

Meta-DNA: A DNA-Based Approach to Synthetic Biology

Harish Chandran¹
harish@cs.duke.edu

Nikhil Gopalkrishnan¹
nikhil@cs.duke.edu

Bernard Yurke²
bernardyurke@boisestate.edu

John Reif^{1,3}
reif@cs.duke.edu

Abstract

The goal of synthetic biology is to design and assemble synthetic systems that mimic biological systems. One of the most fundamental challenges in synthetic biology is to synthesize artificial biochemical systems, which we will call meta-biochemical systems, that provide the same functionality as biological nucleic acids-enzyme systems, but that use a very limited number of types of molecules. The motivation for developing such synthetic biology systems is to enable a better understanding of the basic processes of natural biology, and also to enable re-engineering and programmability of synthetic versions of biological systems.

One of the key aspects of modern nucleic acid biochemistry is its extensive use of protein enzymes that were originally evolved in cells to manipulate nucleic acids, and then later adapted by man for laboratory use. This practice provided powerful tools for manipulating nucleic acids, but also limited the extent of the programmability of the available chemistry for manipulating nucleic acids, since it is very difficult to predictively modify the behavior of protein enzymes. Meta-biochemical systems offer the possible advantage of being far easier to re-engineer and program for desired functionality.

The approach taken here is to develop a biochemical system which we call meta-DNA (abbreviated as mDNA), based entirely on strands of DNA as the only component molecules. Our work leverages prior work on the development of self-assembled DNA nanostructures (see Amin et al. (2009); LaBean et al. (2007); Seeman (2004); Deng et al. (2006); Lund et al. (2006); Bath and Turberfield (2007); Winfree (2003) for excellent reviews of the field). Each base of a mDNA is a DNA nanostructure. Our mDNA bases are paired similar to DNA bases, but have a much larger alphabet of bases, thereby providing increased power of base addressability. Our mDNA bases can be assembled to form flexible linear assemblies (single stranded mDNA) analogous to single stranded DNA, and can be hybridized to form stiff helical structures (duplex mDNA) analogous to double stranded DNA, and also can be denatured back to single stranded mDNA.

Our work also leverages the abstract activatable tile model developed by Majumder et al. (2007) and prior work on the development of enzyme-free isothermal protocols based on DNA hybridization and sophisticated strand displacement hybridization reactions (see Reif and Majumder (2008); Sakamoto et al. (1999); Dirks and Pierce (2004); Zhang et al. (2007); Tian et al. (2006); Sherman and Seeman (2004); Yin et al. (2004)). We describe various isothermal hybridization reactions that manipulate our mDNA in powerful ways analogous to DNA-DNA re-

¹ Department of Computer Science, Duke University, Durham, NC, USA

² Department of Materials Science and Engineering and Department of Electrical and Computer Engineering, Boise State University, Boise, ID, USA

³ Adjunct Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

This chapter is an extensively revised version of Harish Chandran, Nikhil Gopalkrishnan, Bernard Yurke, John Reif, Meta-DNA: Synthetic Biology via DNA Nanostructures and Hybridization Reactions, Journal of the Royal Society Interface, (2012). published online 11 January 2012, pp. 1742-5662

actions and the action of various enzymes on DNA. These operations on mDNA include (i) hybridization of single strand mDNA (ssmDNA) into a double strand mDNA (dsmDNA) and heat denaturation of a dsmDNA into its component ssmDNA (analogous to DNA hybridization and denaturation), (ii) strand displacement of one ssmDNA by another (similar to strand displacement in DNA), (iii) restriction cuts on the backbones of ssmDNA and dsmDNA (similar to the action of restriction enzymes on DNA), (iv) polymerization chain reactions that extend ssmDNA on a template to form a complete dsmDNA (similar to the action of polymerase enzyme on DNA), (v) isothermal denaturation of a dsmDNA into its component ssmDNA (like the activity of helicase enzyme on DNA) and (vi) an isothermal replicator reaction which exponentially amplifies ssmDNA strands (similar to the isothermal PCR reaction). We provide a formal model to describe the required properties and operations of our mDNA, and show that our proposed DNA nanostructures and hybridization reactions provide these properties and functionality.

Keywords: DNA self-assembly; synthetic biology; DNA nanostructures

1 Introduction

1.1 Synthetic Biology

A major goal of synthetic biology is to produce synthetic biochemical systems which have functions similar to the biochemical functions of living organisms. Considerable related work has been done in the fields of artificial and synthetic life, see Luisi (2006); Bedau et al. (2000); Packard and Bedau (2003) for an overview of these fields. Two predominant approaches by researchers in this area have been to design protein-based or RNA-based biochemical systems. However, both proteins and RNA are very difficult to predictively design for given functions, and behavior of the resulting protein-based or RNA-based biochemical systems can be very complex, making their engineering highly challenging. An alternative approach we propose here is to produce synthetic biochemical systems based on a very well understood, relatively non-reactive molecule, such as DNA (DNA strands and DNA nanostructures). DNA-DNA interactions like hybridization and strand displacement are reasonably well understood and a vast literature exists that studies, models, predicts and even controls such interactions (see Zhang and Winfree (2009); Koval et al. (1999)). This literature enables us to program simple DNA systems by controlling experimental conditions like pH, salt concentrations, temperature and DNA concentration.

1.2 Overview of Our Work on meta-DNA

In this work, we engineer synthetic biochemical systems, called meta-DNA, consisting only of DNA that capture the properties and structure of DNA in biological systems. Our work is reductive: we use simple DNA chemistry to emulate more complex enzyme based DNA chemistry through ingenious use of DNA hybridization, strand displacement and hairpin systems. From a computer science perspective, our work can be thought of as using a lower level programming language to simulate programs encoded in a higher level programming language. This kind of simulation often leads to a better abstract understanding of the phenomena being simulated. From a synthetic biology perspective, we are building novel biochemical systems to emulate useful, well known natural biological systems and providing alternatives to protein enzymes. From an engineering perspective, our work is a minimalist approach to designing biochemical systems from simple, predictable yet powerful modules. Our systems are largely isothermal and autonomous which suggest that they may have applications for in vitro biochemical systems like transport devices, molecular motors, detection, signalling and computing systems.

1.3 Prior Work on Synthetic Replicators

Driven by the important role that replication plays in Biology, many self-replicating systems have been proposed, starting with Von Neumann. Early self-replicating systems were designed by von Kiedrowski (1986); Tjivikua et al. (1990). For a review of various artificial replicators see Vidonne and Philp (2009); von Kiedrowski and Patzke (2007); Dadon et al. (2008). A DNA-based artificial replicator was proposed by Zhang and Yurke (2006). Schulman and Winfree (2008) study growth and evolution of simple crystals using DNA. Smith et al. (2002) have independently developed abstractions for self-replication systems that can be thought of as tile-based and also rely on the idea of activation.

The major departure from this prior work on self-replicating systems and our meta-DNA approach, is that meta-DNA allows for a wide variety of key operations far beyond merely the operation of replication, enabling a much more extensive and complex set of synthetic biochemical systems.

1.4 Organization of this Chapter

In Section 2 we review the key properties of DNA, including its structure and enzymatic reactions. In Section 3 we list the desirable properties of meta-DNA and give a design for a meta-nucleotide, single and double stranded meta-DNA, along with its secondary structure. Section 4 states the assumptions we make while designing protocols for meta-DNA manipulation. In Section 5 we discuss various meta-DNA reactions such as meta-hybridization, meta-denaturation, meta-strand displacement, meta-polymerization, meta-restriction, meta-helicase denaturation and exponential amplification using a meta-DNA replicator. Section 6 discusses open problems and future work.

2 Review of DNA Structure and Reactions

DNA is considered to be an ideal material to construct nanoscale structures and devices and has been used as scaffolding material for complex shapes, fuel for molecular motors and aptamers for various organic and inorganic molecules. The key properties of DNA that enable these functionalities are programmability, predictable chemical interaction and secondary structure and simple laboratory protocols for its manipulation. Synthetic DNA is also cheaply and readily available from a variety of commercial sources. At the most abstract level, fabricating structures and devices with DNA is akin to working with smart bricks that fit together in a specific predefined way and then putting them in a bag, shaking it and waiting for the bricks to self-assemble. However, the process is probabilistic and correcting errors is a fundamental challenge for the field.

Most readers will be familiar with the basic structure and reactions of DNA. The following discussion will be useful to even these readers as it frames DNA and its reactions in the context of our work and also specifies the granularity at which we consider DNA, with particular focus on the biochemical properties that we wish to simulate using our synthetic mDNA systems.

2.1 Key Properties of DNA

Here we review the key properties of natural DNA that we wish to mimic in mDNA. Single stranded DNA is a linear polymer made from repeating units called *nucleotides*. The nucleotide repeats contain both a segment of the *backbone* of the molecule, which holds the chain together, and a *base*. Each nucleotide has directionality and hence a DNA strand is asymmetric. The asymmetric ends of DNA strands are called the 5' (five prime) and 3' (three prime) ends. Single stranded DNA is flexible and has a small persistence length and is generally modeled as a freely-jointed chain. In living organisms, DNA does not usually exist as a single molecule, but instead as a pair of molecules entwined like vines, in the shape of a double helix. In a double helix the direction of the nucleotides in one strand is opposite to their direction in the other strand: the strands are *antiparallel*. Double stranded DNA has much greater persistence length than single stranded DNA and is generally modeled as a worm-like chain. The DNA double helix is stabilized by hydrogen bonds between the bases attached to the two strands. The four bases are classified into two types, purines (A and G) and pyrimidines (T and C). Each type of base on one strand overwhelmingly prefers a bond with just one type of base on the other strand. This is called *complementary base pairing*. Here, purines form hydrogen bonds to pyrimidines, with A bonding preferentially to T, and C bonding preferentially to G. This arrangement of two nucleotides binding together across the double helix is called a *base pair*. As hydrogen bonds are not covalent, they can be broken and rejoined relatively easily. The two strands of DNA in a double helix can therefore be pulled apart like a zipper, either by a mechanical force or kinetic energy due to high temperature. Given a pH value and salt (Mg^{++} , Na^+) concentrations, the *melting temperature* is defined as the temperature at which half the DNA strands are in a double helical conformation

while the rest are denatured. Melting temperature depends on the length of the DNA strand as well as its sequence, with longer strands having a higher melting temperature.

2.2 Key Reactions Involving DNA

We view DNA reactions as biochemical programs and attempt to emulate them. Hence, it is important to understand and abstract out these DNA reaction properties. With this goal in mind, we review a few key DNA reactions.

2.2.1 DNA-DNA Reactions

- *Hybridization* is the attachment of a pair of single strands of DNA via hydrogen bonds along their complementary regions to form a double helix. Note that the two strands only attach in an anti-parallel manner and hence are *reverse complements* of each other.
- The reverse process, separation of a double helix into its constituent single strands by the breaking of hydrogen bonds is DNA *denaturation*. This can be achieved via mechanical shear forces, high temperature or presence of denaturing agents like urea or formamide.
- DNA *strand displacement* is the displacement of a single strand of DNA from a double helix by an incoming strand with a longer complementary region to the template strand. The incoming strand has a *toehold*, an empty single stranded region on the template strand complementary to a subsequence of the incoming strand, to which it binds initially. It eventually displaces the outgoing strand via a kinetic process modeled as a one dimensional random walk.

