[Supporting Information to accompany:]

Hydrophobic Organic Linkers in the Self-assembly of Small Molecule-DNA hybrid Dimers: A Computational-Experimental Study of the Role of Linkage Direction in Product Distributions and Stabilities

Ilyas Yildirim,[†,‡,§] Ibrahim Eryazici,[†,§] SonBinh T. Nguyen,[†,*] George C. Schatz[†,‡,*]

[§]These authors contributed equally

[†]*Department of Chemistry and* [‡] *International Institute for Nanotechnology, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3113*

Author contributions audit. I. Y., I. E., and S. T. N. conceived the experiments presented herein. The materials are prepared and characterized by I. E., who carried out all the experiments. I. Y. performed the molecular dynamic simulations, and analyzed the data. I. Y. and I. E drafted the paper together. S. T. N. and G. C. S. supervised the project and finalized the manuscript.

Table of Contents

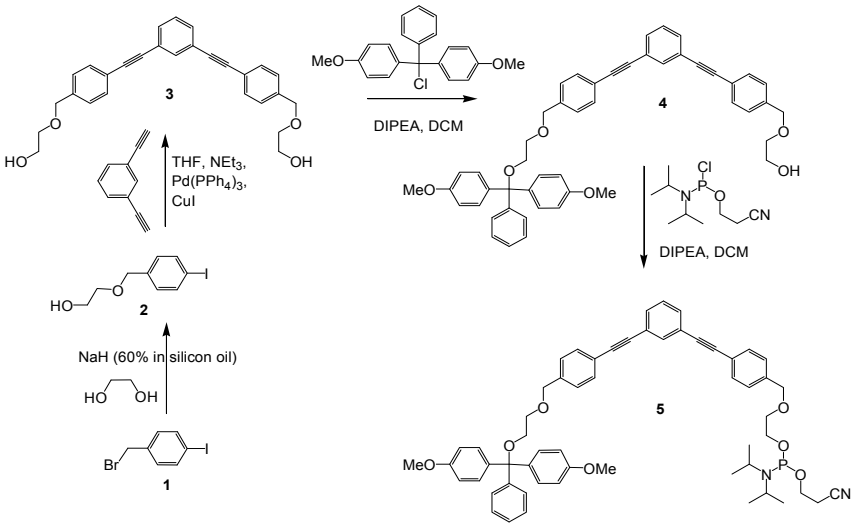
Experimental details

Section S1. General procedures, materials, and instrumentation.

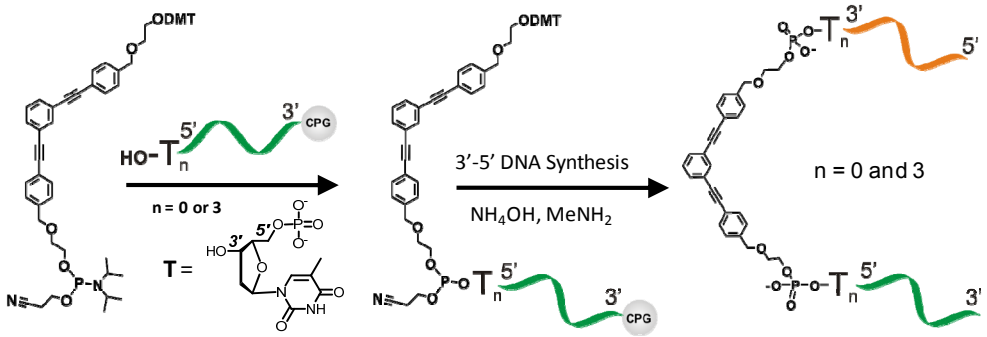
Unless otherwise stated, all reagents and reagent-grade solvents were purchased from Acros Organics (Geel, Belgium), Strem Chemicals, Inc. (Newburyport, MA), Aldrich Chemical Company (Milwaukee, WI), or Glen Research (Sterling, VA), and used as received. Ultrapure deionized H_2O (18.2 MΩ cm resistivity) was obtained from a Milipore system (Milli-Q Biocel). Phosphoramidite **5** (Scheme S1) was synthesized using a previously published procedure.[S1] General procedures for the synthesis and purification of single-stranded (ss) $SMDH_2$s (Table 1) were published in our earlier work.[S2]

