Design and analysis of linear cascade DNA hybridization chain reactions using DNA hairpins

Hieu Bui¹, Sudhanshu Garg², Vincent Miao¹, Tianqi Song¹, Reem Mokhtar² and John Reif¹,³

¹ Department of Computer Science, Duke University Durham, NC 27708, USA
² Department of Biomedical Engineering, Duke University Durham, NC 27708, USA
³ Author to whom any correspondence should be addressed.

E-mail: reif@cs.duke.edu

Keywords: DNA self-assembly, triggered self-assembly, DNA hairpins, metastable DNA, linear cascade reactions, strand displacement, DNA nanotechnology

Abstract
DNA self-assembly has been employed non-conventionally to construct nanoscale structures and dynamic nanoscale machines. The technique of hybridization chain reactions by triggered self-assembly has been shown to form various interesting nanoscale structures ranging from simple linear DNA oligomers to dendritic DNA structures. Inspired by earlier triggered self-assembly works, we present a system for controlled self-assembly of linear cascade DNA hybridization chain reactions using nine distinct DNA hairpins. NUPACK is employed to assist in designing DNA sequences and Matlab has been used to simulate DNA hairpin interactions. Gel electrophoresis and ensemble fluorescence reaction kinetics data indicate strong evidence of linear cascade DNA hybridization chain reactions. The half-time completion of the proposed linear cascade reactions indicates a linear dependency on the number of hairpins.

Introduction
Deoxyribonucleic acids (DNAs) are one of the major building blocks of biological systems. Recently, researchers have been exploiting DNA self-assembly for constructing nanostructures as well as programming molecular devices [1–13]. Yurke et al first observed the phenomenon of DNA strand displacement in their nanomachine [14]. Based on DNA strand displacement thermodynamics, the construction of molecular systems [15], logic circuits [16, 17], sensors [18], amplifiers [19–21], molecular machines [22–25], walkers [21, 22, 26–53], robots [27, 37, 38], chemical controllers [54], and neural networks [55] were realized. The use of DNA hybridization as the primary driving energy source has been the central theme in many DNA systems. An alternative energy source could also come from metastable DNA [56]. For instance, Turberfield et al used unhybridized DNA hairpin loops as the fuel source to operate DNA motors [57]. In addition, energy from DNA hairpins has been used to assist with catalytic hybridization [38] and triggered self-assembly [21, 59]. Inspired by early work on triggered self-assembly of DNA nanostructures [21, 59–61], we present a system for the controlled self-assembly of linear cascade DNA hybridization chain reactions (LCR) using DNA hairpins. Previous studies have demonstrated the formation of linear DNA oligomers from open hairpins hybridized in a staggered two-repeat monomer pattern [59, 61], with or without the assistance of a few auxiliary strands. Yin et al extended the triggered self-assembly from linear growth oligomers to quadratic growth branched oligomers and exponential growth dendritic oligomers [21]. Our system lies in the regime of linear growth oligomers with the following advantages: (i) our system has 9 distinct hairpins which can potentially form a longer chain, whereas prior works [59–62] used only two to four hairpins, (ii) our system allows fluorescence detection at any given oligomer length, (iii) our system enables quantification of chain reaction success rate for any given length, which can be tailored to design larger systems as and when required, (iv) the original system by Dirks and Pierce [59] had to be carefully tailored to have a stem of 18 nt and a toehold/hairpin loop region of 6 nt. Our system however,
introduces clamp domains, removing this constraint, thus allowing us to design systems with variable stem (≥12 nt) and toehold lengths, which considerably increases the sequence space and applicability of these systems and (v) because our system has 9 distinct hairpins (rather than the 2 hairpins of Dirks’ system [59]) it can be extended to do localized reactions on a surface [63, 64]. Our experimental results indicate strong evidence of linear cascade reaction responses based on gel electrophoretic analysis and reaction kinetics.