2.2.2 Enzymatic Reactions on DNA

- DNA *restriction* is the cleaving of the backbone at a sequence specific recognition site by a restriction enzyme.
- DNA *polymerases* are a class of enzymes that catalyze the *polymerization* of nucleotides into a DNA strand. The polymerase “reads” an intact DNA strand as a template and uses it to synthesize the new strand. The newly polymerized molecule is complementary to the template strand. DNA polymerases lack the ability to do *de novo* polymerization and can only extend a DNA strand already attached to the template strand, called a *primer*.
- Isothermal denaturation can be achieved by *helicases* which are motor proteins that move directionally along a DNA backbone, denaturing the double helix.
- DNA *ligation* is the rejoining of nicked double stranded DNA by repairing the backbone break. Enzymes belonging to this class are known as *ligases*.

3 Desired Properties of meta-DNA

We wish to abstract the structure and reactions of DNA described in Section 2.2 and emulate them using only DNA-DNA interactions. In doing so, we would have circumvented DNA-enzyme chemistry with a synthetic biochemical system that uses only DNA hybridization. First we state the desirable properties of meta-DNA (mDNA).

3.1 List of Desirable Properties for mDNA

- We desire a set of $2k$ meta-nucleotides (where k is an integer ≥ 2). Each meta-nucleotide must be directional and have a $3'$ and $5'$ end. There must be two types of meta-nucleotides: meta-purines and meta-pyrimidines, each k in number. There must exist a pairing between meta-purines and meta-pyrimidines where each edge indicates an overwhelming complementary preference of meta-base linking between the meta-purine and meta-pyrimidine it connects.
- The active $5'$ end of any meta-nucleotide should be able to bind to the active $3'$ end of any other meta-nucleotide and vice versa via a meta-backbone link.
- The meta-backbone link must be a *strong* bond (implemented as a long sequence of hybridized DNA) while the meta-base link must be a *weak* bond (implemented as a short sequence of hybridized DNA).
- ssmDNA should be a linear directional polymer chain of meta-nucleotides bound by meta-backbone linkages and has $3'$ and $5'$ ends.
- Two ssmDNA that are reverse complementary to each other must have the ability to meta-hybridize to form a dsmDNA. The strands of any dsmDNA must be anti-parallel.
- When an ensemble of identical dsmDNA is heated to a temperature known as its melting temperature, half of the ensemble must denature into its constituent ssmDNA. The melting temperature must depend on the sequence of meta-nucleotides of the mDNA and also on its length, with longer strands having a higher melting temperature.
- To support strand displacement protocols, the weak meta-base bonds must be continuously broken and remade in a kinetic process called breathing. The rate of breathing should be positively correlated with temperature.
- ssmDNA must be flexible and have a small persistence length when compared to dsmDNA of comparable length. ssmDNA must be like a freely-jointed chain while dsmDNA must be like a worm-like chain.
- Meta-nucleotides should not spontaneously form meta-backbone bonds in the absence of catalyst.

3.2 Abstract Description of mDNA

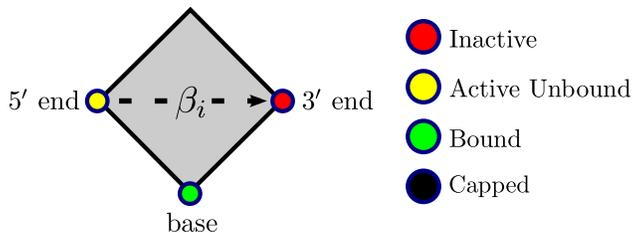


Figure 1: Abstraction of the nucleotide

We model a meta-nucleotide as an activatable tile Majumder et al. (2007) having three activatable pads: a $5'$ pad, a $3'$ pad and a base pad and represent it by a square tile as illustrated in Figure 1. The tile has directionality as indicated by an arrow from $5'$ to $3'$ which is imposed by the sequence in which the pads are activated, with $5'$ always activated before $3'$. Tiles bind to

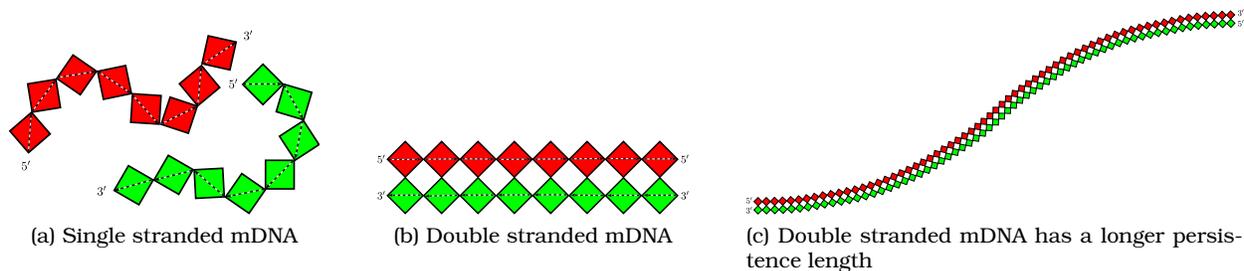


Figure 2: Single and double stranded mDNA

each other via symmetric pad interactions called *binding* or *linking*. Each binding has a *strength* associated with it (1, 2 or 3) that depends only on the type of pads involved in the binding. The strength of a binding models the energy required to break the bond. Base pads can only interact with other base pads through strength 1 bindings, and are called *meta-base bindings*. 5' pads can only interact with 3' pads through strength 3 bindings and these are called *meta-backbone bindings*. The pads exist in one of four states: inactive, active unbound, bound and capped. Inactive pads do not bind with other pads. Any active unbound 5' pad of a meta-nucleotide can bind to any active unbound 3' pad of another meta-nucleotide via a bond of strength 3 after which these pads go to the bound state. There are $2k$ different base pads, each corresponding to a meta-nucleotide type. These are split into two complementary perfectly matched sets. Let $\mathfrak{P} = \{\beta_1, \beta_2, \dots, \beta_k, \bar{\beta}_1, \bar{\beta}_2, \dots, \bar{\beta}_k\}$ be the set of pads. For all i , pads β_i and $\bar{\beta}_i$ bind with each other. The face label on any tile indicates the base type of the meta-nucleotide. Any active unbound base pad of a meta-nucleotide can bind in an antiparallel manner to an active unbound complementary base pad of another meta-nucleotide by a strength 1 bond after which these pads go to the bound state.

A linear chain of these tiles held together by 5'-3' pad bindings forms a directional polymer of meta-nucleotides and hence is ssmDNA. The 3' base pad of the tile at the 3' end of the ssmDNA and the 5' base pad of the tile at the 5' end are always in the capped state. A double stranded mDNA (*dsmDNA*) is a dimer of two ssmDNA held together by base pad bindings. ssmDNA is flexible with a low persistence length while *dsmDNA* is stiff and has a higher persistence length. These properties arise out of the geometric constraints imposed by the tile base structure of mDNA as illustrated in Figure 2.

3.3 Strand Design for mDNA

Figure 3 shows an implementation of a meta-nucleotide tile as a DNA nanostructure. In Figure 3(a) we have the tile with no *protection* and in Figure 3(b) we have the tile with *protection* strands. The purpose of the protecting strands is to prevent spontaneous aggregation of tiles into mDNA. The protection mechanism is designed to impose the requirement that the 5' pad is activated before the 3' pad (this property is used in the polymerization and replicator protocols). Each letter denotes a DNA sequence and a bar atop a letter indicates reverse complement of the sequence that the letter denotes. The red strand ($fgdcba$) contains the 5' pad fg and also the base pad ba and the light green strand ($\bar{g}\bar{f}\bar{e}$) contains the 3' pad $\bar{g}\bar{f}$. The blue strand ($\bar{c}\bar{q}\bar{d}$) and the dark green strand (qe) are bridging strands that hold the nanostructure together and give it the required geometry. We will have a detailed discussion of the secondary structure of the meta-nucleotide tile and mDNA in Section 3.4. The sequence of reactions that occur when a tile is deprotected are as follows. The pink strand $\bar{b}x\bar{f}$ protects (renders inactive) the 5' pad. When ba binds to its complement on another tile, $\bar{b}x\bar{f}$ is ripped away from the 5' pad by the invasion of the strand $f\bar{x}b$ thus activating the 5' pad. The strand $fy\bar{g}$ protects (renders inactive) the 3' pad. When the 5' pad binds to its complement on another tile, $fy\bar{g}$ is ripped away from the 3' pad by the invasion of the strand $g\bar{y}f$ thus activating the 3' pad. Figure 5 illustrates all possible states a meta-nucleotide can exist in. We model *weak* base bonds and *strong* backbone bonds by making the sequences corresponding to the 5' and 3' pads much longer than the sequences corresponding to the base pads. The internal structure of both single and double stranded mDNA are shown in Figures 4a and 4b. Note the black protection strands on the tiles at the ends of the mDNA to implement capping.

3.4 Secondary Structure of mDNA

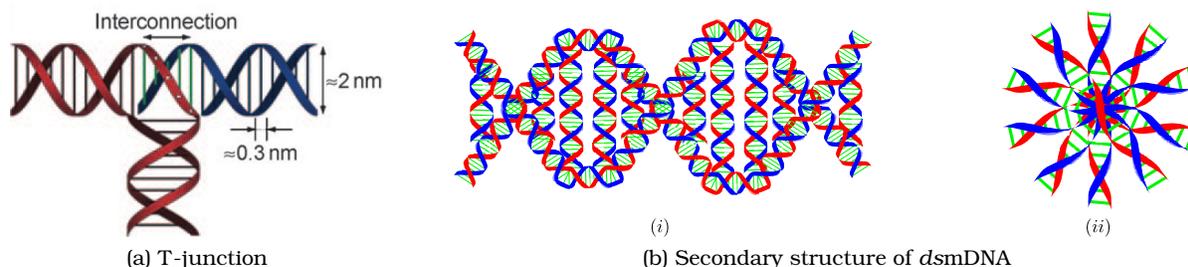


Figure 6: (a): Design of rigid T-junction self-assembled from DNA. Figure from Hamada and Murata (2009). (b (i)): Secondary structure of *dsDNA* forming a double helix (ratio of bases per axial rotation is programmable via strand design). Not to scale. (b (ii)): Axial view of *dsDNA*. Only the rungs of the twisted ladder structure are shown.