Synthesis of DNA sequences were performed on an Expedite 8909 Nucleic Acid system. Unmodified DNAs and $SMDH_2$s were purified on an Agilent 1100 HPLC equipped with semi-preparative (Dynamax, 250×10 mm, Microsorb 300 Å/10 μm/C18, Agilent # R083213C10) and characterized using an analytical column (Dynamax, 100×4.6 mm, Microsorb 100 Å/3 μm/C18, Agilent # R0080200E3). Absorption spectra and melting analyses of DNA materials were recorded on a Varian Cary 300 Bio UV-Vis spectrophotometer using a masked quartz cell (path length = 10 mm, Starna cells Inc., catalog # 29B-Q-10-MS, Atascadero, CA). Gel electrophoresis experiments were carried out using a model V16 polyacrylamide gel electrophoresis system (Labrepco, catalog # E21070010, Horsham, PA) and Bio-Rad PowerPac 1000 power supply. Imaging of the gel was performed in a Kodak Gel Logic 200 Imaging System. PAGE-Gel images were analyzed using ImageJ software, version 1.46r (Wayne Rasband, National Institute of Health, USA, http://imagej.nih.gov/ij). Matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) data was collected on a Bruker (Billerica, MA) Daltonics AutoFlex III MALDI-ToF mass spectrometer. Data from Agilent HPLC and Bruker MALDI-ToF instruments were process using MestreNova

software version 8.1.1-11591. Annealing of the mixtures were performed using either Eppendorf® Thermomixer® R (Eppendorf, # 022670107) or Veriti® 96-Well Fast Thermal Cycler (Applied Biosystems®, # 4375305).



Scheme S1. Synthesis of the core phosphoramidite **5**.[S1]



Scheme S2. Synthesis of the $SMDH_2$ building blocks (Table 1, entries 1-2, 7-8).

Section S2.

Analytical HPLC analysis of the $SMDH_2$ building blocks. After the $SMDH_2$ building blocks (Table 1) was purified using semi-preparative HPLC (see the SI of our previous works[S2, S3]), each sample (1-3 nmole) was injected into an analytical column (Dynamax, 100 × 4.6 mm, Microsorb 100 Å/3 µm/C18, Agilent # R0080200E3) and eluted using a gradient method beginning with 95:5 v/v 0.1 M TEAA (aq):acetonitrile (TEAA (aq) = triethylammonium acetate, aqueous solution), and increasing to 60:40 v/v 0.1 M TEAA(aq):acetonitrile over 70 minutes, with a flow rate of 1 mL/min. These $SMDH_2$ products displayed only one peak in analytical HPLC traces (Figure S1-S14), confirming their good purity.

MALDI-ToF analysis of the $SMDH_2$ building blocks. The MALDI-ToF matrix was prepared by addition of aqueous ammonium hydrogen citrate (0.6 µL of a solution of 15 mg in 30 µL of H_2O) to a solution of 2-hydoxypicolinic acid (Fluka # 56297, 2 mg in H_2O:MeCN (30 µL, 1:1 v/v). An aliquot (1 µL, 10-100 pmole) of the isolated ssDNA sample was then mixed with a portion of this matrix (5 µL). A small amount of the resulting mixture (0.3-3 µL) was dropped on a steel MALDI-ToF plate and dried at 25 °C before being analyzed on a Bruker (Billerica, MA) Daltonics AutoFlex III MALDI-ToF mass spectrometer as negative or positive ions using the linear mode. The instrument was equipped with Smartbeam™ laser technology operated at 80% power with a sampling speed of 10 Hz. Five hundred spectra were averaged for the corresponding mass spectrum. The instrument was operated using the following parameters: ion source voltage 1 = 20 kV, ion source voltage 2 = 18.5 kV, lens voltage

= 8.5 kV, linear detector voltage = 0.6 kV, deflection mass = 3000 Da. The MALDI-ToF spectra of the $SMDH_2$ building blocks were shown in the inset of Figure S1-S14).