The assembly linear cascade DNA hybridization chain reaction is shown in figure 1. Figure 1(A) shows the components: nine hairpins (H₁, H₂, H₃, H₄, H₅, H₆, H₇, H₈, H₉), and an initiator (I). Each hairpin consists of a stem, a loop, and a sticky end. The stem consists of two domains (Ci and R) which form a stable duplex with 14 base pairs. The loop consists of two clamp domains (Ci−1 and Ci+1), a spacer domain (L), and a sequestered domain (Si−1). The sticky end consists of an external toehold domain (Si) which is complementary to the loop domain (Si−1) of hairpin (Hi−1). Hairpins are distinguishable by two single-stranded toehold domains (Si and Si−1)—one toehold is external (a single-stranded sticky end, readily available for hybridization) and one is sequestered within the hairpin loop. The stem (R) and spacer (L) domains of all hairpins comprise the same sequences. The spacer domain (L) is used to offset potential geometrical constraints. The clamp domain (C) is implemented to minimize the breathing effect at the ends of the DNA duplex and to ensure that the cascade reaction proceeds in the forward direction [65].

Initially, all hairpin components are mixed together but do not hybridize. The cascade reaction is initiated by the initiator (I) which interacts with H₁ by hybridizing to the external toehold (Si) as illustrated in figure 1(B). The initiator displaces the stem of H₁ by branch migration, opening the loop and revealing its sequestered toehold domain (Sᵢ). The opened loop of H₁ can now bind to the external toehold of H₂ and displace the stem of H₂ by branch migration, opening in turn the loop of H₂. This reaction cascades downstream as each additional hairpin anneals to the growing structure (which displaces the stem and opens the loop, making the previously sequestered toehold accessible for hybridization to initiate the next stage of chain growth). The designed product of the assembly is a linear chain formed by the staggered hybridization of nine hairpins. A reporter complex was used to determine the rate of the chemical reaction. Hairpin H₉ triggered the opening of the reporter complex, as shown in figure 5(A).

Materials and methods

Sequence design

NUPACK [66] is open-source software with built-in functions for the design and optimization of DNA sequences based on thermodynamic and kinetic criteria. To ensure minimal crosstalk among nine hairpins,
multiple design techniques were implemented in NUPACK. (i) A 3-letter code was used on the stem region of the hairpin. One of the arms of the stem is ‘opened’ up when an incoming strand displaces it. Note that the other arm always stays double stranded pre- and post- (and during) branch migration process. The arm that stays double stranded was constrained to contain only A, G, and T. This ensures that the arm containing the ‘G’ is always in double stranded form. This was so designed because guanine is more promiscuous (binds to both cytosine and thymine), and thus can create undesirable secondary structures when in single stranded state. (ii) Clamp domains were introduced to minimize the breathing effect at both ends of the duplex. (iii) Negative design was used to minimize off-target binding between hairpins during intermediate states and constrained NUPACK to return sequences that maximize the probability of a sub-strand (the region of the hairpin that opens the next hairpin) staying single-stranded in intermediate states. The lengths of toehold, stem, and loop domains were optimized using prior reported systems [21, 59, 65]. NUPACK provides the yield of the optimized sequences and defects at equilibrium and in addition, the software avoids improper hybridization of hairpins using Dirks and Pierce’s algorithm [67]. The yield of defect is updated after each optimization procedure (changing the sequence composition of stem, loop, toehold, and clamp) and the optimal set of DNA sequences is selected based on the least amount of defect in the yield output. A detailed description of the sequence design is in SI S1.