Apart from simulating the reaction properties of DNA, we also wish to simulate its secondary structure, in particular:

- The flexible nature of ssDNA characterized by shorter persistence length
- The rigid nature of dsDNA characterized by a longer persistence length
- The double helical structure of *dsDNA*

Recently, Hamada and Murata (2009) reported a novel self-assembled rigid T-shaped interconnected junction where each arm is a DNA double helix (Figure 6a). They synthesized tile based structures like $1D$ linear ladders, $1D$ ringed structures and $2D$ lattices using the T-junction geometry. This suggests that their T-junction motif is a useful widget for designing stable, rigid, well behaved self-assembled objects. We use the T-junction as the key motif in achieving the

required structural properties for mDNA. Each meta-nucleotide is a T-junction (Figure 3(a)). ssmDNA (Figure 4a) is a linear polymer of these T-junctions with consecutive base pad sections (the double helical structure $c \equiv \bar{c}$) not in the same plane, but slightly rotated so that they stick out of the plane of the paper. This rotation is controlled by carefully choosing the number of bases that make up the horizontal double helical section between consecutive vertical helical sections (Figure 4a). This secondary structure for ssmDNA induces a helical twist for dsmDNA (imagined in 6b). We can think of the secondary structure of dsmDNA as a twisted 1D ladder. The 1D ladder design in Hamada and Murata (2009) can be easily modified by adding or deleting a single base pair from the side rungs to induce a twist to get a double helix structure with approximately 10.5 meta-bases per turn of the double helix, mimicking the twist of dsmDNA. In particular, we choose the length of the repeating DNA double helical unit $efgdq \equiv \bar{q}\bar{d}\bar{g}\bar{f}\bar{e}$ to be either 41 (one less than the number of bases in 4 full turns of a DNA double helix) or 43 (one greater than the number of bases in 4 full turns of a DNA double helix). The following choice of lengths for the subsequences would potentially give us the required geometries and at the same time preserve the thermodynamic and kinetic properties that would allow our subsequent mDNA protocols to succeed ($|x|$ is the length of the DNA sequence represented by x): $|a_i| = |b_i| = |\bar{a}_i| = |\bar{b}_i| = 4, |c| = |\bar{c}| = |d| = |\bar{d}| = |q| = |\bar{q}| = 6, |f| = |\bar{f}| = |g| = |\bar{g}| = 12$ and $|e| = |\bar{e}| = 5$ or 7 .

4 Assumptions for mDNA Reactions

Before describing our protocols involving mDNA we give here our assumptions under which these reactions proceed. We also describe some simple rules of DNA chemistry which are repeatedly used like subroutines in our mDNA protocols.

- Our systems are maintained at only 3 different temperatures characterized by a parameter τ . At room temperature or $\tau = 1$, both the meta-backbone bonds and the meta-base bonds are stable but breathing still occurs. Recall that breathing is the phenomenon of meta-base bonds spontaneously breaking and forming. At the melting temperature, $\tau = 2$, the meta-backbone bonds are stable but the base bonds are broken. At freezing temperature or $\tau = 0$, breathing does not occur.
- The pH, salt concentrations and other factors that affect hybridization are set to levels such that spontaneous hybridization between a DNA sequence and its complement can occur.
- Strand displacement, as defined earlier, always occurs and proceeds to completion. The strand that gets displaced out remains in the solution.
- We assume that each subsequence, denoted by a letter in the figures, only interacts with its perfect and full reverse complements and no other spurious interactions occur.

5 Reactions in mDNA

We set out protocols for mDNA that mimic DNA-DNA and DNA-enzyme interactions. We have two kinds of figures in the discussions that follow. The abstraction diagrams illustrate the protocols in the abstract activatable tile model while the internal structure diagrams illustrate the protocols in greater detail.

5.1 Meta-hybridization and Meta-denaturation in mDNA

The simplest mDNA reactions are meta-hybridization and meta-denaturation, which are reverse reactions of each other. In meta-hybridization, at temperature $\tau = 1$ two complementary ssmDNA

strands bind via complementary base pad bindings to give *dsmDNA* (Figure 7a). Heating *dsmDNA* to temperature $\tau = 2$ meta-denatures the structure into its two component *ssmDNA* (Figure 7b).

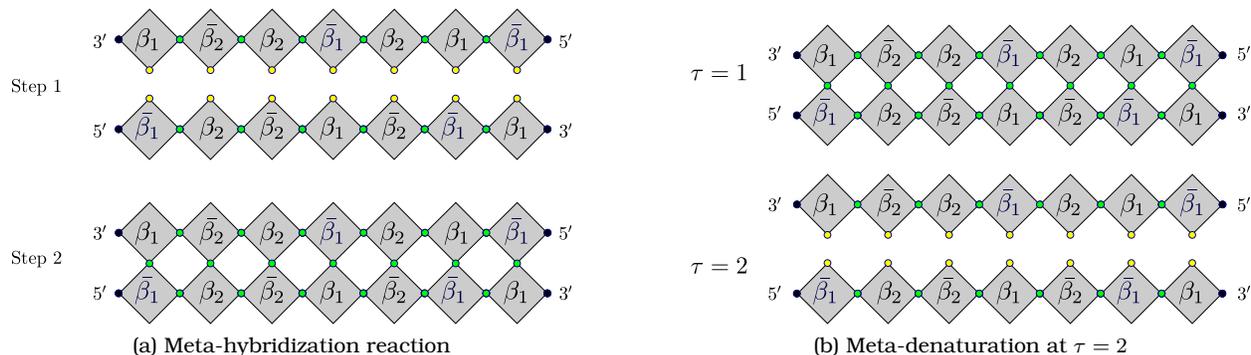


Figure 7: Meta-hybridization and meta-denaturation by heating

5.2 Meta-strand Displacement Process in mDNA

Meta-strand displacement for mDNA is defined as displacement of a *ssmDNA* from a meta-double helix by an incoming *ssmDNA* with a longer complementary region to the template strand. This reaction occurs at temperature $\tau = 1$. Figure 8 gives a high level view of meta-strand displacement using the activatable tile model. Two meta-strands compete to hybridize with a single *ssmDNA*. The shorter of the two is completely meta-hybridized to the template while the longer one comes in by gaining a toehold. Now, breathing of the meta-bases of the short strand gives an opportunity to the meta-bases of the competing incoming strand to meta-hybridize with the template. Note that the intermediate steps are reversible. However, once the incoming meta-strand completely displaces the outgoing meta-strand, the reaction stops as the outgoing meta-strand is extremely unlikely to come back in as it lacks a toehold.

Note that this reaction is made possible because the weak meta-base bonds can breathe at $\tau = 1$. Breathing in mDNA is expected to occur at a slower rate than in DNA because we require multiple bases to spontaneously denature for a single meta-base to denature. Also, contiguous bases in DNA are more immediately local than contiguous meta-bases in mDNA and hence the rate at which a meta-base occupies an empty spot on a complementary meta-base is also expected to be slower than for the corresponding process in DNA. Due to these reasons, we would expect meta-strand displacement in mDNA to proceed slower than strand displacement in DNA.

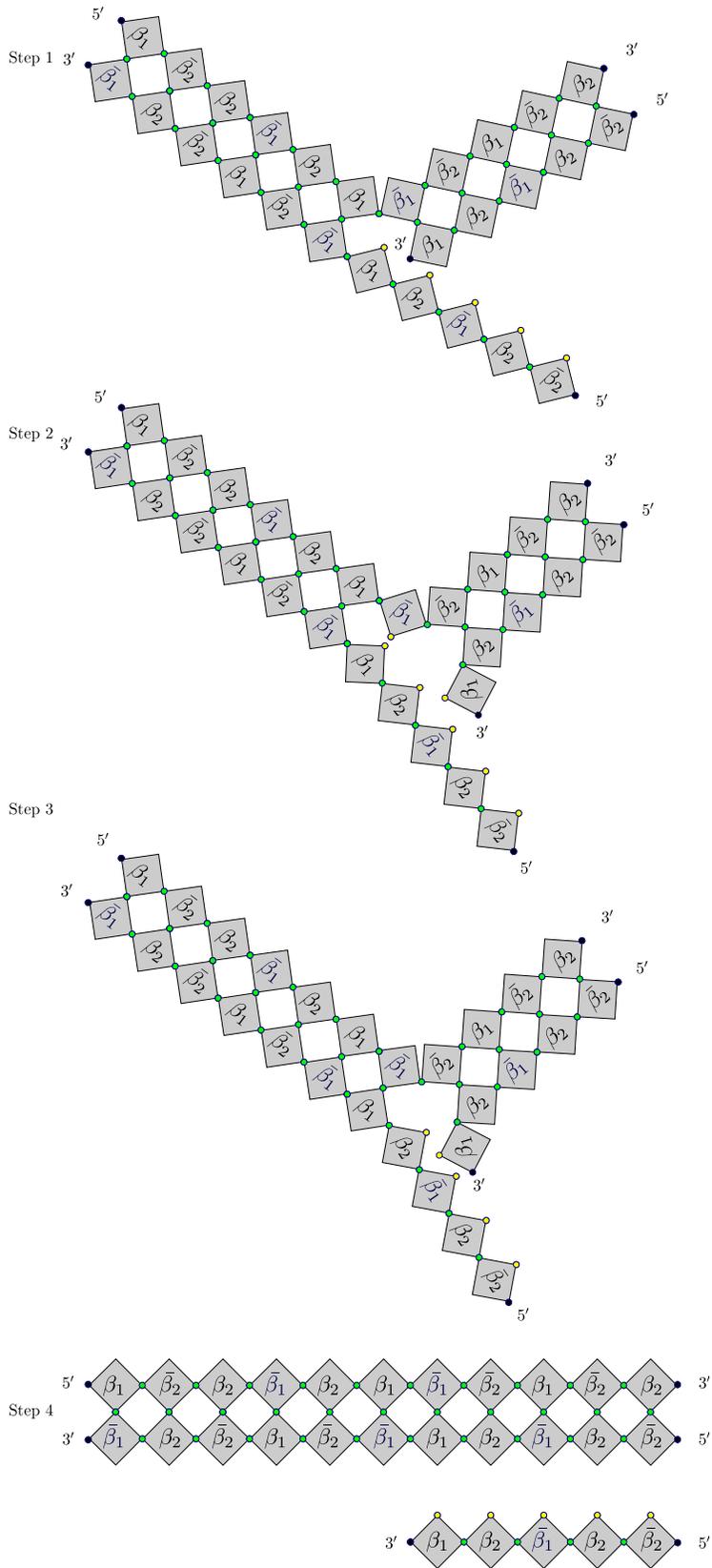


Figure 8: Meta-strand displacement due to breathing at $\tau = 1$

5.3 Meta-polymerization in mDNA

Meta-polymerization in mDNA occurs by the extension of a ssmDNA, called a primer, by additions of free meta-nucleotides in the solution to the 3' end via meta-backbone bonds. The extension cannot occur *de novo*, it requires a primer meta-hybridized to a template ssmDNA. Neither can meta-nucleotides spontaneously aggregate, because of the protection strands (see Figure 1). The meta-nucleotides added have base pads complementary to the corresponding base pads of the template strand.

Figure 9 gives an activatable tile model view of a single step in mDNA meta-polymerization. Each base pad of the template strand that is not bound to its complement is in the active unbound state. The 3' pad of the tile at the 3' end of the primer is in the capped state (step 1). It transitions to the active unbound state when its black capping strand is displaced by an initiator strand. The incoming meta-nucleotide (step 2) has its 5' and 3' pads in the inactive state, while its base pad is in the active unbound state. The complementary base pads bind (step 3), activating the 5' pad of the incoming nucleotide (step 4). The 5' pad then binds to the 3' pad of the previous meta-nucleotide (step 5) causing the 3' pad of the incoming nucleotide to transition to the active unbound state (step 6). This process occurs repeatedly till either no further free meta-nucleotides are available or the end of the template strand is reached. There is an alternate mechanism to stop the polymerization, which involves adding to the solution a black capper strand to transition the 3' pad of growing 3' end to a capped state, thus stopping further additions. Note that in a solution with an ensemble of these nanostructures, in general the stopping point of meta-polymerization cannot be carefully controlled and such an attempt would probably lead to various length subsequences of the fully complementary ssmDNA. However, we can exclude certain meta-base types from the meta-nucleotide mix in the solution ensuring that the polymerization halts when the complementary meta-base is encountered on the template strand. Since we have access to a larger alphabet of bases in our mDNA systems we can set aside a set of bases for such purposes, analogous to *stop codons* in translation.