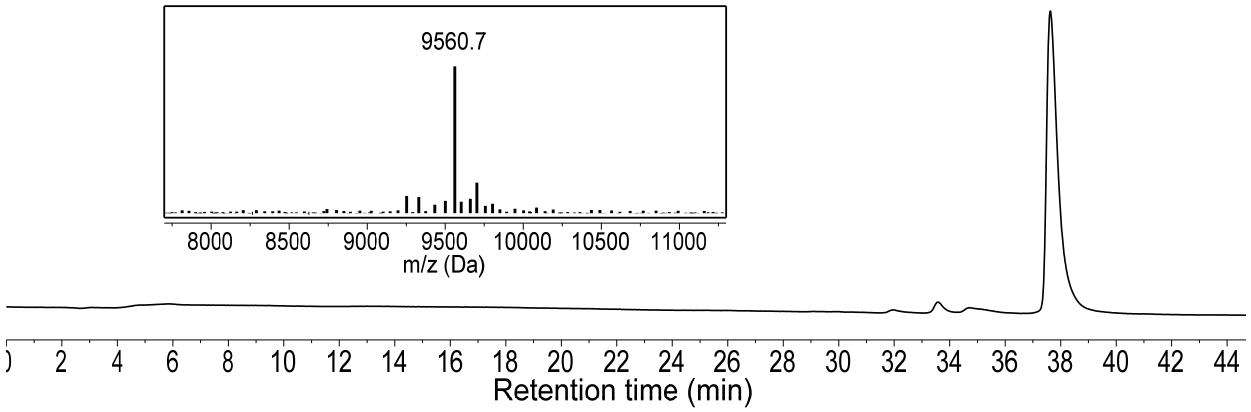


Figure S1. Analytical HPLC trace of 3′-**X**-5′-***C***-3′-**Y**-5′ sequence (Table 1, entry 1). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9560.7 (9552.1 theoretical)).

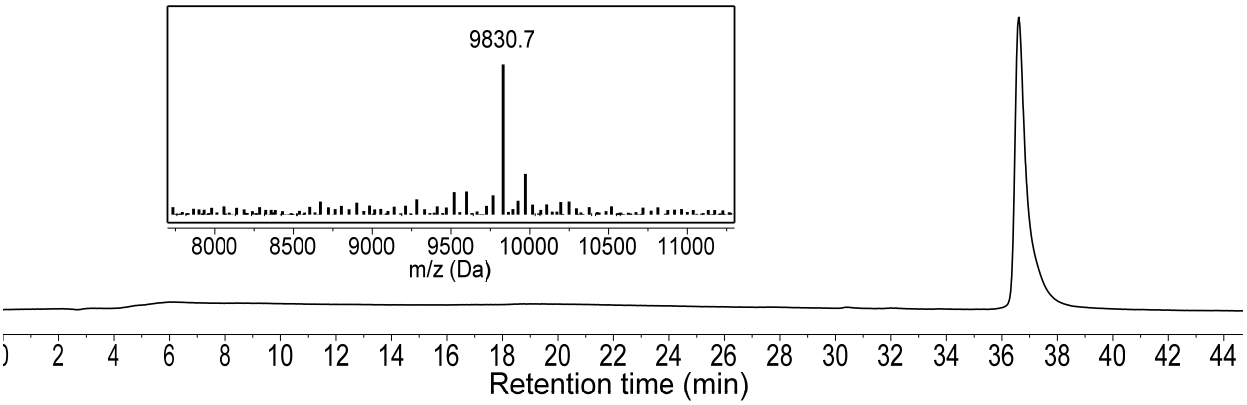


Figure S2. Analytical HPLC trace of 3′-**X′**-5′-***C***-3′-**Y′**-5′ sequence (Table 1, entry 2). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9830.7 (9828.3 theoretical)).

