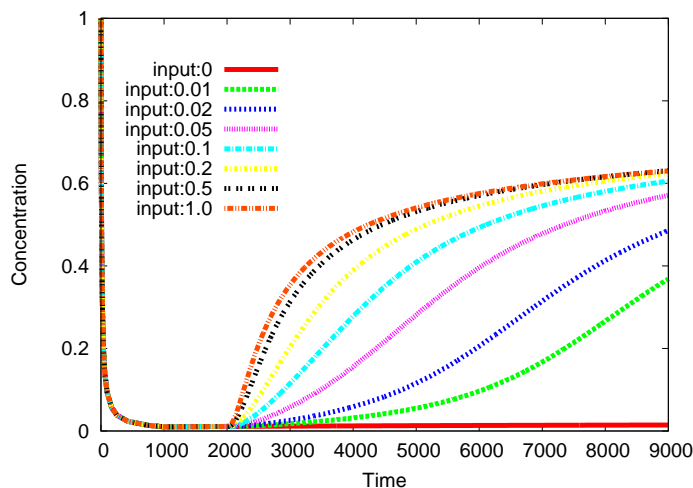
Figure 4.5: Abstract simulation of the catalytic gate [14]. Axes and legends are equivalent to those of Figure 3.4(a)

As can be expected from the expressions of R_2 and R_3, the distribution of hybridized states between complementary domains is assumed to be proportional to the concentrations of both local domains. I call this assumption *ratio assumption*. By these functions, the multi-molecular reactions can be treated as unimolecular reactions. On the basis of the definition, it is possible to simulate the concentration changes by solving the ODE by a numerical analysis. For a technical reason, I coupled hybridizations and denaturation that multiple domains can bind or unbind in a row.

4.2.5 Abstract Simulation of Catalytic Gate

To verify that ability of abstract simulation, I performed the simulation of the catalytic gate [14]. The result of the simulation is shown in Figure 4.5. Unlike the result of naïve simulation, the concentration of the output did increase in abstract simulations when the concentration of input was 0.0.1. The result indicates that all the possible structures are implicitly simulated by the abstract model.

The value to which the concentrations of the output converges was smaller than that of naïve simulation. The result indicates that the concentration of the output is precisely simulated compared with that of naïve model. Remaining concentration of the output is considered to be distributed among entangled structures that are not representable in the naïve model. Further discussion for the exactness of the abstraction appears in the next chapter.

Applying the abstract method directly to a system that utilizes enzymatic reactions is not possible since the size of the set of single strands is variable. In

48

other words, to apply the abstract method, the number of single strands that appear in a system has to be a constant value. Enzymatic reactions that change the phosphate backbone of the nucleic acids violate the requirement. Another abstraction technique for the simulation of RNAi is explained in Chapter 6.

Chapter 5

Exactness and Approximation of Abstract Simulation

5.1 Exactness of Abstraction

The concept of exact and approximate simulations is formally and briefly discussed in this chapter. For example, the abstract model for general hybridization reaction explained in Section 4.2 gives an approximate simulation. In other word, temporal evolution of concentration by the abstract model is not exactly simulating the same phenomena compared with naïve simulation.

The difference between exact and approximate abstractions is explained formally by the following diagram.

$$
\begin{array}{ccc}
N & \xrightarrow{\ R\ } & N \\
\downarrow \alpha & & \downarrow \alpha \\
A & \xrightarrow{\ Q\ } & A
\end{array}
$$

N and A represent the set of naïve models and abstract models, respectively, both of which are sets of maps from molecular species to their concentration. To be more formal, N is the set of maps from G to $\mathbb{R}$, where G denotes the set of global structures and $\mathbb{R}$ the set of non-negative real numbers, while A is the set of maps from L to $\mathbb{R}$, where L denotes the set of local structures.

R and Q denote temporal evolutions by ODE of naïve and abstract models, respectively. I prepare a function $c(g, l)$ for $g \in G$ and $l \in L$ which denotes the number of occurrences of local structure l in global structure g. The abstraction α, which is a map from N to A, is then defined as $\alpha(n)(l) = \sum_{g \in G} c(g, l) n(g)$, where $n \in N$ and $l \in L$.

The abstraction α is called exact if and only if $\alpha(R(n)) = Q(\alpha(n))$. Otherwise, the abstraction is called approximate. Since temporal evolution of local structures precisely corresponds to that of global structures by exact abstraction, it is also possible to say that exact α implies Q from N and R. On the other hand, approximate abstraction is rather defined by case-by-case manner in accordance with the efficiency of simulation and the locality of chemical reactions.

In the case of HCR, where G and L correspond to structures explained in Section 4.1, α was exact. The concentrations of local structures are exactly calculated because of the definition of C_{odd} and C_{even}. Since the local structures with C_{odd} and C_{even} appear only once at the right most position of each global structure in the naïve model (Figure 4.1(b)), they satisfy the following expressions:

$$C_{\text{odd}} = \sum_{i \in \mathbb{N}} C_{2i+1}$$

$$C_{\text{even}} = \sum_{i \in \mathbb{N}} C_{2i+2},$$

which can be regarded as a function to exactly calculate the concentration of local structures from global structures.

As already explained in the begging of the section, the abstract model for general hybridization reaction was approximate mainly due to the violation of the ratio assumption. Other reasons to derive approximate abstraction are the restriction of graph in naïve model and the definition of hybridization in abstract model. If all of these factors were avoided by increasing the size of local structures, it could naturally define an exact abstraction for general hybridization reaction, which is similar to the study of fragmentation. By increasing the size, however, the combinatorial explosion can happen in the number of local structures, which leads to an inefficient simulation. Eventually, local structures become equivalent to global structures as far as I have tried to define an exact abstraction for general hybridization reaction.

5.2 Exactness of Recovered Concentration

Recovering the concentration of global structure from the distributed concentrations of local structures can be further discussed by introducing an inverse function of abstraction. The diagram in previous section can be extended as follow.

$$
\begin{array}{ccc}
N & \xrightarrow{\ R\ } & N \\
\beta\uparrow\ \downarrow\alpha & & \beta\uparrow\ \downarrow\alpha \\
A & \xrightarrow{\ Q\ } & A
\end{array}
$$

The recovery function β is a map from A to N, which satisfies $\alpha(\beta(a)) = a$ where $a \in A$. The function β is called exact if and only if $\beta(Q(\alpha(n))) = R(n)$. Otherwise, the function is called approximate. Note that exact β is only defined when α is exact because it is not possible to simulate the same systems by approximate abstraction.

In the case of HCR, the recovery function β is explained in Section 4.1.3. The function was approximate because of the assumption of Possion distribution. As far as I have considered (such as applying normal distribution and binomial distribution), it is not possible to obtain an exact β because the information of global structures is lost in the local structures.

For the general hybridization reaction, recovery function is always approximate for any kind of β due to the approximate abstraction. One of the possible inverse functions to recover the concentration of global structure from that of local structures can be defined by adopting proportional assumption on all hydrogen bond connections. For example, the concentration of a global structure with three local structures is recovered by R_3 for the branch migration reaction in table 4.1.

In the next chapter, the diagrams and the discussions for the exactness of abstraction and recovery function are utilized from Section 6.2.2.

Chapter 6

Abstraction of Enzymatic Reaction

6.1 Efficient and Approximate Abstraction of RNAi

6.1.1 Assumption on Polymerization and Combinatorial Explosion

In this chapter, I apply optimized version of the abstraction technique to RNAi that involves enzymatic reactions. Although the viewpoint to focus on the locality of a structure is the same, the optimized abstract model is different from that of Chapter 4. Note that the model is explained in the my paper [262], while the aspect of graph representation of RNA is much emphasized in this chapter.

The mechanism and graph based model of RNAi are already explained in Section 2.3.3 and 3.3.3. For simplicity, I ignore RISC, decompose, aberrant, and primer-independent RNA-directed RNA polymerization for the rest of this thesis. Simulation of the expected distribution of ssRNAs [256, 23] adopting large number to the parameter n can be addressed by the abstract model.

When simulating the dynamic behavior of RNAi, assumptions on the polymerization reaction are important factor because they have a strong influence on the number of molecular species. In the model of RNAi, I assume that polymerization reaction extends the primer strand for only one domain though the reaction takes place repeatedly. I also assume that polymerization reaction has neither exonclease nor strand displacement activities at this moment. These assumptions lead to an exponential number of partially double helical intermediate structures as illustrated in Figure 6.1. Because for this exponential increase of the number of structures, the parameter n was limited to 9 under the naïve simulation.

6.1.2 Local and Refined Abstraction

To avoid the exponential increase of the number of structures, I propose an efficient abstraction that is based on the locality of RNA structures. I divide global structures of the naïve model into small local RNA structures according to their connectivity to their neighbor structures. In other word, each global
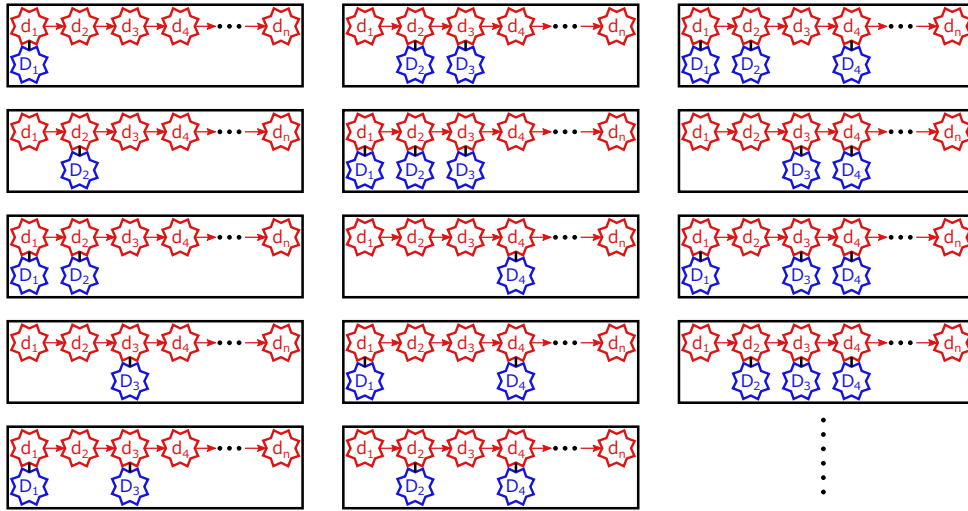
Figure 6.1: Combinatorial explosion of global RNA structures in RNAi. As the parameter n increases, the number of intermediate structures, which are partially hybridized RNA, can glow exponentially. The structures listed in the figure further produce other intermediate structures by primier-dependent polymerization.

structure in the naïve model is represented by a set of local RNA structures. A local RNA structure contains the information about the position of its domain, the presence of a double helix, and the connections with neighbor local RNA structures. This local RNA structure is formally modeled as a subgraph with annotation that represents such connectivities.

If I list all the possible local structures, there are 26 types for each position (except for the leftmost and rightmost positions) as in the Figure 6.2(a). I index the position of a domain from the leftmost position of the dsRNA. The presence of a double helix denotes whether a local structure has complementary RNA hybridized to its position or not. The connections with neighbor structures restrict the local structures in the previous or next position, because information about the presence of a double helix and the presence of nick of lower RNA are included. For example, the global structure in the top Figure 6.2(b) is represented by four local structures at the bottom of the figure.

In order to simulate the concentration of dsRNA accurately, I further refine the abstract model. Some of the local structures are divided into two structures each (Figure 6.3), depending on the position of the primer of polymerization reaction. If the double helix of a local structure is extended from the rightmost primer, the local structure is distinguished from the others. This refined abstraction has 29 types of local structures for each position with the exception of the leftmost and rightmost positions that have 9 and 8 local structures, respectively.
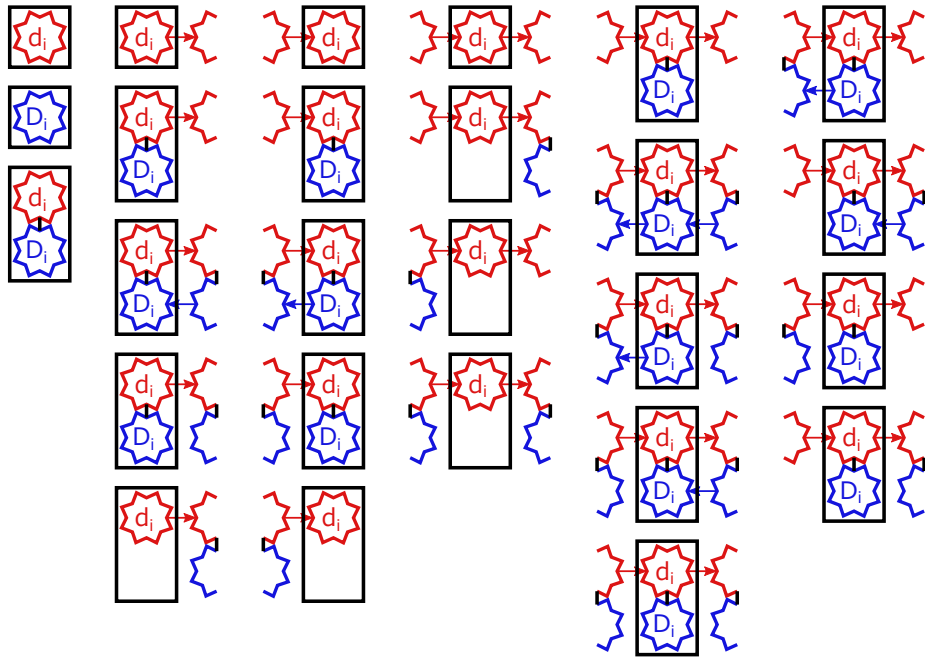
(a) Local RNA structures of RNAi
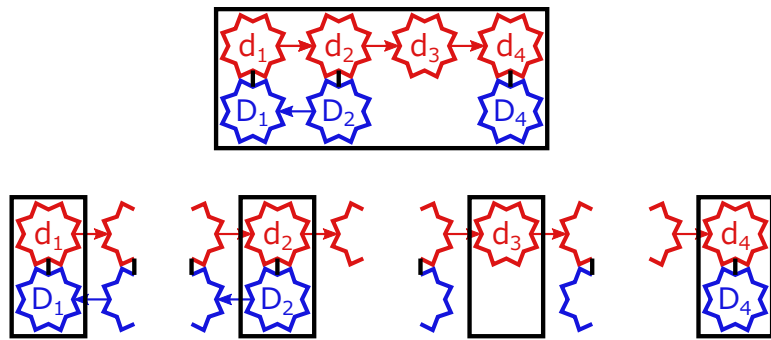
(b) Example of the abstraction of RNAi

Figure 6.2: Local RNA structures and an example of abstraction of RNAi. (a) All 26 local structures of RNAi where the index is i are shown. When the index is either 1 or n, which indicate the leftmost or rightmost position of the RNA, only 8 local structures are defined due to the lack of neighboring structures. A local structure is surrounded by a rectangle, while the annotation about its connectivity is represented outside of the rectangle using edges and incomplete nodes. (b) Global structure in the top, where n is equal to 4, can be represented as a set of four local structures in the bottom.

By this refined abstract model, I can distinguish the local structures that are parts of the complete dsRNA.

(a) Three refined local structures

(b) Refined local structure in the leftmost position

(c) Example of refined abstraction
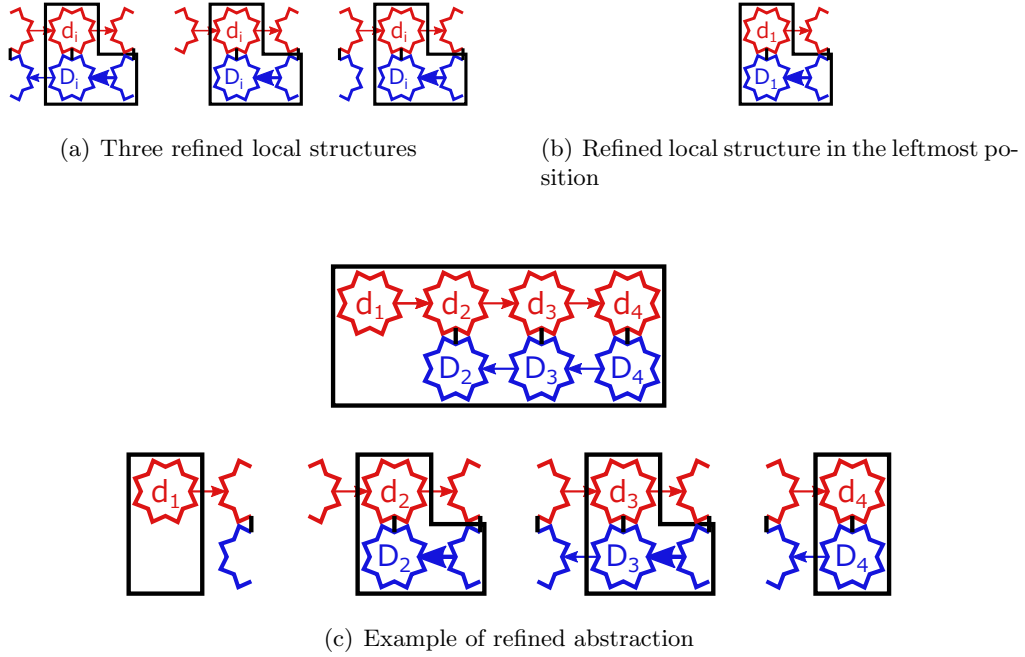
Figure 6.3: Refined local structures that can be a part of complete dsRNA. (a) Three local structures are split from the corresponding local structures. To distinguish the refined local structures, surrounding rectangle and direct edge in the right bottom are different from the original model. (b) When the index is 1, only one refined local structure is defined. The structure is equivalent to a complete dsRNA. (c) Global structure that can produce complete dsRNA in the top, where n is equal to 4, is represented as a set of four local structures including refined ones in the bottom.

6.1.3 Local and Refined Reaction

As global structures are abstracted based on local structures, all reaction rules are also abstracted. Table 6.1 and 6.2 summarize reaction rules for abstract RNAi model. Polymerization of the refined abstract model is carefully defined to preserve the structure of complete dsRNA.
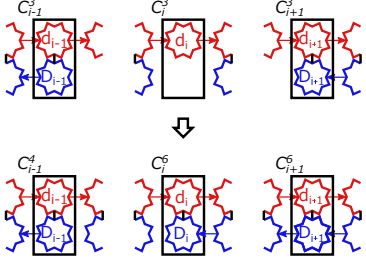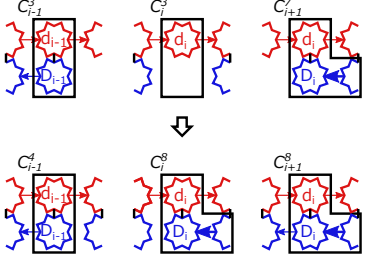
Similar to the assumption of Section 4.2.4, I assume the proportional connections between local structures to imitate a unimolecular reactions of the naïve model. Though the abstract model is different from hybridization reaction system, this assumption is the same as the ratio assumption. The concept can be explained by the example of denaturation reaction using the Figure 6.4. In the figure, some of the possible neighbor local structures of $(4, i-1)$ and $(4, i)$ are listed. For convenience, I denote a local structure with the variable C_i^x just as local structure (x, i).

The ratio of each connection is calculated by dividing the concentration of

Table 6.1: Transition by elemental reactions except for primer-dependent polymerization of abstract model for RNAi. The column is the same as in Table 3.1. Rate constants are omitted since all of them are exactly the same to those of naïve model in Section 3.3.3. Time variations of several reactions are expressed by Δ.

Reaction name	Schematic example	Equations
Dicer		$\Delta = k_{dicer} R_2(C_i^1 C_{i+1}^1)$ $\frac{d}{dt} C_i^1 = -\Delta$ $\frac{d}{dt} C_{i+1}^1 = -\Delta$ $\frac{d}{dt} C_i^2 = \Delta$ $\frac{d}{dt} C_{i+1}^2 = \Delta$
Hybridization		$\Delta = k_h R_3(C_{i-1}^3 C_i^3 C_{i+1}^3)C_i^d$ $\frac{d}{dt} C_{i-1}^3 = -\Delta$ $\frac{d}{dt} C_i^3 = -\Delta$ $\frac{d}{dt} C_{i+1}^3 = -\Delta$ $\frac{d}{dt} C_i^d = -\Delta$ $\frac{d}{dt} C_{i-1}^4 = \Delta$ $\frac{d}{dt} C_i^4 = \Delta$ $\frac{d}{dt} C_{i+1}^4 = \Delta$
Denaturation		$\Delta = k_d R_3(C_{i-1}^4 C_i^4 C_{i+1}^4)$ $\frac{d}{dt} C_{i-1}^4 = -\Delta$ $\frac{d}{dt} C_i^4 = -\Delta$ $\frac{d}{dt} C_{i+1}^4 = -\Delta$ $\frac{d}{dt} C_{i-1}^3 = \Delta$ $\frac{d}{dt} C_i^3 = \Delta$ $\frac{d}{dt} C_{i+1}^3 = \Delta$ $\frac{d}{dt} C_i^d = \Delta$
Decay		$\frac{d}{dt} C_i^u = -k_{decay} C_i^u$
Transcription		$\frac{d}{dt} C_i^5 = k_{idr}$ $(1 \le i \le n)$

Table 6.2: Transition by primer-dependent polymerization of abstract model for RNAi. The column is the same as in Table 3.1. Rate constants are omitted since it is exactly the same to those of naïve model in Section 3.3.3. Local structure of the refined model can propagate from right to left by only primer-dependent polymerization.