We will now examine this protocol in greater detail by looking at the internal strand structure (Figures 10 and 11). The 3' pad of the tile at the 3' end of the primer is capped by $fgt_{3'}$ (step 1). The initiator strand $t_{3'}\bar{g}\bar{f}$ binds to the capper strand and activates the 3' pad $\bar{g}\bar{f}$ (step 2). Now, the protected meta-nucleotide comes in (step 3) and binds to the complementary base sequence $\bar{a}_2\bar{b}_2$, displacing the strand $\bar{b}_2x\bar{f}$ (step 4). The toehold \bar{b}_2 is exposed on $\bar{b}_2x\bar{f}$ allowing $f\bar{x}b_2$ to bind (step 5). This strand displaces $\bar{b}_2x\bar{f}$ exposing f on the 5' end of the incoming meta-nucleotide (step 6). Now, the 5' end of the incoming meta-nucleotide binds to the 3' end of the previous nucleotide, displacing $fy\bar{g}$ (step 7). The strand $fy\bar{g}$ is stripped away by its complement through the toehold \bar{g} , activating the 3' end ($\bar{g}\bar{f}$) of the incoming meta-nucleotide (step 8). This brings the 3' end of the growing ssmDNA back to the same state as in step 2 and thus the reaction can repeat till the end of the template strand is reached (step 9). At this point, we introduce the 3' capper sequence $fgt_{3'}$ which binds to the 3' pad of the last meta-nucleotide added. This terminates the polymerization reaction. Note that in a solution with an ensemble of these nanostructures, we must wait for each copy of the reaction to proceed to step 9 before introducing the capper sequence or we must program a stopper sequence, analogous to *stop codons* in translation. DNA polymerases like $\Phi 29$ possess exceptional strand displacing capability which aids in isothermal amplification of DNA. However, our meta-polymerization protocol lacks this ability. Thus, if in the course of extending the primer an already bound ssmDNA is encountered then the meta-polymerization process halts. In Section 5.6, we show how to achieve isothermal amplification of mDNA.