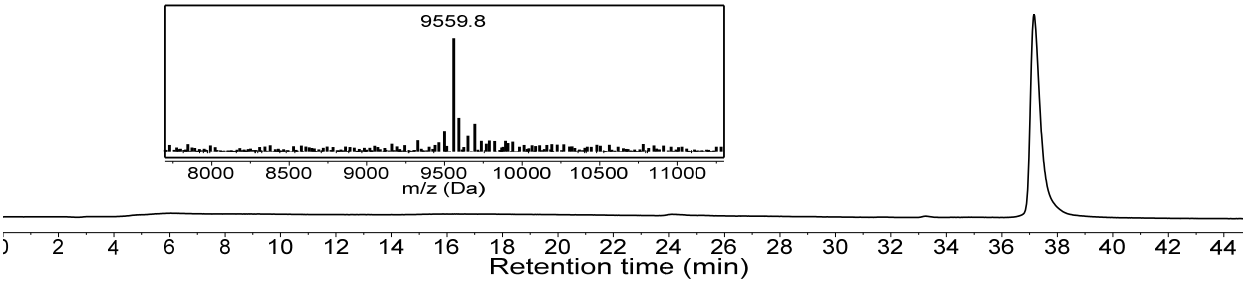


Figure S3. Analytical HPLC trace of 3′-**X**-5′-***C***-5′-**Y**-3′ sequence (Table 1, entry 3). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9559.8 (9552.1 theoretical)).

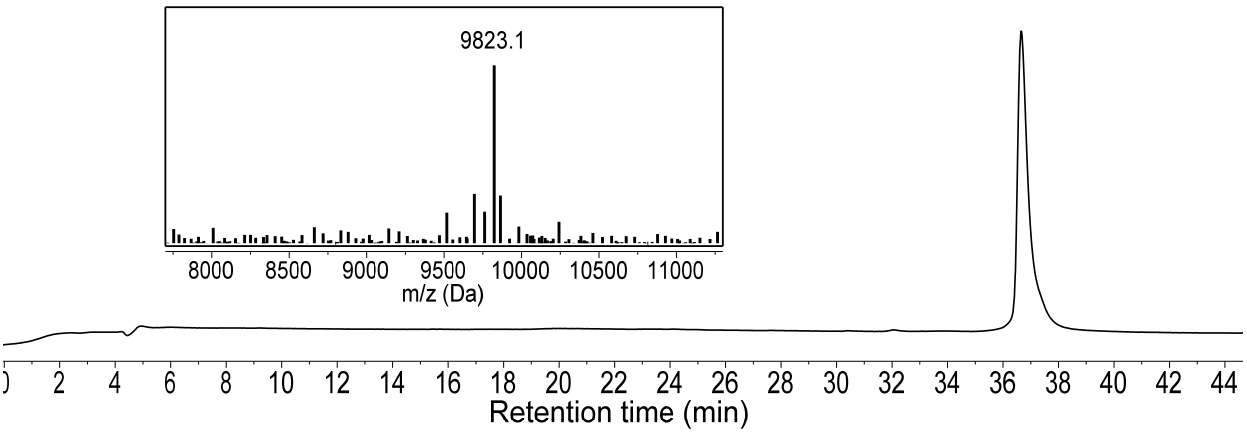


Figure S4. Analytical HPLC trace of 3′-**X′**-5′-***C***-5′-**Y′**-3′ sequence (Table 1, entry 4). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9823.1 (9828.3 theoretical)).

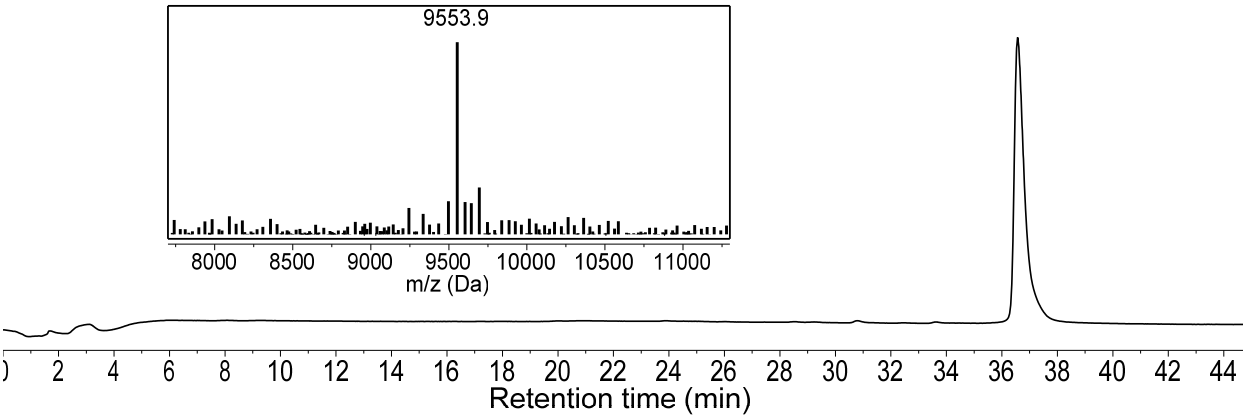


Figure S5. Analytical HPLC trace of 5′-**X**-3′-***C***-3′-**Y**-5′ sequence (Table 1, entry 5). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9553.9 (9552.1 theoretical)).