Reaction name	Schematic example	Equations
Primer-dependent RNA-directed RNA polymerization		$\Delta = k_{drr} R_3(C_{i-1}^3 C_i^3 C_{i+1}^3)$ $\frac{d}{dt}C_{i-1}^3 = -\Delta$ $\frac{d}{dt}C_i^3 = -\Delta$ $\frac{d}{dt}C_{i+1}^3 = -\Delta$ $\frac{d}{dt}C_{i-1}^4 = \Delta$ $\frac{d}{dt}C_i^6 = \Delta$ $\frac{d}{dt}C_{i+1}^6 = \Delta$
Refined Primer-dependent RNA-directed RNA polymerization		$\Delta = k_{drr} R_3(C_{i-1}^3 C_i^3 C_{i+1}^7)$ $\frac{d}{dt}C_{i-1}^3 = -\Delta$ $\frac{d}{dt}C_i^3 = -\Delta$ $\frac{d}{dt}C_{i+1}^7 = -\Delta$ $\frac{d}{dt}C_{i-1}^4 = \Delta$ $\frac{d}{dt}C_i^8 = \Delta$ $\frac{d}{dt}C_{i+1}^8 = \Delta$

local structure by the sum of all the concentrations of possible neighbor local structures. Let $NEIGHBOR_{right}$ a map from local structure to a power set of the local structures, by which all possible neighbor local structures to the right can be obtained. The results of the map in the figure are

$$NEIGHBOR_{right}(4, i-1) = \{(4,i), (6,i), (9,i), \cdots\}$$

$$NEIGHBOR_{right}(4, i) = \{(4,i+1), (6,i+1), (9,i+1), \cdots\}.$$

This map is adopted in a similar manner to the map $CONNECTED$ in Section 4.2.4 .

By the map, R_2 and R_3 are defined as

$$R_2(C_i^4 C_{i+1}^4) = \frac{C_i^4 C_{i+1}^4}{\displaystyle\sum_{l \in NEIGHBOR_{right}(4,i)} C_l}$$

$$R_3(C_{i-1}^4 C_i^4 C_{i+1}^4) = \frac{C_{i-1}^4 C_i^4 C_{i+1}^4}{\displaystyle\sum_{l \in NEIGHBOR_{right}(4,i-1)} C_l \times \sum_{l \in NEIGHBOR_{right}(4,i)} C_l}.$$

By the ratio assumption, it is also possible to define a similar map, namely $NEIGHBOR_{left}$, to calculate R_2 and R_3 in opposite direction. The same kind of
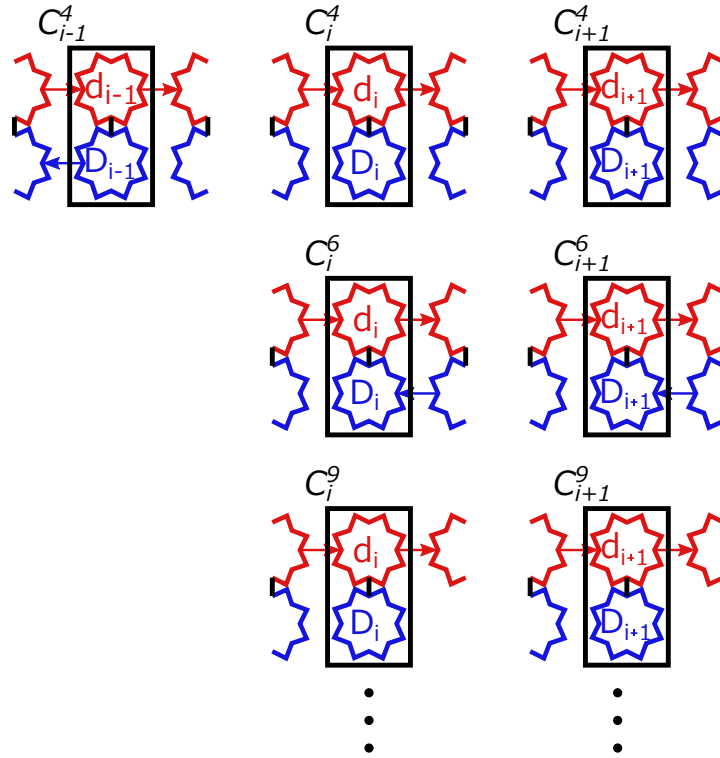
Figure 6.4: Possible neighbor local structures. Because of the annotation, the structure $(4, i-1)$ can connect to $(4, i), (6, i), (9, i), \cdots$. Similarly, the structure $(4, i)$ can connect to $(4, i+1), (6, i+1), (9, i+1), \cdots$.

discussion is possible for general R_i $(2 \le i \le n)$, by which the concentration of global structure can be recovered.

6.1.4 Simulation Result of Abstract Model of RNAi

By the abstract simulation, I obtained an expected concentration distribution of ssRNA (Figure 6.5). For each n, I simulated for 500 time units, and obtained the three-dimensional figure by plotting all distributions. Concentrations of ssRNAs are used because the distribution of ssRNAs are experimentally observed as a clue of siRNAs [23]. Note that obtaining such distribution was extremely difficult with the naïve model, because the number of domains that can be simulated was limited by the combinatorial explosion.

6.2 Efficiency and Exactness of Abstract Models

6.2.1 Efficiency of Simulation

I compared naïve and abstract simulations from the aspect of the number of structures and reactions. The number of global structures and reactions increased
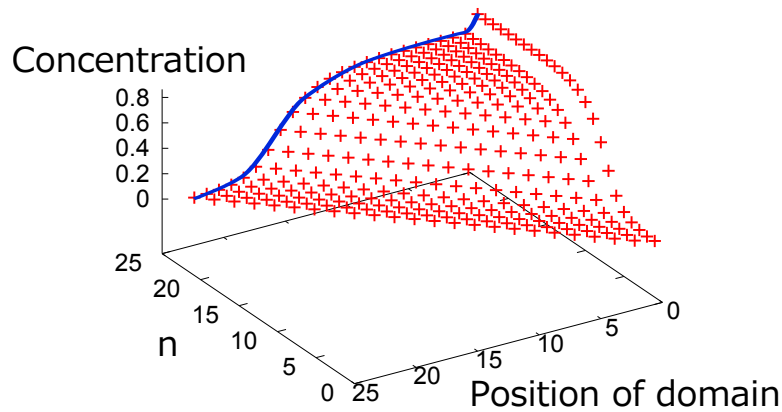
Figure 6.5: Distribution of ssRNA by abstract model simulation. X, y, and z axes of the figure correspond to the parameter n, the position of ssRNA, and the concentration of lower ssRNAs, respectively. Gradient distribution of the concentration of ssRNAs in the case of $n = 24$ is emphasized.

exponentially for the naïve model (Figure 6.6(a) and Figure 6.6(b)). Simulation with more than 9 domains was impossible because of an out-of-memory error (the number of structures and that of reactions exceed 10000 and 800000, respectively). This was the main reason why simulation became intractable by combinatorial explosion. In contrast, the number of local structures and that of reactions increased linearly with both refined and unrefined abstract models, which made the abstract simulation much more efficient. Computation time also decreased dramatically in both abstract models (Figure 6.6(c)).

6.2.2 Approximate Abstraction of RNAi

Property of the abstraction is discussed by the diagram of Chapter 5. In RNAi, G and L correspond to structures explained in Section 3.3.3 and 6.1.2, respectively. The abstraction α was exact for n smaller than or equal to 3. The recovery function β was also exact in the case. This was because each local structure became equivalent to one of the global structures due to the annotation.

For n greater than 3, however, both abstraction and recovery function was approximate. The property can be understood from the following experimental results. I actually compared the concentration change of dsRNA among the three models (Figure 6.7).

Although the unrefined abstract model had a slight difference from the naïve model, the refined abstract model had almost no difference. This result suggests that the refined abstract model is more exactly calculating the concentration of dsRNA than the unrefined abstract model. Though I can empirically conclude
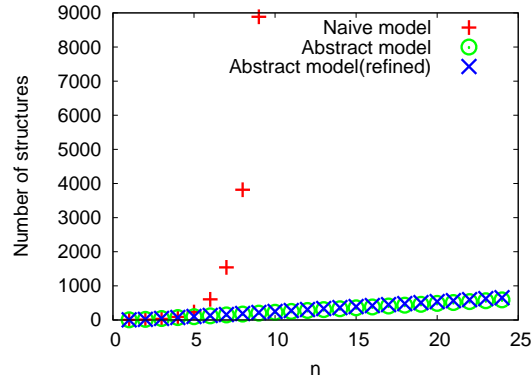
(a) Number of structures
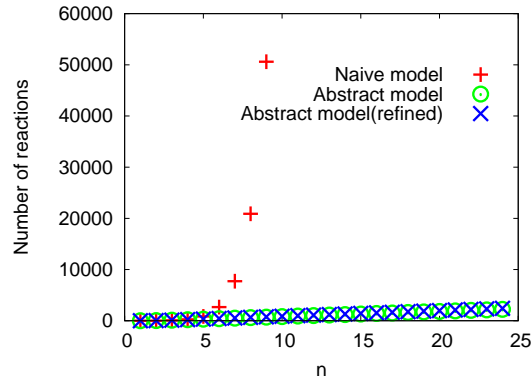
(b) Number of reaction rules
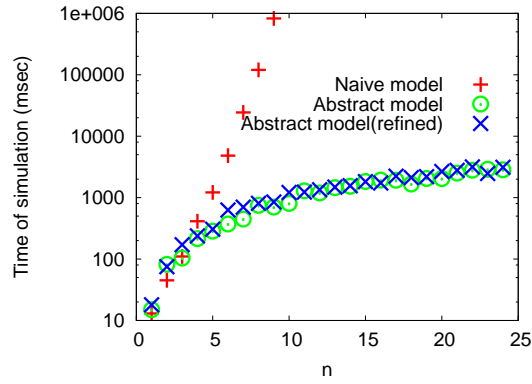
(c) Computation time

Figure 6.6: Comparison of simulations of RNAi. X axis corresponds to the parameter n for all figures. (a) Y axis indicates the number of global and local structures in the simulations. (b) Y axis indicates the number of reaction rules in the simulations. (c) Y axis indicates the computation time of the simulations in logarithmic scale.

that both abstract models are calculating the concentration of dsRNA accurately, their results are theoretically approximate due to the ratio assumption.

By calculating the time change of a ratio, which appears in R_2, I examined
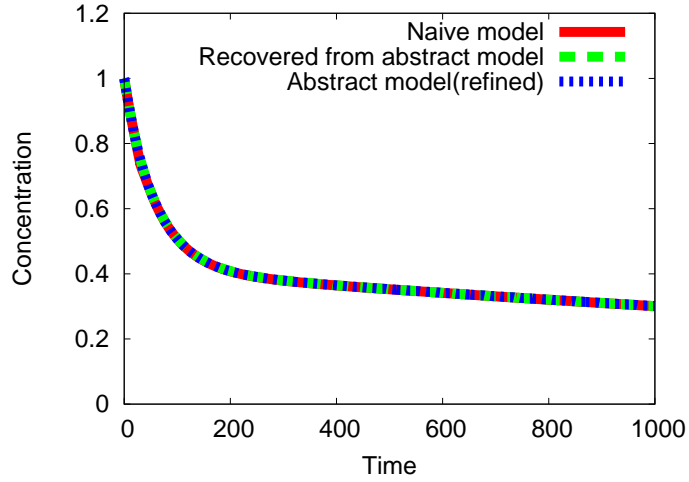
(a) Concentration change of dsRNA when n is 3
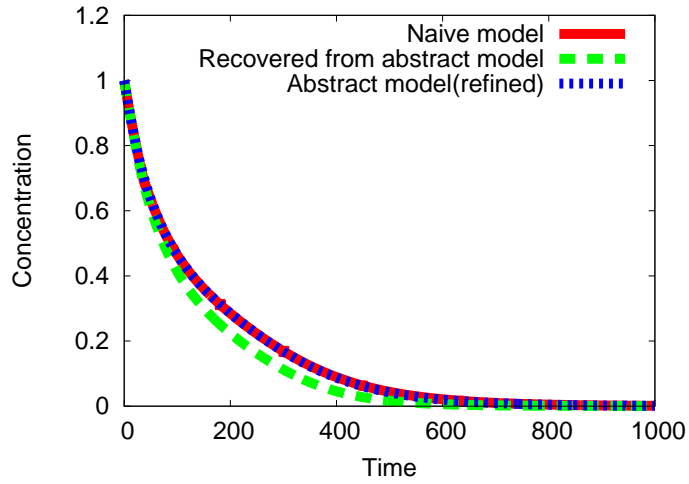
(b) Concentration change of dsRNA when n is 9

Figure 6.7: Concentration changes of dsRNA for (a) $n = 3$ and (b) $n = 9$. X and y axes are time and concentration of dsRNA, respectively. When n is 3, the results of three simulations coincide.

the correctness of the ratio assumption for abstract models. For example, the ratio of a local structure (x, i) is calculated as

$$\frac{C_i^x}{\displaystyle\sum_{l \in NEIGHBOR_{left}(y, i+1)} C_l},$$

where $(y, i + 1)$ can be randomly chosen from $NEIGHBOR_{right}(x, i)$. The value indicates the ratio of the concentration of selected local structure among all local structures that have the same annotations about right neighbor. Note that the ratio is independent from the choice of $(y, i + 1)$ due to the definition of the

(a) A local structure

(b) Time evolution of the ratio of (a)

(c) Another local structure

(d) Time evolution of the ratio of (c)
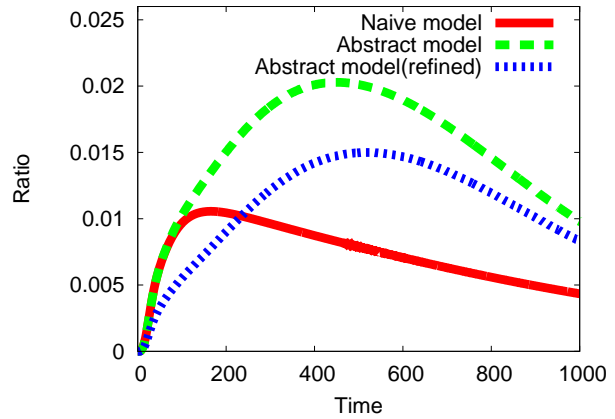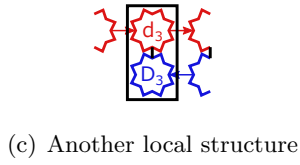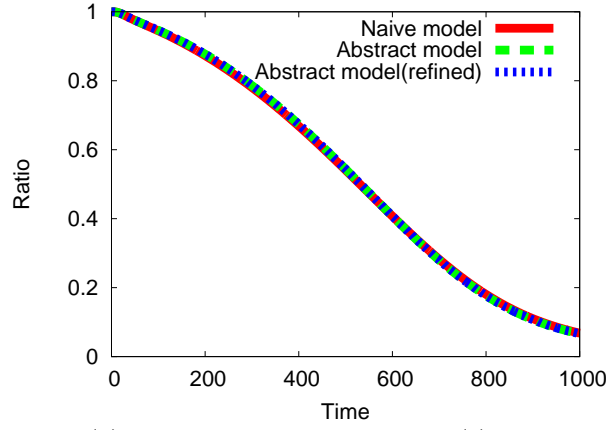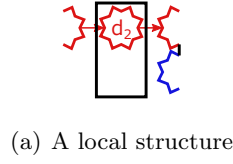
Figure 6.8: Selected local structures and the time evolution of their ratio. Two local structures represented in (a) and (c) are selected to trace its ratio. X and y axes of the graphs of (b) and (d) are time and the ratio, respectively. Results are obtained from the simulation of $n = 5$, and the same ratio was computed from the naïve model for comparison. Index of the local structures of (a) and (c) are 2 and 3, respectively.

annotation of local structures.

Selected local structures and the time evolution of their ratio are shown in Figure 6.8. The ratio assumption seemed to be satisfied in the case of Figure 6.8(a). In the next example of Figure 6.8(c), however, the ratio assumption was not satisfied because the ratio was significantly different between naïve and unrefined model. Note that the result of refined model is not comparable because the structure is divided into two local structures. The result exhibits that both abstract models are based on approximate abstraction.
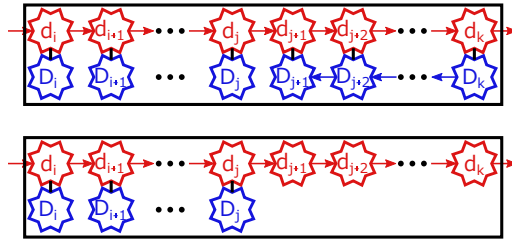
Figure 6.9: Model of RNAi for exact abstraction. Only local structures with upper domains are shown, while the length of local structures of lower domain is always one. There are three degree of freedom to determine a local structure, which are denoted by variables i, j, and k .

6.2.3 Exact Abstraction of RNAi

As far as I have considered, it seems impossible to define an exact abstraction in the current model of RNAi. An important aspect to define an exact abstraction is to make each reaction satisfy the ratio assumption before and after the reaction. The ratio assumption is easily broken when a reaction requires particular connection of local structures to occur (polymerization and denaturation are typical examples). In contrast, if a reaction is unimolecular in the abstract model, it does not break the ratio assumption (decay is a typical example). Moreover, if a bimolecular reaction occurs between any combinations of local structures, the reaction does not break the assumption (hybridization is a typical example).

To define an exact abstraction, some assumptions that restrict the polymerase reaction are required. One assumption is that the reaction speed of polymerization is very fast compared to others. By this assumption, I can define bigger local structures (Figure 6.9), which give a new and exact abstraction for RNAi. In this model, all the local structures have one of the two forms in the figure, both of which have multiple (or no) short (one domain) lower RNAs from the left part of the structure. One form of local structure has a long (more than one domain) lower RNAs at the right part. The other does not have any complementary RNA at the right part. A global structure of the naïve model is then represented as a set of these local structures.

Reaction rules are also redefined for this model. Examples of denaturation reactions are shown in Figure 6.10. Since the reactions are unimolecular, they do not require nor violate the ratio assumption.

It is also possible to take into account exonuclease or displacement activities as other assumptions. By these assumptions, only one lower RNA can hybridize to mRNA because polymerization reaction immediately extends the lower RNA

(a) A transition by denaturation   (b) Another transition by denaturation
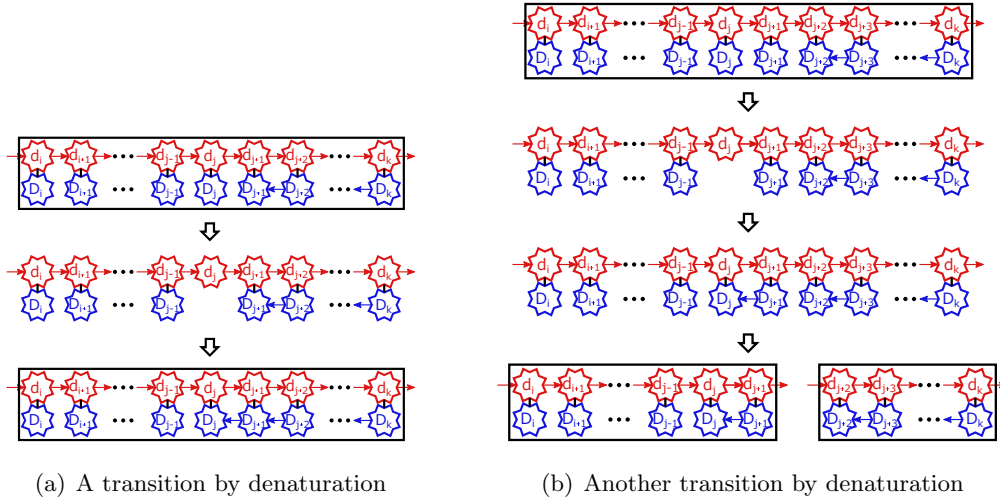
Figure 6.10: Examples of transition by denaturation for exact abstraction of RNAi. In both of the cases, one domain lower ssRNA represented as D_j is released from the structure. Even though intermediates structures, which are not surrounded by rectangles, are produced, they are neglected because the polymerization reaction occurs immediately. (a) The intermediate structure returns to the form of new local structure by polymerization. (b) After polymerization, the intermediate structure will be regarded as two split local structures. As a consequence, those reactions are defined as transitions between new local structures.

toward the leftmost end of mRNA. With such assumption, the number of structures in the naïve model is strongly restricted because combination of lower RNAs is eliminated completely.

6.2.4 Classification of RNAi System

The number of local structures in exact abstraction can be a measure of computational complexity of a reaction system. According to the abstractions which are already explained, I classified the complexity of RNAi in Table 6.3. As shown in the table, there are three groups of systems of RNAi depending on the assumption on polymerization.

Without any assumption (first class), it seems impossible to define an exact abstraction to allow simulation with polynomial complexity. By approximate abstraction, however, it was possible to efficiently simulate such systems with reasonable ratio assumption. Even when the speed of polymerization reaction is fast enough (second class), it is still impossible to efficiently simulate the system by a naïve model because the number of structures is exponentially large. By the exact abstraction in the previous subsection, however, such system is efficiently

Table 6.3: Three classes of RNAi. First column indicates the assumption on polymerization reaction. Second and third columns correspond to the number of global and local structures in naïve and exact abstract models, respectively.