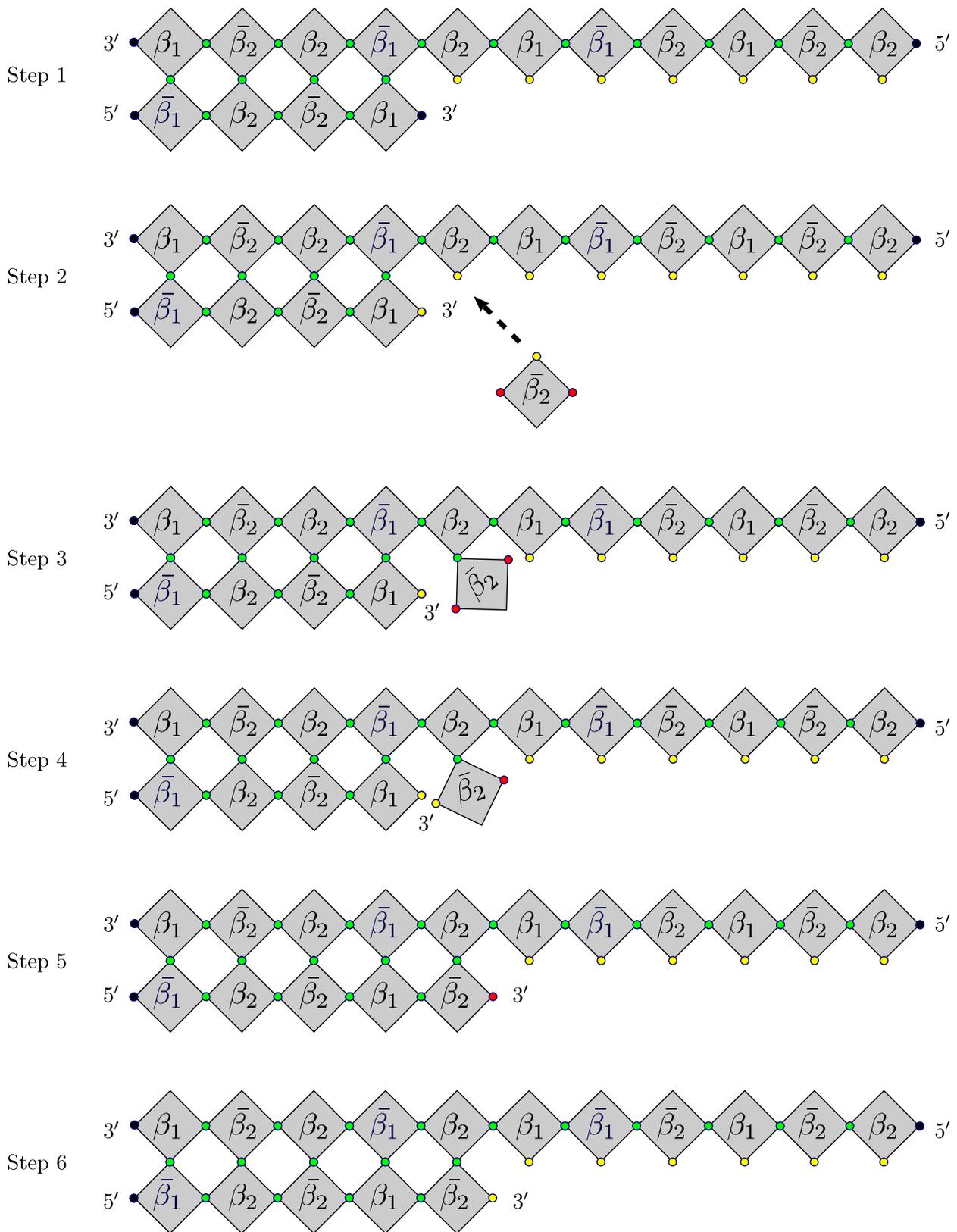


Figure 9: Meta-polymerization reaction

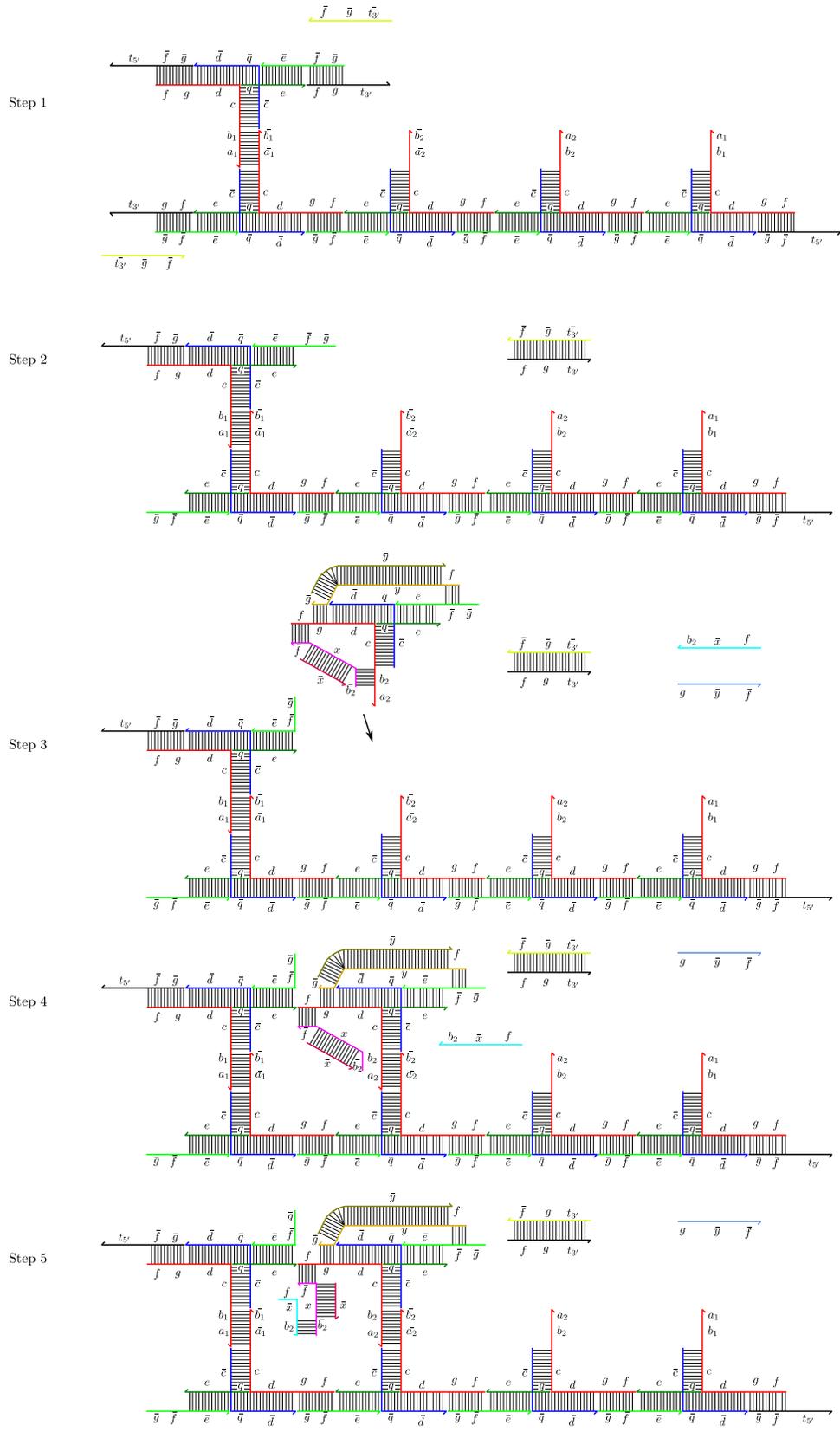


Figure 10: Internals of meta-polymerization reaction - I

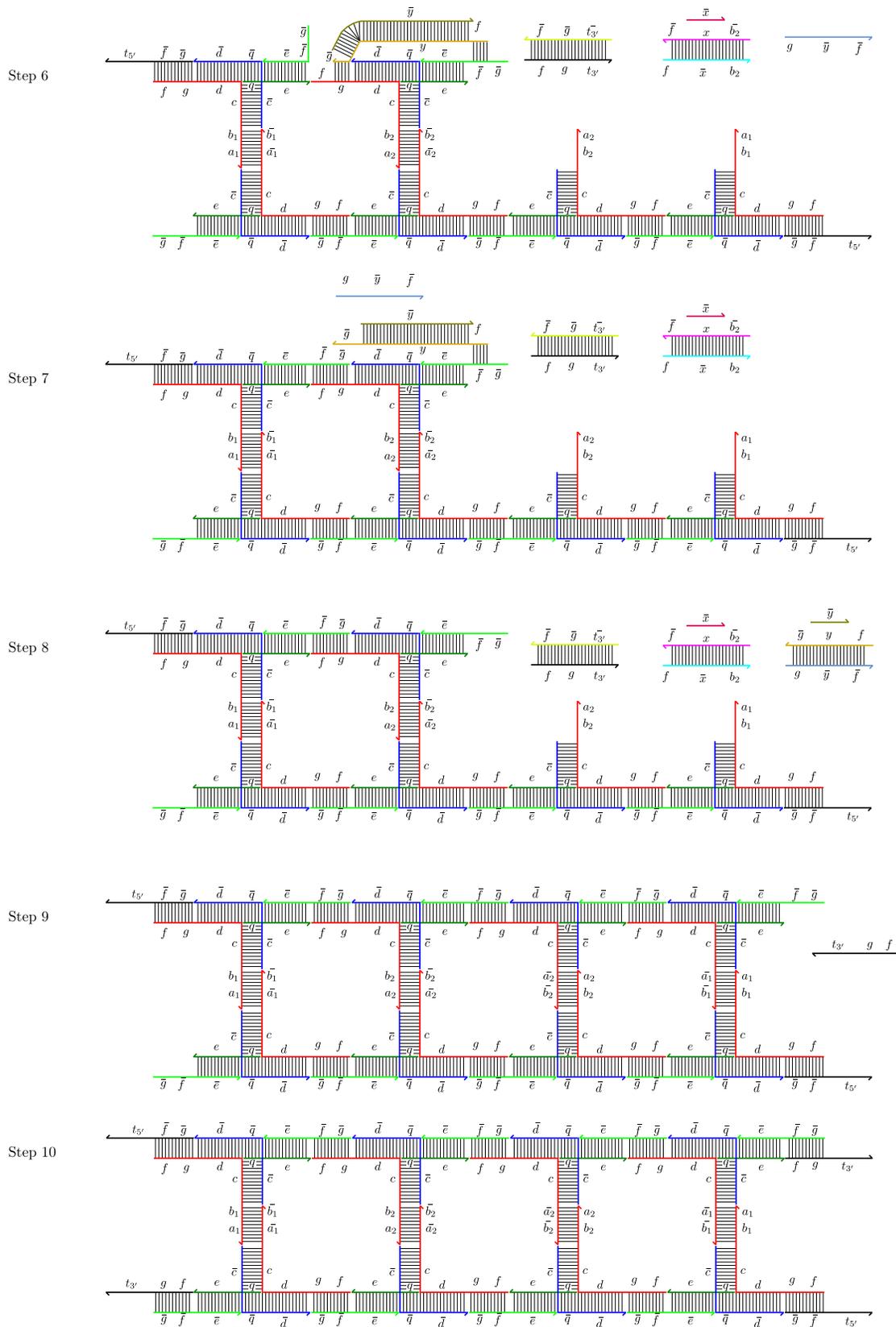


Figure 11: Internals of meta-polymerization reaction - II

5.4 Meta-restriction Cuts in mDNA

We can achieve site specific meta-restriction in both *ssmDNA* and *dsmDNA* by slightly modifying the internal structure of a nucleotide. This modification does not significantly affect the secondary structure of the mDNA and is compatible with all the other mDNA protocols described in this chapter. We prefix a sequence h_i to the 5' pad of the meta-nucleotide with base pad β_i . This is the sequence that will be recognized by cleaving strands that break the meta-backbone bonds. We illustrate restriction for *dsmDNA*. The protocol for *ssmDNA* is very similar and can be thought of as a special case of the meta-restriction of *dsmDNA*.

The abstract activatable tile model of meta-restriction is illustrated in Figure 12. The 5' end of the meta-nucleotide β_1 is recognized, cut and sent to the capped state. In Figure 13 the site h_2 associated with the meta-base b_2a_2 is recognized by the cleaver strand $\bar{g}\bar{f}\bar{h}_2r_2$. There is a single recognition site h_2 on each of the strands of the *dsmDNA* (step 1). The cleaver strands bind to the toehold h_2 on each strand (step 2) and break the meta-backbone bond by strand displacing the 3' pad $\bar{g}\bar{f}$ (step 3). We now introduce the 3' capper sequence $f\bar{g}t_{3'}$ to cap the exposed 3' pads (step 4). We prevent interaction between the cleaver strand $\bar{g}\bar{f}\bar{h}_2r_2$ and the 3' capper sequence $f\bar{g}t_{3'}$ by executing step 4 only after we are reasonably certain that step 3 is complete. This means that meta-restriction is not autonomous. The strand r_2h_2fg is introduced to strip away the cleaver strand $\bar{g}\bar{f}\bar{h}_2r_2$ from the 5' pad by exploiting the toehold r_2 (step 5). This exposes the 5' pads (step 6) which are then capped by introducing the 5' capper sequence $\bar{g}\bar{f}t_{5'}$ (step 7), completing the process of meta-restriction. Again, we prevent interaction between the strand r_2h_2fg and the 5' capper sequence $\bar{g}\bar{f}t_{5'}$ by executing step 7 only after we are reasonably certain that step 6 is complete. We do not require that these extraneous strands do not interact at all. Rather, it is sufficient that even after interacting among themselves there are a sufficient concentration of them to perform the tasks described in Figures 13 and 14.

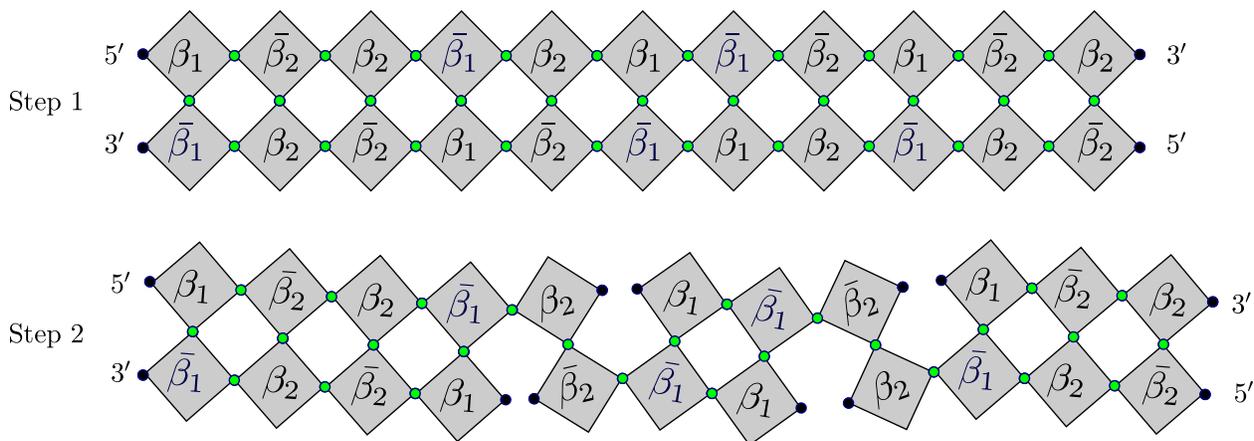


Figure 12: Abstract description of meta-restriction at site β_1

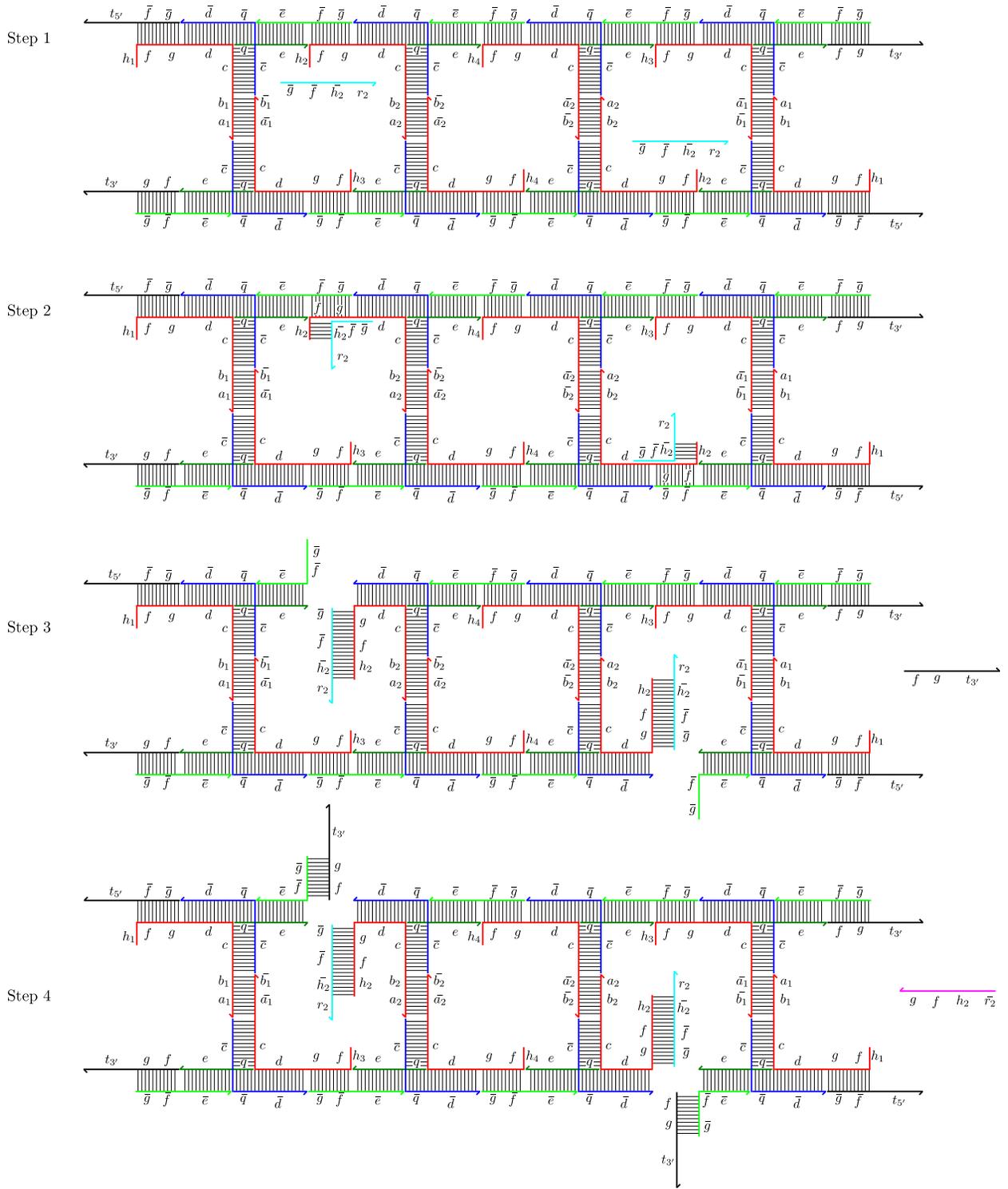


Figure 13: Internals of meta-restriction at h_2 associated with meta-base $b_2 a_2$ - I