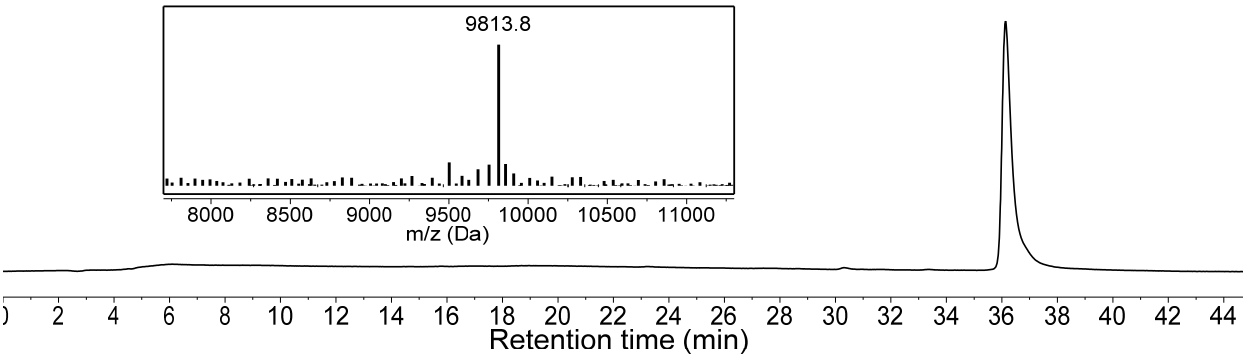


Figure S6. Analytical HPLC trace of 5′-**X′**-3′-***C***-3′-**Y′**-5′ sequence (Table 1, entry 6). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9813.8 (9828.3 theoretical)).

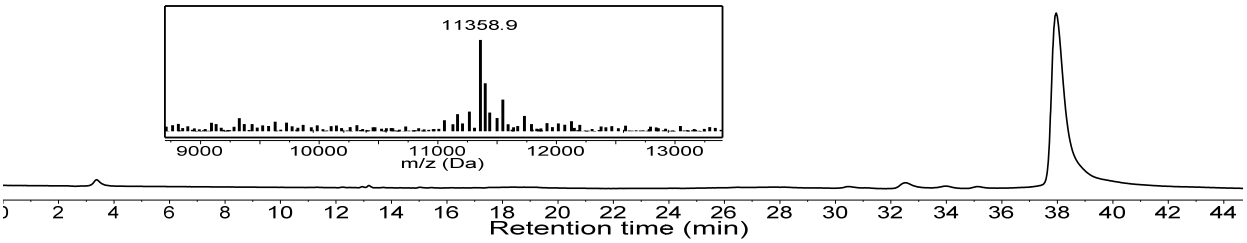


Figure S7. Analytical HPLC trace of 3′-**X**-5′-***T*$_3$*CT*$_3$-3′-**Y**-5′ sequence (Table 1, entry 7). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 11358.9 (11377.3 theoretical)).