Bimolecular reaction model of LCR

Each DNA sequence is treated as an individual molecule. In the absence of the target initiator molecule, the LCR system remains inactive. Only when the target initiator molecule is present is the LCR activated and the cascade chain reaction initiated, resulting in a linear duplex chain formation. The rate of the linear cascade chain reaction depends on the toehold-binding and branch migration processes. To simulate the formation of various linear duplex lengths, previously reported rate constants were used [19, 68]. For instance, the rate constant for displacement of the reporter complex was $1.3 \times 10^6 \text{M}^{-1} \text{s}^{-1}$. The rate constant used for toehold-mediated strand displacement was $10^5 \text{M}^{-1} \text{s}^{-1}$. Because the toehold binding and branch migration domains between a pair of hairpins have the same length, the strand displacement rate constant is assumed to be equal across all 9 hairpins. Other factors such as the G–C bond having a higher binding energy than the A–T bond, defects in strand synthesis, and non-functional hairpins were not taken into account during simulations. In addition, all hairpin–hairpin interactions were assumed to proceed forward with insignificant reverse reactions. Simulation results of different linear cascade reactions with 9, 6, 4, 2, and 1 hairpin(s) are shown figure 2. As expected, the completion rate of the linear cascade reaction depends on the number of hairpins—less hairpins correspond to a faster completion rate. A description of simulation scripts is in the SI.

All DNA strands were purchased lyophilized from Integrated DNA Technologies and purified using PAGE. In brief, a denaturing PAGE gel was run for 90 min at 300 V constant voltage in 1× TBE buffer. Purified DNA bands were identified, excised, crushed and eluted in elution buffer (NaCl) overnight. The solution containing purified DNA oligonucleotide was centrifuged and the liquid was extracted without disturbing the crushed gel pieces. The solution was washed with butanol. DNA oligonucleotides were precipitated in pure ethanol and

Figure 2. Simulation results of various linear cascade reaction lengths. The length depends on the number of hairpins involved in each reaction. Linear cascade reactions when the rate constant is $10^7 \text{M}^{-1} \text{s}^{-1}$ (A) versus $1.3 \times 10^6 \text{M}^{-1} \text{s}^{-1}$ (B). Simulation is performed based on the assumption that all hairpin–hairpin pairs follow the second order (bimolecular) chemical reaction. Simulation conditions are specified as the following: 5 nM concentration of hairpins, 6.5 nM concentration of reporter complex, 10 nM concentration of initiator, and the rate constant of displacing the reporter complex is common across all simulation experiments.
washed with 70%/30% ethanol/water. Excess ethanol was removed and the remaining ethanol was removed by vacuum centrifugation. Purified DNA was then re-suspended in 1× TE buffer pH 8.0. A list of all DNA strands is shown in table 1. Native PAGE analysis was used to identify the effects of the linear cascade DNA hybridization reactions. In brief, each native PAGE was run for 3.5 h at a constant 170 V. Each well was loaded with 5–10 picomoles of DNA sample. Ethidium bromide was used to stain DNA for visualizing DNA bands. Ensemble fluorescence spectroscopy was used to monitor the linear cascade chain reaction kinetics of DNA hybridization in solution. A reporter complex consisting of a fluorophore (Tetrachlorofluorescein TET) and a quencher molecule (Iowablack FQ), was adopted from Zhang et al. [19]. Only the last hairpin, when activated, binds to the reporter complex to trigger the fluorescence emission. An equimolar solution of hairpins was prepared. The fluorophore was excited at 524 nm and emission was measured at 541 nm. The excitation and emission slits were set at 10 nm and 5 nm, respectively. For each experiment 80 μl of the sample (5 nM of DNA hairpins, 6.5 nM of reporter complex with and without initiator) was combined in quartz cuvettes. In order to prevent DNA loss in the pipetting process (by adhesion to the pipette tips), a non-reactive 20 nt poly-T strand was added to all fluorescence experiments at a concentration of 1 μM [19]. All measurements were performed at 20 °C. Rapid mixing was achieved by carefully but quickly pipetting the whole volume up and down for 10 s without generation of air bubbles or loss of material. All experiments were performed in TAE/12.5 mM Mg²⁺ buffer solution.