Speed of polymerase reaction and its activity	Number of global structures in naïve model	Number of local structures by exact abstraction
One-by-one	Exponential	Seems exponential
Immediate	Exponential	$O(n^3)$
Immediate Exonuclease/ displacement activity	$O(n^3)$	$O(n^3)$

simulated because the number of structures becomes a polynomial size. If the assumption on the polymerization is too much (third class), the system becomes simple enough to simulate by a naïve model.

Chapter 7

Design Automation

7.1 Strategy of Automatic Design Method

7.1.1 Algorithm of Evolutionary Computation

As an application of analyzing a system by graph based model, I explain the method and the results of automatic design of nucleic acid reaction systems in this chapter. Such automatic design is performed by *in silico* evolution. The method is expected to be applied for designing robust and sophisticated topology of nucleic acid reaction system. Preliminary results of this chapter are included in my paper [263].

In fact, the method tries to maximizes a fitness value, which is a typical combinatorial optimization problem. For the purpose, I employ a heuristic algorithm called simulated annealing (SA) described in Figure 7.1. To carry the algorithm, a model to represent genotype, a method to predict its phenotype, an evaluation function to calculate the fitness, and an operation to mutate a genotype are necessary. All of them are described step-by-step in following sub-sections, while brief explanation of the algorithm is explained in the next paragraph.

The method searches for a topology with bigger fitness value by iterating local searches which starts from a randomly generated candidate. The fitness value can be calculated by predicting the dynamical behavior of the candidate. By a mutation, slightly different candidate is produced and the fitness values is calculated again. By comparing the fitness values before and after mutation, the method decides whether to transit to the mutated candidate. After iterating the processes, the method outputs a system with the biggest fitness value among produced candidates as a semi-optimal solution.

7.1.2 Genotype of Nucleic Acid Reaction System

To represent the genotype of a candidate nucleic acid reaction system, sequences of letters are used, which defines the search space of the algorithm. An alphabet of
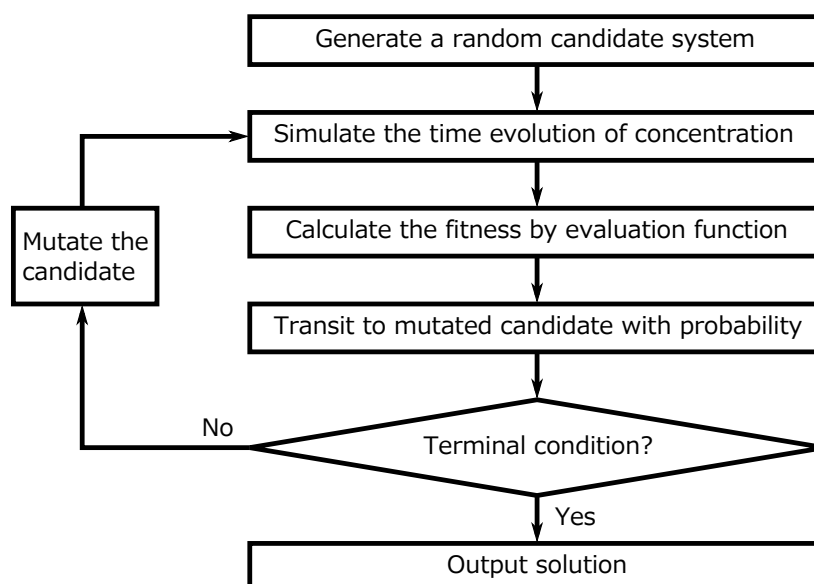
Figure 7.1: Flowchart of automatic design algorithm. After random generation of initial candidate, the algorithm iterates the process of simulation, evaluation, transition, and mutation until the terminal condition is satisfied.

the letters is defined as a set of thirteen letters, which consists of letters from 'a' to 'f', from 'A' to 'F', and null letter. In general, a genotype of a gate is represented as a sequence of sixteen letters, which are divided into four four-letter sequences. The alphabet is regarded as a set of domains for nucleic acid reactions, while each four-letter sequence corresponds to a single stranded nucleic acids. Once a system is represented by a sequence of letters, it is possible to chemically implement the system by adopting sequence design method to each domain.

The null letter is included in the alphabet to represent a strand that is composed of less than four letters and a gate that has less than four strands. These parameters such as sixteen and thirteen that restrict the search space are imposed to eliminate an ineffective search. The parameters may be adjusted depending on the function of interest, which is an important procedure to optimize the search space for the automatic design.

For example, a genotype represented by the following sequence of letters:

a, b, null, null; f, c, b, null; null, null, null, null; E, D, C, B

corresponds to the substrate of the catalytic gate whose phenotype is represented by a graph based model in Figure 3.1(a). Thanks to the flexibility of the graph based model, the search space comprise a system that utilize various secondary

68

structures such as hairpin loop, bulge loop, and pseudoknot. For another example, the genotype of HCR hairpins (Figure 4.1(a)) is represented by following sequence:

a, b, c, B; B, A, b, C; null, null, null, null; null, null, null, null .

Since HCR has only two hairpins, all of the last two four-letters are null. To simplify the notation, the genotype of HCR is described as "abCB", "BAbC" by ignoring null letters.

To represent inputs to the gate, additional sequences are introduced to the genotype depending on the target. I target two types of systems, that have functions of combinatorial circuit (logic gate) and state transition machine (automaton). For the case of logic gate that utilizes single stranded nucleic acid as inputs, two four-letter sequences are added to the genotype to represent the input strands. In contrast, two letters are added int the case of automaton to represent domains that are stimulated by inputs.

By those definitions, a random candidate is generated by choosing random 24 (for logic gate) and 18 (for automaton) letters from the alphabet. The length parameter of each domain is also uniformly chosen from 3 to 34 for each trial of evolutionary computation. A mutation is defined as an operation that flips randomly chosen letter into other random letter. More precisely, the mutated candidate is produced by repeating such flip for random times (five times at most).

7.1.3 Simulation of Phenotype and Probabilistic Transition

By adopting continuous simulation of the graph based model, it is possible to simulate the dynamic behavior of given genotype. A gate is constituted by hierarchical simulations, in which four strands are simulated separately in the beginning. Then, the results of the first two simulations are mixed for another round of simulations, while same simulation is carried for the last two. Finally, both results are mixed and simulated to constitute a gate, to which inputs can cause successive reactions.

From the result of the simulation, a fitness value is calculated through an evaluation function. The evaluation functions differ from target to target depending on the function of interest. Concrete definitions of the evaluation functions of logic gate and automaton are explained in Section 7.1.4 and 7.1.5, respectively.

To achieve a probabilistic transition of candidates, temperature parameter T

is provided in SA. The parameter T gradually decreases as the method iterates the process, which contribute to an efficient optimization. Concretely speaking, T is 1.0 at the beginning of the algorithm, and 0.96 is multiplied for each iteration.

The provability of transition is 1 if the mutated candidate has bigger fitness value than current candidate. Otherwise, the provability of transition is calculated by

$$\exp(\frac{v_{dif}}{T}),$$

where v_{dif} is a difference of fitness values between the current and mutated candidates. Terminal condition is defined as a time restriction that the algorithm terminates after calculating for one hour.

7.1.4 Evaluation Function of Logic Gate

To calculate the fitness value to determine how well a system behaves as a desired system, evaluation functions are required. I define the evaluation function of a logic gate by three criteria. The first criterion v_1 is calculated as

$$v_1 = \max \left\{ \min(T(s)) - \max(F(s)) \mid s \in S \right\},$$

where S is a set of single stranded structures in a system, and $T(s)$ and $F(s)$ are sets of concentrations of single strand s after simulations by adding specific combinations of inputs to the gate. $T(s)$ contains only the results from the combinations of inputs that satisfy the logic, while $F(s)$ contains only the results from the combinations of inputs that do not satisfy the logic.

In an AND gate for example, $T(s)$ has only one element that is the concentration of s after simulating with both inputs. In contrast, $F(s)$ has three elements that are the concentrations of s after simulations with remaining three combination. This value indicates the largest difference between true states and false states, which means how well the system work as desired logic gate.

The second criterion v_2 is given as

$$v_2 = \max \left\{ average(T(s)) - average(F(s)) \mid s \in S \right\},$$

where S, $T(s)$ and $F(s)$ are the same sets in the first criterion, and *average* calculates an average value from the elements of a given set. This criterion is introduced to make the evaluation function smoother.

The last criterion v_3 is calculated as

$$v_3 = \frac{1}{|S|} \sum_{s \in S} (\max(C(s)) - \min(C(s))),$$

where $C(s)$ is a union of $T(s)$ and $F(s)$ $(C(s) = T(s) \cup F(s))$. This value indicates how well each strand is involved in the reaction, which is an important factor to minimize unnecessary strands.

Finally, the fitness value of a candidate is calculated as a weighted average of these criteria by a function

$$0.8v_1 + 0.1v_2 + 0.1v_3.$$

The output of the gate is dynamically selected in the process of evaluation by

$$output = \underset{s \in S}{argmax}(\min(T(s)) - \max(F(s))),$$

which is a strand that decides v_1.

7.1.5 Evaluation Function of Automaton

Another target is an automaton in which the order of inputs can control the output. I assume that the automaton is a gate composed of DNA that has the following features.

1. The gate has two inputs and one output.

2. Inputs to the gate are small molecule and light that block the hydrogen bond between complementary domains.

3. Other state transitions of the gate are driven by hybridization reactions.

4. The final states of the automaton can be different if the orders to add inputs are different.

5. The gate emits the output when the state changes to one of the final states.

To apply the automatic design method to the automaton, I extended the simulator to support the second feature.

Extended version of the simulator has an operation to change the rate constants of hybridization and denaturation of specific domain at specific time units. If the operation is executed, the rate constants of hybridization and denaturation turn to 0.0 and 1.0, respectively, which leads to the separation of the domains. For chemical implementation of such feature, the strategies of ATP and UV sensors (Figure 2.2) can be adopted.

To evaluate a system as an automaton, I defined an evaluation function by the following expression,

$$v = \max \left\{ |C_{1,2}(s) - C_{2,1}(s)| \mid s \in S \right\},$$

where $C_{1,2}(s)$ and $C_{2,1}(s)$ are the concentrations of the structure s after simulations that adds inputs in the orders denoted by the indexes. This fitness value indicates the difference between the two ordering of inputs, and I can distinguish the two states when this fitness value is large. Similar to the logic gate, the output of the gate is dynamically selected during the evaluation process by

$$output = \underset{s \in S}{argmax}(|C_{1,2}(s) - C_{2,1}(s)|).$$

7.2 Results of automatic Design

7.2.1 Logic Gates Driven by Hybridization Reaction

As a result of executing the automatic design method, logic gates with novel topologies were successfully designed. Examples of OR and AND gates utilizing only hybridization reaction are shown in Figures 7.3 and 7.2, respectively. Note that letters and its order of the genotype are rearranged to make them intuitively understandable. Because unused domains do not effect the evaluation function, such domains can be preserved during the evolution process.

In the designed OR gate, both input 1 and input 2 can release the output by sequential hybridization and branch migration. The design of OR gate was too simple so that all trials of evolutionary computation converged to a system that has the same topology.

In the designed AND gate, the gate and input 1 first hybridize to temporarily release the output which will hybridize to the remaining gate immediately. Eventually, input 2 can completely release the output. The design of AND gate did not converge to a single topology, which indicates that AND operation can be achieved by multiple topologies. Empirically, candidates were trapped in local solutions for about half trials of evolutionary computations.

7.2.2 AND Gate Driven by Enzymatic Reaction

To demonstrate the ability of the kinetic simulator and the automatic design method, a logic gate that utilizes enzymatic reaction was also designed. The gate consists of DNA while inputs and output of the gate are RNA, which is inspired from RTRACS. The enzymatic reactions involved in the system are also the same to RTRACS, that are decay, primer-dependent DNA polymerization, and transcription.

The designed AND gate is shown in Figure 7.4, in which the gate first hybridizes to input 1 and then, the 3' end of DNA strand is extend by polymerization. After a decay reaction of input 1, input 2 hybridizes to the extended
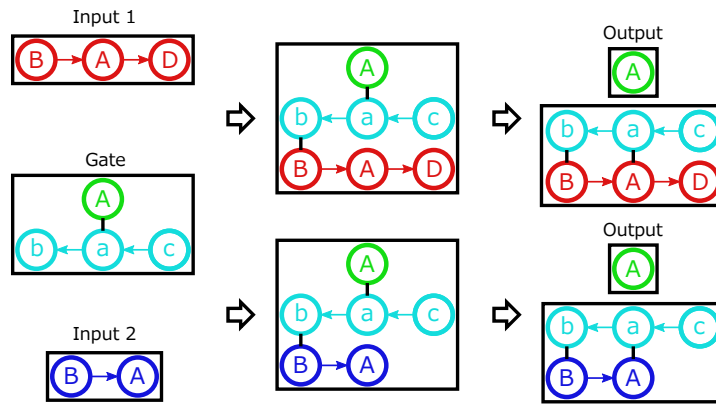
(a) Transition of designed OR gate
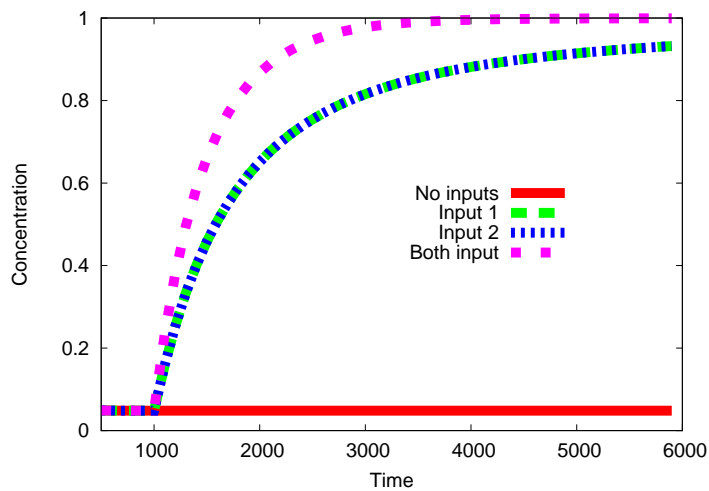
(b) Simulation result of the OR gate

Figure 7.2: Designed enzyme-free OR gate. Obtained genotype of the gate was "cab","A", while that of inputs were "BA" and "BAD". (a) Two transitions to produce the output with either of inputs are illustrated. (b) The simulation result of the OR gate by naïve model is shown. X and y axes correspond to time and concentration of the output, respectively. The concentration of output increased if at least one input was added at 1000 time units.

domain. By extending hybridized input 2, the promoter region denoted by 'T' become double stranded DNA. The downstream of the promoter is transcribed to produce the output RNA.

Although the mechanism of the behavior of the gate was similar to that of RTRACS, the topology of the system was not the same. The result indicates that the methodology in this thesis can apply to design a system using not only DNA but also RNA and enzymatic reactions. The evolutions of fitness values in
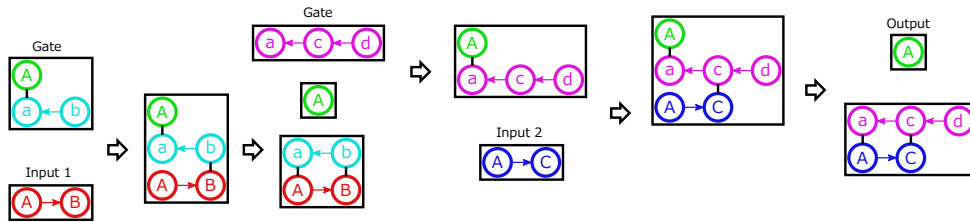
(a) Transition of designed AND gate
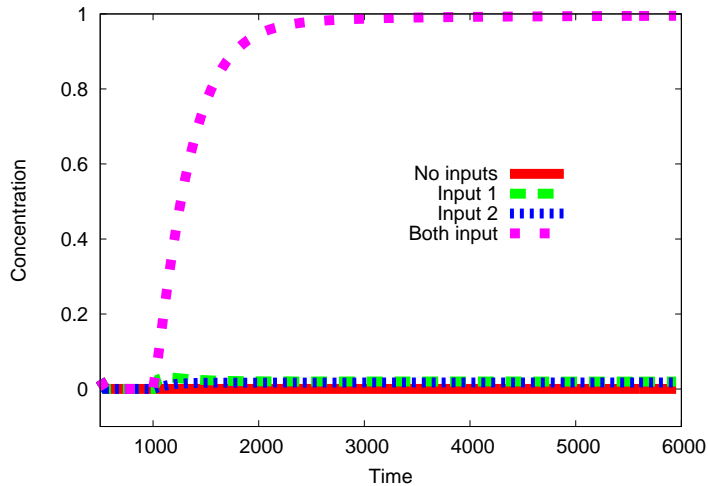
(b) Simulation result of the AND gate

Figure 7.3: Designed enzyme-free AND gate. Obtained genotype of the gate was "A","ba", "dca", while that of inputs were "AB" and "AC". (a) Transitions to produce the output with both inputs are illustrated. (b) The simulation result of the AND gate by naïve model is shown. X and y axes correspond to time and concentration of the output, respectively. The concentration of the output increases if and only if both inputs are added at 1000 time units.

respect to the number of iterations are shown in Figure 7.5. Unlike the design of enzyme-free AND gate, fitness values for all trials converged to relatively high value which were near 1.0.

7.2.3 Automaton Sensing the Stimuli from Environment

The design of logic gates was rather regarded as a proof of concept, because important criteria to cascade such gates was not included. For example, one of the inputs of the OR gate or enzymatic AND gate was a subsequence of the other input. Such properties are undesirable to implement a logic circuit by integrating multiple modules in a real system.

The design of the automaton does not need to care such undesirable phe-
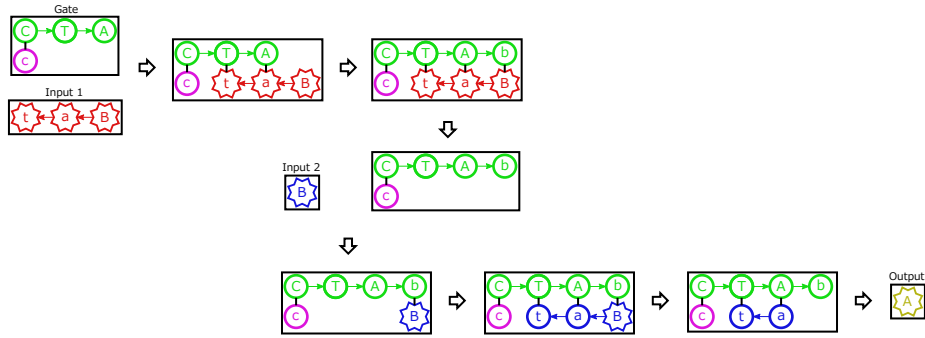
(a) Transition of designed AND gate
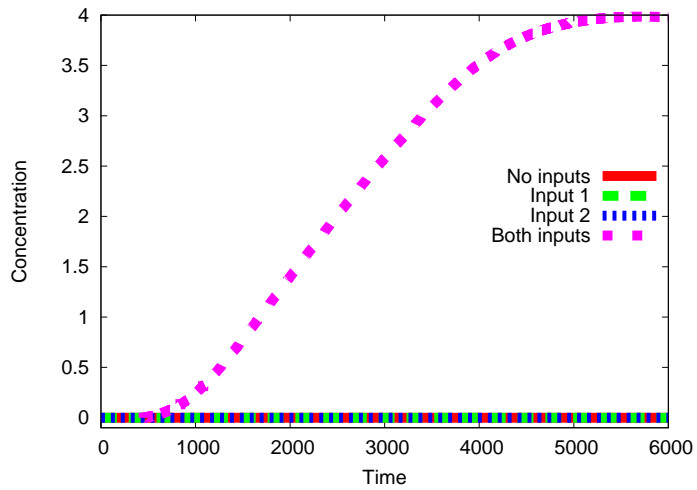
(b) Simulation result of the AND gate

Figure 7.4: Designed AND gate that utilizes enzyamtic reactions. Obtained genotype of the gate was "CTA","c", while that of inputs were "Bat" and "B". (a) The gate and its transitions to produce the output with both inputs are illustrated. (b) The simulation result of the AND gate by naïve model is shown. X and y axes correspond to time and concentration of the output, respectively. The concentration of output increased if and only if both inputs were added at 1000 time units.

nomenon. This is because the inputs of the automaton are assumed to be irrelevant to each other. Furthermore, the output of the automaton is rather a dynamic control of nucleic acids than a cascadable single strand.