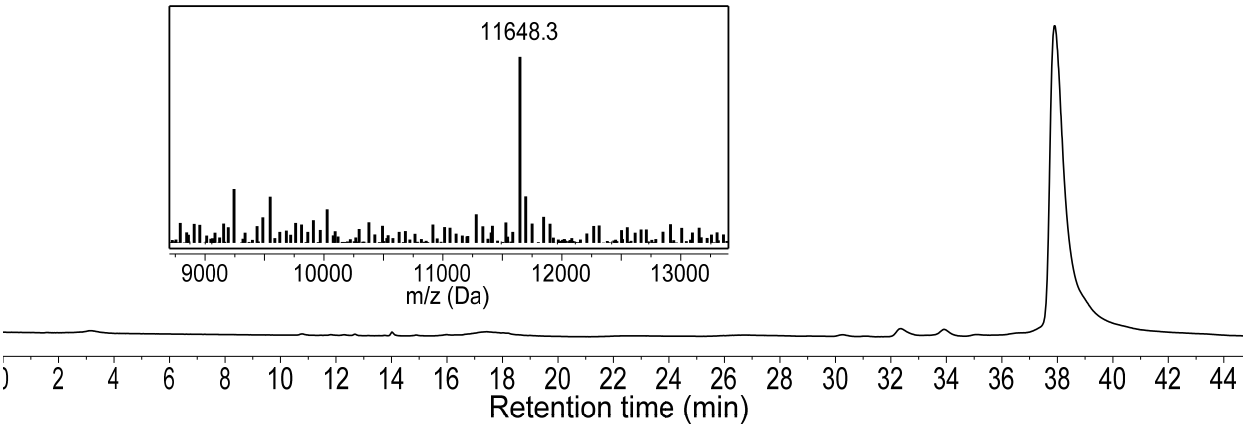


Figure S8. Analytical HPLC trace of 3′-**X′**-5′-***T*$_3$*CT*$_3$-3′-**Y′**-5′ sequence (Table 1, entry 8). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 11648.3 (11653.5 theoretical)).

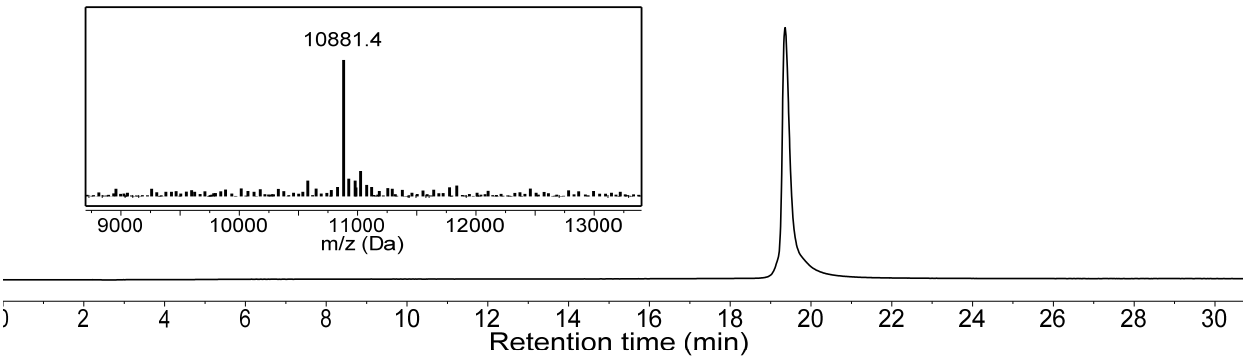


Figure S9. Analytical HPLC trace of 3′-**X**-5′-***T*$_6$-3′-**Y**-5′ sequence (Table 1, entry 9). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 10881.4 (10889.2 theoretical)).