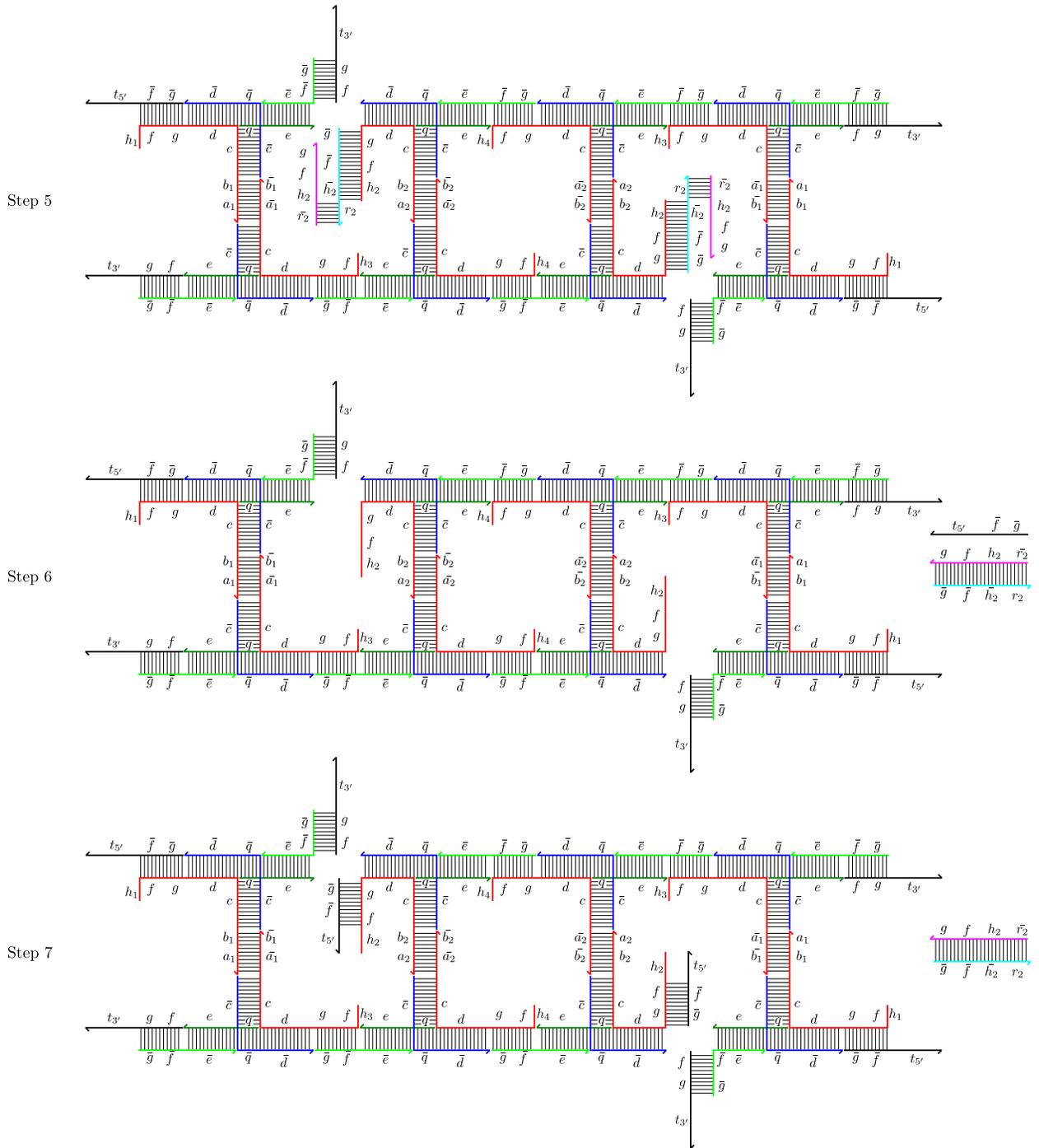


Figure 14: Internals of meta-restriction at h_2 associated with meta-base b_2a_2 - II

5.5 Meta-helicase Denaturation Reaction in mDNA

We have previously described meta-denaturation for mDNA using temperature. The same result can be achieved isothermally, which we call meta-helicase denaturation. Consider the abstract view of meta-helicase denaturation given in Figure 15. Denaturation is the breaking of meta-base bonds of a *dsmDNA*. When all the meta-base bonds are broken, the meta-strands float apart. Initially all the meta-base pads are in the bound state (step 1). Meta-helicase activity breaks the bonds and sends the base pads to the inactive state. This meta-helicase reaction doesn't necessarily act contiguously. Some meta-base bonds are broken before others (steps 2 and 3). When all the meta-base pads are broken the meta-strands float apart (step 4). We can reactivate the meta-base pads by transitioning them to the active unbound state (step 5) at which point the *ssmDNA* can recombine to form a *dsmDNA*.

Meta-helicase activity in mDNA is performed by a host of strands (colored pink in Figure 16), two for each type of meta-base bond that must be broken. In Figure 16 there are two types of meta-base bonds and hence we have four strands to perform helicase activity (step 1). We introduce a slight modification in the internal strand structure by appending a sequence p to the base pad of each meta-nucleotide. This modification does not significantly affect the secondary structure of mDNA and neither does it interfere with any other mDNA protocol described in this chapter. This sequence (p) will act as a toehold for the strands involved in the meta-helicase process. A pair of strands invade the meta-base pad bond and break them (step 2). Note that half of the meta-base bond is broken by one of these strands and half by the other, ensuring symmetry. Once each meta-base bond is broken the meta-strands drift apart (step 3) and meta-helicase activity is complete. We can strip off the protecting strand, for example $\bar{m}\bar{p}b_1$, by adding in its complement, say b_1pm (step 4). The *ssmDNA* can now recombine into *dsmDNA*. It is possible to repeat this process of meta-denaturation and meta-renaturation, however the protocol would not be autonomous.

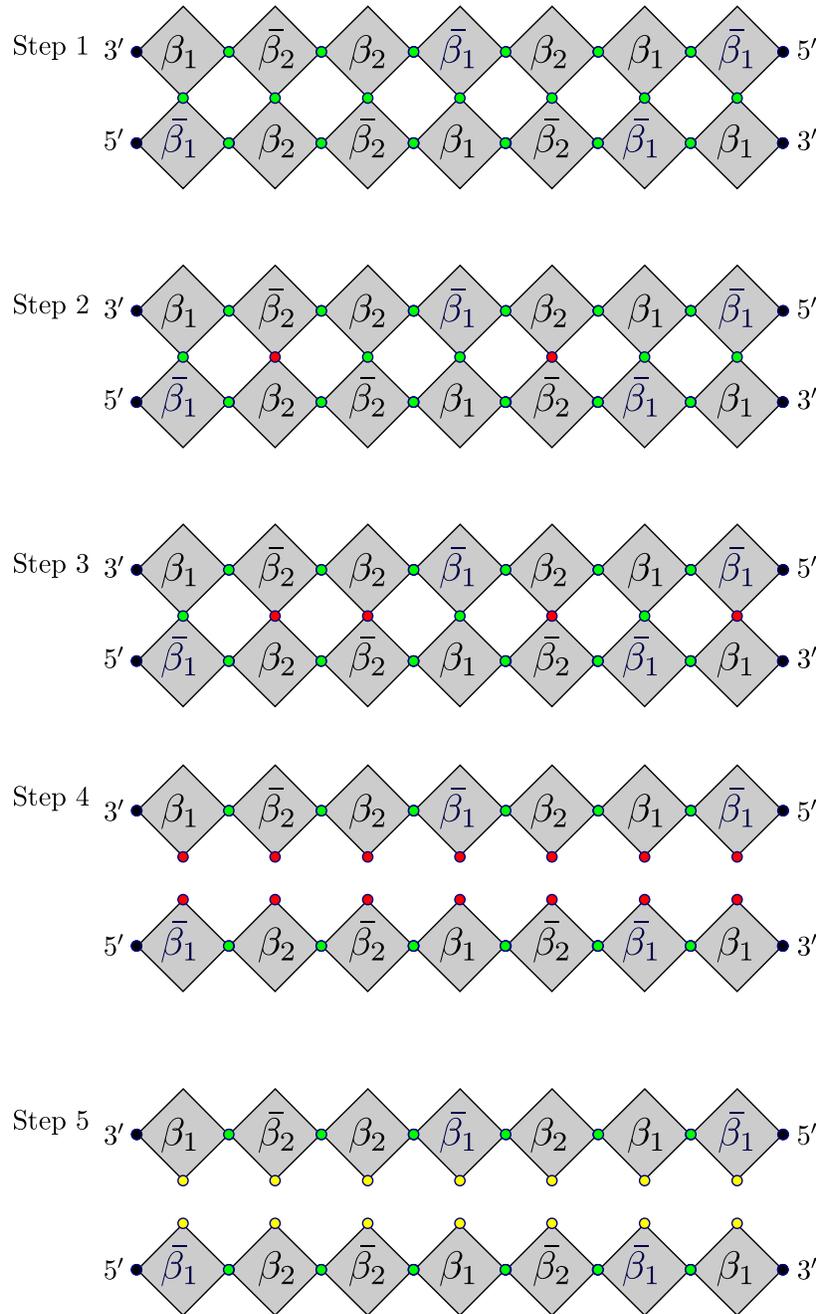


Figure 15: Abstract description of meta-denaturation by helicase action

5.6 Exponential Amplification in mDNA Using a Replicator

Our protocol for meta-polymerization in mDNA lacked meta-strand displacement capabilities and hence could not be used for isothermal PCR-like amplification. In this section we describe a method to get isothermal exponential amplification using a replicator mechanism. The protocol is similar to meta-polymerization, it involves linear contiguous extension of a primer by addition of meta-nucleotides, with the newly polymerized ssmDNA having the complementary sequence to that of the template. The key difference in the two protocols is a mechanism to isothermally dissociate the newly synthesized ssmDNA from the template. We achieve this using a new idea, not present in any previous protocol described in this work, of active DNA sequences sequestered in hairpins that are released by strand displacement reactions. The release of the “hidden” DNA sequence inside the hairpin structure can be thought of as an activation step, setting off another strand displacement reaction. This idea of sequestering sequences within hairpins has been demonstrated previously by Zhang et al. (2007) and Dirks and Pierce (2004).

For the purposes of this protocol, we define a new state of the base pad in the activatable model of mDNA, which we call *semi active unbound* (shaded purple in Figure 17(c)). If two base pads are in the semi active unbound state, they cannot bind to each other. However, a semi active unbound base pad can bind with an active unbound base pad, provided their sequences are complementary. There are also modifications (Figures 17 (a),(b)) to how the other pad states are implemented, however the properties of the state do not change. Note the introduction of the hairpin structure $g_2\bar{p}\bar{a}l\bar{g}_2$ at the 3' end of a meta-nucleotide (Figure 17 (a)) which will be used to cleave the growing strand from the template (Figure 17 (c)), sending the base pad to the semi active unbound state.