One of the nontrivial designed automaton is illustrated in Figure 7.6. In the initial state, the gate is in one of the two states which can switch back and forth by branch migration reactions. When either of inputs is added, the systems promotes the reaction irreversibly.
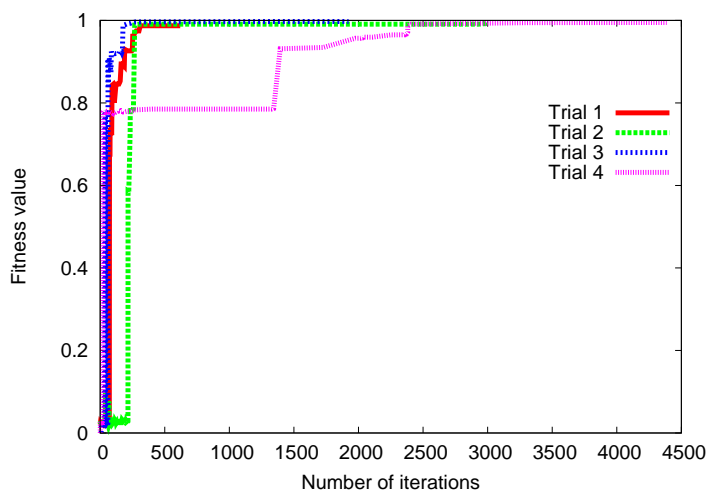
Figure 7.5: Evolutions of fitness values of the enzymatic AND gate. X and y axes correspond to the number of iterations and fitness value, respectively. Each line connects a sequence of plots, which reflects the transitions of candidates during the evolutionary computation.

By input 1, hybridization between domains 'a' and 'A' are blocked, which keeps the output connected to the gate. Subsequent input 2 blocks the hybridization between 'b' and 'B', which has nothing to do with the output. In contrast, when input 2 is added first, connection between 'b' and 'B' is blocked, which leads to the release of the output by a branch migration reaction. Subsequent input 1 does not affect the output.

The evolutions of fitness values among several trials are shown in Figure 7.7. Though the execution time was same among all trials, the number of iterations differs from each other, which indicates that simulation time differs dramatically from trial to trial. Unlike the design of enzymatic AND gate, most of the trials to design automata did not obtained high fitness values. This may be duet to the complexity of the function.

7.3 Validation by Chemical Experiment

7.3.1 Experimental Result of Enzyme-free AND Gate

To verify the effectiveness of proposed method, the designed enzyme-free AND gate and the automaton were implemented by a real nucleic acids. For the AND gate, unnecessary domains were omitted from the system and fluorophore and quencher molecules were attached to the output and gate, respectively. Thus, activity of FRET was observed to trace the concentration of single stranded output.

76
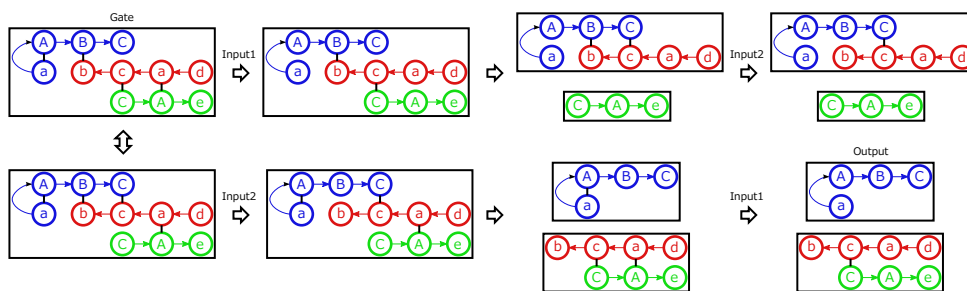
(a) Transition of designed automaton
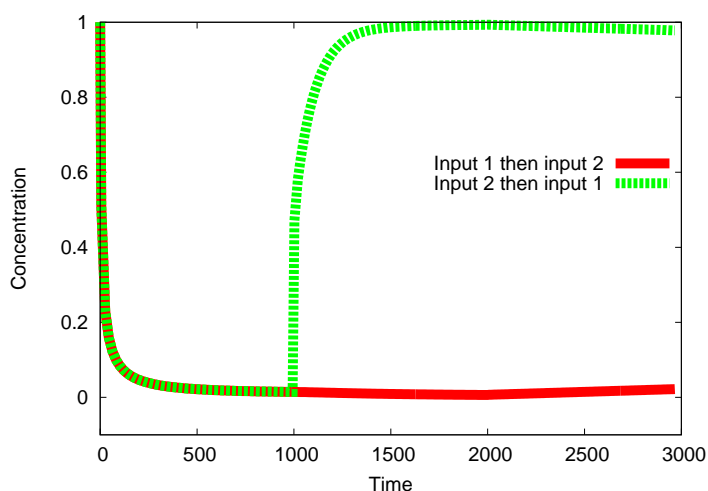
(b) Simulation result of the automaton

Figure 7.6: Designed automaton. Obtained genotype of the gate was "aABC","dacb", "CAe", while the domains to be blocked by inputs were 'a' and 'b'. (a) Transitions to produce the output with two ordering of inputs are illustrated. (b) The simulation result of the automaton by naïve model is shown. X and y axes correspond to time and concentration of the output, respectively. First and second inputs are added in 1000 and 2000 time units. The concentration of output increased if input 2 was added before input 1.

The result of the chemical experiment is shown in Figure 7.8. The fluorescent intensity increased if and only if both inputs were added. Because input 1 can temporally release the output, the intensity slightly grew when only input 1 was added.
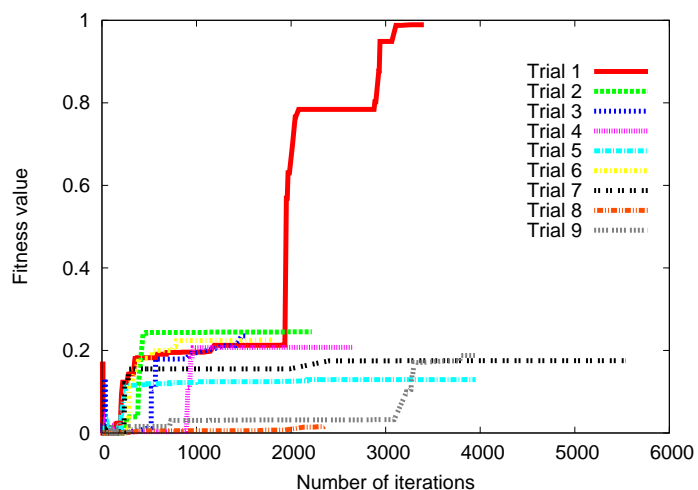
Figure 7.7: Evolutions of fitness values of the automaton. X and y axes correspond to the number of iterations and fitness value, respectively. Each line connects a sequence of plots, which reflects the transitions of candidates during the evolutionary computation.
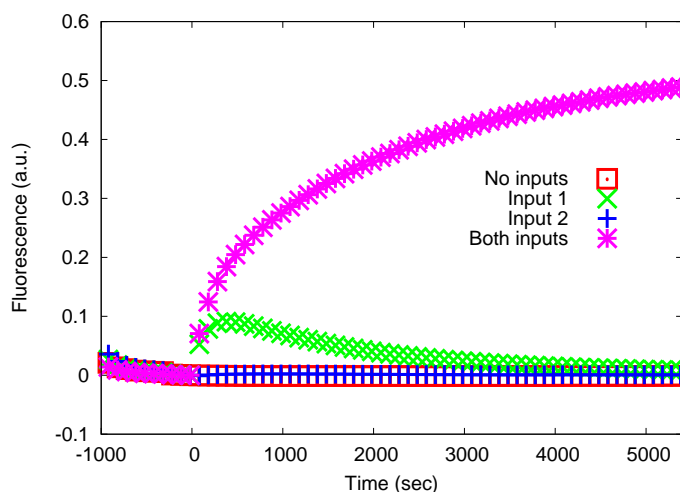
Figure 7.8: Experimental result of designed enzyme-free AND gate. X and y axes correspond to time in second and fluorescent intensity in arbitrary unit, respectively. All the intensity were shifted to match zero before adding inputs. Inputs were added at 0 second. Fluorescent intensity was normalized by the maximum intensity recorded in advance.

7.3.2 Experimental Results of Automaton

Preliminary Implementation

The activity of the automaton was also examined in two systems. One responds to DNA inputs, while the other responds to ATP and UV as inputs. Though
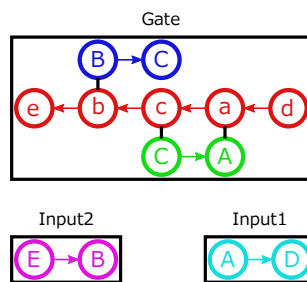
(a) Preliminary implementation of the automaton that responds to DNA inputs
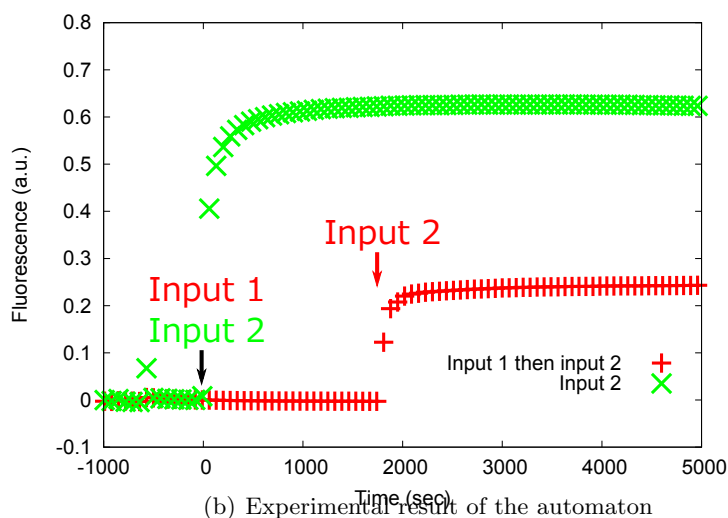
(b) Experimental result of the automaton

Figure 7.9: Preliminary implementation and experimental result of the designed automaton. (a) The modified system that responds to DNA inputs are shown. Toehold domains 'e' and 'd' were added to initiate branch migration reaction by DNA inputs. (b) The experimental results of the automaton is shown. X and y axes correspond to time in second and relative fluorescent intensity in arbitrary unit, respectively. Both intensity were shifted to match zero before adding inputs. First and second inputs were added at 0 and about 1800 seconds, respectively. Because the fluorescence intensity increased just by adding input 2 as expected, input 1 was not added afterward.

some of the domains were removed or added and the direction of the phosphate backbones were altered, the topology of the automaton was preserved.

As a preliminary implementation, I modified the automaton to respond to DNA inputs by adding new toehold domains (Figure 7.9(a)). Input DNAs displace the connections of target domains by sequential hybridization and branch
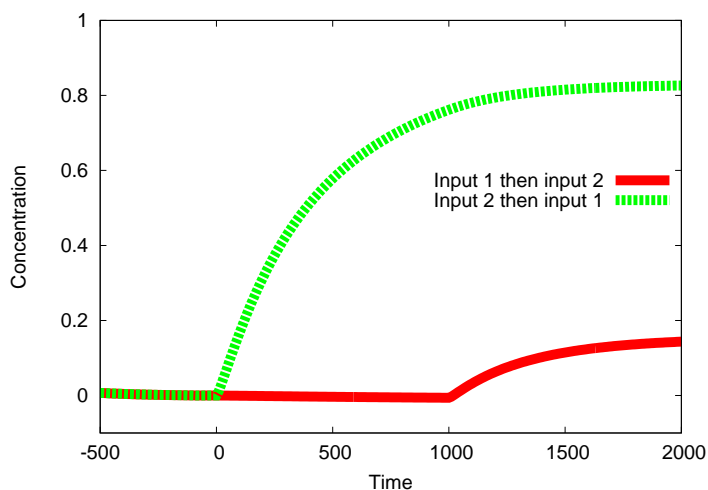
Figure 7.10: Simulation result of the automaton that responds to DNA inputs by naïve model. X and y axes correspond to time and concentration of the output, respectively. First and second inputs were added at 0 and 1000 time units. The concentration of the output slightly increased by adding input 2 after input 1.

migration reactions. Fluorescence and quencher molecules were attached to the output and gate, respectively, to observe the state transition of the automaton by FRET.

As a result of the chemical experiments (Figure 7.9(b)), it is reasonable to claim that the correct order of inputs can emit the output and wrong order cannot emit except for slight leakage. Note that I added only input 2 for one of the ordering to investigate the increase of output. This was because the operation was enough to emit the output, which was expected by the state transitions of the system. The leakage seems partly due to the change of the mechanism to respond to DNA inputs. Simulation result of the modified version of the automaton (Figure 7.10) indicates that the hydrogen bonds were not completely displaced by the branch migration reactions.

Implementation Using ATP Aptamer and Azobenzene

Because the original target was an automaton that responds to small molecules or light, I applied ATP aptamer switch [149] and azobenzene molecules switch [200] to block hydrogen bonds between specific domains. The modified version of the system and experimental results are shown in Figure 7.11. The result of the spectrophotometry (Figure 7.11(b)) indicates that the state transitions of the automaton took place to some extent. Since the dissociation constant of ATP aptamer was not efficient and the conformation change of the ATP aptamer was
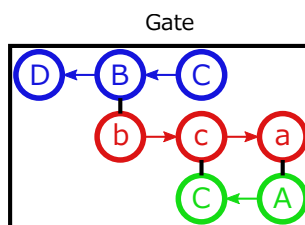
(a) Modified automaton that responds to ATP and UV
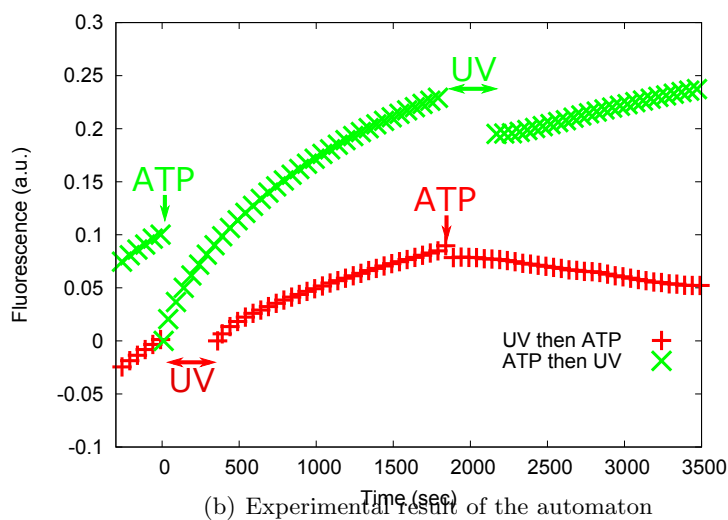
(b) Experimental result of the automaton
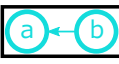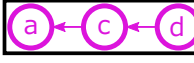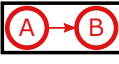
Figure 7.11: Implementation and experimental result of the automaton that responds to ATP and UV. (a) The modified automaton that responds to ATP and UV. The nucleotide sequence of ATP aptamer was contained in domains 'B' and 'D'. Azobenzene molecules were inserted in the sequence of domain 'A'. (b) The experimental result of the automaton by adding ATP and irradiating UV is shown. X and y axes correspond to time in second and relative fluorescent intensity in arbitrary unit, respectively. Both intensity were shifted to match zero after adding the first inputs. Because of the shift, the fluorescence intensities before adding inputs does not match in the figure. First and second inputs were added at 0 and about 1800 seconds.

larger than expected, the result contained non-negligible leakage.

7.4   Material and Method of Chemical Experiment

To monitor the fluorescent of the AND gate and the automaton, fluorescence spectrometer from Hitach was utilized. Wavelengths of excitation and emission for the fluorophore called FAM were 494nm and 512nm, respectively.

81

Table 7.1: Sequences of oligonucleotides from 5' end to 3' ends. Fluorophore called FAM was attached to the 5' end of the output strand. Quencher molecules called BHQ were attached to the 3' end of remaining strands of the gate. Sequence of the unnecessary domain 'd' was omitted in the implementation. In fact, the sequence of Input1 was "CCAAACTACTTACGTTGAACATACACCGAG-GTTTAGTGAACACTTCTAAGCAACTAA" because the DNA was borrowed from other system of the laboratory. Only the essential sub-sequences are shown in the table.

(A)	FAM-TGAACATACACCGAGGTTTAG
(a ← b)	TGTTCACTAAACCTCGGTGTATGTTCA-BHQ
(a ← c ← d)	TTTGGACTAAACCTCGGTGTATGTTCA-BHQ
(A → B)	TGAACATACACCGAGGTTTAGTGAACA
(A → C)	TGAACATACACCGAGGTTTAGTCCAAA

For the experiments of the AND gate and the automaton responding to DNA, nucleotides were dissolved in a standard TAE buffer. Final concentrations of all DNA strands were approximately 125 nM.

In contrast, the solution of the automaton responding to UV and ATP contained 300 mM NaCl, 5 mM $MgCl_2$, and 20 mM Tris·HCl. Final concentration of all DNA strands were approximately 125nM and the solution was kept at 37 °C. The wavelength of UV light was 365 nm and I irradiated for 300 seconds using an instrument from OMRON (ZUV-C2-H). 25mM of ATP was added as an input.

For the experiment of the enzyme-free AND gate, oligonucleotides in Table 7.1 were purchased from Sigma-Aldrich Japan. Similarly, sequence in Table 7.1 were purchased from Sigma-Aldrich Japan and Tsukuba-Oligo-Service (for azobenzene modified DNA) for the automaton responding to UV and ATP.

Table 7.2: Sequences of oligonucleotides from 5' end to 3' ends. Fluorophore called FAM was attached to the 5' end of the second strand. Quencher molecules called BHQ were attached to the 3' end of the first strand. Domain 'D' and 'B' contained a sequence of ATP aptamer which is indicated by underline. X in domain 'A' represents azobenzene molecule.

D←B←C	AGCCATCTACCAACGACTCAAGCATG<u>ACCTGG</u>- <u>GGGAGTATTGCGGAGGAAGGT</u>CATGCTTG-BHQ
b→c→a	FAM-TCCCCCAGGTCATGCTTGAGTCGT- TGGTAGATGGCTCTGGTAACAATCACG
C←A	CGXTGXATXTGXTTXACXC- AGAGCCATCTACCAACGACT

Chapter 8

Discussion and Conclusion

8.1 Difference from Conventional Technique

8.1.1 Model and Simulation of Nucleic Acid Reaction System

In this chapter, methods of this thesis and related researches are first compared. Limitation, future work, and contribution of this thesis are then mentioned.

The differences among models to represent and predict the behavior of nucleic acid reaction systems are summarized in Table 8.1. Details of them are explained in the rest of this section.

One of the strongest points of the proposed graph based model was the ability to represent various structures of nucleic acids. Flexible representation of nucleic acid was important since conventional modules have taken the advantage of various secondary structures such as hairpin loop [32], non-linear structure [36], and kissing structure [35]. The model was capable of analyzing and designing such modules, due to a flexible representation by graph. In contrast, conventional techniques to analyze and design reaction systems limited the structure of nucleic acids to hairpin loop [34], linear structure [30], and simple gate [2].

The flexibility of graph based model allowed the representation of various chemical reactions as rewriting rules. In general, graph rewriting has been commonly applied in chemistry [48], molecular biology [46], and biological pathway [47]. In this thesis, I proposed a set of transition rules for nucleic acid reactions, which were directed to analyze and design nucleic acid reaction systems with hybridization and enzymatic reactions.

The simulation technique in this thesis was applied to predict the dynamic behavior of a system by solving ODE based on chemical kinetics. Above conventional techniques, however, were further applied to a model checking of chemical system [31], integrating large scale circuit [57], and the implementation of arbitrary chemical reactions [223] with the aid of limitations of those models. In contrast, the naïve simulation of this thesis cannot be directly applied to such purposes since the combinatorial explosion became a crucial problem.

Table 8.1: Comparison among graph based model and conventional models. Graph based model, programming language, hairpin pathway, and see-saw gate correspond to the models of this thesis, the papers [30], [34], and [2], respectively

.

	Graph based model	Programming language	Hairpin pathway	See-saw gate
Target nucleic acid	DNA and RNA	DNA	DNA	DNA
Representation of structure	Graph data structure	String notation	Nodes with sites	Node with wires
Example structures	Linear duplex hairpin loop bulge loop pseudoknot	Only linear duplex	Only hairpin loop	Only duplex
Representation of reaction	Graph rewriting rules	Reduction rules	Transition between open and solid states	Network of nodes
Elementary reactions	Hybridization and enzymatic reaction	Hybridization reaction	Hybridization reaction (opening of hairpin)	Hybridization reaction
Enumeration and simulation	Both are automated	Both are automated	Both are manual	Enumeration is manual simulation is automated

Table 8.2: Comparison among model reduction methods. Abstraction, κ-calculs, and hypergraph approach correspond to the model reduction methods by this thesis, the papers [51, 49, 52], and [56], respectively.

	Abstraction	κ-calculus	hypergraph appraoch
Target system	Nucleic acid system with hybridization and enzyamtic reaction	Signalling cascade of phosphorylation	Hybridization reaction system in terms of secondary structure
Simulation (in most of the case)	Continuous and kinetics	Stochastic and kinetics	Continuous and equilibrium
Abstraction	Both exact and approximate (but efficient)	Exact	Exact

8.1.2 Exact and Approximate Abstraction

The combinatorial explosion was a fundamental problem when enumerating all possible structures by the graph based model. Imposing restrictions such as threshold and structures of nucleic acid did not eliminate the problem and introduced the possibility of incorrect simulation. An abstraction method that focused on the locality of a global structure avoided the problem by introducing local structures. The strongest aspect of the model was the ability to express any kind of structure at the expense of losing some part of the information, even in the case where unbounded number of structures were involved.

