Design of an Autonomous DNA Nanomechanical Device Capable of Universal Computation and Universal Translational Motion

Peng Yin [*]   Andrew J. Turberfield [†]   Sudheer Sahu [‡]   John H. Reif [§]

Abstract

Intelligent nanomechanical devices that operate in an autonomous fashion are of great theoretical and practical interest. Recent successes in building large scale DNA nano-structures, in constructing DNA mechanical devices, and in DNA computing provide a solid foundation for the next step forward: designing autonomous DNA mechanical devices capable of arbitrarily complex behavior. One prototype system towards this goal can be a DNA mechanical device that is capable of universal computation, by mimicking the operation of a universal Turing machine. Building on our prior theoretical designs and a prototype experimental construction of autonomous unidirectional DNA walking devices that move along linear tracks, we present in this paper the design of a nanomechanical DNA device that autonomously mimics the operation of a 2-state 5-color universal Turing machine. Our autonomous nanomechanical device, which we call an Autonomous DNA Turing Machine, is thus capable of universal computation and hence complex translational motion which we define as *universal translational motion*.

1 Introduction

1.1 Previous and current work

DNA has been explored as an excellent material for building large scale nano-structures, constructing individual nanomechanical devices, and performing computations [25]. Recent years have seen substantial progress in these three fields and this progress provides a solid foundation for the next step forward: designing (and implementing) autonomous nanomechanical devices capable of arbitrarily complex behavior, e.g. motion and computation. A major challenge in nanoscience is to design a nanomechanical device that is capable of *universal translational motion*, which we define as the motion determined by the head of a universal Turing machine. To meet this challenge, we describe the construction of a nanomechanical device embedded in a DNA lattice that mimics the operation of a universal Turing machine in an autonomous fashion. We call this prototype device an Autonomous DNA Turing Machine. The design of Autonomous DNA Turing Machine benefits from recent progress in the three aforementioned fields, can be viewed as an exciting synergistic point of the three fields, and in turn contributes to the advance of each of these fields: it can be viewed as an autonomous intelligent nano-lattice, an autonomous intelligent nanorobotics device, and a compact autonomous universal computing device.

Large scale periodic lattices have been made from a diverse set of branched DNA molecules, such as the DX molecules [34], TX molecules [10], rhombus molecules [16], and 4x4 molecules [37]. In addition, researchers have also successfully constructed aperiodic programmable DNA lattices [36]. These self-assembled lattices provide a structural base for our construction of Autonomous DNA Turing Machine, whose main structure can be implemented as two parallel arrays of DNA molecules embedded in a one-dimensional DNA lattice.

A variety of DNA nanomechanical devices have been previously constructed that demonstrate motions such as open/close [27, 28, 41], extension/contraction [2, 8, 12], and rotation [17, 38], mediated by external environmental changes such as the addition and removal of DNA fuel strands [2, 8, 12, 27, 28, 38, 41] or the change of ionic strength of the solution [17]. A desirable improvement of these devices is the construction of DNA nanomechanical devices that achieve motions beyond the non-autonomous and localized non-extensible motions exhibited by the above devices. There have already been some exciting work in this direction. Turberfield and colleagues have designed a free running DNA machine [30] which uses DNA as fuels and demonstrates autonomous motion. Mao's group recently demonstrated an autonomous DNA motor powered by a DNA enzyme [6]. Seeman's group has constructed a DNA walking device controlled by DNA fuel strands [26]. Reif has designed an autonomous DNA walking device and an autonomous DNA rolling device that move in a random bidirectional fashion along DNA tracks [22]. The authors have recently de-

[*]Department of Computer Science, Duke University, Box 90129, Durham, NC 27708-0129, USA. py@cs.duke.edu

[†]University of Oxford, Department of Physics, Clarendon Laboratory, Parks Road, Oxford OX 1 3PU, UK. a.turberfield@physics.ox.ac.uk

[‡]Department of Computer Science, Duke University, Box 90129, Durham, NC 27708-0129, USA. sudheer@cs.duke.edu

[§]Department of Computer Science, Duke University, Box 90129, Durham, NC 27708-0129, USA. reif@cs.duke.edu

signed autonomous DNA walking devices capable of autonomous programmable unidirectional motions along linear tracks [39, 40]. One of these prototype unidirectional DNA walking devices has also been experimentally constructed in our lab [40]. However, these devices only exhibit simple unidirectional motion.

A rich family of DNA computation schemes have been proposed and implemented [7, 11, 13, 14, 15, 18, 19, 20, 21, 23, 24, 29, 33] following Adleman's seminal report in 1994 [1]. Among them, the most relevant work that has inspired the construction here is the universal DNA Turing machine design by Rothemund [23] and the autonomous 2-state 2-color finite state automata constructed by Shapiro's group [3, 4, 5]. In Rothemund's innovative design, the transition table of a universal Turing machine is encoded in a circular DNA and the encoded transitions are carried out by enzymatic cleavages and ligations. However, these reactions need to be carried out manually for each transition. In contrast, the Autonomous DNA Turing Machine described here operates in an autonomous fashion with no external environmental mediation. In the inspiring construction by Shapiro's group, a duplex DNA encoding the sequence of input symbols is digested sequentially by an endonuclease in a fashion mimicking the processing of input data by a finite state automaton. Some of their encoding schemes are used in the construction described in this paper, i.e. using DNA sticky end to encode the combination of state and symbol and using a class II restriction enzyme to effect state change. A limitation of the finite state automata construction is that the data are destroyed as the finite state automaton proceeds. Though this feature does not affect the proper operation of a finite state automaton, it poses a barrier to further extending the finite state automaton to more powerful computing devices such as Turing machines.

In this work, we encode computational power into a DNA walking device embedded in a DNA lattice and thus accomplish the design for an autonomous nanomechanical device capable of universal computation, by mimicking the operation of a 2-state 5-color universal Turing machine. In the process of computation, the device can also demonstrate universal translational motion as defined above, i.e. the motion demonstrated by the head of universal Turing machine.

1.2 Universal Turing machine

A Turing machine is a theoretical computational device invented by Turing for performing mechanical or algorithmic mathematical calculations [31, 32]. Though the construction and operational rules of a Turing machine may seem beguilingly simple and rudimentary, it has been shown that any computational process that can be done by present computers can be carried out by a Turing machine.

A Turing machine consists of two parts, a *read-write head* and a linear tape of *cells* encoding the input data. The head has an internal state q and each cell has a color (or
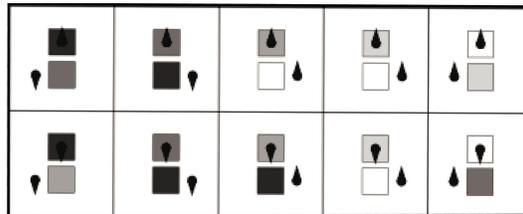
Figure 1: A universal Turing machine with two internal states and five colors. Figure excerpted from Wolfram [35].

data) c (as described in [35]). The sets of the admissible states and colors are denoted as $\mathcal{Q}$ and $\mathcal{C}$, respectively. At any step, the head resides on top of one cell, and the color of that cell and the state of the head together determines a transition: (i) the current cell may change to another color; (ii) the head may take a new state; (iii) the head may move to the cell immediately to the left or the right of the current cell.

A universal Turing machine is a Turing machine that can simulate the operation of any other Turing machine. Let m and n be the number of possible states and the number of possible colors of a Turing machine, respectively. The Turing machine with a proven universal computation capacity and the smallest $m \times n$ value is a 2-state 5-color Turing machine described in [35]. The operation of this Turing machine is shown in Figure 1. The directions of the arrow (up and down) indicate the states of the Turing machine; the background color of the cell indicates the color of the Turing machine. In each case, the upper level shows the current state and color of the Turing machine as well the position of the head; the lower level shows the state, color and head position after the transition. The table in Figure 1 is also known as the *transition table* of a Turing machine.

In this paper, we describe the design of a DNA nanomechanical device that simulates the general operation of an arbitrary 2-state 5-color Turing machine whose head moves to either its left or right neighbor in every transition, and in particular, the universal Turing machine depicted in Figure 1. We note that the head of this Turing machine has to move either to the left or right during each transition instead of staying in the same position. This is a limitation of our device, but it does not hinder its simulation of the universal Turing machine in Figure 1.

The rest of the paper is organized as follows. We give a structural overview and introduce definitions and notations in Section 2. An overview of the operation of the Autonomous DNA Turing Machine is given in Section 3, followed by a detailed step-by-step molecular implementation of the operation in Section 4. Section 5 describes futile reactions that occur in the background during the operation of the Autonomous DNA Turing Machine. In Section 6, we discuss a major challenge in designing the Autonomous DNA Turing Machine – encoding in limited space multiple layers of information. We close in Section 7 with discus-

sions of future work.

2 Definitions and structural overview

Some standard terms known to the DNA computing community are briefly reviewed in Appendix A. These notions are used in the description of the Autonomous DNA Turing Machine. We then describe below the structural overview of the Autonomous DNA Turing Machine and, in this context, introduce more definitions specific to the construction of the Autonomous DNA Turing Machine.

2.1 Structural overview of the Autonomous DNA Turing Machine
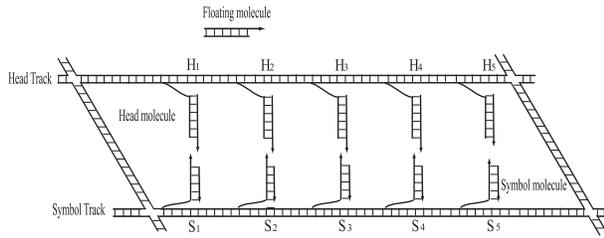
Figure 2: Schematic drawing of the structure of the Autonomous DNA Turing Machine. H_i and S_i denote Head-Molecule and Symbol-Molecule, respectively. The backbones of DNA strands are depicted as line segments. The short bars represent base pairing between DNA strands.