**Experimental results and discussion**

Analysis of the cascade reaction by polyacrylamide gel electrophoresis (PAGE) is shown in figure 3. In figure 3(A), lane 1 contains a 20 bp DNA ladder. Lane 2 contains all hairpins but no initiator: products of high molecular weight are almost completely absent, indicating a very low background rate of LCR. Lane 3 contains all hairpins but no initiator (the sample was incubated at room temperature overnight): some evidence of assembly is visible, indicating leak of LCR (leak is defined as undesired reaction in the absence of the initiator strand). Lanes 4–8 contain all hairpins with different concentrations of initiator as indicated (samples were incubated at room temperature overnight): strong evidence of assembly occurred, as designed. Presence of residual hairpins in those lanes could be due to (i) stoichiometric concentration differences between DNA strands and (ii) erroneous DNA strands that did not take part in the reaction. By increasing the initiator concentration from 1× to 2× excess (lanes 4 and 5), the products of high molecular weight form two distinct bands, indicating that the assembly of 8 hairpins coexists with the assembly of 9 hairpins (evidence in lanes 8 and 9 in figure 3(B)). Two distinct high molecular weight products are visible even at higher excess initiator concentration (lanes 6–8). In figure 3(B), lanes 1–9 contain different numbers of hairpins as indicated: products of higher molecular weight are successively visible, indicating that each cascade reaction is achieved as designed. Residual hairpins are visible toward the lower end of the gel. Lane 10 contains a 20 bp DNA ladder. Note that the difference between peaks 8 and 9 hairpin lanes in figure 3(B) appears smaller.
than the difference between the two bands in figure 3(A). This discrepancy appears to be due to one or both of the following: (i) two different comb sizes were used to prepare the gels (i.e. a larger comb size was used in figure 3(A) than figure 3(B)), (ii) each lane in figure 3(A) was loaded with 10 pmoles whereas each lane in figure 3(B) was loaded with 5 pmoles.

Control analysis of the linear cascade reaction with different number of hairpins is shown in figure 4. In gel A, only hairpins 8 and 9 are characterized with and without the initiator (I8). Lane 2 contains a mixture of hairpins 8 and 9; products of high molecular weight are almost absent. Lane 3 contains a mixture of hairpins 8 and 9 subjected to a quick 90 °C anneal for 10 min; products of high molecular weight are visible but with little evidence that the cascade reaction occurred. Lane 4 contains hairpin 8 with the initiator; products of high molecule weight are visible, indicating that the hairpin and initiator formed as anticipated. Lane 5 contains both hairpins 8 and 9 with the initiator; products of high molecular weight are visible, indicating that the linear cascade reaction occurred. Residual hairpins in lanes 4–5 indicate a mismatch between hairpin concentrations. In gel B, only hairpins 6, 7, 8, and 9 are characterized with and without the initiator (I9). Lanes 2, 3, and 4 contain a mixture of two hairpins (H6,7), three hairpins (H6,7,9), and four hairpins (H6,7,9,11), respectively without the initiator; products of high molecular weight are slightly visible, indicating evidence of some leak. Lanes 5, 6, 7 contain a mixture of H6,7, H6,9, and H6,9, respectively with the initiator; products of high molecular weight are visible, indicating strong evidence of linear cascade reactions. A mismatch of hairpin concentrations is also visible in lines 7 and 8. In gel C, only hairpins 4, 5, 6, 7, 8, and 9 are characterized with and without the initiator (I9). Lanes 2–6 contain a mixture of two hairpins (H4,5), three hairpins (H4,5,6), four hairpins (H4,5,6,7), five hairpins (H4,5,6,7,9) and six hairpins (H4,5,6,7,9,11) without the initiator; most of hairpins stay as monomers and little evidence of cascade reaction is visible. Lanes 7–11 contain a mixture of H4,5, H4,6, H4,7, H4,6,9, and H4,9, respectively with the initiator; products of high molecular weight are visible, indicating that strong linear cascade reactions are initiated by the initiator. Lanes 5, 8, and 11 in gels A, B, and C show high molecular weight of linear oligomer products formed, indicating strong evidence of linear cascade DNA hybridization chain reactions as anticipated.