The abstraction in this thesis turns out to have close relationships to other conventional model reduction techniques. Comparison among the abstraction, κ-calculus [51, 49, 52], and hypergraph approach [56] is summarized in Table 8.2. Model reductions by κ-calculus and hypergraph approach are already introduced in Section 1.3.4.

Since a local structure of the abstract model was based on sub-graphs with annotations, some case studies in this thesis can be directly encoded by fragmentation of κ-calculus. In the case of exact abstraction of HCR, it is possible to define fragments from local structures, which can be regarded as an abstraction by κ. In the framework of κ, a fragment should overlap several local structures to satisfy the ratio assumption. Local structures that do not satisfy the assumption needs to be enlarged so that all reactions of fragments become independent. On the other hand, defining fragments from the exact abstract model of RNAi seems impossible because two local structures was produced from one local structure by a denaturation reaction.

Another noticeable abstraction is the enumeration technique using hypergraph [56]. Representing global structure as a path of hypergraph was the unique idea of the paper compared with the abstract model of this thesis and κ. Although the author of the paper mentioned about the application to kinetic simulation, the techniques was invented rather for efficiently computing a thermodynamic equilibrium of hybridization reaction system. In contrast, the efficiency was more important than accuracy when the simulation technique was applied for evolutionary computation. Evaluation process to calculate the fitness value of given candidate topology was repeated again and again in the algorithm of this thesis.

Fragmentation of κ and hyper-graph approach were carefully developed for exact simulation. Similarly, simulation by the abstract model of thesis was exact when the ratio assumption was satisfied in all the rewriting rules. Otherwise abstract model violated the ratio assumption for the sake of obtaining smaller

local structures, which led to an approximate but efficient simulation.

I also pointed out that the computational complexity of bio-molecular reaction systems can be classified using the number of structures by exact abstraction as a measure. It was supplemental but original result of this thesis.

8.1.3 Automatic Design by Evolutionary Computation

Applying simulation technique to design a new topology of nucleic acid reaction by evolutionary computation was an original idea of this thesis. In the area of genetic network, such evolutionary computations were commonly applied to design an interesting topology (e.g. oscillatory behavior). Some of them were based on graph models [63, 64]. Those researches, however, did not implement the designed topology into a real gene regulatory network since inverse problem to construct genes from given topology was a challenging task to solve. On the other hand, I demonstrated by *in vitro* experiments that two of the designed systems actually behaved as expected to some extent.

The number of single strands is a simple scale to measure the complexity of nucleic acid modules. For example, the modules of hairpin pathway [34], see-saw gate [2], and enzyme-free logic gate [1] consist of 1, 5, and 3 strands. It may be possible to state that the ability of the automatic design method catches up with a design by human trial and error in some functions, because there are not so many differences between the numbers of molecular species.

8.2 Limitations of Proposed Techniques

The target structures of the graph based model were nucleic acids with Watson-Crick complementary hydrogen bonds and normal phosphate backbone. Nodes of the graph based model have at most one for each undirected, incoming directed, and outgoing directed edges. Obviously, the model does not cover other structures such as triplex with Hoogsteen base pairs, i-motif, and so on. Note that ad-hoc representation was introduced in the case of RISC and aberrant RNA. It was also not supported whether the end of phosphate backbone has hydroxy group or not, although the information is important for some enzymatic reactions such as ligation.

The naïve simulation did not thoroughly apply all possible rewriting rules due to the threshold and the restriction on graph imposed in order to minimize the number of structures. The threshold was introduced to eliminate unimportant structures, which made the simulation incorrect when the concentration of important structure cannot exceed the threshold. Without the restriction on graph, the

simulation can spontaneously produce unbounded number of structures, which made the simulation intractable.

The abstractions of the graph based model were developed by rather ad-hoc definitions in this thesis. The definition arises a simple question whether it is possible to apply the method to general graph rewriting systems. Since the abstract model in this thesis extremely depended on the locality of the elemental chemical reactions, the question does not seem so simple. Unlike κ, the model was developed by a trade-off between the efficiency and the exactness of the simulation.

The limits to design much complex systems by evolutionary computation lied in the search space of the topology. The combinatorial optimization method cannot escape from the trade-off between the broadness of the search space and the efficiency of the heuristic algorithm. Because the targets of the automatic design were gates that output single stranded nucleic acid, using abstract model may contribute to the efficiency.

The design of DNA automaton was a challenging task because molecular machine that detects the ordering of inputs was nontrivial for human. The state transitions of the designed automaton were not so simple to understand intuitively even though there were only four steps to release the output. Since the search spaces were the same among evolutionary computations, difference of the target function affected the efficiency of the design. It suggests that the evaluation function to calculate the fitness value requires further optimization to design much optimal solution.

8.3 Future Work

8.4 Reaction Diffusion System for Molecular Robotics

Kinetic simulation in this thesis focused only on a temporal evolution of the concentration of nucleic acids. However, much complicated reaction-diffusion system are reported in both natural and artificial systems. In such systems, predicting spatiotemporal evolution of chemical species is necessary. I think that further adopting such simulation to an automatic design method is also possible to design a nontrivial reaction-diffusion system.

More concretely, one of the targets for the application is constructing a slime mold molecular robot [229]. It is defined as rather a macro-scale system compared with conventional nano-scale device. As a result of both diffusion and reaction of molecular species, the dynamics of the system becomes heterogeneous. In such system, time and position where reactions of nucleic acids take place can be

controlled.

For the purpose, gel and sol are suitable as a filed of reaction. I am actually aiming to control the transitions between gel and sol. Further complicated molecular robot may be developed by the aid of computational techniques.

8.5 Abstraction and Design Automation

Defining an exact abstraction of RNAi with the assumption of slow polymerization remains an open question. I assumed the ratio assumption, even though the polymerization violated it in the abstract model due to the annotation of local structures. Since any attempts to define better abstract models were successless, I believe that question may be negatively proved.

Only naïve simulation was utilized for the automatic design method as a proof of concept of the design automation. To design further complex systems, applying abstract simulation may be possible. For that purpose, abstract simulation for general hybridization systems can be exploited. It is because the concentration of single stranded nucleic acids can be traced in the simulation. In contrast, the abstract model of the enzymatic reactions was optimized for the simulation of RNAi, which seems not suitable for the purpose.

8.6 Summary of Contribution

8.6.1 Short Overview

This thesis contributes to the computer aided modeling, analysis, and design of nucleic acid reaction systems. Approaches for the contribution are listed as follow:

1. Proposing novel and flexible graph rewriting model and demonstrating continuous simulation based on chemical kinetics.

2. Avoiding combinatorial explosion of global structures by defining abstract model that focuses on the locality of structures.

3. Conducting efficient simulations by employing approximate abstraction and giving a discussion for exact abstraction.

4. Minimizing human trial and error by applying continuous simulation to automatic design based on evolutionary computation.

5. Verifying the automatically designed systems by chemical experiments.

They are briefly explained in the last subsection.

8.6.2　Detail Contribution

A novel graph based model to predict the dynamics of nucleic acid reaction systems is proposed since analyzing and designing the systems become rapidly important from biological and engineering aspect. Employing graph brings about a flexible representation of nucleic acid compared with other conventional methods to design or program a system driven by hybridization reactions. The graph based model includes not only hybridization reactions but also enzymatic reactions as graph rewriting rules. As a consequence, chemical (rather than phenomenological) analysis of biological pathway becomes possible. Concretely speaking, simulations of the dynamic behavior of enzyme-free catalytic gate, enzymatic logic gate, and RNA interference (RNAi) are carried out. As a result, the behavior of those systems are successfully explained in terms of the time evolution of concentration.

Elemental chemical reactions spontaneously give rise to a combinatorial explosion of molecular species, which is a fundamental problem in graph based model. To overcome the problem, exact and approximate but efficient abstraction are introduced. The abstractions are based on locality of the global structures of nucleic acids, which turns out to have a significant relationship to a technique called fragmentation. The abstraction method are adopted to two types of nucleic acid reaction systems (hybridization reaction systems and RNAi), while the fragmentation targets signalling cascade of phosphorylation.

In the case of hybridization reaction, it is impossible to formalize a differential equations without abstraction due to the unbounded number of structures. By abstract simulation, the problem is overcome and the concentration change of each global structure is recovered using a reasonable assumption. In RNAi, the distribution of small RNA cannot be predicted without abstraction because of the combinatorial explosion of the intermediate structures. By adopting abstract simulation, the distribution can be successfully predicted.

The performance of abstraction is discussed from the aspect of exactness of the simulation. To achieve an exact abstraction, additional assumption on the elementary chemical reaction is required for RNAi simulation. In contrast, approximate abstraction is capable of efficiently simulating a chemical reaction.

As an application of kinetic simulation, design automation of the topology of nucleic acid reaction system by evolutionary computation is addressed. By defining appropriate search space and evaluation function, logic gate and state transition machine are successfully designed. To evaluate the correctness of the method, two selected gate are implemented in actual chemical molecules and the

functions of the gates are verified by chemical experiments. Expected applications of this research include analysis of huge bio-molecular reaction systems that can be represented by graph, and the development of nontrivial nucleic acid reaction systems that can serve as a module of molecular device.

References

[1] Seelig, G., Soloveichik, D., Zhang, D. Y. and Winfree, E.: Enzyme-Free Nucleic Acid Logic Circuits., *Science*, Vol. 314, No. 5805, pp. 1585–1588 (2006), DOI: `10.1126/science.1132493`

[2] Qian, L. and Winfree, E.: A simple DNA gate motif for synthesizing large-scale circuits, *Journal of the Royal Society Interface*, Vol. 8, No. 62, pp. 1281–1297 (2011), DOI: `10.1098/rsif.2010.0729`

[3] Elbaz, J., Lioubashevski, O., Wang, F., Remacle, F., Levine, R. D. and Willner, I.: DNA computing circuits using libraries of DNAzyme subunits, *Nature Nanotechnology*, Vol. 5, No. 6, pp. 417–422 (2010), DOI: `10.1038/nnano.2010.88`

[4] Huizenga, D. E. and Szostak, J. W.: A DNA Aptamer That Binds Adenosine and ATP, *Biochemistry*, Vol. 34, No. 2, pp. 656–665 (1995), DOI: `10.1021/bi00002a033`

[5] Bock, L. C., Griffin, L. C., Latham, J. A., Vermaas, E. H. and Toole, J. J.: Selection of single-stranded DNA molecules that bind and inhibit human thrombin, *Nature*, Vol. 355, No. 6360, pp. 564–566 (1992), DOI: `10.1038/355564a0`

[6] Omabegho, T., Sha, R. and Seeman, N. C.: A Bipedal DNA Brownian Motor with Coordinated Legs, *Science*, Vol. 324, No. 5923, pp. 67–71 (2009), DOI: `10.1126/science.1170336`

[7] Bath, J., Green, S. J. and Turberfield, A. J.: A Free-Running DNA Motor Powered by a Nicking Enzyme, *Angewandte Chemie International Edition*, Vol. 44, No. 28, pp. 4358–4361 (2005), DOI: `10.1002/anie.200501262`

[8] Tombelli, S., Minunni, M. and Mascini, M.: Analytical applications of aptamers, *Biosensors and Bioelectronics*, Vol. 20, No. 12, pp. 2424–2434 (2005), DOI: `10.1016/j.bios.2004.11.006`

[9] Mairal, T., Özalp, V. C., Sánchez, P. L., Mir, M., Katakis, I. and O'Sullivan, C. K.: Aptamers: molecular tools for analytical applications, *Analytical and Bioanalytical Chemistry*, Vol. 390, No. 4, pp. 989–1007 (2008), DOI: `10.1007/s00216-007-1346-4`

[10] Cho, E. J., Lee, J.-W. and Ellington, A. D.: Applications of Aptamers as Sensors, *Annual Review of Analytical Chemistry*, Vol. 2, pp. 241–264 (2009), DOI: `10.1146/annurev.anchem.1.031207.112851`

[11] Douglas, S. M., Bachelet, I. and Church, G. M.: A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads, *Science*, Vol. 335, No. 6070, pp. 831–834 (2012), DOI: `10.1126/science.1214081`

[12] Venkataraman, S., Dirks, R. M., Ueda, C. T. and Pierce, N. A.: Selective cell death mediated by small conditional RNAs, *Proceedings of the National Academy of Sciences*, Vol. 107, No. 39, pp. 16777–16782 (2010), DOI: `10.1073/pnas.1006377107`

[13] Qian, L., Winfree, E. and Bruck, J.: Neural network computation with DNA strand displacement cascades, *Nature*, Vol. 475, No. 7356, pp. 368–372 (2011), DOI: `10.1038/nature10262`

[14] Zhang, D. Y., Turberfield, A. J., Yurke, B. and Winfree, E.: Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA., *Science*, Vol. 318, No. 5853, pp. 1121–1125 (2007), DOI: `10.1126/science.1148532`

[15] Watson, J. D. and Crick, F. H. C.: MOLECULAR STRUCTURE OF NUCLEIC ACIDS, *Nature*, Vol. 171, No. 4356, pp. 737–738 (1953).

[16] Carter, E. S. and Tung, C.-S.: NAMOT2 - a redesigned nucleic acid modeling tool: construction of non-canonical DNA structures, *Computer applications in the biosciences: CABIOS*, Vol. 12, No. 1, pp. 25–30 (1996), DOI: `10.1093/bioinformatics/12.1.25`

[17] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. and Ferrin, T. E.: UCSF Chimera - A visualization System for Exploratory Research and Analysis, *Journal of Computational Chemistry*, Vol. 25, No. 13, pp. 1605–1612 (2004), DOI: `10.1002/jcc.20084`

[18] Nishikawa, A., Yamamura, M. and Hagiya, M.: DNA computation simulator based on abstract bases, *Soft Computing - A Fusion of Founda-*

tions, Methodologies and Applications, Vol. 5, No. 1, pp. 25–38 (2001), DOI: `10.1007/s005000000062`

[19] Cardelli, L.: Strand algebras for DNA computing, *Natural Computing*, Vol. 10, No. 1, pp. 407–428 (2011), DOI: `10.1007/s11047-010-9236-7`

[20] Takinoue, M., Kiga, D., Shohda, K.-I. and Suyama, A.: Experiments and simulation models of a basic computation element of an autonomous molecular computing system, *Physical Review E*, Vol. 78, No. 4 Pt 1, p. 041921 (2008), DOI: `10.1103/PhysRevE.78.041921`

[21] Brodersen, P. and Voinnet, O.: The diversity of RNA silencing pathways in plants, *TRENDS in Genettics*, Vol. 22, No. 5, pp. 268–280 (2006), DOI: `10.1016/j.tig.2006.03.003`

[22] Baulcombe, D. C.: Amplified Silencing, *Science*, Vol. 315, No. 5809, pp. 199–200 (2007), DOI: `10.1126/science.1138030`

[23] Pak, J. and Fire, A.: Distinct Populations of Primary and Secondary Effectors During RNAi in C. elegans., *Science*, Vol. 315, No. 5809, pp. 241–244 (2007), DOI: `10.1126/science.1132839`

[24] Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P.: *Molecular Biology of THE CELL*, Garland Science, 5 edition (2008).

[25] Okazaki, R., Okazaki, T., Sakabe, K., Sugimoto, K. and Sugino, A.: MECHANISM OF DNA CHAIN GROWTH, I. POSSIBLE DISCONTINUITY AND UNUSUAL SECONDARY STRUCTURE OF NEWLY SYNTHESIZED CHAINS, *Proceedings of the National Academy of Sciences*, Vol. 59, No. 2, pp. 598–605 (1968).

[26] Takinoue, M., Kiga, D., Shohda, K.-I. and Suyama, A.: RNA Oscillator: Limit Cycle Oscillations based on Artificial Biomolecular Reactions, *New Generation Computing*, Vol. 27, No. 2, pp. 107–127 (2009), DOI: `10.1007/s00354-008-0057-5`

[27] Kan, A., Shohda, K.-I. and Suyama, A.: A DNA Based Molecular Logic Gate Capable of a Variety of Logical Operations, *Lecture Notes in Computer Science*, Vol. 7433, pp. 86–97 (2012), DOI: `10.1007/978-3-642-32208-2_7`

[28] Gillespie, D. T.: Exact Stochastic Simulation of Coupled Chemical Reactions, *The Journal of Physical Chemistry*, Vol. 81, No. 25, pp. 2340–2361 (1977), DOI: `10.1021/j100540a008`

[29] Medema, M. H., van Raaphorst, R., Takano, E. and Breitling, R.: Computational tools for the synthetic design of biochemical pathways, *Nature Reviews Microbiology*, Vol. 10, No. 3, pp. 191–202 (2012), DOI: `10.1038/nrmicro2717`

[30] Phillips, A. and Cardelli, L.: A programming language for composable DNA circuits, *Journal of the Royal Society Interface*, Vol. 6 Suppl 4, pp. S419–S436 (2009), DOI: `10.1098/rsif.2009.0072.focus`

[31] Lakin, M. R., Parker, D., Cardelli, L., Kwiatkowska, M. and Phillips, A.: Design and analysis of DNA strand displacement devices using probabilistic model checking, *Journal of the Royal Society Interface*, Vol. 9, No. 72, pp. 1470–1485 (2012), DOI: `10.1098/rsif.2011.0800`

[32] Sakamoto, K., Gouzu, H., Komiya, K., Kiga, D., Yokoyama, S., Yokomori, T. and Hagiya, M.: Molecular Computation by DNA Hairpin Formation, *Science*, Vol. 288, No. 5469, pp. 1223–1226 (2000), DOI: `10.1126/science.288.5469.1223`

[33] Dirks, R. M. and Pierce, N. A.: Triggered amplification by hybridization chain reaction, *Proceedings of the National Academy of Sciences*, Vol. 101, No. 43, pp. 15275–15278 (2004), DOI: `10.1073/pnas.0407024101`

[34] Yin, P., Choi, H. M. T., Calvert, C. R. and Pierce, N. A.: Programming biomolecular self-assembly pathways, *Nature*, Vol. 451, No. 7176, pp. 318–322 (2008), DOI: `10.1038/nature06451`

[35] Seelig, G., Yurke, B. and Winfree, E.: Catalyzed Relaxation of a Metastable DNA Fuel, *Journal of the American Chemical Society*, Vol. 128, No. 37, pp. 12211–12220 (2006), DOI: `10.1021/ja0635635`

[36] Li, W., Yang, Y., Yan, H. and Liu, Y.: Three-input Majority Logic Gate and Multiple Input Logic Circuit Based on DNA Strand Displacement, *Nano Letters*, Vol. 13, No. 6, pp. 2980–2988 (2013), DOI: `10.1021/nl4016107`

[37] Benenson, Y., Paz-Elizur, T., Adar, R., Keinan, E., Livneh, Z. and Shapiro, E.: Programmable and autonomous computing machine made of biomolecules, *Nature*, Vol. 414, No. 6862, pp. 430–434 (2001), DOI: `10.1038/35106533`

[38] Benenson, Y., Adar, R., Paz-Elizur, T., Livneh, Z. and Shapiro, E.: DNA molecule provides a computing machine with both data and fuel, *Proceed-*

ings of the National Academy of Sciences, Vol. 100, No. 5, pp. 2191–2196 (2003), DOI: `10.1073/pnas.0535624100`

[39] Benenson, Y., Gil, B., Ben-Dor, U., Adar, R. and Shapiro, E.: An autonomous molecular computer for logical control of gene expression, *Nature*, Vol. 429, No. 6990, pp. 423–429 (2004), DOI: `10.1038/nature02551`

[40] Sakamoto, K., Kiga, D., Komiya, K., Gouzu, H., Yokoyama, S., Ikeda, S., Sugiyama, H. and Hagiya, M.: State transitions by molecules, *Biosystems*, Vol. 52, No. 1-3, pp. 81–91 (1999), DOI: `10.1016/S0303-2647(99)00035-0`

[41] Rose, J. A., Komiya, K., Yaegashi, S. and Hagiya, M.: Displacement Whiplash PCR: Optimized Architecture and Experimental Validation, *Lecture Notes in Computer Science*, Vol. 4287, pp. 393–403 (2006), DOI: `10.1007/11925903_31`

[42] Montagne, K., Plasson, R., Sakai, Y., Fujii, T. and Rondelez, Y.: Programming an in vitro DNA oscillator using a molecular networking strategy, *Molecular System Biology*, Vol. 7, p. 466 (2011), DOI: `10.1038/msb.2010.120`

[43] Kitano, H.: Systems Biology: A Brief Overview., *Science*, Vol. 295, No. 5560, pp. 1662–1664 (2002), DOI: `10.1126/science.1069492`

[44] Dematté, L., Priami, C. and Romanel, A.: The Beta Workbench: a computational tool to study the dynamics of biological systems, *Briefings in Bioinformatics*, Vol. 9, No. 5, pp. 437–449 (2008), DOI: `10.1093/bib/bbn023`

[45] Fisher, J. and Henzinger, T. A.: Executable cell biology, *Nature Biotechnology*, Vol. 25, No. 11, pp. 1239–1249 (2007), DOI: `10.1038/nbt1356`

[46] Rosselló, F. and Valiente, G.: Graph Transformation in Molecular Biology, *Lecture Notes in Computer Science*, Vol. 3393, pp. 116–133 (2005), DOI: `10.1007/978-3-540-31847-7_7`

[47] Danos, V., Feret, J., Fontana, W. and Krivine, J.: Scalable Simulation of Cellular Signaling Networks, *Lecture Notes in Computer Science*, Vol. 4807, pp. 139–157 (2007), DOI: `10.1007/978-3-540-76637-7_10`

[48] Mann, M., Ekker, H. and Flamm, C.: The Graph Grammar Library - a generic framework for chemical graph rewrite systems, *Lecture Notes in Computer Science*, Vol. 7909, pp. 52–53 (2013), DOI: `10.1007/978-3-642-38883-5_5`

[49] Harmer, R., Danos, V., Feret, J., Krivine, J. and Fontana, W.: Intrinsic Information Carriers in Combinatorial Dynamical Systems., *Chaos*, Vol. 20, No. 3, p. 037108 (2010), DOI: `10.1063/1.3491100`

[50] Weng, G., Bhalla, U. S. and Iyengar, R.: Complexity in Biological Signaling Systems, *Science*, Vol. 284, No. 5411, pp. 92–96 (1999), DOI: `10.1126/science.284.5411.92`

[51] Feret, J., Danos, V., Krivine, J., Harmer, R. and Fontana, W.: Internal coarse-graining of molecular systems, *Proceedings of the National Academy of Sciences*, Vol. 106, No. 16, pp. 6453–6458 (2009), DOI: `10.1073/pnas.0809908106`

[52] Danos, V., Feret, J., Fontana, W., Harmer, R. and Krivine, J.: Abstracting the differential semantics of rule-based models: exact and automated model reduction, in *Proceedings of the Twenty-Fifth Annual IEEE Symposium on Logic in Computer Science*, pp. 362–381IEEE (2010).