Figure 2 illustrates the structure of the Autonomous DNA Turing Machine. The Autonomous DNA Turing Machine operates in a solution system. The major components of the Autonomous DNA Turing Machine are two parallel arrays of *dangling molecules* tethered to two *rigid* tracks. The two rigid tracks can be implemented as rigid DNA lattices as described in Section 1.1, for example, the rhombus lattice [16] as shown in Figure 2. A dangling-molecule is a duplex DNA fragment with one end linked to the track via a flexible single strand DNA and the other end possessing a single strand DNA extension (the sticky end). Due to the flexibility of the single strand DNA linkage, a dangling-molecule moves rather freely around its joint at the track. The upper and lower arrays of dangling-molecules are called *Head-Molecules*, denoted as H, and *Symbol-Molecules*, denoted as S, respectively. We require that the only possible interactions between two dangling-molecules are either a reaction between a Head-Molecule and the Symbol-Molecule immediately below it or a reaction between two neighboring dangling-molecules along the same track. This requirement can be ensured by the

rigidity of the tracks and the properly spacing of dangling-molecules along the rigid tracks. In addition to the two arrays of dangling-molecules, there are *floating-molecules*. A floating-molecule is a free floating (unattached to the tracks) duplex DNA segment with a single strand overhang at one end (sticky end). A floating-molecule floats freely in the solution and thus can interact with another floating-molecule or a dangling-molecule provided that they possess complementary sticky ends. There are two kinds of floating-molecules: the Rule-Molecules and the Assisting-Molecules. The Rule-Molecules specify the computational rules and are the programmable part of the Autonomous DNA Turing Machine while the Assisting-Molecules assist in the carrying out the operations of the Autonomous DNA Turing Machine, as described in detail later. The array of Symbol-Molecules represent the data tape of a Turing Machine; the array of Head-Molecules represent the moving head of a Turing Machine (more specifically, at any time, only one Head-Molecule is active, and its position indicates the position of the head of a Turing Machine); the Rule-Molecules collectively specify the transition rules for the Autonomous DNA Turing Machine; the Assisting-Molecules are auxiliary molecules that assist in maintaining the operation of the Autonomous DNA Turing Machine.

The duplex portion and/or the sticky end of a DNA molecule may encode the following information: (i) *state*, the Turing machine state; (ii) *color*, the color (data) encoded in a symbol molecule; (iii) *position*, the position type of a Head-Molecule. The state, color, and position information are denoted as q, c, and p, respectively, where $q \in \{Q_A = LONG, Q_B = SHORT\}$, $c \in \{C_A, C_B, C_C, C_D, C_E\}$, $p \in \{P_A, P_B, P_C\}$ for the 2-state 5-color Autonomous DNA Turing Machine. The position information p indicates the position type of a Head-Molecule. This information is essential for dictating the bidirectional motion of the head. An information encoding molecule is denoted as

$$X^a [y]^b$$

where X is its duplex portion, $[y]$ is its sticky end portion; a is the state/color/position information encoded in X, and b is the state/color/position information encoded in $[y]$. A complementary sticky end of $[y]$ is denoted as $[\bar{y}]$.

The array of Head-Molecules is denoted as $(H_1, H_2, H_3, \dots)$; the array of Symbol-Molecules is denoted as $(S_1, S_2, S_3, \dots)$. To specify the motion of the Autonomous DNA Turing Machine head, we have Head-Molecules arranged in periodic linear order, $(H_1^{P_A}, H_2^{P_B}, H_3^{P_C}, H_4^{P_A}, H_5^{P_B}, H_6^{P_C} \dots)$, along the Head-track.

3 Operational overview

At the beginning of a transition operation of the Autonomous DNA Turing Machine, all the Symbol-
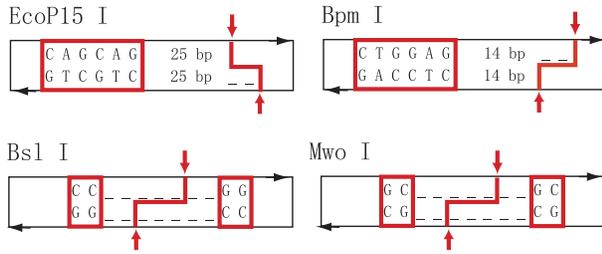
Figure 3: Four endonucleases used in the molecular implementation of the Autonomous DNA Turing Machine. The recognition site of an enzyme is bounded by a box and the cleavage site indicated with a pair of bold arrows. The symbol "−" indicates the position of a base that does not affect recognition.

Figure 4: Operational overview of the Autonomous DNA Turing Machine. The dangling Head-Molecules and Symbol-Molecules are depicted as dark line fragments. The floating Rule-Molecules and Assisting-Molecules are depicted as light colored light segments. H, S, R, and A denote Head-Molecule, Symbol-Molecule, Rule-Molecule, and Assisting-Molecule, respectively. The triangle indicates the *active* Head-Molecule. (i) indicates the ligation event that occurs in Step i as will be described in the detailed step-by-step implementation of the Autonomous DNA Turing Machine (Section 4).

Molecules possess sticky ends $[\bar{s}]$. A Symbol-Molecule with a sticky end $[\bar{s}]$ is referred to as in its *default* configuration; the $[\bar{s}]$ sticky end is referred to as a *default sticky end*. One of the Head-Molecules encodes the current state of the Autonomous DNA Turing Machine in its duplex portion and possesses an *active* sticky end $[s]$ that is complementary to the sticky end $[\bar{s}]$ of the Symbol-Molecule just below it. This Head-Molecule is referred to as the *active Head-Molecule*. In contrast, all other Head-Molecules (with sticky ends other than $[s]$) are in *default* or *inactive* configuration.

Figure 4 gives a high level description of the events that occur during one transition of Autonomous DNA Turing Machine. For the ease of exposition, we describe the operation in 4 stages. The 8 types of ligation events that correspond to the detailed 8-step implementation of the Autonomous DNA Turing Machine (Section 4) are also marked in the figure to assist the reader in relating the high level description in this section to detailed step-by-step implementation in Section 4.

In Stage 1, the active Head-Molecule (labeled with a triangle, H_3 in the example shown in Figure 4) is ligated to the Symbol-Molecule (S_3 in Figure 4) directly below it, creating an endonuclease recognition site in the ligation product (event (1) in Figure 4). The ligation product is subsequently cleaved into two molecules by an endonuclease. The sticky end of each of the two newly generated molecules encodes the current state and the current color of the Autonomous DNA Turing Machine.

In Stage 2, both the new Symbol-Molecule and the new Head-Molecule are ligated to floating Rule-Molecules (events (2) and (4) in Figure 4), which possess complementary sticky ends to them and correspond to one entry in the Turing machine transition table. The ligation product between the Symbol-Molecule and the Rule-Molecule is in turn cleaved, generating a new Symbol-Molecule dictated by the current state and color information as well as the transition rule. The new Symbol-Molecule encodes the new color in its sticky end. Similarly, the ligation product between the Head-Molecule and the Rule-Molecule is

cleaved, generating a new Head-Molecule whose duplex portion encodes information of Turing machine's next state and whose sticky end encodes the moving direction of the head.

In Stage 3, the newly generated Symbol-Molecule is further modified by an Assisting-Molecule so that it will encode the new color in its duplex portion (rather than sticky end) and possess an $[\bar{s}]$ sticky end (event (3) in Figure 4). The sticky end of the Head-Molecule will dictate it to hybridize with either the Head-Molecule to its left or to its right, depending on which of its neighbors possesses a complementary sticky end (event (5) in Figure 4, H_3 is ligated with its left neighbor H_2). Next, the ligation product between these two Head-Molecules is cleaved.

In Stage 4, the two Head-Molecules are modified by floating Assisting-Molecules (events (6) and (7, 8) in Figure 4) so that the first Head-Molecule is restored to its inactive configuration (with a default sticky end) and the second Head-Molecule encodes the state information in its duplex part and possesses an active sticky end $[s]$ and thus becomes an active Head-Molecule, ready to interact with the Symbol-Molecule located directly below it.

This finishes a transition and the operation can thus go on in an inductive way. We emphasize that we describe the events in stages only for the ease of exposition. The proper operation of the Autonomous DNA Turing Machine does not require the synchronization of the events as described above. For example, event (3) in Stage 3 can (though not necessarily) occur either before event (4) or after event (8).

4 Step-by-step implementation

We next give a detailed 8-step description of the operation of the Autonomous DNA Turing Machine. Each step consists of ligation and cleavage events. The ligation events are marked in Figure 4 with (i), where $i = 1, 2, \ldots, 8$. To demonstrate the practicality of our design, we give full DNA sequence for the reactions of each step. In addition to the A, T, C, and G bases, we also occasionally require another pair of unnatural bases which we denote as E and F. The reason to use E and F is to minimize the futile reactions as described later in Appendix C and hence increase the efficiency of our Autonomous DNA Turing Machine. The practicality of use of E and F is justified by the existing technology to make such bases and incorporate them into DNA strands. For a recent survey on unnatural bases, see [9].

At the start of the operation of the Autonomous DNA Turing Machine, the configuration of the Head-Molecules array along the Head-track is

$$(\hat{H}_1^{p_1 q}[s])([\bar{h}]^{p_2} H_2^{p_2})([\bar{h}]^{p_3} H_3^{p_3}) \ldots$$

where $p_i = P_A$ for $i = 3k + 1$, $p_i = P_B$ for $i = 3k + 2$, $p_i = P_C$ for $i = 3k + 3$ for $k = 0, 1, 2, \ldots$. The first Head-Molecule is special: it is the active Head-Molecule and represents the current position of the active head. We use the symbol ˆ to denote the active configuration of a Head-Molecule. H_1 has the unique sticky end $[s]$, which is complementary to the sticky end $[\bar{s}]$ of a Symbol-Molecule in default configuration (in particular, the Symbol-Molecule directly below it). Thus, H_1 can hybridize and be ligated with Symbol-Molecule S_1, and this will start the operation of the Turing machine. Recall that p encodes the position type information of a Head-Molecule. This position type information is encoded both in the sticky end portion and in the duplex portion of a Head-Molecule. As we will see below, the sticky end encoding of p is necessary for dictating the appropriate motion of an active head; the duplex portion encoding is necessary for restoring a Head-Molecule to its default configuration after it turns from an active to an inactive state.

The Symbol-Molecules array along the Symbol-track is

$$([\bar{s}]S_1^{c_1})([\bar{s}]S_2^{c_2})([\bar{s}]S_3^{c_3}) \ldots$$

All the Symbol-Molecules have the same sticky end $[\bar{s}]$. As such, whenever a Head-Molecule directly above a Symbol-Molecule becomes active, this Symbol-Molecule can interact with the active Head-Molecule. Note that $[\bar{s}]$ encodes no color information – the color information c_i is instead encoded completely in the duplex portion of a Symbol-Molecule.