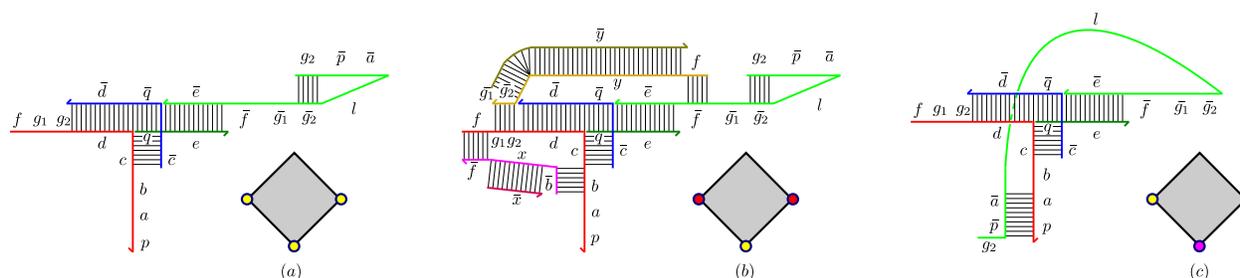


Figure 17: Replicator: new states

Figure 18 illustrates the replicator in the abstract activatable tile model of mDNA. The top strand is the template and the bottom strand grows from a primer. For initiating and terminating replication, we require the use of a pair of special tiles with complementary base pads, labeled $\beta_{3'}$ and $\bar{\beta}_{3'}$ in Figure 18. These are capped at one end and occur at the terminal ends of the template and hence in each replicated mDNA. $\bar{\beta}_{3'}$ acts as the primer, initiating the replication. Consider the replication process after a few meta-nucleotides have been added (step 1). The growing strand is attached via a single base pad (at the 3' terminal meta-nucleotide) to the template. The base pads of the rest of the meta-nucleotides are in the semi active unbound state and hence cannot bind with each other. A new meta-nucleotide comes in (step 1) with its base pad in the active unbound state and binds to the template (step 2). This activates its 5' end (step 3) and allows the 3' end of the growing strand to bind (step 4). This activates the 3' end of the meta-nucleotide added and also cleaves the meta-base pad binding between the template and the previously attached meta-nucleotide of the growing strand, sending the cleaved meta-base pads to the semi active unbound state (step 5). The process (step 1 to step 5) repeats till the last meta-nucleotide $\beta_{3'}$ attaches. At this point, we want the two strands to separate. We achieve this by making the meta-base pad bond between $\beta_{3'}$ and $\bar{\beta}_{3'}$ relatively weak, allowing the strands to separate spontaneously due to breathing of the base pad bond. This is implemented by choosing a very short length sequence

for the base pads of $\beta_{3'}$ and $\bar{\beta}_{3'}$. Both the template and the newly synthesized ssmDNA can now act as templates for further replication and hence we can achieve exponential amplification. Note that we can easily introduce mutations in the replicator mechanism by designing DNA sequences for certain base pads that do not have any exact complementary base pads but rather several partial complements. This allows us to probabilistically evolve a diverse sequence population of ssmDNA.

Let us examine the replicator protocol in greater detail, paying attention to the strand diagram (Figures 19 and 20). We have a four meta-base mDNA sequence, with the first and last bases being the special terminator bases (step 1). Note that the sequence $g = g_1g_2$ and $\bar{g} = \bar{g}_2\bar{g}_1$. A meta-nucleotide with an active unbound base pad (step 1) comes in and binds to the template via the sequence b_2a_2 (step 2), strand displacing out the sequences $\bar{p}b_2$ (part of the light green strand) and \bar{b}_2 (part of the pink strand). The strand $f\bar{x}b_2$ now strips away $\bar{b}_2x\bar{f}$, activating the 5' pad of the incoming meta-nucleotide (step 3). This allows the 5' pad fg_1g_2 to bind with the 3' pad $\bar{g}_2\bar{g}_1\bar{f}$, displacing $\bar{g}_2\bar{g}_1$ and opening up the hairpin structure $\bar{p}\bar{a}_1l$ (step 4). This allows the strand $fy\bar{g}_2\bar{g}_1$ to be stripped away through the toehold $\bar{g}_2\bar{g}_1$, activating the 3' end (step 5). The released hairpin from step 4 can now cleave a_1 from \bar{a}_1 using the toehold p , which is half of the base pad binding, while the other half \bar{b}_1 of the base pad binding is cleaved from b_1 via the toehold p by the sequence $\bar{p}b_1$ (step 6). Note that cleaving the bond between b_1 and \bar{b}_1 is actually a reversible process, and could have occurred after step 1 itself. Only when it is combined with the cleaving of the bond between a_1 and \bar{a}_1 does it get biased towards the configuration depicted in step 6. This completes the addition of a single meta-nucleotide accompanied by cleaving the previous meta-nucleotide. The process repeats till the last meta-nucleotide is added at which point the short base sequence of $a_3\bar{b}_3$ allows the ssmDNA to separate due to breathing. We note that both the separate ssmDNA are in a configuration that allows them to act as template strands for further replication, allowing exponential amplification of mDNA.