[53] Conzelmann, H., Saez-Rodriguez, J., Sauter, T., Kholodenko, B. N. and Gilles, E. D.: A domain-oriented approach to the reduction of combinatorial complexity in signal transduction networks, *BMC Bioinformatics*, Vol. 7, p. 34 (2006), DOI: `10.1186/1471-2105-7-34`

[54] Borisov, N. M., Chistopolsky, A. S., Faeder, J. R. and Kholodenko, B. N.: Domain-oriented reduction of rule-based network models, *IET System Biology*, Vol. 2, No. 5, pp. 342–351 (2008), DOI: `10.1049/iet-syb:20070081`

[55] Radulescu, O., Gorban, A. N., Zinovyev, A. and Noel, V.: Reduction of dynamical biochemical reactions networks in computational biology, *Frontiers in Genetics*, Vol. 3, pp. 1–17 (2012), DOI: `10.3389/fgene.2012.00131`

[56] Kobayashi, S.: Enumeration approach to computing chemical equilibria, *Theoretical Computer Science*, Vol. 499, pp. 51–87 (2013), DOI: `10.1016/j.tcs.2013.05.017`

[57] Qian, L. and Winfree, E.: Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades., *Science*, Vol. 332, No. 6034, pp. 1196–1201 (2011), DOI: `10.1126/science.1200520`

[58] Gardner, T. S., Cantor, C. R. and Collins, J. J.: Construction of a genetic toggle switch in Escherichia coli, *Nature*, Vol. 403, No. 6767, pp. 339–342 (2000), DOI: `10.1038/35002131`

[59] Elowitz, M. B. and Leibler, S.: A synthetic oscillatory network of transcriptional regulators, *Nature*, Vol. 403, No. 6767, pp. 335–338 (2000), DOI: `10.1038/35002125`

[60] Cao, H., Romero-Campero, F. J., Heeb, S., Cámara, M. and Krasnogor, N.: Evolving cell models for systems and synthetic biology, *Systems and Synthetic Biology*, Vol. 4, No. 1, pp. 55–84 (2010), DOI: `10.1007/s11693-009-9050-7`

[61] Kobayashi, Y., Shibata, T., Kuramoto, Y. and Mikhailov, A. S.: Robust network clocks: Design of genetic oscillators as a complex combinatorial optimization problem, *Physical Review*, Vol. 83, No. 6 Pt 1, p. 060901 (2011), DOI: `10.1103/PhysRevE.83.060901`

[62] Drennan, B. and Beer, R. D.: Evolution of repressilators using a biologically-motivated model of gene expression, *Proceedings of the Tenth International Conference on the Simulation and Synthesis of Living Systems*, Vol. 10, pp. 22–27 (2006).

[63] Chu, D.: Evolving Genetic Regulatory Networks for Systems Biology, in *Proceedings of IEEE Congress on Evolutionary Computation 2007 (CEC2007)*, pp. 875–882 (2007).

[64] Kobayashi, Y., Shibata, T., Kuramoto, Y. and Mikhailov, A. S.: Evolutionary design of oscillatory genetic networks, *The European Physical Journal B*, Vol. 76, No. 1, pp. 167–178 (2010), DOI: `10.1140/epjb/e2010-00200-9`

[65] Marchisio, M. A. and Stelling, J.: Automatic Design of Digital Synthetic Gene Circuits, *PLoS Computational Biology*, Vol. 7, No. 2, p. e1001083 (2011), DOI: `10.1371/journal.pcbi.1001083`

[66] Feynman, R. P.: There's Plenty of Room at the Bottom, *Engineering and Science*, Vol. 23, No. 5, pp. 22–36 (1960).

[67] Channon, K., Bromley, E. H. C. and Woolfson, D. N.: Synthetic biology through biomolecular design and engineering, *Current Opinion in Structual Biology*, Vol. 18, No. 4, pp. 491–498 (2008), DOI: `10.1016/j.sbi.2008.06.006`

[68] Seeman, N. C.: Nucleic Acid Junctions and Lattices, *Journal of Theoretical Biology*, Vol. 99, No. 2, pp. 237–247 (1982), DOI: `10.1016/0022-5193(82)90002-9`

[69] Chen, J. and Seeman, N. C.: Synthesis from DNA of a molecule with the connectivity of a cube, *Nature*, Vol. 350, No. 6319, pp. 631–633 (1991), DOI: `10.1038/350631a0`

[70] Goodman, R. P., Schaap, I. A. T., Tardin, C. F., Erben, C. M., Berry, R. M., Schmidt, C. F. and Turberfield, A. J.: Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication, *Science*, Vol. 310, No. 5754, pp. 1661–1665 (2005), DOI: `10.1126/science.1120367`

[71] Zhang, Y. and Seeman, N. C.: Construction of a DNA-Truncated Octahedron, *Journal of the American Chemical Society*, Vol. 116, No. 5, pp. 1661–1669 (1994), DOI: `10.1021/ja00084a006`

[72] Mao, C., Sun, W. and Seeman, N. C.: Assembly of Borromean Rings from DNA, *Nature*, Vol. 386, No. 6621, pp. 137–138 (1997), DOI: `10.1038/386137b0`

[73] Wang, X. and Seeman, N. C.: Assembly and Characterization of 8-Arm and 12-Arm DNA Branched Junctions, *Journal of the American Chemical Society*, Vol. 129, No. 26, pp. 8169–8176 (2007), DOI: `10.1021/ja0693441`

[74] Bath, J. and Turberfield, A. J.: DNA nanomachines, *Nature Nanotechnology*, Vol. 2, No. 5, pp. 275–284 (2007), DOI: `10.1038/nnano.2007.104`

[75] Seeman, N. C.: Nanomaterials based on DNA, *Annual Review of Biochemistry*, Vol. 79, pp. 65–87 (2010), DOI: `10.1146/annurev-biochem-060308-102244`

[76] Li, H., LaBean, T. H. and Leong, K. W.: Nucleic acid-based nanoengineering: novel structures for biomedical applications, *Interface Focus*, Vol. 1, No. 5, pp. 702–724 (2011), DOI: `10.1098/rsfs.2011.0040`

[77] Pinheiro, A. V., Han, D., Shih, W. M. and Yan, H.: Challenges and opportunities for structural DNA nanotechnology, *Nature Nanotechnology*, Vol. 6, No. 12, pp. 763–772 (2011), DOI: `10.1038/nnano.2011.187`

[78] Tørring, T., Voigt, N. V., Nangreave, J., Yan, H. and Gothelf, K. V.: DNA origami: a quantum leap for self-assembly of complex structures, *Chemical Society Reviews*, Vol. 40, No. 12, pp. 5636–5646 (2011), DOI: `10.1039/c1cs15057j`

[79] Zadegan, R. M. and Norton, M. L.: Structural DNA Nanotechnology: From Design to Applications, *International Journal of Molecular Sciences*, Vol. 13, No. 6, pp. 7149–7162 (2012), DOI: `10.3390/ijms13067149`

[80] Linko, V. and Dietz, H.: The enabled state of DNA nanotechnology, *Current Opinion in Biotechnology*, Vol. 24, pp. 1–7 (2013), DOI: `10.1016/j.copbio.2013.02.001`

[81] He, Y., Ye, T., Su, M., Zhang, C., Ribbe, A. E., Jiang, W. and Mao, C.: Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra, *Nature*, Vol. 452, No. 7184, pp. 198–201 (2008), DOI: `10.1038/nature06597`

[82] Zhang, C., Su, M., He, Y., Zhao, X., Fang, an P., Ribbe, A. E., Jiang, W. and Mao, C.: Conformational flexibility facilitates self-assembly of complex DNA nanostructures, *Proceedings of the National Academy of Sciences*, Vol. 105, No. 31, pp. 10665–10669 (2008), DOI: `10.1073/pnas.0803841105`

[83] Hamada, S. and Murata, S.: Substrate-Assisted Assembly of Interconnected Single-Duplex DNA Nanostructures, *Angewandte Chemie International Edition*, Vol. 48, No. 37, pp. 6820–6823 (2009), DOI: `10.1002/anie.200902662`

[84] Yin, P., Hariadi, R. F., Sahu, S., Choi, H. M. T., Park, S. H., Labean, T. H. and Reif, J. H.: Programming DNA Tube Circumferences, *Science*, Vol. 321, No. 5890, pp. 824–826 (2008), DOI: `10.1126/science.1157312`

[85] Yang, Y., Zhao, Z., Zhang, F., Nangreave, J., Liu, Y. and Yan, H.: Self-Assembly of DNA Rings from Scaffold-Free DNA Tiles, *Nano Letters*, Vol. 13, No. 4, pp. 1862–1866 (2013), DOI: `10.1021/nl400859d`

[86] Zheng, J., Birktoft, J. J., Chen, Y., Wang, T., Sha, R., Constantinou, P. E., Ginell, S. L., Mao, C. and Seeman, N. C.: From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal, *Nature*, Vol. 461, No. 7260, pp. 74–77 (2009), DOI: `10.1038/nature08274`

[87] Wei, B., Dai, M. and Yin, P.: Complex shapes self-assembled from single-stranded DNA tiles, *Nature*, Vol. 485, No. 7400, pp. 623–626 (2012), DOI: `10.1038/nature11075`

[88] Ke, Y., Ong, L. L., Shih, W. M. and Yin, P.: Three-Dimensional Structures Self-Assembled from DNA Bricks, *Science*, Vol. 338, No. 6111, pp. 1177–1183 (2012), DOI: `10.1126/science.1227268`

[89] Park, S. H., Yin, P., Liu, Y., Reif, J. H., LaBean, T. H. and Yan, H.: Programmable DNA Self-Assemblies for Nanoscale Organization of Lig-

ands and Proteins, *Nano Letters*, Vol. 5, No. 4, pp. 729–733 (2005), DOI: `10.1021/nl050175c`

[90] Li, X., Zhang, C., Hao, C., Tian, C., Wang, G. and Mao, C.: DNA Polyhedra with T-Linkage, *ACS Nano*, Vol. 6, No. 6, pp. 5138–5142 (2012), DOI: `10.1021/nn300813w`

[91] Zhang, C., Tian, C., Guo, F., Liu, Z., Jiang, W. and Mao, C.: DNA-Directed Three-Dimensional Protein Organization, *Angewandte Chemie International Edition*, Vol. 51, No. 14, pp. 3382–3385 (2012), DOI: `10.1002/anie.201108710`

[92] Rothemund, P. W. K., Papadakis, N. and Winfree, E.: Algorithmic Self-Assembly of DNA Sierpinski Triangles, *PLoS Biology*, Vol. 2, No. 12, p. e424 (2004), DOI: `10.1371/journal.pbio.0020424`

[93] Tanaka, F., Mochizuki, T., Liang, X., Asanuma, H., Tanaka, S., Suzuki, K., Kitamura, ichi S., Nishikawa, A., Ui-Tei, K. and Hagiya, M.: Robust and Photocontrollable DNA Capsules Using Azobenzenes, *Nano Letters*, Vol. 10, No. 9, pp. 3560–3565 (2010), DOI: `10.1021/nl101829m`

[94] Evans, C. G., Hariadi, R. F. and Winfree, E.: Direct Atomic Force Microscopy Observation of DNA Tile Crystal Growth at the Single-Molecule Level, *Journal of the American Chemical Society*, Vol. 134, No. 25, pp. 10485–10492 (2012), DOI: `10.1021/ja301026z`

[95] Winfree, E.: *Algorithmic Self-Assembly of DNA*, PhD thesis, California Institute of Technology (1998).

[96] Ma, X. and Lombardi, F.: Synthesis of Tile Sets for DNA Self-Assembly, *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 27, pp. 963–967 (2008).

[97] Dibrov, S. M., McLean, J., Parsons, J. and Hermann, T.: Self-assembling RNA square, *Proceedings of the National Academy of Sciences*, Vol. 108, No. 16, pp. 6405–6408 (2011), DOI: `10.1073/pnas.1017999108`

[98] Severcan, I., Geary, C., Chworos, A., Voss, N., Jacovetty, E. and Jaeger, L.: A polyhedron made of tRNAs, *Nature Chemistry*, Vol. 2, No. 9, pp. 772–779 (2010), DOI: `10.1038/nchem.733`

[99] Chworos, A., Severcan, I., Koyfman, A. Y., Weinkam, P., Oroudjev, E., Hansma, H. G. and Jaeger, L.: Building Programmable Jigsaw Puzzles

with RNA, *Science*, Vol. 306, No. 5704, pp. 2068–2072 (2004), DOI: `10.1126/science.1104686`

[100] Rothemund, P. W. K.: Folding DNA to create nanoscale shapes and patterns, *Nature*, Vol. 440, No. 7082, pp. 297–302 (2006), DOI: `10.1038/nature04586`

[101] Han, D., Pal, S., Liu, Y. and Yan, H.: Folding and cutting DNA into reconfigurable topological nanostructures, *Nature Nanotechnology*, Vol. 5, No. 10, pp. 712–717 (2010), DOI: `10.1038/nnano.2010.193`

[102] Andersen, E. S., Dong, M., Nielsen, M. M., Jahn, K., Subramani, R., Mamdouh, W., Golas, M. M., Sander, B., Stark, H., Oliveira, C. L. P., Pedersen, J. S., Birkedal, V., Besenbacher, F., Gothelf, K. V. and Kjems, J.: Self-assembly of a nanoscale DNA box with a controllable lid, *Nature*, Vol. 459, No. 7243, pp. 73–76 (2009), DOI: `10.1038/nature07971`

[103] Douglas, S. M., Dietz, H., Liedl, T., Högberg, B., Graf, F. and Shih, W. M.: Self-assembly of DNA into nanoscale three-dimensional shapes, *Nature*, Vol. 459, No. 7245, pp. 414–418 (2009), DOI: `10.1038/nature08016`

[104] Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y. and Yan, H.: DNA Origami with Complex Curvatures in Three-Dimensional Space, *Science*, Vol. 332, No. 6027, pp. 342–346 (2011), DOI: `10.1126/science.1202998`

[105] Wang, R., Liu, W. and Seeman, N. C.: Prototyping Nanorod Control: A DNA Double Helix Sheathed within a DNA Six-Helix Bundle, *Chemistry & Biology*, Vol. 16, No. 8, pp. 862–867 (2009), DOI: `10.1016/j.chembiol.2009.07.008`

[106] Fu, Y., Zeng, D., Chao, J., Jin, Y., Zhang, Z., Liu, H., Li, D., Ma, H., Huang, Q., Gothelf, K. V. and Fan, C.: Single-Step Rapid Assembly of DNA Origami Nanostructures for Addressable Nanoscale Bioreactors, *Journal of the American Chemical Society*, Vol. 135, No. 2, pp. 696–702 (2013), DOI: `10.1021/ja3076692`

[107] Woo, S. and Rothemund, P. W. K.: Programmable molecular recognition based on the geometry of DNA nanostructures, *Nature Chemistry*, Vol. 3, No. 8, pp. 620–627 (2011), DOI: `10.1038/nchem.1070`

[108] Ke, Y., Voigt, N. V., Gothelf, K. V. and Shih, W. M.: Multilayer DNA Origami Packed on Hexagonal and Hybrid Lattices, *Journal of the Amer-*

ican Chemical Society, Vol. 134, No. 3, pp. 1770–1774 (2012), DOI: `10.1021/ja209719k`

[109] Langecker, M., Arnaut, V., Martin, T. G., List, J., Renner, S., Mayer, M., Dietz, H. and Simmel, F. C.: Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures, *Science*, Vol. 338, No. 6109, pp. 932–936 (2012), DOI: `10.1126/science.1225624`

[110] Hernández-Ainsa, S., Bell, N. A. W., Thacker, V. V., Göpfrich, K., Misiunas, K., Fuentes-Perez, M. E., Moreno-Herrero, F. and Keyser, U. F.: DNA Origami Nanopores for Controlling DNA Translocation, *ACS Nano*, Vol. 7, No. 7, pp. 6024–6030 (2013), DOI: `10.1021/nn401759r`

[111] Bell, N. A. W., Engst, C. R., Ablay, M., Divitini, G., Ducati, C., Liedl, T. and Keyser, U. F.: DNA Origami Nanopores, *Nano Letters*, Vol. 12, No. 1, pp. 512–517 (2012), DOI: `10.1021/nl204098n`

[112] Wei, R., Martin, T. G., Rant, U. and Dietz, H.: DNA Origami Gatekeepers for Solid-State Nanopores, *Angewandte Chemie International Edition*, Vol. 51, No. 20, pp. 4864–4867 (2012), DOI: `10.1002/anie.201200688`

[113] Johnson-Buck, A., Nangreave, J., Jiang, S., Yan, H. and Walter, N. G.: Multifactorial Modulation of Binding and Dissociation Kinetics on Two-Dimensional DNA Nanostructures, *Nano Letters*, Vol. 13, No. 6, pp. 2754–2759 (2013), DOI: `10.1021/nl400976s`

[114] Lin, C., Jungmann, R., Leifer, A. M., Li, C., Levner, D., Church, G. M., Shih, W. M. and Yin, P.: Submicrometre geometrically encoded fluorescent barcodes self-assembled from DNA, *Nature Chemistry*, Vol. 4, No. 10, pp. 832–839 (2012), DOI: `10.1038/nchem.1451`

[115] Ko, S. H., Du, K. and Liddle, J. A.: Quantum-Dot Fluorescence Lifetime Engineering with DNA Origami Constructs, *Angewandte Chemie International Edition*, Vol. 52, No. 4, pp. 1193–1197 (2013), DOI: `10.1002/anie.201206253`

[116] Kuzuya, A., Koshi, N., Kimura, M., Numajiri, K., Yamazaki, T., Ohnishi, T., Okada, F. and Komiyama, M.: Programmed Nanopatterning of Organic/Inorganic Nanoparticles Using Nanometer-Scale Wells Embedded in a DNA Origami Scaffold, *Small*, Vol. 6, No. 23, pp. 2664–2667 (2010), DOI: `10.1002/smll.201001484`

[117] Fu, J., Liu, M., Liu, Y., Woodbury, N. W. and Yan, H.: Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures, *Journal of the American Chemical Society*, Vol. 134, No. 12, pp. 5516–5519 (2012), DOI: `10.1021/ja300897h`

[118] Acuna, G. P., Möller, F. M., Holzmeister, P., Beater, S., Lalkens, B. and Tinnefeld, P.: Fluorescence Enhancement at Docking Sites of DNA-Directed Self-Assembled Nanoantennas, *Science*, Vol. 338, No. 6106, pp. 506–510 (2012), DOI: `10.1126/science.1228638`

[119] Shen, X., Asenjo-Garcia, A., Liu, Q., Jiang, Q., García de Abajo, F. J., Liu, N. and Ding, B.: Three-Dimensional Plasmonic Chiral Tetramers Assembled by DNA Origami, *Nano Letters*, Vol. 13, No. 5, pp. 2128–2133 (2013), DOI: `10.1021/nl400538y`

[120] Lan, X., Chen, Z., Dai, G., Lu, X., Ni, W. and Wang, Q.: Bifacial DNA Origami-Directed Discrete, Three-Dimensional, Anisotropic Plasmonic Nanoarchitectures with Tailored Optical Chirality, *Journal of the American Chemical Society*, Vol. 135, No. 31, pp. 11441–11444 (2013), DOI: `10.1021/ja404354c`

[121] Mangalum, A., Rahman, M. and Norton, M. L.: Site-Specific Immobilization of Single-Walled Carbon Nanotubes onto Single and One-Dimensional DNA Origami, *Journal of the American Chemical Society*, Vol. 135, No. 7, pp. 2451–2454 (2013), DOI: `10.1021/ja312191a`

[122] Sobczak, J.-P. J., Martin, T. G., Gerling, T. and Dietz, H.: Rapid Folding of DNA into Nanoscale Shapes at Constant Temperature, *Science*, Vol. 338, No. 6113, pp. 1458–1461 (2012), DOI: `10.1126/science.1229919`