4.1 Reaction between a Head-Molecule and a Symbol-Molecule
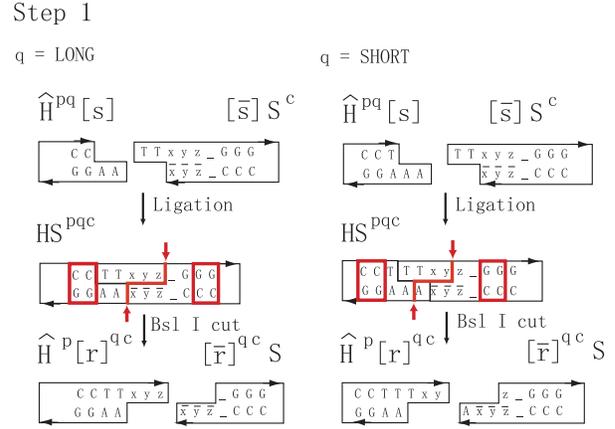
Figure 5: Step 1 of the operation of Autonomous DNA Turing Machine. The current state q and color c are initially encoded in the duplex portion of the Head-Molecule and the Symbol-Molecule, respectively. After the ligation and cleavage, both the sticky ends of the new Head-Molecule and Symbol-Molecule encode the current state q and the current color c. The encoding scheme of c is described in Table 1. Bsl I recognition sites and cleavage sites are indicated with boxes and pairs of bold arrows, respectively.

Step 1. In step 1, the active state-encoding Head-Molecule is first ligated with the color-encoding Symbol-Molecule below it, and then the ligation product is cut into a new Head-Molecule and a new Symbol-Molecule, the sticky ends of which both encode the current state and color information.

Let $\hat{H}_i^{pq}[s]$ be the current active head (encoding position type p and current state q); let $[\bar{s}]S_i^c$ be the Symbol-Molecule below it (encoding current color c). $\hat{H}_i$ and S_i has complementary sticky ends and hence these two are ligated into $(H_i S_i)^{pqc}$. An endonuclease recognizes the newly formed recognition site in the ligation product and cuts the ligation product into $\hat{H}_i^p[r]^{qc}$ and $[\bar{r}]^{qc} S_i$. Now the sticky ends of both $\hat{H}_i$ and S_i encode the current color and state. Step 1 can be described by the following equation,

$$\hat{H}_i^{pq}[s] + [\bar{s}]S_i^c \to (H_i S_i)^{pqc} \to \hat{H}_i^p[r]^{qc} + [\bar{r}]^{qc} S_i$$

The first part of the equation is the ligation of Head-Molecule $\hat{H}_i^{pq}[s]$ with Symbol-Molecule $[\bar{s}]S_i^c$ into $(H_i S_i)^{pqc}$; the second part is the cleavage of the ligation product into Head-Molecule $\hat{H}_i^p[r]^{qc}$ and Symbol-Molecule $[\bar{r}]^{qc} S_i$. Note that now both the sticky ends of

	c	C_A	C_B	C_C	C_D	C_E
$q = LONG$	xyz	TTA	CTT	CAA	AEA	CEA
$q = SHORT$	Txy	TTT	TCT	TCA	TAE	TCE

Table 1: The molecular implementation of the color encoding scheme of a Symbol-Molecule. c is the color; xyz is the sticky $[r]$ exposed when state $q = LONG$; Txy is the sticky end $[r]$ when state $q = SHORT$. Note that all the ten sticky end sequences are different from each other.

	c'	C_A	C_B	C_C	C_D	C_E
	$\bar{e}$	CA	AC	CT	TT	TG
$q = LONG$	l	8	7	6	5	4
$q = SHORT$	l	7	6	5	4	3

Table 2: The relation between the length of the spacer, the sequence of sticky end $[\bar{e}]$ and the new color of a Symbol-Molecule. l is the spacer length; $\bar{e}$ is the sticky end sequence; c' is the new color.

the Head-Molecule and the Symbol-Molecule are encoding the current state and color. This encoding scheme is in the same spirit as the one used in [5].

Figure 5 gives the molecular implementation of Step 1. For simplicity, only the relevant end sequences are given. The encoded information p is not shown. Both the case when $q = SHORT$ and the case when $q = LONG$ are depicted. xyz is the color encoding region for Symbol-Molecule S. The encoding scheme used is shown in Table 1.

4.2 Color change of a Symbol-Molecule

After Step 1, the sticky end of $[\bar{r}]^{qc} S_i$ encodes the current state and color. This sticky end is subsequently detected by a Rule-Molecule $\tilde{R}[r]^{qc}$, which has a complementary sticky end. $\tilde{R}[r]^{qc}$ corresponds to one entry in the transition table for Autonomous DNA Turing Machine, and determines the next color c' that will be encoded in S_i. This color transition occurs in Step 2 and S_i is modified to possess a sticky end $[\bar{e}]^{c'}$ that encodes the new color c'. In Step 3, S_i is restored to a default configuration with a sticky end $[\bar{s}]$, and the new color c' encoded in its duplex portion. We next describe the reactions in detail.

Step 2. In Step 2, Rule-Molecule $\tilde{R}[r]^{qc}$ hybridizes and is ligated with Symbol-Molecule $[\bar{r}]^{qc} S_i$. The ligation product is cut into $\tilde{R}_w[e]^{c'}$ (a waste molecule that diffuses away) and $[\bar{e}]^{c'} S_i$. The sticky end $[\bar{e}]$ encodes the new color c'. Schematically, we have,

$$\tilde{R}[r]^{qc} + [\bar{r}]^{qc} S_i \to (RS_i)^{qcc'} \to \tilde{R}_w^{qc}[e]^{c'} + [\bar{e}]^{c'} S_i$$
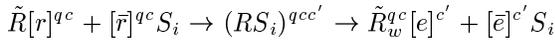
Figure 6 describes the molecular implementation of Step 2 for the case when current state is $q = LONG$, and the new color is $c' = C_B$. The case for $q = SHORT$ is similar, except that sticky end $[\bar{r}]$ of S is $C\bar{x}\bar{y}$ instead of $\bar{x}\bar{y}\bar{z}$. The Rule-Molecule $\tilde{R}[r]^{qc}$ consists of three parts, in the terminology of [5], Bpm I recognition site, spacer region, and <state,color> detector. The <state, color> detector is the sticky end $[r]^{qc}$, which hybridizes with and thus detects the sticky end $[\bar{r}]^{qc}$ of the symbol molecule. The Rule-Molecule and the Symbol-Molecule are ligated and Bpm I cuts the ligation product into a waste Rule-Molecule $\tilde{R}_w^{qc}[e]^{c'}$ (w for waste), which diffuses away, and a new Symbol-Molecule $[\bar{e}]^{c'} S$, effecting the color change of the Symbol-Molecule from c to c'. The length of the spacer of $\tilde{R}$ (see Figure 6) determines the position of the

cut in the ligation product and hence the sticky end $[\bar{e}]$ and the new color c' encoded in it. See Table 2 for the relation between the length of the spacer, the sequence of sticky end $[\bar{e}]$ and the new color c'.

Figure 6: Step 2 of the operation of Autonomous DNA Turing Machine. In this example, the current state is $q = LONG$; the current color q and state c are encoded in the Symbol-Molecule's sticky end $[\bar{r}]$ whose sequence is $\bar{x}\bar{y}\bar{z}$; the new color, in this case, will be $c' = C_B$, encoded in sticky end $[\bar{e}]$ whose sequence is "TG". Bpm I recognition site and cleavage site are indicated with a box and a pair of bold arrows, respectively.

Step 3. The Symbol-Molecule $[\bar{e}]^{c'} S_i$ obtained from Step 2 needs to be restored to its default configuration $[\bar{s}]S_i^{c'}$ so that it can interact with the Head-Molecule H_i above it when H_i becomes active again. Note that this reusability of the Symbol-Molecule is essential for the proper functioning of Autonomous DNA Turing Machine. After the restoration, the new color c' is encoded in the duplex portion of S_i, whose sticky end is the default sticky end $[\bar{s}]$ for a Symbol-Molecule: it encodes no color, but is ready to interact with an active Head-Molecule. The reaction of Step 3 is,

$$E^{c'}[e]^{c'} + [\bar{e}]^{c'} S_i \to (ES)^{c'} \to E_w[s] + [\bar{s}]S_i^{c'}$$
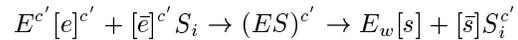
Figure 7 gives a molecular implementation of Step 3 for the case $c' = C_B$. Color c' is encoded both in the sticky end portion and the duplex portion of Assisting-Molecule $E^{c'}[e]^{c'}$. Assisting-Molecule $E^{c'}[e]^{c'}$ detects the color encoding sticky end of Symbol-Molecule S_i and transfers its color encoding duplex portion to S_i via ligation and
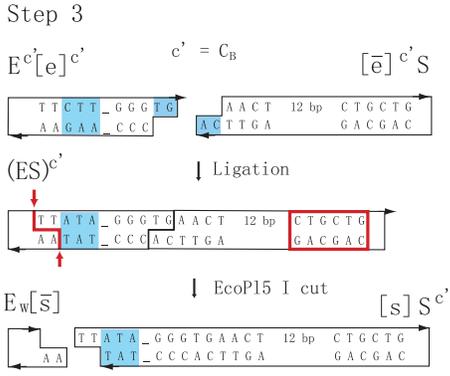
Figure 7: Step 3 of the operation of Autonomous DNA Turing Machine. In this example, the new color $c' = C_B$. See Figure 8 for the complete set of Assisting-Molecules $E^{c'}e^{c'}$. The color encoding regions are indicated with light gray background. EcoPl5 I recognition site and cleavage site are indicated with a box and a pair of bold arrows, respectively.

Figure 9: Step 4 of the operation of Autonomous DNA Turing Machine. The motion encoding regions are indicated with light gray background. l is the length of the spacer region of Rule-Molecule R. EcoPl5 I recognition site and cleavage site are indicated with a box and a pair of bold arrows, respectively.

Figure 8: The complete set of Assisting-Molecules $E^{c'}[e]^{c'}$. The color encoding regions are indicated with light gray background.

subsequent cleavage. This step generates a waste product $E_w[s]$ that diffuses away. Note that $E_w[s]$ may hybridize and be ligated with some other $[\bar{s}]$ end of a Symbol-Molecules, say $[\bar{s}]S_k$, however, as this only represents some futile reactions that will not block, reverse or alter the operation of the Autonomous DNA Turing Machine, since $E_w[s]$ will be cut subsequently away from $[\bar{s}]S_k$ by EcoPl 5I. However, this does decrease the efficiency of the Autonomous DNA Turing Machine and as the concentration of $E_w[s]$ increases, the negative effect on the efficiency becomes more prominent. For a complete set of Assisting-Molecules $E^{c'}[e]^{c'}$, see Figure 8.