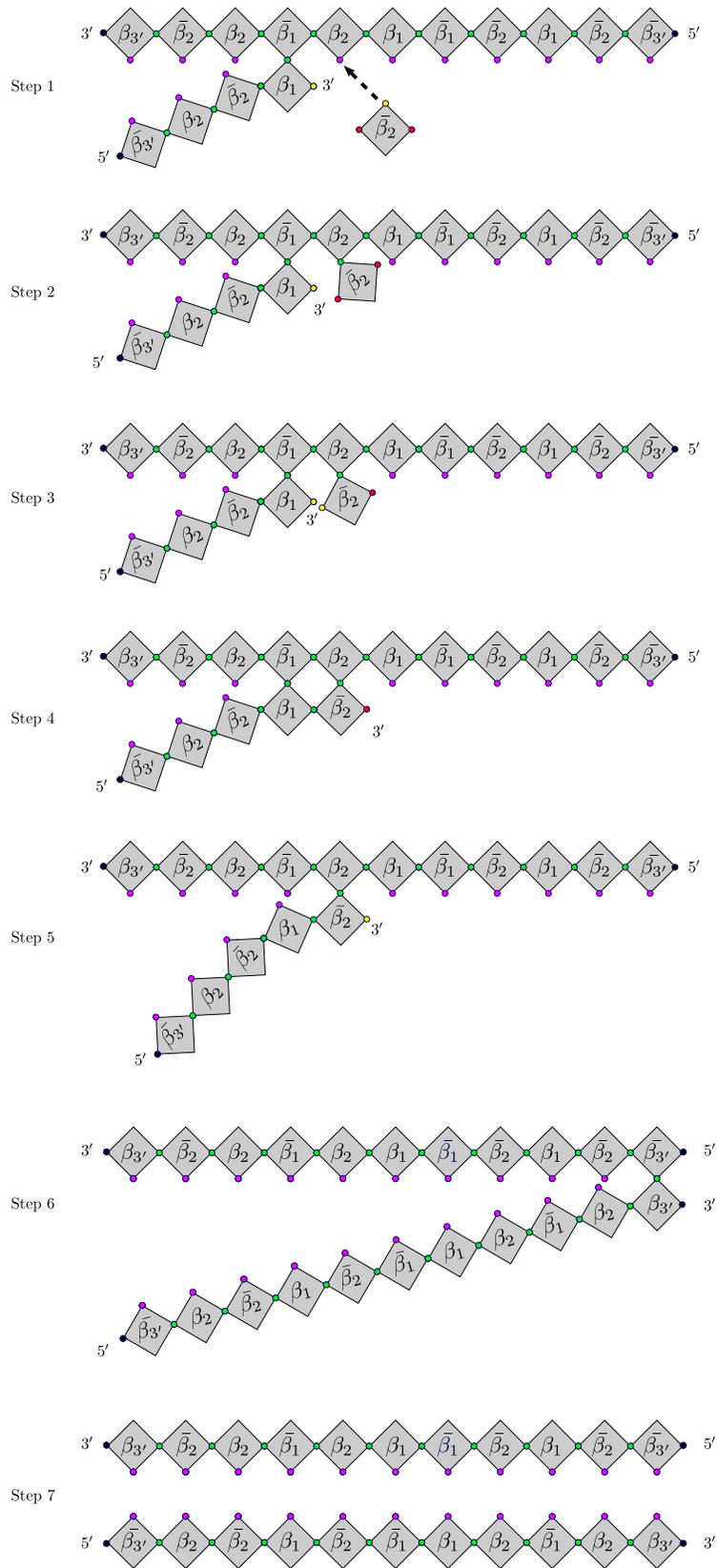


Figure 18: Replicator abstraction

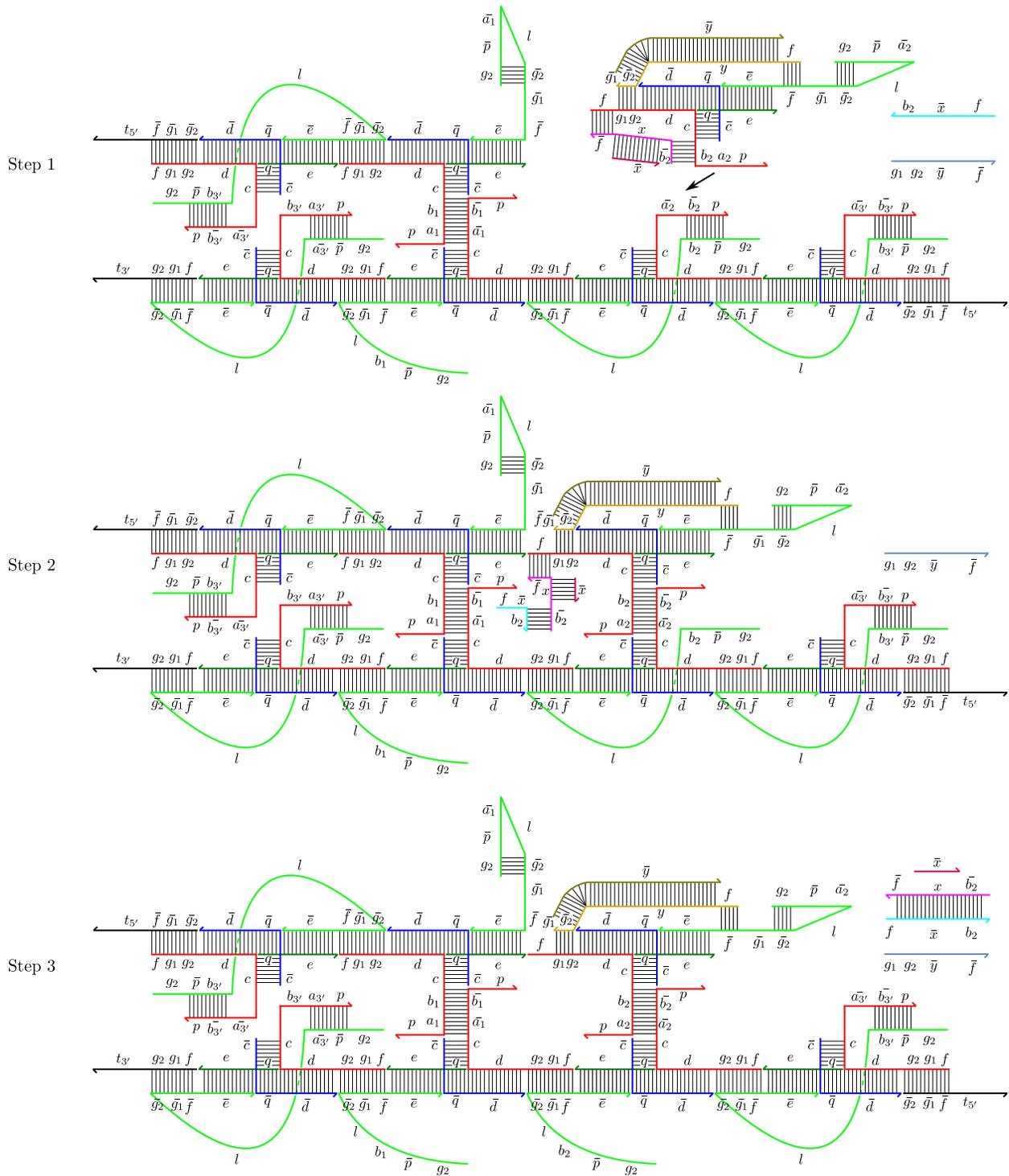


Figure 19: Internals of exponential amplification by replication - I

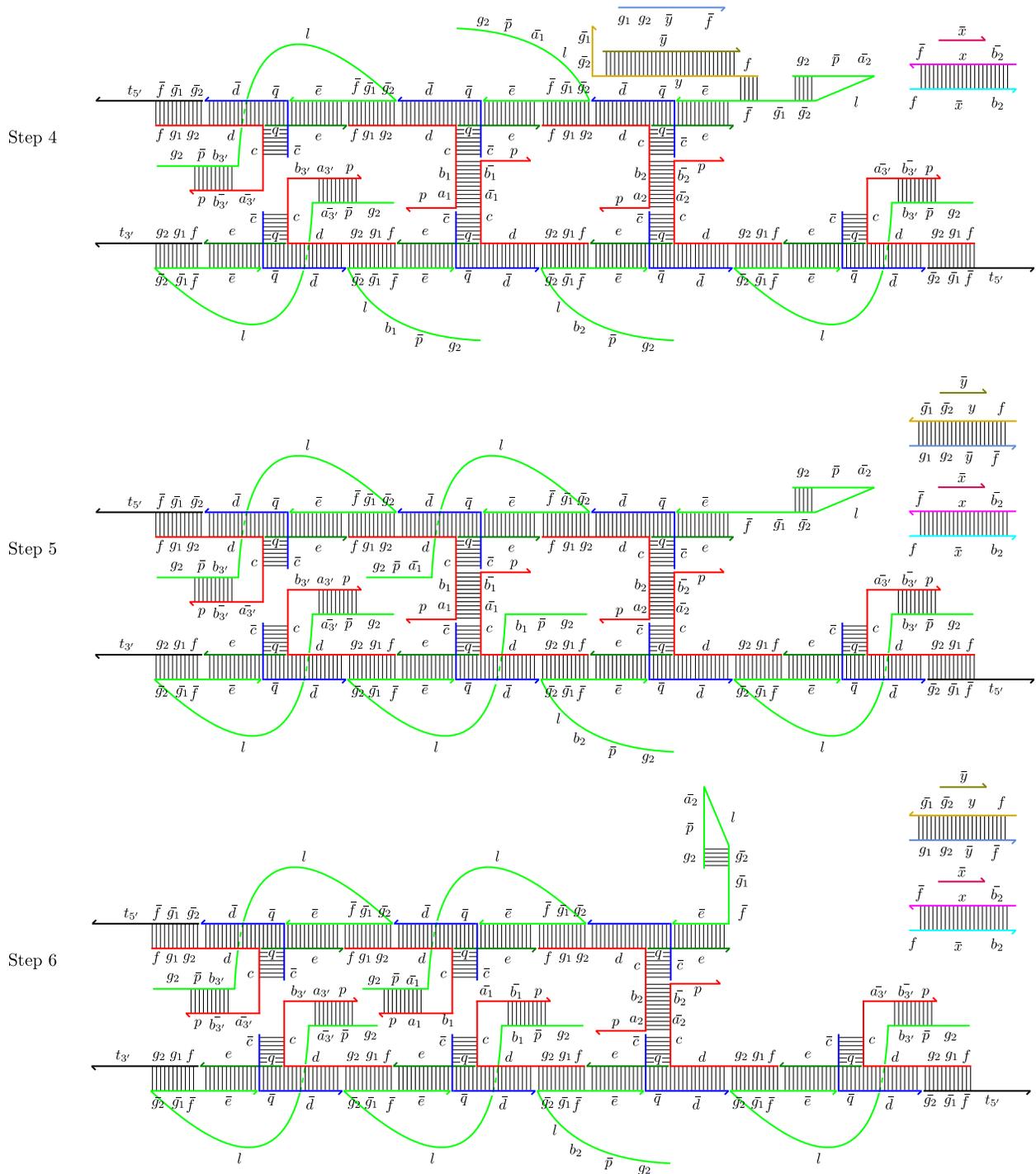


Figure 20: Internals of exponential amplification by replication - II

6 Lessons Learnt

In this work we have outlined a synthetic biochemical system made purely from DNA strands that simulates the behavior of various protein enzymes acting on DNA. We believe mDNA systems are a powerful programming paradigm for designing complex biochemical systems. In Section 5.6 we provided a protocol for exponential self-replication of a meta-DNA sequence and pointed out how to introduce mutations during replication and thus achieving sequence diversity. Incorporating environmental selection pressure in mDNA to evolve functional biosystems is a major open challenge. The next step in our investigations of mDNA systems is development of a strand level kinetic simulation of mDNA protocols. We hope to use data from such simulations to enhance and fine tune our protocols and also perform *in vitro* experiments to validate them. A major facilitating factor in the field of DNA nanotechnology has been technology to efficiently synthesize synthetic DNA strands of defined sequence *de novo* from individual nucleotides. Analogously, we require methods to efficiently (high throughput with low error rates) synthesize mDNA strands of defined sequence from synthetic DNA. The current protection-deprotection DNA synthesis protocols can be adapted for the use of synthesizing mDNA from pre-formed meta-nucleotides. Note that meta-nucleotides can be synthesized using standard hierarchical assembly techniques from DNA self-assembly developed by Park et al. (2006). Other important protocols for mDNA systems are meta-polymerization with meta-strand displacing capability and meta-ligation of mDNA strands.

7 Acknowledgments

We wish to acknowledge support by NSF grants CCF-0829797, CCF-0829798, CCF-1217457 and CCF-1141847.

References

- Amin, R., Kim, S., Park, S. H., and LaBean, T. (2009). Artificially Designed DNA Nanostructures. *NANO: Brief Reports and Reviews*, 4(3):119–139.
- Bath, J. and Turberfield, A. (2007). DNA Nanomachines. *Nature Nanotechnology*, 2:275–284.
- Bedau, M., McCaskill, J., Packard, N., Rasmussen, S., Adami, C., Green, D., Ikegami, T., Kaneko, K., and Ray, T. (2000). Open Problems in Artificial Life. *Artificial Life*, 6(4):363–376.
- Dadon, Z., Wagner, N., and Ashkenasy, G. (2008). The Road to Non-Enzymatic Molecular Networks. *Angewandte Chemie International Edition*, 47(33):6128–6136.
- Deng, Z., Chen, Y., Tian, Y., and Mao, C. (2006). A Fresh Look at DNA Nanotechnology. *Nanotechnology: Science and Computation*, pages 23–34.
- Dirks, R. and Pierce, N. (2004). Triggered Amplification by Hybridization Chain Reaction. *Proceedings of the National Academy of Sciences of the United States of America*, 101(43):15275–15278.
- Hamada, S. and Murata, S. (2009). Substrate-Assisted Assembly of Interconnected Single-Duplex DNA Nanostructures. *Angewandte Chemie International Edition*, 48(37):6820–6823.
- Koval, V., Gnedenko, O., Ivanov, Y., Fedorova, O., Archakov, A., and Knorre, D. (1999). Real-Time Oligonucleotide Hybridization Kinetics Monitored by Resonant Mirror Technique. *IUBMB Life*, 48(3):317–320.

- LaBean, T., Gothelf, K., and Reif, J. (2007). Self-Assembling DNA Nanostructures for Patterned Molecular Assembly. *Nanobiotechnology II*, pages 79–97.
- Luisi, P. L. (2006). *The Emergence of Life - From Chemical Origins to Synthetic Biology*. Cambridge University Press.
- Lund, K., Williams, B., Ke, Y., Liu, Y., and Yan, H. (2006). DNA Nanotechnology: A Rapidly Evolving Field. *Current Nanoscience*, 2:113–122.
- Majumder, U., LaBean, T., and Reif, J. (2007). Activatable Tiles: Compact, Robust Programmable Assembly and Other Applications. *DNA Computing*.
- Packard, N. and Bedau, M. (2003). Artificial Life. *Encyclopedia of Cognitive Science*, 1:209–215.
- Park, S. H., Pistol, C., Ahn, S. J., Reif, J., Lebeck, A., and LaBean, C. D. T. (2006). Finite-Size, Fully Addressable DNA Tile Lattices Formed by Hierarchical Assembly Procedures. *Angewandte Chemie International Edition*, 45(5):735–739.
- Reif, J. and Majumder, U. (2008). Isothermal Reactivating Whiplash PCR for Locally Programmable Molecular Computation. *DNA Computing*, pages 41–56.
- Sakamoto, K., Kiga, D., Momiya, K., Gouzu, H., Yokoyama, S., Ikeda, S., Sugiyama, H., and Hagiya, M. (1999). State Transitions by Molecules. *Biosystems*, pages 81–91.
- Schulman, R. and Winfree, E. (2008). How Crystals that Sense and Respond to their Environments could Evolve. *Natural Computing*, 7(2):219–237.
- Seeman, N. (2004). Nanotechnology and the Double Helix. *Scientific American*, 290(6):64–75.
- Sherman, W. and Seeman, N. (2004). A Precisely Controlled DNA Biped Walking Device. *Nano Letters*, 4:1203–1207.
- Smith, A., Turney, P., and Ewaschuk, R. (2002). JohnnyVon: Self-Replicating Automata in Continuous Two-Dimensional Space. *Computing Research Repository*.
- Tian, Y., He, Y., and Mao, C. (2006). Cascade Signal Amplification for DNA Detection. *ChemBioChem*, 7(12):1882–1864.
- Tjivikua, T., Ballester, P., and Rebek, J. (1990). A Self-Replicating System. *Journal of the American Chemical Society*, 112(2):1249–1250.
- Vidonne, A. and Philp, D. (2009). Making Molecules Make Themselves - the Chemistry of Artificial Replicators. *European Journal of Organic Chemistry*, 5:593–610.
- von Kiedrowski, G. (1986). A Self-Replicating Hexadeoxynucleotide. *Angewandte Chemie International Edition*, 25(10):932–935.
- von Kiedrowski, G. and Patzke, V. (2007). Self Replicating Systems. *ARKIVOK*, pages 293–310.
- Winfree, E. (2003). DNA Computing by Self-Assembly. *NAE's The Bridge*, 33:31–38.
- Yin, P., Yan, H., Daniell, X., Turberfield, A., and Reif, J. (2004). A Unidirectional DNA Walker Moving Autonomously Along a Linear Track. *Angewandte Chemie International Edition*, 116(37):5014–5019.
- Zhang, D., Turberfield, A., Yurke, B., and Winfree, E. (2007). Engineering Entropy-Driven Reactions and Networks Catalyzed by DNA. *Science*, 318:1121–1125.

Zhang, D. Y. and Winfree, E. (2009). Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *Journal of the American Chemical Society*, 131(48):17303–17314.

Zhang, D. Y. and Yurke, B. (2006). A DNA Superstructure-based Replicator without Product Inhibition. *Natural Computing*, 5(2):183–202.