[123] Zhang, Z., Song, J., Besenbacher, F., Dong, M. and Gothelf, K. V.: Self-Assembly of DNA Origami and Single-Stranded Tile Structures at Room Temperature, *Angewandte Chemie International Edition*, Vol. 52, No. 35, pp. 9219–9223 (2013), DOI: `10.1002/anie.201303611`

[124] Jungmann, R., Liedl, T., Sobey, T. L., Shih, W. and Simmel, F. C.: Isothermal Assembly of DNA Origami Structures Using Denaturing Agents, *Journal of the American Chemical Society*, Vol. 130, No. 31, pp. 10062–10063 (2008), DOI: `10.1021/ja8030196`

[125] Eskelinen, A.-P., Rosilo, H., Kuzyk, A., Törmä, P. and Kostiainen, M. A.: Controlling the Formation of DNA Origami Structures with Exter-

nal Signals, *Small*, Vol. 8, No. 13, pp. 2016–2020 (2012), DOI: `10.1002/smll.201102697`

[126] Lin, C., Perrault, S. D., Kwak, M., Graf, F. and Shih, W. M.: Purification of DNA-origami nanostructures by rate-zonal centrifugation, *Nucleic Acids Research*, Vol. 41, No. 2, p. e40 (2012), DOI: `10.1093/nar/gks1070`

[127] Song, J., Zhang, Z., Zhang, S., Liu, L., Li, Q., Xie, E., Gothelf, K. V., Besenbacher, F. and Dong, M.: Isothermal Hybridization Kinetics of DNA Assembly of Two-Dimensional DNA Origami, *Small*, Vol. 9, No. 17, pp. 2954–2959 (2013), DOI: `10.1002/smll.201202861`

[128] Wei, X., Nangreave, J., Jiang, S., Yan, H. and Liu, Y.: Mapping the Thermal Behavior of DNA Origami Nanostructures, *Journal of the American Chemical Society*, Vol. 135, No. 16, pp. 6165–6176 (2013), DOI: `10.1021/ja4000728`

[129] Bai, X.-C., Martin, T. G., Scheres, S. H. W. and Dietz, H.: Cryo-EM structure of a 3D DNA-origami object, *Proceedings of the National Academy of Sciences*, Vol. 109, No. 49, pp. 20012–20017 (2012), DOI: `10.1073/pnas.1215713109`

[130] Said, H., Schller, V. J., Eber, F. J., Wege, C., Liedl, T. and Richert, C.: M1.3 - a small scaffold for DNA origami, *Nanoscale*, Vol. 5, No. 1, pp. 284–290 (2012), DOI: `10.1039/c2nr32393a`

[131] Han, D., Jiang, S., Samanta, A., Liu, Y. and Yan, H.: Unidirectional Scaffold-Strand Arrangement in DNA Origami, *Angewandte Chemie International Edition*, Vol. 52, No. 34, pp. 9031–9034 (2013), DOI: `10.1002/anie.201302177`

[132] Yang, Y., Han, D., Nangreave, J., Liu, Y. and Yan, H.: DNA Origami with Double-Stranded DNA As a Unified Scaffold, *ACS Nano*, Vol. 6, No. 9, pp. 8209–8215 (2012), DOI: `10.1021/nn302896c`

[133] Marchi, A. N., Saaem, I., Tian, J. and LaBean, T. H.: One-Pot Assembly of a Hetero-dimeric DNA Origami from Chip-Derived Staples and Double-Stranded Scaffold, *ACS Nano*, Vol. 7, No. 2, pp. 903–910 (2013), DOI: `10.1021/nn302322j`

[134] Ouyang, X., Li, J., Liu, H., Zhao, B., Yan, J., Ma, Y., Xiao, S., Song, S., Huang, Q., Chao, J. and Fan, C.: Rolling Circle Amplification-Based

DNA Origami Nanostructrures for Intracellular Delivery of Immunostimulatory Drugs, *Small*, Vol. 9, No. 18, pp. 3082–3087 (2013), DOI: 10.1002/smll.201300458

[135] Han, D., Pal, S., Yang, Y., Jiang, S., Nangreave, J., Liu, Y. and Yan, H.: DNA Gridiron Nanostructures Based on Four-Arm Junctions, *Science*, Vol. 339, No. 6126, pp. 1412–1415 (2013), DOI: 10.1126/science.1232252

[136] Douglas, S. M., Marblestone, A. H., Teerapittayanon, S., Vazquez, A., Church, G. M. and Shih, W. M.: Rapid prototyping of 3D DNA-origami shapes with caDNAno, *Nucleic Acids Research*, Vol. 37, No. 15, pp. 5001–5006 (2009), DOI: 10.1093/nar/gkp436

[137] Castro, C. E., Kilchherr, F., Kim, D.-N., Shiao, E. L., Wauer, T., Wortmann, P., Bathe, M. and Dietz, H.: A primer to scaffolded DNA origami, *Nature Methods*, Vol. 8, No. 3, pp. 221–229 (2011), DOI: 10.1038/nmeth.1570

[138] Zhou, C., Luo, H., Feng, X., Li, X., Zhu, J., He, L. and Li, C.: FOLDNA, a Web Server for Self-Assembled DNA Nanostructure Autoscaffolds and Autostaples, *Journal of Nanotechnology*, Vol. 2012, p. 453953 (2012), DOI: 10.1155/2012/453953

[139] Tuerk, C. and Gold, L.: Systematic Evolution of Ligands by Exponential Enrichment: RNA Ligands to Bacteriophage T4 DNA Polymerase, *Science*, Vol. 249, No. 4968, pp. 505–510 (1990), DOI: 10.1126/science.2200121

[140] Ellington, A. D. and Szostak, J. W.: In vitro selection of RNA molecules that bind specific ligands, *Nature*, Vol. 346, No. 6287, pp. 818–822 (1990), DOI: 10.1038/346818a0

[141] Hermann, T. and Patel, D. J.: Adaptive Recognition by Nucleic Acid Aptamers, *Science*, Vol. 287, No. 5454, pp. 820–825 (2000), DOI: 10.1126/science.287.5454.820

[142] Hamaguchi, N., Ellington, A. and Stanton, M.: Aptamer Beacons for the Direct Detection of Proteins, *Analytical Biochemistry*, Vol. 294, No. 2, pp. 126–131 (2001), DOI: 10.1006/abio.2001.5169

[143] Li, J. J., Fang, X. and Tan, W.: Molecular Aptamer Beacons for Real-Time Protein Recognition, *Biochemical and Biophysical Research Communications*, Vol. 292, No. 1, pp. 31–40 (2002).

[144] Heyduk, E. and Heyduk, T.: Nucleic Acid-Based Fluorescence Sensors for Detecting Proteins, *Analytical Chemistry*, Vol. 77, No. 4, pp. 1147–1156 (2005), DOI: `10.1021/ac0487449`

[145] Choi, J. H., Chen, K. H. and Strano, M. S.: Aptamer-Capped Nanocrystal Quantum Dots: A New Method for Label-Free Protein Detection, *Journal of the American Chemical Society*, Vol. 128, No. 49, pp. 15584–15585 (2006), DOI: `10.1021/ja066506k`

[146] Wang, X., Zhou, J., Yun, W., Xiao, S., Chang, Z., He, P. and Fang, Y.: Detection of thrombin using electrogenerated chemiluminescence based on Ru(bpy)3(2+)-doped silica nanoparticle aptasensor via target protein-induced strand displacement, *Analytica Chimica Actamica Acta*, Vol. 598, No. 2, pp. 242–248 (2007), DOI: `10.1016/j.aca.2007.07.050`

[147] Numnuam, A., Chumbimuni-Torres, K. Y., Xiang, Y., Bash, R., Thavarungkul, P., Kanatharana, P., Pretsch, E., Wang, J. and Bakker, E.: Aptamer-Based Potentiometric Measurements of Proteins Using Ion-Selective Microelectrodes, *Analytical Chemistry*, Vol. 80, No. 3, pp. 707–712 (2008), DOI: `10.1021/ac701910r`

[148] Jhaveri, S. D., Kirby, R., Conrad, R., Maglott, E. J., Bowser, M., Kennedy, R. T., Glick, G. and Ellington, A. D.: Designed Signaling Aptamers that Transduce Molecular Recognition to Changes in Fluorescence Intensity, *Journal of the American Chemical Society*, Vol. 122, No. 11, pp. 2469–2473 (2000), DOI: `10.1021/ja992393b`

[149] Nutiu, R. and Li, Y.: Structure-Switching Signaling Aptamers, *Journal of the American Chemical Society*, Vol. 125, No. 16, pp. 4771–4778 (2003), DOI: `10.1021/ja028962o`

[150] Feng, K., Sun, C., Kang, Y., Chen, J., Jiang, J.-H., Shen, G.-L. and Yu, R.-Q.: Label-free electrochemical detection of nanomolar adenosine based on target-induced aptamer displacement, *Electrochemistry Communications*, Vol. 10, No. 4, pp. 531–535 (2008), DOI: `10.1016/j.elecom.2008.01.024`

[151] Li, N. and Ho, C.-M.: Aptamer-Based Optical Probes with Separated Molecular Recognition and Signal Transduction Modules, *Journal of the American Chemical Society*, Vol. 130, No. 8, pp. 2380–2381 (2008), DOI: `10.1021/ja076787b`

[152] Tang, Z., Mallikaratchy, P., Yang, R., Kim, Y., Zhu, Z., Wang, H. and Tan, W.: Aptamer Switch Probe Based on Intramolecular Displacement, *Journal of the American Chemical Society*, Vol. 130, No. 34, pp. 11268–11269 (2008), DOI: `10.1021/ja804119s`

[153] Stojanovic, M. N., de Prada, P. and Landry, D. W.: Fluorescent Sensors Based on Aptamer Self-Assembly, *Journal of the American Chemical Society*, Vol. 122, No. 46, pp. 11547–11548 (2000), DOI: `10.1021/ja0022223`

[154] Stojanovic, M. N., de Prada, P. and Landry, D. W.: Aptamer-Based Folding Fluorescent Sensor for Cocaine, *Journal of the American Chemical Society*, Vol. 123, No. 21, pp. 4928–4931 (2001), DOI: `10.1021/ja0038171`

[155] Baker, B. R., Lai, R. Y., Wood, M. S., Doctor, E. H., Heeger, A. J. and Plaxco, K. W.: An Electronic, Aptamer-Based Small-Molecule Sensor for the Rapid, Label-Free Detection of Cocaine in Adulterated Samples and Biological Fluids, *Journal of the American Chemical Society*, Vol. 128, No. 10, pp. 3138–3139 (2006), DOI: `10.1021/ja056957p`

[156] Green, L. S., Jellinek, D., Jenison, R., Östman, A., Heldin, C.-H. and Janjic, N.: Inhibitory DNA Ligands to Platelet-Derived Growth Factor B-Chain, *Biochemistry*, Vol. 35, No. 45, pp. 14413–14424 (1996), DOI: `10.1021/bi961544+`

[157] Fang, X., Sen, A., Vicens, M. and Tan, W.: Synthetic DNA Aptamers to Detect Protein Molecular Variants in a High-Throughput Fluorescence Quenching Assay, *Chembiochem*, Vol. 4, No. 9, pp. 829–834 (2003), DOI: `10.1002/cbic.200300615`

[158] Chinnapen, D. J. F. and Sen, D.: Hemin-Stimulated Docking of Cytochrome c to a Hemin-DNA Aptamer Complex, *Biochemistry*, Vol. 41, No. 16, pp. 5202–5212 (2002), DOI: `10.1021/bi015785f`

[159] Xiao, Y., Pavlov, V., Niazov, T., Dishon, A., Kotler, M. and Willner, I.: Catalytic Beacons for the Detection of DNA and Telomerase Activity, *Journal of the American Chemical Society*, Vol. 126, No. 24, pp. 7430–7431 (2004), DOI: `10.1021/ja031875r`

[160] Shangguan, D., Li, Y., Tang, Z., Cao, Z. C., Chen, H. W., Mallikaratchy, P., Sefah, K., Yang, C. J. and Tan, W.: Aptamers evolved from live cells as effective molecular probes for cancer study, *Proceedings of the National*

Academy of Sciences, Vol. 103, No. 32, pp. 11838–11843 (2006), DOI: `10.1073/pnas.0602615103`

[161] Xiong, X., Liu, H., Zhao, Z., Altman, M. B., Lopez-Colon, D., Yang, C. J., Chang, L.-J., Liu, C. and Tan, W.: DNA Aptamer-Mediated Cell Targeting, *Angewandte Chemie International Edition*, Vol. 52, No. 5, pp. 1472–1476 (2013), DOI: `10.1002/anie.201207063`

[162] Conrad, R., Keranen, L. M., Ellington, A. D. and Newton, A. C.: Isozyme-specific Inhibition of Protein Kinase C by RNA Aptamers, *Journal of Biological Chemistry*, Vol. 269, No. 51, pp. 32051–32054 (1994).

[163] Yamamoto, R. and Kumar, P. K. R.: Molecular beacon aptamer fluoresces in the presence of Tat protein of HIV-1, *Genes to Cells*, Vol. 5, No. 5, pp. 389–396 (2000), DOI: `10.1046/j.1365-2443.2000.00331.x`

[164] Grate, D. and Wilson, C.: Laser-mediated, site-specific inactivation of RNA transcripts, *Proceedings of the National Academy of Sciences*, Vol. 96, No. 11, pp. 6131–6136 (1999), DOI: `10.1073/pnas.96.11.6131`

[165] Kolpashchikov, D. M.: Binary Malachite Green Aptamer for Fluorescent Detection of Nucleic Acids, *Journal of the American Chemical Society*, Vol. 127, No. 36, pp. 12442–12443 (2005), DOI: `10.1021/ja0529788`

[166] Jenison, R. D., Gill, S. C., Pardi, A. and Polisky, B.: High-resolution molecular discrimination by RNA, *Science*, Vol. 263, No. 5152, pp. 1425–1429 (1994), DOI: `10.1126/science.7510417`

[167] Lupold, S. E., Hicke, B. J., Lin, Y. and Coffey, D. S.: Identification and Characterization of Nuclease-stabilized RNA Molecules That Bind Human Prostate Cancer Cells via the Prostate-specific Membrane Antigen, *Cancer Research*, Vol. 62, No. 14, pp. 4029–4033 (2002).

[168] Chu, T. C., Shieh, F., Lavery, L. A., Levy, M., Richards-Kortum, R., Korgel, B. A. and Ellington, A. D.: Labeling tumor cells with fluorescent nanocrystal-aptamer bioconjugates, *Biosensors and Bioelectronics*, Vol. 21, No. 10, pp. 1859–1866 (2006), DOI: `10.1016/j.bios.2005.12.015`

[169] Potyrailo, R. A., Conrad, R. C., Ellington, A. D. and Hieftje, G. M.: Adapting Selected Nucleic Acid Ligands Aptamers to Biosensors, *Analytical Chemistry*, Vol. 70, No. 16, pp. 3419–3425 (1998), DOI: `10.1021/ac9802325`

[170] Stojanovic, M. N. and Kolpashchikov, D. M.: Modular aptameric Sensors, *Journal of the American Chemical Society*, Vol. 126, No. 30, pp. 9266–9270 (2004), DOI: `10.1021/ja032013t`

[171] Yoshida, W., Sode, K. and Ikebukuro, K.: Aptameric Enzyme Subunit for Biosensing Based on Enzymatic Activity Measurement, *Analytical Chemistry*, Vol. 78, No. 10, pp. 3296–3303 (2006), DOI: `10.1021/ac060254o`

[172] Ikebukuro, K., Kiyohara, C. and Sode, K.: Novel electrochemical sensor system for protein using the aptamers in sandwich manner, *Biosensors and Bioelectronics*, Vol. 20, No. 10, pp. 2168–2172 (2005), DOI: `10.1016/j.bios.2004.09.002`

[173] Wang, W., Chen, C., Qian, M. and Zhao, X. S.: Aptamer biosensor for protein detection using gold nanoparticles, *Analytical Biochemistry*, Vol. 373, No. 2, pp. 213–219 (2008), DOI: `10.1016/j.ab.2007.11.013`

[174] Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A. and Feigon, J.: Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution, *Proceedings of the National Academy of Sciences*, Vol. 90, No. 8, pp. 3745–3749 (1993), DOI: `10.1073/pnas.90.8.3745`

[175] Padmanabhan, K., Padmanabhan, K. P., Ferrara, J. D., Sadler, J. E. and Tulinsky, A.: The Structure of α-Thrombin Inhibited by a 15-Mer Single-Stranded DNA Aptamer, *Journal of Biological Chemistry*, Vol. 268, No. 24, pp. 17651–17654 (1993).

[176] Lin, C. H. and Patel, D. J.: Structural basis of DNA folding and recognition in an AMP-DNA aptamer complex: distinct architectures but common recognition motifs for DNA and RNA aptamers complexed to AMP, *Chemistry & Biology*, Vol. 4, No. 11, pp. 817–832 (1997), DOI: `10.1016/S1074-5521(97)90115-0`

[177] Nonin-Lecomte, S., Lin, C. H. and Patel, D. J.: Additional Hydrogen Bonds and Base-Pair Kinetics in the Symmetrical AMP-DNA Aptamer Complex, *Biophysical Journal*, Vol. 81, No. 6, pp. 3422–3431 (2001), DOI: `10.1016/S0006-3495(01)75974-7`

[178] Nakamura, I., Shi, A.-C., Nutiu, R., Yu, J. M. Y. and Li, Y.: Kinetics of signaling-DNA-aptamer-ATP binding, *Physical Review*, Vol. 79, No. 3 Pt 1, p. 031906 (2009), DOI: `10.1103/PhysRevE.79.031906`

[179] Chushak, Y. and Stone, M. O.: In silico selection of RNA aptamers, *Nucleic Acids Research*, Vol. 37, No. 12, p. e87 (2009), DOI: `10.1093/nar/gkp408`

[180] Luo, X., McKeague, M., Pitre, S., Dumontier, M., Green, J., Golshani, A., Derosa, M. C. and Dehne, F.: Computational approaches toward the design of pools for the in vitro selection of complex aptamers, *RNA*, Vol. 16, No. 11, pp. 2252–2262 (2010), DOI: `10.1261/rna.2102210`

[181] Li, Y. and Breaker, R. R.: Deoxyribozymes: new players in the ancient game of biocatalysis, *Current Opinion in Structural Biology*, Vol. 9, No. 3, pp. 315–323 (1999), DOI: `10.1016/S0959-440X(99)80042-6`

[182] Velez, T. E., Singh, J., Xiao, Y., Allen, E. C., Wong, O. Y., Chandra, M., Kwon, S. C. and Silverman, S. K.: Systematic Evaluation of the Dependence of Deoxyribozyme Catalysis on Random Region Length, *ACS Combinatorial Science*, Vol. 14, No. 12, pp. 680–687 (2012), DOI: `10.1021/co300111f`

[183] Gu, H., Furukawa, K., Weinberg, Z., Berenson, D. F. and Breaker, R. R.: Small, Highly Active DNAs That Hydrolyze DNA, *Journal of the American Chemical Society*, Vol. 135, No. 24, pp. 9121–9129 (2013), DOI: `10.1021/ja403585e`

[184] Willner, I., Shlyahovsky, B., Zayats, M. and Willner, B.: DNAzymes for sensing, nanobiotechnology and logic gate applications, *Chemical Society Reviews*, Vol. 37, No. 6, pp. 1153–1165 (2008), DOI: `10.1039/b718428j`

[185] Sando, S., Sasaki, T., Kanatani, K. and Aoyama, Y.: Amplified Nucleic Acid Sensing Using Programmed Self-Cleaving DNAzyme, *Journal of the American Chemical Society*, Vol. 125, No. 51, pp. 15720–15721 (2003), DOI: `10.1021/ja0386492`

[186] Zagorovsky, K. and Chan, W. C. W.: A Plasmonic DNAzyme Strategy for Point-of-Care Genetic Detection of Infectious Pathogens, *Angewandte Chemie International Edition*, Vol. 52, No. 11, pp. 3168–3171 (2013), DOI: `10.1002/anie.201208715`

[187] Xiao, Y., Rowe, A. A. and Plaxco, K. W.: Electrochemical Detection of Parts-Per-Billion Lead via an Electrode-Bound DNAzyme Assembly, *Journal of the American Chemical Society*, Vol. 129, No. 2, pp. 262–263 (2007), DOI: `10.1021/ja067278x`

[188] Stojanovic, M. N., Mitchell, T. E. and Stefanovic, D.: Deoxyribozyme-Based Logic Gates, *Journal of the American Chemical Society*, Vol. 124, No. 14, pp. 3555–3561 (2002), DOI: `10.1021/ja016756v`

[189] Kolpashchikov, D. M. and Stojanovic, M. N.: Boolean Control of Aptamer Binding States, *Journal of the American Chemical Society*, Vol. 127, No. 32, pp. 11348–11351 (2005), DOI: `10.1021/ja051362f`

[190] Macdonald, J., Li, Y., Sutovic, M., Lederman, H., Pendri, K., Lu, W., Andrews, B. L., Stefanovic, D. and Stojanovic, M. N.: Medium Scale Integration of Molecular Logic Gates in an Automaton, *Nano Letters*, Vol. 6, No. 11, pp. 2598–2603 (2006), DOI: `10.1021/nl0620684`

[191] Niazov, T., Pavlov, V., Xiao, Y., Gill, R. and Willner, I.: DNAzyme-functionalized Au nanoparticles for the amplified detection of DNA or telomerase activity, *Nano Letters*, Vol. 4, No. 9, pp. 1683–1687 (2004), DOI: `10.1021/nl0491428`

[192] Pei, R., Taylor, S. K., Stefanovic, D., Rudchenko, S., Mitchell, T. E. and Stojanovic, M. N.: Behavior of Polycatalytic Assemblies in a Substrate-Displaying Matrix, *Journal of the American Chemical Society*, Vol. 128, No. 39, pp. 12693–12699 (2006), DOI: `10.1021/ja058394n`

[193] Gehring, K., Leroy, J.-L. and Guéron, M.: A tetrameric DNA structure with protonated cytosine-cytosine base pairs, *Nature*, Vol. 363, pp. 561–565 (1993), DOI: `10.1038/363561a0`

[194] Gellert, M., Lipsett, M. N. and Davies, D. R.: HELIX FORMATION BY GUANYLIC ACID, *Proceedings of the National Academy of Sciences*, Vol. 48, No. 12, p. 2013 (1962).