Note that the existence of endonuclease EcoPl5 I recognition site in the duplex portion of S_i adds extra complication to Step 1 and Step 2: it results in futile reactions (F2) which will be discussed in Appendix C.

4.3 State change of a Head-Molecule

Step 4. In Step 4, the Head-Molecule $H_i^p[r]^{qc}$ generated in Step 2 (with its sticky end encoding the current state and color) is modified by a Rule-Molecule that decides the state transition and the motion of the head. After the modification, the new state information is encoded in the du-

plex portion of the modified Head-Molecule, and the motion direction of the head is encoded in the sticky end of the modified Head-Molecule in the form of a complementary sticky end to one of its neighboring Head-Molecules. The sticky end of the modified Head-Molecule will dictate it to interact with either its left or right neighbors, and thus determines the motion of the head.

More specifically, Head-Molecule $H_i^p[r]^{qc}$ hybridizes and is ligated with a free floating Rule-Molecule $[\bar{r}]^{qc}R$ and the ligation product $(H_iR)^{pqc}$ is cut by endonuclease EcoPl5 I into $H_i^{pq'}[h]^{p'}$ and $[\bar{h}]^{p'}R_w$, a waste molecule that diffuses away. Head-Molecule $H_i^{pq'}[h]^{p'}$ encodes the new state q' in its duplex portion, and the motion direction p' of the head in its sticky end. The reaction of Step 4 is,

$$\hat{H}_i^{\,p}[r]^{qc} + [\bar{r}]^{qc}R \rightarrow (H_iR)^{pqc} \rightarrow \hat{H}_i^{\,pq'}[h]^{p'} + [\bar{h}]^{p'}R_w$$
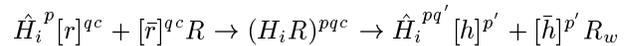
Figure 9 describes the molecular implementation for the case when the current state $q = LONG$; new state $q' = SHORT$; the position type of the current Head-Molecule H_i is $p = P_A$; the position type of the Head-Molecule H_j that it will interact with is $p' = P_B$ (hence $j = i + 1$ in this case). The Rule-Molecule $[\bar{r}]^{qc}R$ consists of three parts: the detector sticky end $[\bar{r}]^{qc}$ that encodes the current state and color; the spacer, whose length determines the transition results (new state and motion direction of the head); and recognition site for endonuclease EcoPl5 I. The Rule-Molecule $[\bar{r}]^{qc}R$ detects the current state q and color c encoded in sticky end $[r]^{qc}$ of H_i and is ligated to H_i. After ligation, endonuclease EcoPl5 I cuts into the motion encoding region of the Head-Molecule and exposes a new sticky end that encodes the position type information p' (and hence determines the motion direction). Cleavages at motion encoding regions I and II result in new states $q' = LONG$ and $q' = SHORT$, respectively. Table 3 describes the complete set of transitions for all the com-

binations of different p, q, d, and q'. Note that l is not dependent on p: in each case of $p = P_A$, $p = P_B$, and $p = P_C$, l is the same. This is an essential property since the end $[r]^{qc}$ of H does not encode the p information.

4.4 Reaction between two adjacent Head-Molecules

Head-Molecule $\hat{H}_i^{pq'}[h]^{p'}$ produced in Step 4 will next interact with one of its neighboring Head-Molecules, $[\bar{h}]^{p'} H'_j$, where $j = i-1$ for its left neighbor and $j = i+1$ for its right neighbor (Step 5). Then H_j becomes an active Head-Molecule encoding the new state q' (Step 6) while H_i is restored to its default inactive configuration (Steps 7 and 8).

Step 5. In Step 5, Head-Molecule $\hat{H}_i^{pq'}[h]^{p'}$ is ligated to either its left neighbor or its right neighbor $[\bar{h}]^{p'} H'_j$, where $j = i-1$ or $i+1$, as dictated by the p' information encoded in its sticky end. The ligation product $(H_i H_j)^{q'}$ is cut into $H_i^p[\bar{t}]^{pp'q'}$ and $[t]^{pp'q'} \hat{H}_j$. The reaction of Step 5 is,

$$\hat{H}_i^{pq'}[h]^{p'} + [\bar{h}]^{p'} H'_j \rightarrow (H_i H_j)^{q'} \rightarrow H_i[\bar{t}]^{pp'q'} + [t]^{pp'q'} \hat{H}'_j$$
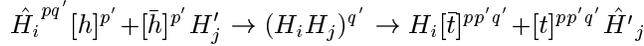
Note that now both the sticky ends of H_i and H_j encode position type p of H_i, position type p' of H_j and the new state q'.

Figure 10 gives a molecular implementation for this step. Panel I depicts an example case in full details; Panel II and III show all the cases in a simplified way. Note that the sticky end $[\bar{t}]$ (and $[t]$) encodes all the information for position type p of H_i, position type p' of H_j, and the new state q', we hence have $3 \times 2 \times 2 = 12$ different sticky ends $[\bar{t}]$.

Step 6. In Step 6, Head-Molecule $\hat{H}'_j$ is modified into a Head-Molecule ready to interact with a Symbol-Molecule; in other words, it becomes an active head. The reaction of Step 6 is,

$$\hat{H}'_j[t]^{pp'q'} + [\bar{t}]^{pp'q'} T \rightarrow H'_j T \rightarrow \hat{H}'^{q'}_j[s] + [\bar{s}] T_w$$
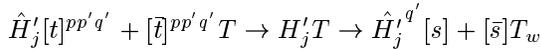
Figure 11 describes the molecular implementation for Step 6. The mechanism of this step is very similar to Step 4, and hence we omit the details of its description and refer the reader to Figure 11.

Steps 7. and 8. In Step 7, the sticky end $[\bar{t}]^{pp'q'}$ of Head-Molecule H_i is modified by an Assisting-Molecule $\tilde{T}[t]^{pp'q'}$ to a new sticky end $[\bar{g}]^{pp'q'}$. In Step 8, the sticky end $[\bar{g}]^{pp'q'}$ initiates a sequential "growing-back" process which restores H_i to its default (inactive) configuration $[\bar{h}]^p H_i^p$. The reaction of Step 7 is,

$$\tilde{T}[t]^{pp'q'} + [\bar{t}]^{pp'q'} H_i \rightarrow \tilde{T} H_i \rightarrow \tilde{T}_w[g]^{pp'q'} + [\bar{g}]^{pp'q'} H_i$$

The reaction of Step 8 is,

$$[\bar{g}]^{pp'q'} H_i \rightarrow [\bar{h}]^p H_i^p$$
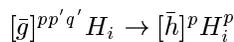
p	$[\bar{h}]$	m	q	d	p'	q'	l	$[h]$	$[h]^R$
P_A	AC	ATC TAG	S	$\rightarrow$	P_B	S	17	AG	GA
P_A	AC	ATC TAG	S	$\rightarrow$	P_B	L	6	AG	GA
P_A	AC	ATC TAG	S	$\leftarrow$	P_C	S	16	TA	AT
P_A	AC	ATC TAG	S	$\leftarrow$	P_C	L	5	TA	AT
P_A	AC	ATC TAG	L	$\rightarrow$	P_B	S	17	AG	GA
P_A	AC	ATC TAG	L	$\rightarrow$	P_B	L	6	AG	GA
P_A	AC	ATC TAG	L	$\leftarrow$	P_C	S	16	TA	AT
P_A	AC	ATC TAG	L	$\leftarrow$	P_C	L	5	TA	AT
P_B	CT	CAT GTA	S	$\rightarrow$	P_C	S	17	TA	AT
P_B	CT	CAT GTA	S	$\rightarrow$	P_C	L	6	TA	AT
P_B	CT	CAT GTA	S	$\leftarrow$	P_A	S	16	GT	TG
P_B	CT	CAT GTA	S	$\leftarrow$	P_A	L	5	GT	TG
P_B	CT	CAT GTA	L	$\rightarrow$	P_C	S	17	TA	AT
P_B	CT	CAT GTA	L	$\rightarrow$	P_C	L	6	TA	AT
P_B	CT	CAT GTA	L	$\leftarrow$	P_A	S	16	GT	TG
P_B	CT	CAT GTA	L	$\leftarrow$	P_A	L	5	GT	TG
P_C	TA	TCA AGT	S	$\rightarrow$	P_A	S	17	GT	TG
P_C	TA	TCA AGT	S	$\rightarrow$	P_A	L	6	GT	TG
P_C	TA	TCA AGT	S	$\leftarrow$	P_B	S	16	AG	GA
P_C	TA	TCA AGT	S	$\leftarrow$	P_B	L	5	AG	GA
P_C	TA	TCA AGT	L	$\rightarrow$	P_A	S	17	GT	TG
P_C	TA	TCA AGT	L	$\rightarrow$	P_A	L	6	GT	TG
P_C	TA	TCA AGT	L	$\leftarrow$	P_B	S	16	AG	GA
P_C	TA	TCA AGT	L	$\leftarrow$	P_B	L	5	AG	GA