The kinetics of linear cascade DNA hybridization chain reactions (LCR) can be explored using fluorescence emissions. The reporter complex is initially quenched due to the close proximity of the fluorophore and quencher, which are constrained by duplex formation (figure 5(A)). Upon hybridization, the reporter complex is displaced by the open arm of hairpin 9, resulting in an increase in fluorescence emission. Figure 5(B) shows real-time kinetic characterization of the linear cascade DNA hybridization chain reaction of nine distinct hairpins. All fluorescence emissions were normalized to the concentration of the output of the last hairpin (H9); the normalization technique was adopted from prior studies [19]. In this case, each opened hairpin 9 is assumed to bind to the reporter complex. In figure 5(B), no fluorescence emission is observed while monitoring the kinetic...
of LCR in the absence of initiator (black curve) for more than 10 h, indicating that the linear cascade chain reaction does not occur in the absence of the initiator (I). Addition of I to the hairpin mixture at 1× concentration leads to rapid fluorescence emission for up to 200 min which slowly increases as equilibrium is reached, provide strong evidence of the linear cascade chain reaction (red curve). In the presence of excess initiator, the LCR system yields minimal difference in fluorescence signal compared to 1× concentration of initiator (green curve), demonstrating that the linear cascade chain reaction is sufficient at 1× concentration, although equilibrium gel analysis (figure 3(A) lane 5) indicates minimal leftover hairpins in the presence of 2× initiator concentration. Figure 5(B) shows the effect of a leak (an undesired reaction) in the absence of the initiator as a function of hairpin concentrations—a higher concentration causes more leak. This result confirms the challenge of leak when constructing DNA systems to operate at high concentrations. Even though DNA strands were purified using a denaturing PAGE method, leaks are difficult to eliminate. At 5 nM concentration, our LCR system leak occurs at a slower rate than the actual reaction as shown in figure 5(A). An alternative approach to conquer leak is (i) to tether DNA systems on a surface of DNA nanostructures [63, 64, 69–72], or (ii) to incorporate double or triple length domain motifs [73]. Note that the fluorescence data for leaks at 50 nM indicates quite significant leaks as compared to the gel data for leaks at 500 nM of the same system at equilibrium in figure 3(A) lane 3. This discrepancy appears to be due to any of the following: (i) the fluorescence method relied on the signal readout from the reporter complex whereas the gel method relied on the nanostructure formation directly due to just hairpins, (ii) improper synthesis of the last hairpin could cause interaction directly with the reporter complex, and thus produce a false positive fluorescence signal in addition to the actual fluorescence leak resulting from triggering LCR in the absence of the initiator strand, or (iii) the gel images in figures 3(B) and 4 do show a detectable amount of leak, while the gel image in figure 3(A) does not show many intermediate bands as part of the leak, just DNA bands at the top of the lane. It is possible that at higher concentration, the intermediate products shifted/aggregated to form larger products that can be seen as bands at the top of lane 3.
The effect of the number of hairpins on the rate of the linear cascade reaction is shown in figure 6(A). The dotted horizontal line indicates the half-life of the linear cascade reaction (i.e. the time required for the reaction to reach 50% completion). All experimental data was collected at 5 nM concentration. Initiating the linear cascade chain reaction with a single hairpin results in a rapid fluorescence emission which slowly increases as equilibrium is reached (black curve). The linear cascade chain reaction reaches 50% completion within 1.47 min. The linear cascade chain reaction involving nine hairpins (cyan curve) takes approximately 41.97 min to reach 50% completion, roughly 28 times slower than the reaction involving a single hairpin. To reach 50% completion, the linear cascade reactions involving two (red curve), four (green curve), and six (blue curve) hairpins were measured to be 8.84, 18.78, and 27.98 min, respectively. As expected from the simulation data in figure 2 and the experimental results in figure 6(A), a lower number of hairpins results in a faster linear cascade reaction and vice versa. Experimental results indicate that the half-time completion of our LCR system follows a linear behavior as shown in figure 6(B). As more hairpins are involved in the linear cascade reaction, more time is required to reach the half-time completion. The linear behavior provides an additional criterion for the design and scaling of new DNA circuits. In addition, the experimental results in figure 6(A) were then fitted using the least squares method which was adopted from Zhang et al [74]. In particular, data from figure 6(A) were fitted to
the corresponding cascade simulations and the fitting results were detailed in SI figure S3. The bimolecular rate constant of the proposed system was within the same order of magnitude as of the prior studies [19, 74]. Although linear behavior has been observed in our systems, the yield of the cascade reaction is difficult to quantitatively obtain from our gel data or fluorescence data. One possible approach to roughly estimate the yield of the cascade reaction is to compare the raw fluorescence data among various linear cascade reaction chains (figure S2). Ideally, the maximum fluorescence intensity at thermal equilibrium of different numbers of hairpins participating in the linear cascade reaction would reach the same level. However, the actual fluorescence data indicate that the maximum fluorescence intensity at thermal equilibrium of a linear cascade reaction consisting of a single hairpin is higher than that of a linear cascade reaction consisting of nine hairpins. Assuming that (i) the maximum fluorescence intensity at thermal equilibrium is correlated with the yield of hairpins participating in the linear cascade reaction and (ii) the yield of a last hairpin participating in the linear cascade reaction is 100%; then the yield of the cascade reaction consisting of 9 hairpins is approximated to be 79% (figure S2). An alternative approach to estimate the yield is to employ a quantitative calibrated DNA ladder via gel electrophoresis or to directly measure the final linear structures via scanning probe microscopy. To avoid non-completion reactions, stoichiometry among DNA strands must be preserved and erroneous DNA strands must not participate in the reaction. In addition to these proposed approaches, different versions of hairpins can be tuned to determine the causes of additional peaks in the gel data.