[195] Zhou, J., Amrane, S., Korkut, D. N., Bourdoncle, A., He, H.-Z., Ma, D.-L. and Mergny, J.-L.: Combination of i-Motif and G-Quadruplex Structures within the Same Strand: Formation and Application, *Angewandte Chemie International Edition*, Vol. 52, No. 30, pp. 7742–7746 (2013), DOI: `10.1002/anie.201301278`

[196] Liu, D. and Balasubramanian, S.: A proton-fuelled DNA nanomachine, *Angewandte Chemie International Edition*, Vol. 42, No. 46, pp. 5734–5736 (2003), DOI: `10.1002/anie.200352402`

[197] Shu, W., Liu, D., Watari, M., Riener, C. K., Strunz, T., Welland, M. E., Balasubramanian, S. and McKendry, R. A.: DNA Molecular Motor Driven

Micromechanical Cantilever Arrays, *Journal of the American Chemical Society*, Vol. 127, No. 48, pp. 17054–17060 (2005), DOI: `10.1021/ja0554514`

[198] Bochman, M. L., Paeschke, K. and Zakian, V. A.: DNA secondary structures: stability and function of G-quadruplex structures, *Nature Reviews Genetics*, Vol. 13, No. 11, pp. 770–780 (2012), DOI: `10.1038/nrg3296`

[199] Asanuma, H., Ito, T., Yoshida, T., Liang, X. and Komiyama, M.: Photoregulation of the Formation and Dissociation of a DNA Duplex by Using the cis-trans Isomerization of Azobenzene, *Angewandte Chemie International Edition*, Vol. 38, No. 16, pp. 2393–2395 (1999), DOI: `10.1038/nprot.2006.465`

[200] Liang, X., Nishioka, H., Takenaka, N. and Asanuma, H.: A DNA Nanomachine Powered by Light Irradiation, *Chembiochem*, Vol. 9, No. 5, pp. 702–705 (2008), DOI: `10.1002/cbic.200700649`

[201] Nishioka, H., Liang, X., Kato, T. and Asanuma, H.: A Photon-Fueled DNA Nanodevice that Contains Two Different Photoswitches, *Angewandte Chemie International Edition*, Vol. 51, No. 5, pp. 1165–1168 (2012), DOI: `10.1002/anie.201106093`

[202] Hamad-Schifferli, K., Schwartz, J. J., Santos, A. T., Zhang, S. and Jacobson, J. M.: Remote electronic control of DNA hybridization through inductive coupling to an attached metal nanocrystal antenna, *Nature*, Vol. 415, No. 6868, pp. 152–155 (2002), DOI: `10.1038/415152a`

[203] Stehr, J., Hrelescu, C., Sperling, R. A., Raschke, G., Wunderlich, M., Nichtl, A., Heindl, D., Kürzinger, K., Parak, W. J., Klar, T. A. and Feldmann, J.: Gold Nanostoves for Microsecond DNA Melting Analysis, *Nano Letters*, Vol. 8, No. 2, pp. 619–623 (2008), DOI: `10.1021/nl073028i`

[204] Liu, J. and Lu, Y.: A Colorimetric Lead Biosensor Using DNAzyme-Directed Assembly of Gold Nanoparticles, *Journal of the American Chemical Society*, Vol. 125, No. 22, pp. 6642–6643 (2003), DOI: `10.1021/ja034775u`

[205] Hazarika, P., Ceyhan, B. and Niemeyer, C. M.: Reversible Switching of DNA-Gold Nanoparticle Aggregation, *Angewandte Chemie International Edition*, Vol. 43, No. 47, pp. 6469–6471 (2004), DOI: `10.1002/anie.200461887`

[206] Elbaz, J., Cecconello, A., Fan, Z., Govorov, A. O. and Willner, I.: Powering the programmed nanostructure and function of gold nanoparticles with catenated DNA machines, *Nature Communications*, Vol. 4, p. 2000 (2013), DOI: `10.1038/ncomms3000`

[207] Yurke, B., Turberfield, A. J., Mills, A. P., Jr., Simmel, F. C. and Neumann, J. L.: A DNA-fuelled molecular machine made of DNA, *Nature*, Vol. 406, No. 6796, pp. 605–608 (2000), DOI: `10.1038/35020524`

[208] Shin, J.-S. and Pierce, N. A.: A Synthetic DNA Walker for Molecular Transport, *Journal of the American Chemical Society*, Vol. 126, No. 35, pp. 10834–10835 (2004), DOI: `10.1021/ja047543j`

[209] Wang, C., Tao, Y., Song, G., Ren, J. and Qu, X.: Speeding Up a Bidirectional DNA Walking Device, *Langmuir*, Vol. 28, No. 41, pp. 14829–14837 (2012), DOI: `10.1021/la303332s`

[210] Wickham, S. F. J., Bath, J., Katsuda, Y., Endo, M., Hidaka, K., Sugiyama, H. and Turberfield, A. J.: A DNA-based molecular motor that can navigate a network of tracks, *Nature Nanotechnology*, Vol. 7, No. 3, pp. 169–173 (2012), DOI: `10.1038/nnano.2011.253`

[211] Green, S. J., Bath, J. and Turberfield, A. J.: Coordinated Chemomechanical Cycles: A Mechanism for Autonomous Molecular Motion, *Physical Review Letters*, Vol. 101, No. 23, p. 238101 (2008), DOI: `10.1103/PhysRevLett.101.238101`

[212] Yin, P., Yan, H., Daniell, X. G., Turberfield, A. J. and Reif, J. H.: A Unidirectional DNA Walker That Moves Autonomously along a Track, *Angewandte Chemie International Edition*, Vol. 43, No. 37, pp. 4906–4911 (2004), DOI: `10.1002/anie.200460522`

[213] Tian, Y., He, Y., Chen, Y., Yin, P. and Mao, C.: A DNAzyme That Walks Processively and Autonomously along a One-Dimensional Track, *Angewandte Chemie International Edition*, Vol. 44, No. 28, pp. 4355–4358 (2005), DOI: `10.1002/anie.200500703`

[214] Lu, C.-H., Cecconello, A., Elbaz, J., Credi, A. and Willner, I.: A Three-Station DNA Catenane Rotary Motor with Controlled Directionality, *Nano Letters*, Vol. 13, No. 5, pp. 2303–2308 (2013), DOI: `10.1021/nl401010e`

[215] Mao, C., Sun, W., Shen, Z. and Seeman, N. C.: A nanomechanical device based on the B-Z transition of DNA, *Nature*, Vol. 397, No. 6715, pp. 144–146 (1999), DOI: 10.1038/16437

[216] Zhang, D. Y. and Winfree, E.: Control of DNA Strand Displacement Kinetics Using Toehold Exchange, *Journal of the American Chemical Society*, Vol. 131, No. 47, pp. 17303–17314 (2009), DOI: 10.1021/ja906987s

[217] Genot, A. J., Zhang, D. Y., Bath, J. and Turberfield, A. J.: Remote Toehold: A Mechanism for Flexible Control of DNA Hybridization Kinetics, *Journal of the American Chemical Society*, Vol. 133, No. 7, pp. 2177–2182 (2011), DOI: 10.1021/ja1073239

[218] Adleman, L. M.: Molecular Computation of Solutions to Combinatorial Problems, *Science*, Vol. 266, No. 5187, pp. 1021–1024 (1994), DOI: 10.1126/science.7973651

[219] Benenson, Y.: Biomolecular computing systems: principles, progress and potential, *Nature Reviews Genetics*, Vol. 13, No. 7, pp. 455–468 (2012), DOI: 10.1038/nrg3197

[220] Li, T. and Famulok, M.: I-Motif-Programmed Functionalization of DNA Nanocircles, *Journal of the American Chemical Society*, Vol. 135, No. 4, pp. 1593–1599 (2013), DOI: 10.1021/ja3118224

[221] Jiang, Q., Wang, Z.-G. and Ding, B.: Programmed Colorimetric Logic Devices Based on DNA-Gold Nanoparticle Interactions, *Small*, Vol. 9, No. 7, pp. 1016–1020 (2013), DOI: 10.1002/smll.201201760

[222] Orbach, R., Remacle, F., Levine, R. D. and Willner, I.: Logic reversibility and thermodynamic irreversibility demonstrated by DNAzyme-based Toffoli and Fredkin logic gates, *Proceedings of the National Academy of Sciences*, Vol. 109, No. 52, pp. 21228–21233 (2012), DOI: 10.1073/pnas.1219672110

[223] Soloveichik, D., Seelig, G. and Winfree, E.: DNA as a universal substrate for chemical kinetics, *Proceedings of the National Academy of Sciences*, Vol. 107, No. 12, pp. 5393–5398 (2010), DOI: 10.1073/pnas.0909380107

[224] Kim, J. and Winfree, E.: Synthetic in vitro transcriptional oscillators, *Molecular System Biology*, Vol. 7, p. 465 (2011), DOI: 10.1038/msb.2010.119

[225] Liu, D., Cheng, E. and Yang, Z.: DNA-based switchable devices and materials, *NPG Asia Materials*, Vol. 3, No. 12, pp. 109–114 (2011), DOI: `10.1038/asiamat.2011.147`

[226] Zhang, D. Y. and Seelig, G.: Dynamic DNA nanotechnology using strand-displacement reactions, *Nature Chemistry*, Vol. 3, No. 2, pp. 103–113 (2011), DOI: `10.1038/nchem.957`

[227] Wang, Z.-G., Song, C. and Ding, B.: Functional DNA Nanostructures for Photonic and Biomedical Applications, *Small*, Vol. 9, No. 13, pp. 2210–2222 (2013), DOI: `10.1002/smll.201300141`

[228] Song, C., Wang, Z.-G. and Ding, B.: Smart Nanomachines Based on DNA Self-Assembly, *Small*, Vol. 9, No. 14, pp. 2382–2392 (2013), DOI: `10.1002/smll.201300824`

[229] Murata, S., Konagaya, A., Kobayashi, S., Saito, H. and Hagiya, M.: Molecular Robotics: A New Paradigm for Artifacts, *New Generation Computing*, Vol. 31, pp. 27–45 (2013), DOI: `10.1007/s00354-012-0121-z`

[230] Gu, H., Chao, J., Xiao, S.-J. and Seeman, N. C.: A proximity-based programmable DNA nanoscale assembly line, *Nature*, Vol. 465, No. 7295, pp. 202–205 (2010), DOI: `10.1038/nature09026`

[231] Lund, K., Manzo, A. J., Dabby, N., Michelotti, N., Johnson-Buck, A., Nangreave, J., Taylor, S., Pei, R., Stojanovic, M. N., Walter, N. G., Winfree, E. and Yan, H.: Molecular robots guided by prescriptive landscapes, *Nature*, Vol. 465, No. 7295, pp. 206–210 (2010), DOI: `10.1038/nature09012`

[232] Derr, N. D., Goodman, B. S., Jungmann, R., Leschziner, A. E., Shih, W. M. and Reck-Peterson, S. L.: Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold, *Science*, Vol. 338, pp. 662–665 (2012), DOI: `10.1126/science.1226734`

[233] Ke, Y., Lindsay, S., Chang, Y., Liu, Y. and Yan, H.: Self-Assembled Water-Soluble Nucleic Acid Probe Tiles for Label-Free RNA Hybridization Assays, *Science*, Vol. 319, No. 5860, pp. 180–183 (2008), DOI: `10.1126/science.1150082`

[234] Kuzuya, A., Sakai, Y., Yamazaki, T., Xu, Y. and Komiyama, M.: Nanomechanical DNA origami'single-molecule beacons' directly imaged by atomic force microscopy, *Nature Communications*, Vol. 2, p. 449 (2011), DOI: `10.1038/ncomms1452`

[235] Chen, S. X., Zhang, D. Y. and Seelig, G.: Conditionally fluorescent molecular probes for detecting single base changes in double-stranded DNA, *Nature Chemistry*, Vol. 5, pp. 782–789 (2013), DOI: `10.1038/nchem.1713`

[236] Lee, H., Lytton-Jean, A. K. R., Chen, Y., Love, K. T., Park, A. I., Karagiannis, E. D., Sehgal, A., Querbes, W., Zurenko, C. S., Jayaraman, M., Peng, C. G., Charisse, K., Borodovsky, A., Manoharan, M., Donahoe, J. S., Truelove, J., Nahrendorf, M., Langer, R. and Anderson, D. G.: Molecularly self-assembled nucleic acid nanoparticles for targeted in vivo siRNA delivery, *Nature Nanotechnology*, Vol. 7, No. 6, pp. 389–393 (2012), DOI: `10.1038/nnano.2012.73`

[237] Zhao, Y.-X., Shaw, A., Zeng, X., Benson, E., Nyström, A. M. and Högberg, B.: DNA Origami Delivery System for Cancer Therapy with Tunable Release Properties, *ACS Nano*, Vol. 6, No. 10, pp. 8684–8691 (2012), DOI: `10.1021/nn3022662`

[238] Moradi, M., Babin, V., Roland, C. and Sagui, C.: Reaction path ensemble of the B-Z-DNA transition: a comprehensive atomistic study, *Nucleic Acids Research*, Vol. 41, No. 1, pp. 33–43 (2012), DOI: `10.1093/nar/gks1003`

[239] Edens, L. E., Brozik, J. A. and Keller, D. J.: Coarse-Grained Model DNA: Structure, Sequences, Stems, Circles, Hairpins, *Journal of Physical Chemistry B*, Vol. 116, No. 51, pp. 14735–14743 (2012), DOI: `10.1021/jp3009095`

[240] Knotts IV, T. A., Rathore, N., Schwartz, D. C. and de Pablo, J. J.: A coarse grain model for DNA, *The Journal of Chemical Physics*, Vol. 126, No. 8, p. 084901 (2007), DOI: `10.1063/1.2431804`

[241] Cragnolini, T., Derreumaux, P. and Pasquali, S.: Coarse-Grained Simulations of RNA and DNA Duplexes, *Journal of Physical Chemistry B*, Vol. 117, No. 27, pp. 8047–8060 (2013), DOI: `10.1021/jp400786b`

[242] Šulc, P., Romano, F., Ouldridge, T. E., Rovigatti, L., Doye, J. P. K. and Louis, A. A.: Sequence-dependent thermodynamics of a coarse-grained DNA model, *Journal of Chemical Physics*, Vol. 137, p. 135101 (2012), DOI: `10.1063/1.4754132`

[243] Ouldridge, T. E., Louis, A. A. and Doye, J. P. K.: DNA Nanotweezers Studied with a Coarse-Grained Model of DNA, *Physical Review Letters*, Vol. 104, No. 17, p. 178101 (2010), DOI: `10.1103/PhysRevLett.104.178101`

[244] Ouldridge, T. E., Hoare, R. L., Louis, A. A., Doye, J. P. K., Bath, J. and Turberfield, A. J.: Optimizing DNA Nanotechnology through Coarse-Grained Modeling: A Two-Footed DNA Walker, *ACS Nano*, Vol. 7, No. 3, pp. 2479–2490 (2013), DOI: 10.1021/nn3058483

[245] Nussinov, R. and Jacobson, A. B.: Fast algorithm for predicting the secondary structure of single-stranded RNA, *Proceedings of the National Academy of Sciences*, Vol. 77, No. 11, pp. 6309–6313 (1980).

[246] Zuker, M.: Mfold web server for nucleic acid folding and hybridization prediction, *Nucleic Acids Research*, Vol. 31, No. 13, pp. 3406–3415 (2003), DOI: 10.1093/nar/gkg595

[247] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M. and Pierce, N. A.: NUPACK: Analysis and design of nucleic acid systems, *Journal of Computational Chemistry*, Vol. 32, No. 1, pp. 170–173 (2011), DOI: 10.1002/jcc.21596

[248] Kiss, G., Ölçüm, Çelebi N., Moretti, R., Baker, D. and Houk, K. N.: Computational Enzyme Design, *Angewandte Chemie International Edition*, Vol. 52, No. 22, pp. 5700–5725 (2013), DOI: 10.1002/anie.201204077

[249] Piccolboni, A. and Mauri, G.: Application of Evolutionary Algorithms to Protein Folding Prediction, *Lecture Notes in Computer Science*, Vol. 1363, pp. 123–135 (1998), DOI: 10.1007/BFb0026595

[250] Venkatesan, A., Gopal, J., Candaveloui, M., Gollapalli, S. and Karthikeyan, K.: Computational Approach for Protein Structure Prediction, *Healthcare Informatics Research*, Vol. 19, No. 2, pp. 137–147 (2013), DOI: 10.4258/hir.2013.19.2.137

[251] Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z. and Foldit-players, : Predicting protein structures with a multiplayer online game, *Nature*, Vol. 466, No. 7307, pp. 756–760 (2010), DOI: 10.1038/nature09304

[252] Khatib, F., Cooper, S., Tyka, M. D., Xu, K., Makedon, I., Popovic, Z., Baker, D. and Foldit-players, : Algorithm discovery by protein folding game players, *Proceedings of the National Academy of Sciences*, Vol. 108, No. 47, pp. 18949–18953 (2011), DOI: 10.1073/pnas.1115898108

[253] Eiben, C. B., Siegel, J. B., Bale, J. B., Cooper, S., Khatib, F., Shen, B. W., Foldit-players, , Stoddard, B. L., Popovic, Z. and Baker, D.: Increased

Diels-Alderase activity through backbone remodeling guided by Foldit players, *Nature Biotechnology*, Vol. 30, No. 2, pp. 190–192 (2012), DOI: `10.1038/nbt.2109`

[254] Shin, S.-Y., Lee, I.-H., Kim, D. and Zhang, B.-T.: Multiobjective Evolutionary Optimization of DNA Sequences for Reliable DNA Computing, *IEEE Transactions on Evolutionary Computation*, Vol. 9, No. 2, pp. 143–158 (2005), DOI: `10.1109/TEVC.2005.844166`

[255] Cuccato, G., Polynikis, A., Siciliano, V., Graziano, M., Bernardo, di M. and di Bernardo, D.: Modeling RNA interference in mammalian cells, *BMC Systems Biology*, Vol. 5, p. 19 (2011), DOI: `10.1186/1752-0509-5-19`

[256] Lipardi, C., Wei, Q. and Paterson, B. M.: RNAi as Random Degradative PCR: siRNA Primers Convert mRNA into dsRNAs that Are Degraded to Generate New siRNAs, *Cell*, Vol. 107, No. 3, pp. 297–307 (2001), DOI: `10.1016/S0092-8674(01)00537-2`

[257] Marshall, W. F.: Modeling Recursive RNA Interference, *PLoS Computational Biology*, Vol. 4, No. 9, p. e1000183 (2008), DOI: `10.1371/journal.pcbi.1000183`

[258] Groenenboom, M. A. C., Marée, A. F. M. and Hogeweg, P.: The RNA Silencing Pathway: The Bits and Pieces That Matter., *PLoS Computational Biology*, Vol. 1, No. 2, p. e21 (2005), DOI: `10.1371/journal.pcbi.0010021`

[259] Bergstrom, C. T., McKittrick, E. and Antia, R.: Mathematical models of RNA silencing: Unidirectional amplification limits accidental self-directed reactions, *Proceedings of the National Academy of Sciences*, Vol. 100, No. 20, pp. 11511–11516 (2003), DOI: `10.1073/pnas.1931639100`

[260] Fehlberg, E.: Klassische Runge-Kutta-Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme, *Computing*, Vol. 6, No. 1, pp. 61–71 (1970).

[261] Kawamata, I., Tanaka, F. and Hagiya, M.: Abstraction of DNA Graph Structures for Efficient Enumeration and Simulation, in *The 2011 International Conference on Parallel and Distributed Processing Techniques and Applications*, pp. 800–806 (2011).

[262] Kawamata, I., Aubert, N., Hamano, M. and Hagiya, M.: Abstraction of Graph-Based Models of Bio-molecular Reaction Systems for Efficient Sim-

ulation, *Lecture Notes in Computer Science*, Vol. 7605, pp. 187–206 (2012), DOI: 10.1007/978-3-642-33636-2_12

[263] Kawamata, I., Tanaka, F. and Hagiya, M.: Automatic Design of DNA Logic Gates Based on Kinetic Simulation, *Lecture Notes in Computer Science*, Vol. 5877, pp. 88–96 (2009), DOI: 10.1007/978-3-642-10604-0_9