Table 3: The transition of Head-Molecule $\hat{H}_i^p[r]^q c$ to $\hat{H}_i^{pq'}[h]^{p'}$, and its subsequent interaction with $[\bar{h}]^{p'} H_j^{p'}$. In the table, p is the position type of Head-Molecule $\hat{H}_i$, encoded in the duplex portion of H_i; $[\bar{h}]$ is the sticky end of Head-Molecule $H_j^p[\bar{h}]^p$ in its default inactive configuration, encoding the position type p of H_j; m is the sequence of the motion encoding region of Head-Molecule $\hat{H}_i$ (see Figure 9); d is the moving direction of the head; p' is the position type information encoded in sticky end $[h]$ of $\hat{H}_i pq'[h]^{p'}$, dictating the moving direction of the head; $[h]^R$ is the reverse sequence of sticky end $[h]$; q and q' are the current state and the new state, respectively; S and L stand for $SHORT$ and $LONG$ states, respectively; l is the length of the spacer region of the Rule-Molecule $[\bar{r}]^{qc} R$ (see Figure 9). Note that $[h]^R$ of H_i is complementary to $[\bar{h}]$ of H_j.
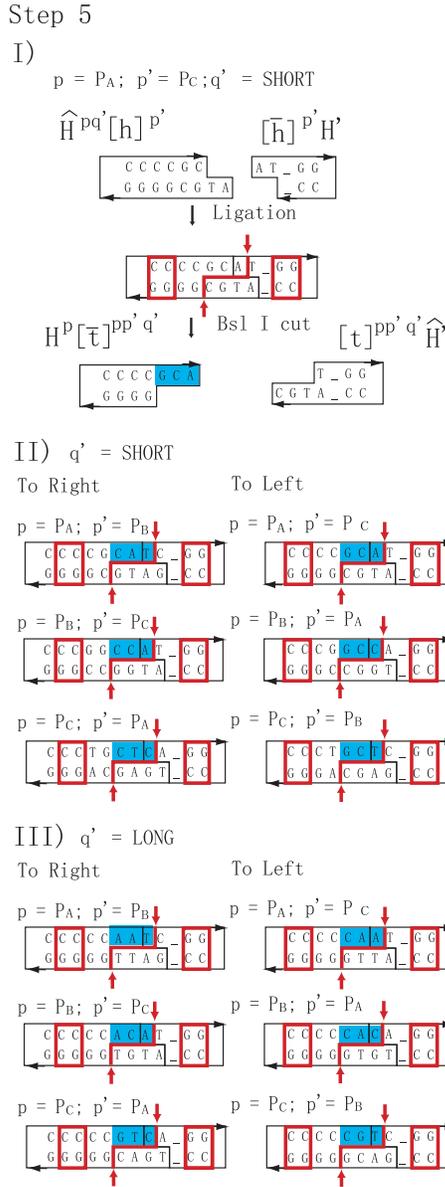
Figure 10: Step 5 of the operation of Autonomous DNA Turing Machine. Panel I depicts the case when $p = P_A$, $p' = P_C$, and $q' = SHORT$. Panel II and III describe all the cases when $q' = SHORT$ and all the cases when $q' = LONG$, respectively. In panel II and III, each case is represented in a simplified fashion that only shows the ligation product before the cleavage. Bsl I recognition sites and cleavage sites are indicated with boxes and pairs of bold arrows, respectively. The unique sticky ends $[\bar{t}]^{pp'q'}$ are shown with gray background.
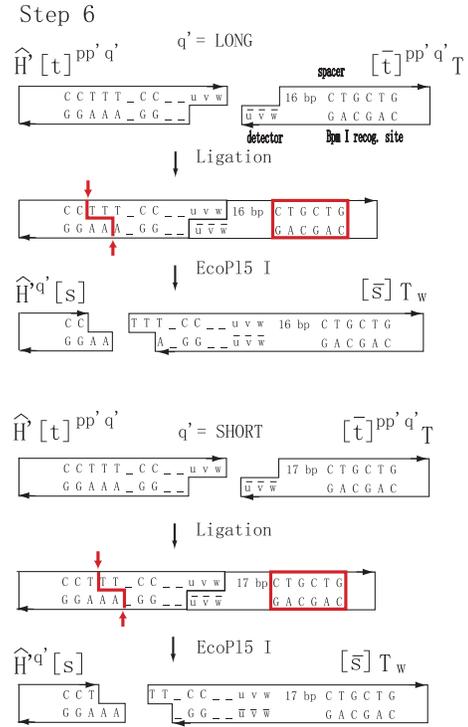
Figure 11: Step 6 of the operation of Autonomous DNA Turing Machine. Bpm I recognition sites and cleavage sites are indicated with boxes and pairs of bold arrows, respectively.

Figure 12 and Figure 13 describe the molecular implementation of Step 7 and Step 8 for the case $p = P_A$, $p' = P_B$, and $q' = LONG$, respectively. The figures are self-explanatory and hence we omit the details of description for brevity. Note that Step 8 is a rather spectacular process which illustrates a precisely controlled elongation mechanism using alternating ligations and cleavages. This mechanism may be of independent interest for designing other molecular devices.

4.5 Overall reaction flow

Putting all the above steps together, we have a schematic drawing for the overall flow of the reactions (Figure 14).

4.6 Molecule set

The complete molecule set for the construction of our Autonomous DNA Turing Machine is described in Appendix B.

5 Technical challenges

Two major technical challenges in designing the Autonomous DNA Turing Machine are to accommodate the *futile* reactions occurring during the operation of the Autonomous DNA Turing Machine and to design the Au-
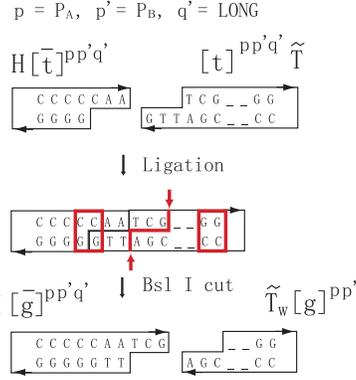
Figure 12: Step 7 of the operation of Autonomous DNA Turing Machine. Bsl I recognition site and cleavage site are indicated with a box and a pair of bold arrows, respectively.

tonomous DNA Turing Machine using limited encoding space dictated by the four (six) letter vocabulary of DNA bases and by the sizes of the recognition, restriction, and spacing regions of endonucleases.

The key technique used here to address the first challenge is to make all the futile reactions fully reversible so that they do not obstruct or alter the operation of the Autonomous DNA Turing Machine. The key technique to address the second challenge is to use overlay technique as shown in Figure 16 and to carefully select the sticky ends to avoid undesirable reactions.

These two issues are discussed in detail in Appendices C and D.

6 Computer simulations

To fully test the technical validity of our complex construction of Autonomous DNA Turing Machine, we performed computer simulation of the Autonomous DNA Turing Machine. The detailed description, software, and simulation results can be accessed at http://www.cs.duke.edu/~py/paper/dnaUTM/, and are omitted here for brevity.

7 Discussions

In this paper, we present the design of a DNA nanomechanical device capable of universal computation and hence universal translational motion. In addition to general design principles, we give detailed molecular implementation of the Autonomous DNA Turing Machine. A next step would be to construct a DNA cellular automata that demonstrates parallel computations.

As a consequence of the universal computation, the Autonomous DNA Turing Machine demonstrates universal
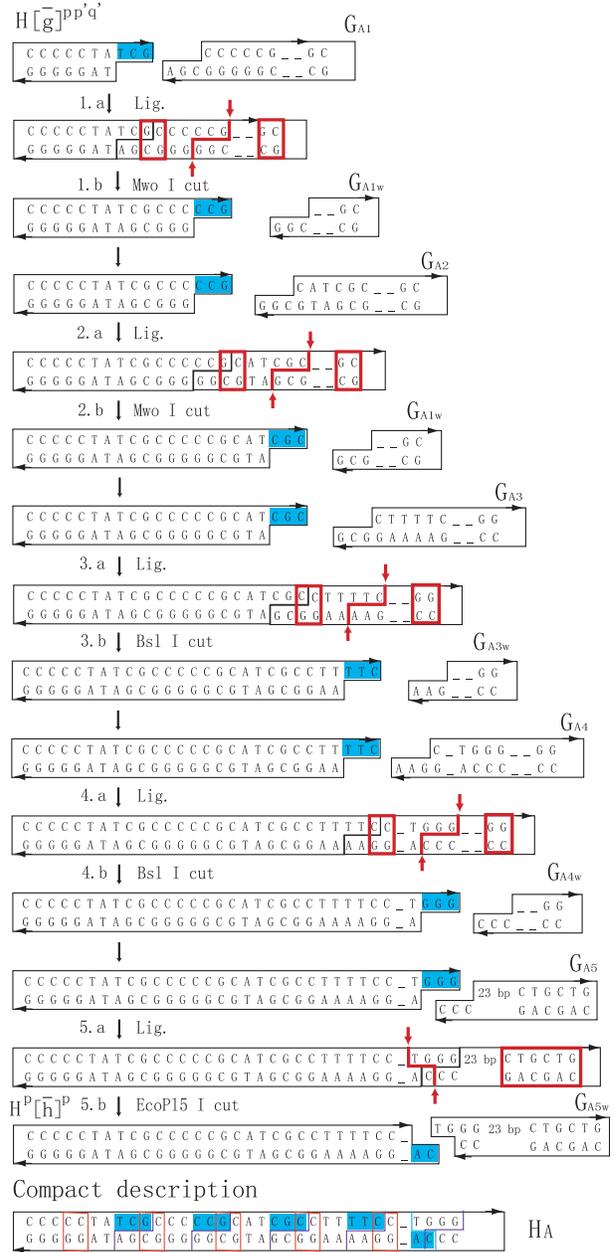
Figure 13: Step 8 of the operation of Autonomous DNA Turing Machine for the case $p = P_A$, $p' = P_B$, and $q' = LONG$. This step consists of a sequence of alternating ligations and cleavages. At each stage k, where $k = 1 - 5$, the Head-Molecule is first ligated to an Assisting-Molecule G_{Ak} (stage $k.a$), then the ligation product is cut by an endonuclease (stage $k.b$). A waste molecule G_{Akw} is generated at each stage. The last panel gives a compact representation of the whole process. The unique sticky end generated at each stage is indicated with gray background. Endonuclease recognition sites and cleavage sites are indicated with boxes and pairs of bold arrows, respectively.
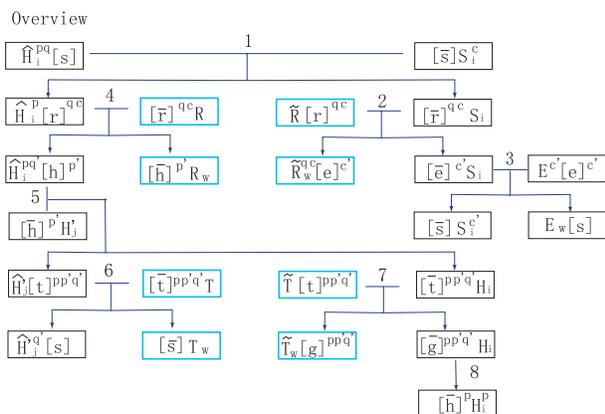
Figure 14: Overview of the operation of the Autonomous DNA Turing Machine.

translational motion. This motion is a symbolic motion in the sense that no physical entity is moved from one location to the other. Instead, the motion is the motion of the active head symbol relative to the tracks. A nanorobotics challenge is to extend the Autonomous DNA Turing Machine to a device that can move a physical entity, probably a DNA fragment, in a universal translational motion fashion. As a first step, it is conceivable that a DNA nanomechanical device that moves a DNA fragment bidirectionally along the track can be designed and possibly experimentally constructed.