Conclusion

We have demonstrated the design, simulation, and synthesis of linear cascade DNA hybridization chain reactions by non-equilibrium assembly. Cascade reactions provide a potential strategy to reduce short conventional step-by-step reactions in a highly efficient and elegant fashion. Cascade reactions have been shown to work in many dynamic DNA systems. We have developed a linear cascade DNA hybridization chain reaction consisting of nine distinct DNA hairpins. Gel analysis and reaction kinetics indicate strong evidence of the designed assembly. For scaling up DNA circuits, empirical data show that half-time completion follows linear behavior as the number of hairpins increases. This method can be used for constructing more complex linear cascade reactions on the surface of DNA nanostructures as well as for realizing nano-breadboard DNA circuits.

Acknowledgments

This work was supported by NSF Grants CCF-1217457 and CCF-1141847. The authors wish to thank Brian Petkov who assisted in the proof-reading of the manuscript.
References

[18] Li B L, Ellington A D and Chen X 2011, 2012 Rational, modular adaptation of enzyme-free DNA circuits to multiple detection methods *Nucleic Acids Res.* 39 e119
[34] Omabegho T, Sha R and Seeman N C 2009 A bipedal DNA brownian motor with coordinated legs *Science* 324 67–71
[37] Lund K et al 2010 Molecular robots guided by prescriptive landscapes *Nature* 465 206–10
Liber M, Tomov T E, Tsukanov R, Berger Y and Nir E 2015 A bipedal DNA motor that travels back and forth between two DNA origami tiles. Small 11 568–75


Chen Y J et al 2013 Programmable chemical controllers made from DNA. Nanotechnol. 8 753–62

Qian L, Winfree E and Bruck J 2011 Neural network computation with DNA strand displacement cascades. Nature 475 368–72


Teichmann M, Kopperger E and Simmel F C 2014 Robustness of localized DNA strand displacement cascades. ACS Nano 8 8487–96


Kopperger E, Pirzer T and Simmel F C 2015 Diffusive transport of molecular cargo tethered to a DNA origami platform. Nano Lett. 15 2693–9


Thachuk C, Winfree E and Soloveichik D (ed) 2015 Leakless DNA strand displacement systems. Int. Workshop on DNA-Based Computers (Berlin: Springer)