Our complex design of Autonomous DNA Turing Machine makes some unconventional physical and chemical assumptions. Two lines of recent work lends partial experimental support to the practicality of this design. The first one is the autonomous DNA finite state automata experimentally constructed by Shapiro's group [3, 4, 5], in which a cascade of cleavages and ligations drive the operation of the machine. A more relevant study is the experimental construction of an autonomous unidirectional DNA walker that moves along a DNA track [40]. In this device, a walker moves unidirectionally over a sequence of three dangling anchorages sites (a structural analog to the dangling-molecules in the Autonomous DNA Turing Machine design) embedded in a DNA track in an autonomous fashion, driven by alternating actions of DNA endonulceases and ligases. In particular, this walking device exploits some very similar enzyme reactions as used in the design of the Autonomous DNA Turing Machine, such as the ligation and cleavages of DNA duplices tethered to another DNA nanostructure and the ligation of DNA fragments with 3-base overhangs at a relatively high temperature ($37\,^{\circ}C$).

Though a full experimental implementation of the Autonomous DNA Turing Machine appears daunting, due to the rich set of molecules, reactions, futile reactions involved, it might be possible to experimentally test a subset of the mechanisms described here. Another challenge

to experimental demonstration of the Autonomous DNA Turing Machine is the design of an output detection mechanism. Many futile reactions happen in the background during the operation of the Autonomous DNA Turing Machine. A key feature of these futile reactions is that they are *fully reversible*. This is critical in ensuring the autonomous operation of the Autonomous DNA Turing Machine as explained below. We *initially* supply the system with sufficiently high concentrations of Rule-Molecules and Assisting-Molecules as well as all the *byproducts* generated in the futile reactions. As such, the futile reactions will reach a dynamic balance and the concentrations of all the components involved in the futile reactions, including both the "active" components essential for the operation of the Autonomous DNA Turing Machine and the "futile" byproducts, will stay relatively constant during the operation of the Autonomous DNA Turing Machine. Note that since the active components will *not* be depleted by the futile reactions (which could have happened should some futile reactions are irreversible), the autonomous operation of the Autonomous DNA Turing Machine will not be disrupted. Though we have shown that the futile reactions are innocuous, they do decrease the efficiency of Autonomous DNA Turing Machine. A desirable improvement of the current design is to decrease the level of futile reactions and thus increase the efficiency and robustness of the Autonomous DNA Turing Machine.

Acknowledgement

References

[1] L. Adleman. Molecular computation of solutions to combinatorial problems. *Science*, 266:1021–1024, 1994.

[2] P. Alberti and J. L. Mergny. DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *PNAS*, 100:1569–1573, 2003.

[3] Y. Benenson, R. Adar, T. Paz-Elizur, E. Keinan, Z. Livneh, and E. Shapiro. DNA molecule provides a computing machine with both data and fuel. *PNAS*, 100:2191–2196, 2003.

[4] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro. An autonomous molecular computer for logical control of gene expression. *Nature*, Advance online publication, 2004.

[5] Y. Benenson, T. Paz-Elizur, R. Adar, E. Keinan, Z. Livneh, and E. Shapiro. Programmable and autonomous computing machine made of biomolecules. *Nature*, 414:430–434, 2001.

[6] Y. Chen, M. Wang, and C. Mao. An autonomous DNA motor powered by a DNA enzyme. *Angew. Int. Ed.*, 43:2–5, 2004.

[7] D. Faulhammer, A. R. Cukras, R. J. Lipton, and L. F. Landweber. Molecular computation: RNA solutions to chess problems. *Proc. Natl Acad. Sci. USA*, 97:1385 – 1389, 2000.

[8] L. Feng, S. H. Park, J. H. Reif, and H. Yan. A two-state DNA lattice switched by DNA nanoactuator. *Angew. Int. Ed.*, 42:4342–4346, 2003.

[9] A. A. Henry and F. E. Romesberg. Beyond A, C, G, and T: augmenting nature's alphabet. *Curr. Opin. Chem. Biol.*, 7:727–733, 2003.

[10] T. H. LaBean, H. Yan, J. Kopatsch, F. Liu, E. Winfree, J. H. Reif, and N. C. Seeman. The construction, analysis, ligation and self-assembly of DNA triple crossover complexes. *Journal of American Chemistry Society*, 122:1848–1860, 2000.

[11] L. F. Landweber, R. J. Lipton, and M. O. Rabin. DNA^2 DNA computations: A potential 'Killer App'? In H. Rubin and D. H. Wood, editors, *DNA Based Computers III: DIMACS Workshop, June 23-27, 1997, University of Pennsylvania*, pages 161–172, Providence, Rhode Island, 1997. American Mathematical Society.

[12] J. Li and W. Tan. A single DNA molecule nanomotor. *Nanoletter*, 2:315–318, 2002.

[13] R. J. Lipton. DNA solution of hard computational problem. *Science*, 268:542–545, 1995.

[14] Q. Liu, L. Wang, A. G. Frutos, A. E. Condon, R. M. Corn, and L. M. Smith. DNA computing on surfaces. *Nature*, 403:175–179, 2000.

[15] C. Mao, T. H. LaBean, J. H. Reif, and N. C. Seeman. Logical computation using algorithmic self-assembly of dna triple-crossover molecules. *Nature*, 407:493–495, 2000.

[16] C. Mao, W. Sun, and N. C. Seeman. Designed two-dimensional DNA holiday junction arrays visualized by atomic force microscopy. *Journal of the American Chemical Society*, 121:5437–5443, 1999.

[17] C. Mao, W. Sun, Z. Shen, and N. C. Seeman. A DNA nanomechanical device based on the B-Z transition. *Nature*, 397:144–146, 1999.

[18] Q. Ouyang, P. D. Kaplan, S. Liu, and A. Libchaber. DNA solution of the maximal clique problem. *Science*, 278:446–449, 1997.

[19] J. H. Reif. Paradigms for biomolecular computation. In *First International Conference on Unconventional Models of Computation, Auckland, New Zealand*, pages 72–93, 1998.

[20] J. H. Reif. Local parallel biomolecular computation. In H. Rubin and D. H. Wood, editors, *DNA-Based Computers*, volume 48 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 217–254. American Mathematical Society, 1999.

[21] J. H. Reif. Parallel molecular computation: Models and simulations. In *Proceedings: 7th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA'95) Santa Barbar,CA*, pages 213–223, 1999.

[22] J. H. Reif. The design of autonomous DNA nanomechanical devices: Walking and rolling DNA. *Lecture Notes in Computer Science*, 2568:22–37, 2003. Published in Natural Computing, DNA8 special issue, Vol. 2, p 439-461, (2003).

[23] P. W. K. Rothemund. A DNA and restriction enzyme implementation of turing machines. In R. J. Lipton and E. B. Baum, editors, *DNA Based Computers: Proceedings of the DIMACS Workshop, April 4, 1995, Princeton University*, pages 75 – 119, Providence, Rhode Island, 1996. American Mathematical Society.

[24] A. J. Ruben and L. F. Landweber. The past, present and future of molecular computing. *Nature Rev. Mol. Cell Biol.*, 1:69–72, 2000.

[25] N. C. Seeman. DNA in a material world. *Nature*, 421:427–431, 2003.

[26] W. B. Sherman and N. C. Seeman. A precisely controlled DNA biped walking device. *Nano. Lett.*, to appear, 2004.

[27] F. C. Simmel and B. Yurke. Using DNA to construct and power a nanoactuator. *Physical Review E*, 63:041913, 2001.

[28] F. C. Simmel and B. Yurke. A DNA-based molecular device switchable between three distinct mechanical states. *Applied Physics Letters*, 80:883–885, 2002.

[29] W. D. Smith. DNA computers in vitro and in vivo. In R. J. Lipton and E. B. Baum, editors, *DNA Based Computers: Proceedings of the DIMACS Workshop, April 4, 1995, Princeton University*, pages 121 – 185, Providence, Rhode Island, 1996. American Mathematical Society.

[30] A. J. Turberfield, J. C. Mitchell, B. Yurke, Jr. A. P. Mills, M. I. Blakey, and F. C. Simmel. DNA fuel for free-running nanomachine. *Phys. Rev. Lett.*, 90:118102, 2003.

[31] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. In *Proc. London Math. Society Ser. II*, volume 42 of 2, pages 230–265, 1936.

[32] A. M. Turing. On computable numbers, with an application to the entscheidungsproblem. In *Proc. London Math. Society Ser. II*, volume 43, pages 544–546, 1937.

[33] E. Winfree. On the computational power of DNA annealing and ligation. In R. J. Lipton and E. B. Baum, editors, *DNA Based Computers*, volume 27 of *DIMACS*, pages 199–221. American Mathematical Society, 1996.

[34] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman. Design and self-assembly of two-dimensional DNA crystals. *Nature*, 394:539–544, 1998.

[35] S. Wolfram. *A new kind of science*. Wolfram Media, Champaign, IL, 2002.

[36] H. Yan, T. H. LaBean, L. Feng, and J. H. Reif. Directed nucleation assembly of barcode patterned dna lattices. *Proceedings of the National Academy of Science*, 100:8103–8108, 2003.

[37] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, and T. H. LaBean. DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science*, 301:1882–1884, 2003.

[38] H. Yan, X. Zhang, Z. Shen, and N. C. Seeman. A robust DNA mechanical device controlled by hybridization topology. *Nature*, 415:62–65, 2002.

[39] P. Yin, A. J. Turberfield, and J. H. Reif. Designs for autonomous unidirectional walking DNA devices. In *DNA Based Computers 10 (to appear)*, 2004.

[40] P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield, and J. H. Reif. A unidirectional DNA walker moving along a track. 2004. In preparation.

[41] B. Yurke, A. J. Turberfield, Jr. A. P. Mills, F. C. Simmel, and J. L. Neumann. A DNA-fueled molecular machine made of DNA. *Nature*, 406:605–608, 2000.

A Basic notations
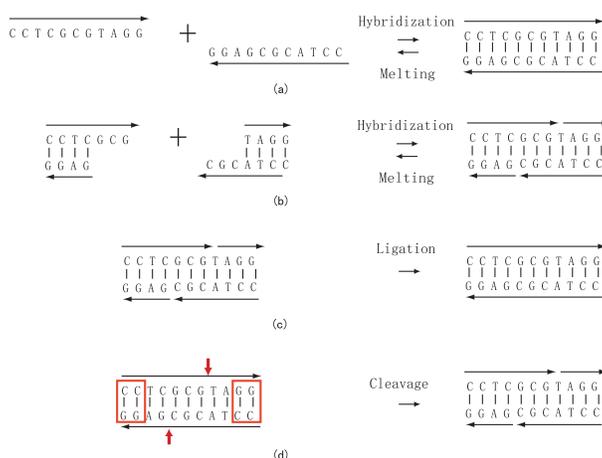
Figure 15: (a) Watson-Crick complementarity and hybridization of complementary DNA strands. Two directed single strands with anti-parallel complementary sequences form a DNA duplex. Each single strand is represented by directed line segment accompanied by the base sequence. The hydrogen bonding interactions between two complementary bases are represented with short vertical lines. (b) Hybridization of sticky ends. (c) Ligation. (d) Cleavage. The recognition site and cleavage site are indicated with boxes and a pair of bold arrows, respectively.

We start with the DNA molecule: it consists of directed strands referred to as *single strand DNA*, which consists of a chain of monomeric units known as *bases*. There are four types of bases: A, T, C, and G. A DNA molecule consisting of two anti-parallel single strands is commonly known as a *duplex* DNA, or the famous double helix DNA. The two anti-parallel single strands are associated with each other via the bonding of corresponding bases of the two strands known as *Watson-Crick complementarity*, or *complementarity* for short: base A prefers to form hydrogen bonds with T; base C with G. Hence A (resp. C) is the complementary base to T (resp. G). Note that a more precise name of the complementarity between two single strands is *reverse* complementarity, since one strand is actually complementary to the reverse of the other strand. However, when the context is clear, we will simply use complementarity. Two single DNA strands with anti-parallel complementary bases can bond to each other and form duplex DNA in a process known as *hybridization*. The reverse process in which two bonded single strand DNA go apart is known as *melting*. Hybridization and melting are in dynamic balance, and the tendency towards melting increases as the temperature of the system increases. Figure 15 (a) illustrates the concepts of complementarity and the formation of duplex DNA by hybridization of two single strand as well as the melting process. A duplex DNA fragment may possess single strand overhang at its end, and this single strand is called a *sticky end*. Two duplex DNA with complementary sticky ends can come together via hybridization of their sticky ends as shown in Figure 15 (b). Again, the hybridization product can dissociate via the melting process.

Two basic enzyme operations on DNA molecules are *ligation* and *cleavage*. Two duplex DNA fragments possessing complementary sticky ends can associate with each other via the hybridization of their sticky ends (Figure 15 (b)). In the presence of *ligase* enzyme, the nicks at either end of the hybridized section are subsequently sealed in a process known as *ligation*, joining the two duplex fragments into one via the formation of covalent bonds in the sugar-phosphate backbone of DNA strands, as shown in Figure 15 (c). As such, two separate DNA duplex fragments with complementary sticky ends can be joined into one in a two step process, hybridization and then ligation. However, when the context is clear, we simply use ligation to refer to this two step process. In *cleavage*, an approximately reverse process to ligation, a duplex DNA fragment is cut into two separate duplex parts (with each usually possessing a complementary sticky end) by enzymes known as restriction *endonucleases*. Cleavage by an endonuclease usually requires that the substrate DNA fragment contains *recognition site(s)* (specific DNA sequence(s)) detectable by the endonuclease and that the cleavage happens at specific *cleavage sites* on the DNA fragment. Note the difference between recognition sites and cleavage sites. Figure 3 depicts some commercially available endonucleases (which are used in the construction of our Autonomous DNA Turing Machine). In contrast, ligation does not require specific recognition sites, though it requires complementarity of the sticky ends of the two parts to be joined.

The hybridization, melting, ligation, and cleavages are all referred to generically as *reactions*. The reactions happen in an aqueous solution system, or *solution* for short, where the DNA components as well as enzymes float freely and "bump" into each other and hence incur reactions.
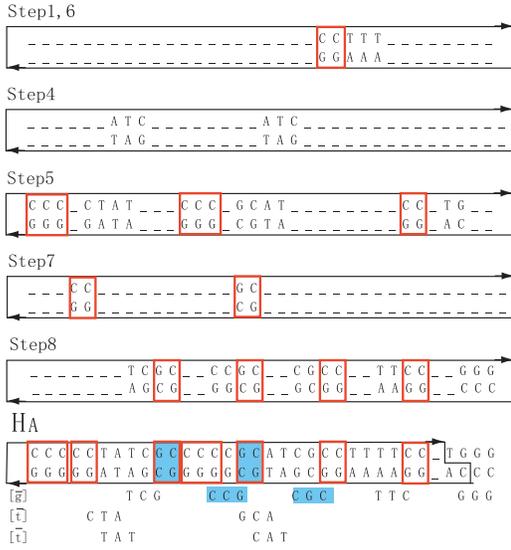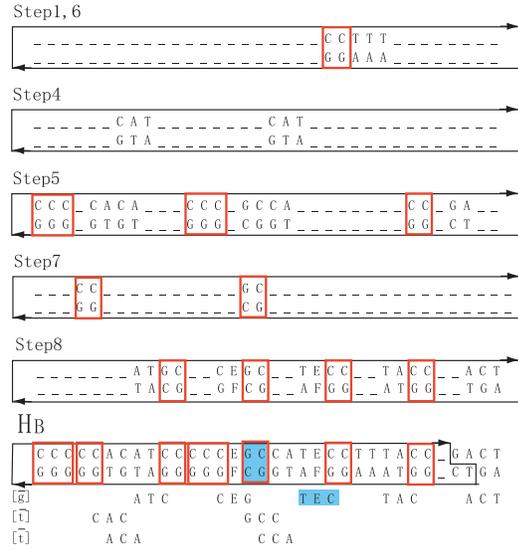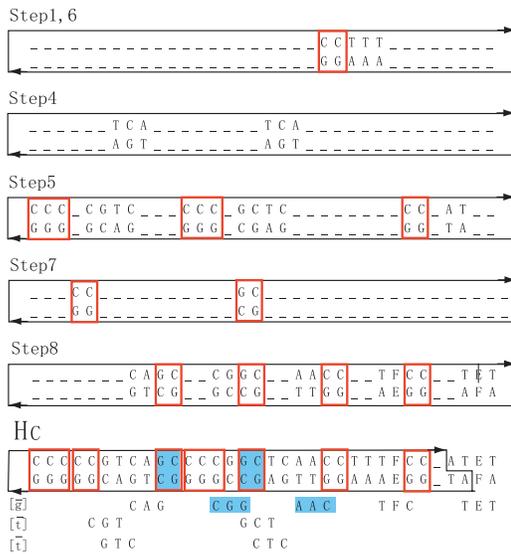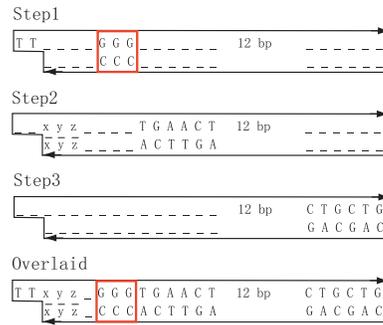
Figure 16: Head-Molecules and Symbol-Molecule used in Autonomous DNA Turing Machine. Panels (a), (b), and (c) describe Head-Molecules of position types P_A, P_B, and P_C, respectively; panel (d) illustrates a Symbol-Molecule. The molecule at the bottom of each panel gives the complete sequence information of a Head-Molecule or Symbol-Molecule while the molecules above the bottom one illustrate the relevant bases for individual steps. The sequence that belongs to an endonuclease recognition site is bounded with a box. The bases whose values are irrelevant to endonuclease recognition sites or unique sticky ends are denoted with "-". In panels (a), (b), and (c), the unique sticky ends generated in Steps 5 and 8 are listed below the bottom molecule, and labeled with $[\bar{t}]$ and $[\bar{g}]$, respectively. Of these sticky ends, those generated by endonuclease Mwo I are further labeled with gray background. See Table 2 for the values of xyz.
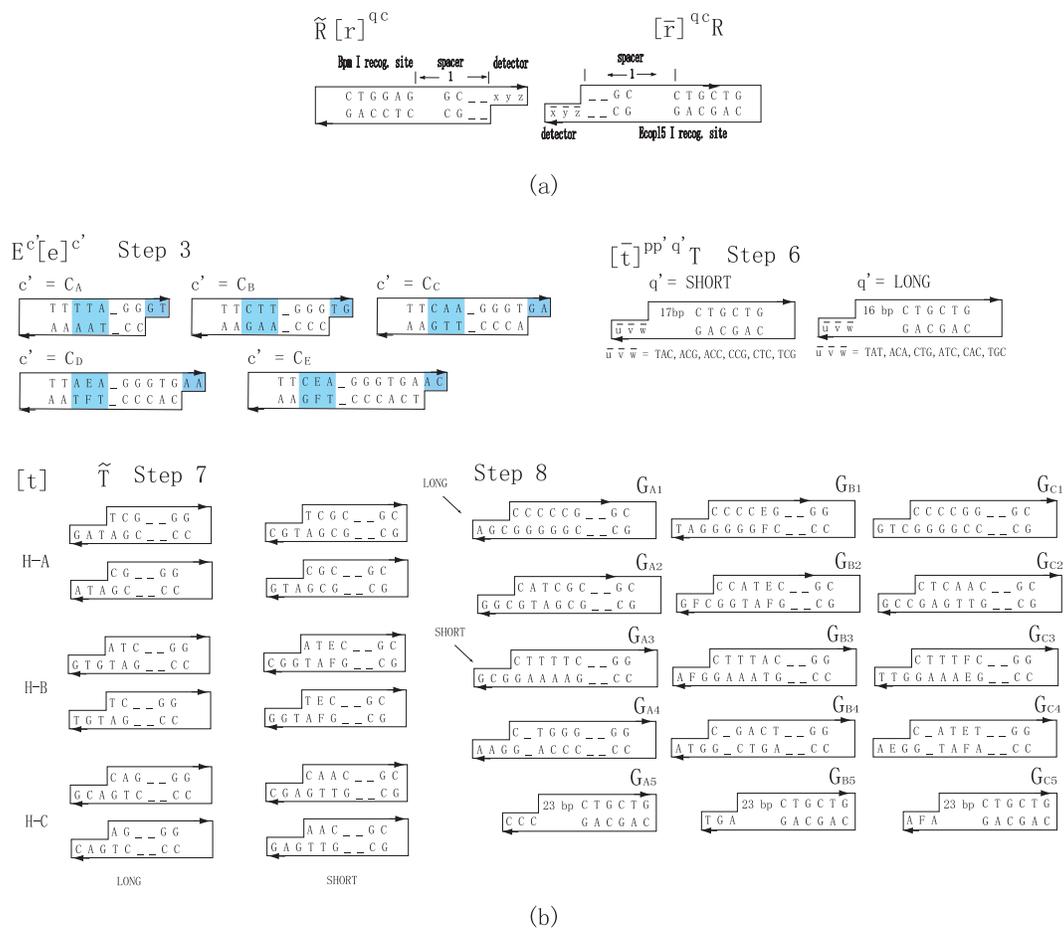
Figure 17: (a) The Rule-Molecules used in Autonomous DNA Turing Machine. xyz values are determined as in Table 1. The l values for R and $\tilde{R}$ are determined as described in Tables 2 and 3, respectively. (b)The Assisting-Molecules used in Autonomous DNA Turing Machine.

B   Molecule set

Figure 16 describes the Head-Molecules and Symbol-Molecules used in the construction of Autonomous DNA Turing Machine, respectively. A Head-Molecule encodes several layers (for Steps 1, 4, 5, 6, 7, and 8) of information in a single molecule. Figure 16 (a), (b), and (c) describe the layout as well as the overlaid scheme of the layers of information for Head-Molecules of type H_A, H_B, and H_C, respectively. The Symbol-Molecule participates in Steps 1, 2, and 3 and its construction is comparatively simple. Figure 16 (d) illustrates the relevant sequences for each of the three steps. The Rule-Molecules and Assisting-Molecules are shown in Figure 17.

C   Reversible futile reactions

We note that there exist reactions other than those already described during the operation of the Autonomous DNA Turing Machine. Upon careful examination, we will see that these reactions do not block, reverse, or alter the proper operation of the Autonomous DNA Turing Machine, although they decrease its efficiency. Therefore, these innocuous reactions are referred to as *futile reactions*.

The first kind of futile reactions (F1) occur between a Rule-Molecule $([\bar{r}]^{qc})R$ and its dual $\tilde{R}([r]^{qc})$ (Figure 18 (a)) or a Rule-Molecule $[\bar{t}]T$ and its dual $\tilde{T}[t]$ (Figure 18 (b)). Note that the GG sticky end in Figure 18 (b) is different from the $[s]$ sticky end of the Symbol-Molecules since it is a protruding $5'$ end instead of a protruding $3'$ end.

The second kind of futile reactions (F2) occur between Symbol-Molecules and Head-Molecules (Figure 18 (c)) with complementary sticky ends. We can not completely avoid the occurrence of these undesirable complementary sticky ends due to the limited encoding space. See Figure 19 for examples. However, the endonuclease Bsl I recognition site in the ligation product makes the undesirable ligation reversible and hence innocuous.

The third kind of futile reactions (F3) occur between S_i and H_i or $\tilde{R}$ – these futile reactions are caused by the endonuclease EcoPl5 I recognition site in the duplex portion of S_i. Figure 18 (d) and (e) illustrate these two cases. We need to pay particular attention to futile reaction F3b. In this case, the molecule $\tilde{R}_f[r_f]$ could diffuse away and hence the ligation for regenerating $\tilde{R}S$ could be blocked. This would disastrously block the operation of the whole Autonomous DNA Turing Machine. To fix this problem, we require that the concentration of $\tilde{R}_f[r_f]$ molecules stays sufficiently high in the system. This additional requirement warrants the regeneration of $\tilde{R}S$ and hence makes reaction F3b an innocuous futile reaction.

In addition, we note that all the cleavage reactions during the operation of Autonomous DNA Turing Machine are reversible, and hence represent additional idling processes or innocuous futile reactions.

D   Encoding space

One major challenge in designing DNA nanomechanical devices is the limited encoding space dictated by the four (six) letter vocabulary of the bases and by the sizes of the recognition, restriction, and spacing regions of endonucleases. Figure 19 (a) gives a table of all the possible permutations of 3-base sequences consisting of A, C, G, and T. Among them, the sequences of 3-base sticky ends used in the construction of Autonomous DNA Turing Machine are labeled with boxes. Figure 19 (b) shows some additional 3-base sticky ends containing the unnatural bases E and F. Figure 19 (a) and (b) together contain all the 3-base sticky ends used in the construction.

The encoding schemes have the following properties. First, each sticky end is unique – this ensures that the transition of state and color, the motion of the head, and the restorations of Symbol-Molecules and Head-Molecules are conducted according to designated rules. In addition, we need to ensure that there are no cross ligations between these sticky ends that can hinder the operation of the Autonomous DNA Turing Machine. Undesirable cross reactions could result from the un-programmed hybridizations between sticky ends of molecules during different stages. Due to the limited encoding space available to 3-base sticky ends, we can not avoid the cross hybridization completely. However, we carefully ensure that such cross hybridizations only result in idling processes and do not block or alter the programmed operation of the Turing Machine. To this end, we require that the cross hybridization only happen either between two sticky ends both generated by endonuclease Bsl I or two sticky ends both generated by endonuclease Mwo I. A ligation product between two molecules resulted from such cross hybridizations will be cut back into the original molecules and hence such ligation only represents an idling process or an innocuous futile reaction (see Figure 18 (c)). By using the expanded 6-base vocabulary of A, C, G, T, E, and F, we only have one such cross ligation (CCG,CGG).
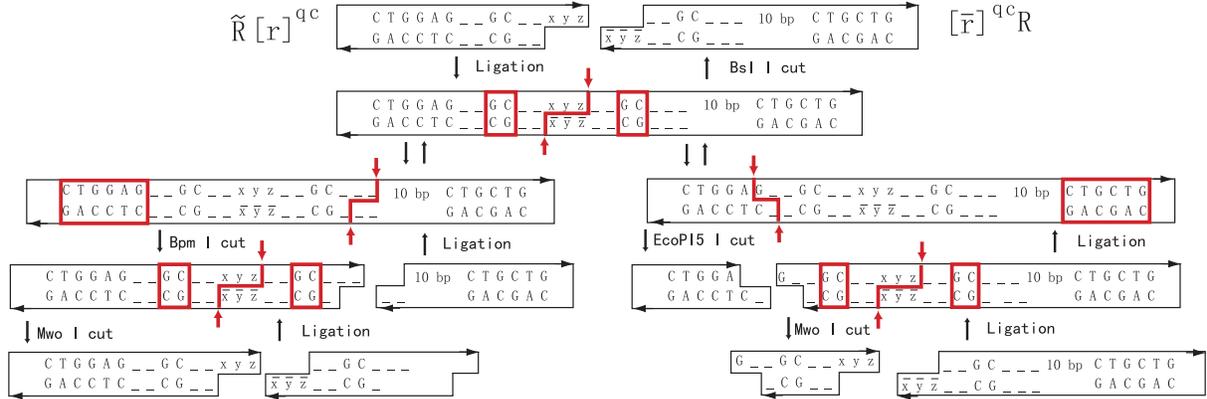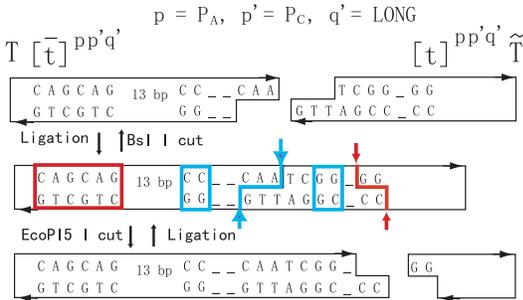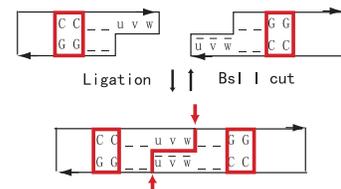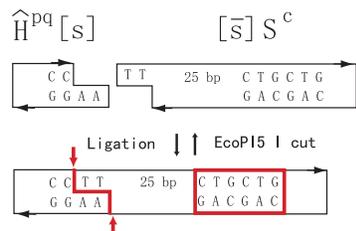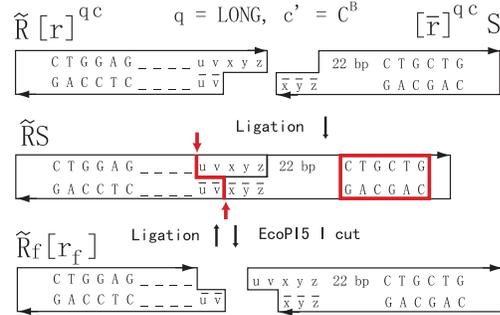
Figure 18: Futile reactions during the operation of Autonomous DNA Turing Machine. Endonuclease recognition sites and cleavage sites are indicated with boxes and pairs of bold arrows, respectively.

(a)

(b)
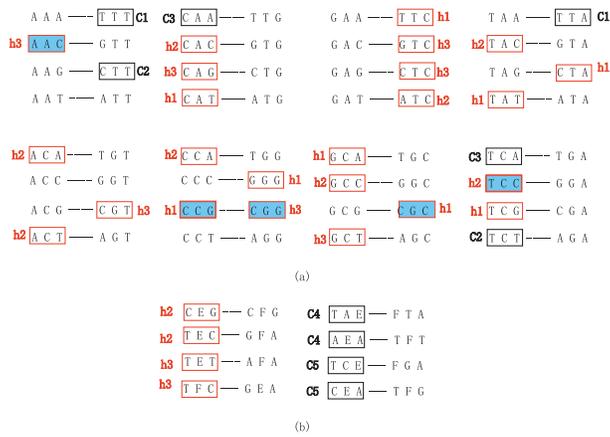
Figure 19: Encoding schemes used for 3-base sticky ends. The 3-base sequences are laid out in pairs such that two sequences in a pair are complementary to each other. Note that each 3-base sequence is written in 5' to 3' direction. During hybridization, the direction of one of the sequences will be reversed to 3' to 5'. For example, AAC is paired with GTT. The reverse of GTT is TTG, and this sequence is complementary to AAC. The sequences used to encode the current state q and current color c of Head-Molecule $\dot{H}_i[r]^{qc}$ and Symbol-Molecule $[\bar{r}]^{qc}\hat{S}$ are bounded with black boxes and labeled with $c1$, $c2$, $c3$, $c4$, and $c5$ (indicating color C_A, C_B, C_C, C_D, and C_E, respectively). The sequences used to encode the position types of Head-Molecules during Step 5 - 8 are bounded with boxes and labeled with $h1$, $h2$, and $h3$ (indicating the position types P_A, P_B, and P_C, respectively). The sticky ends produced by endonuclease Mwo I are further indicated with gray background.