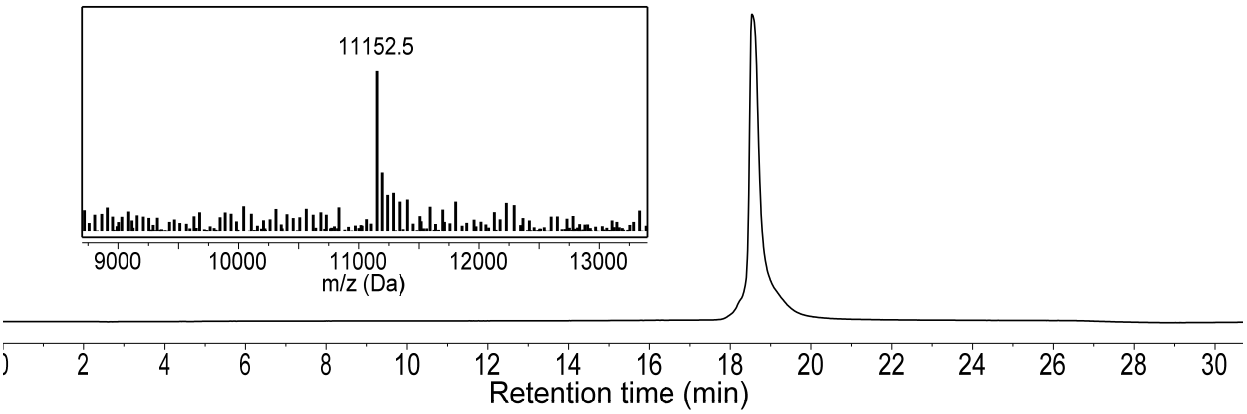


Figure S10. Analytical HPLC trace of 3′-**X′**-5′-***T*$_6$**-3′-**Y′**-5′ sequence (Table 1, entry 10). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 11152.5 (11165.3 theoretical)).

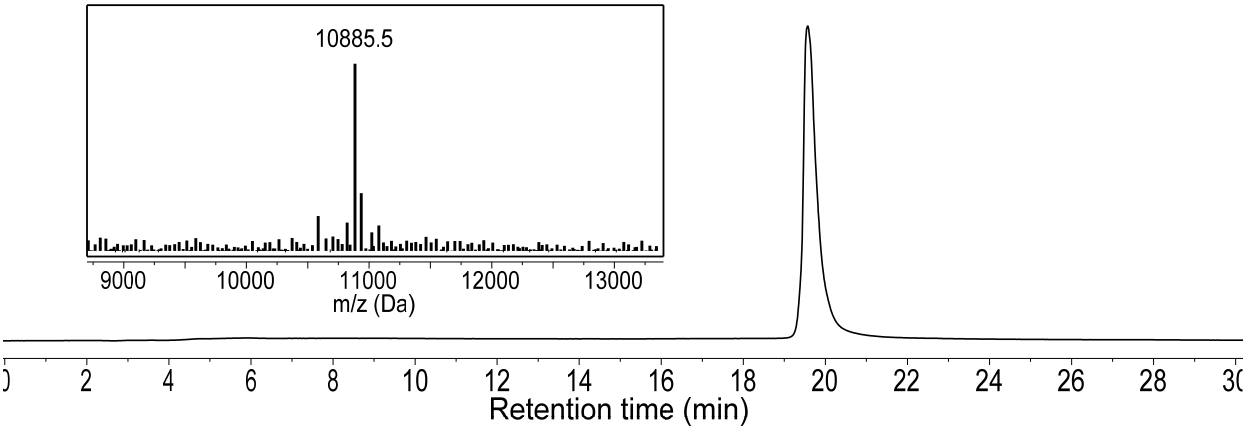


Figure S11. Analytical HPLC trace of 3′-**X**-5′-***T*$_6$**-5′-**Y**-3′ sequence (Table 1, entry 11). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 10885.5 (10889.2 theoretical)).

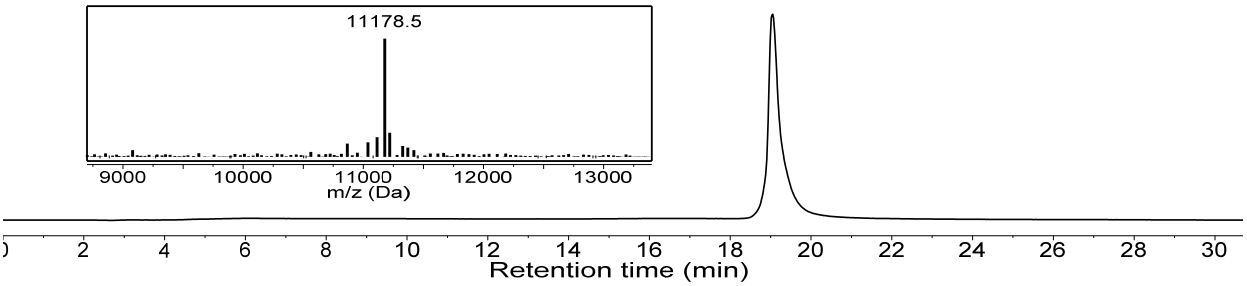


Figure S12. Analytical HPLC trace of 3′-**X′**-5′-***T*$_6$**-5′-**Y′**-3′ sequence (Table 1, entry 12). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 11178.5 (11165.3 theoretical)).

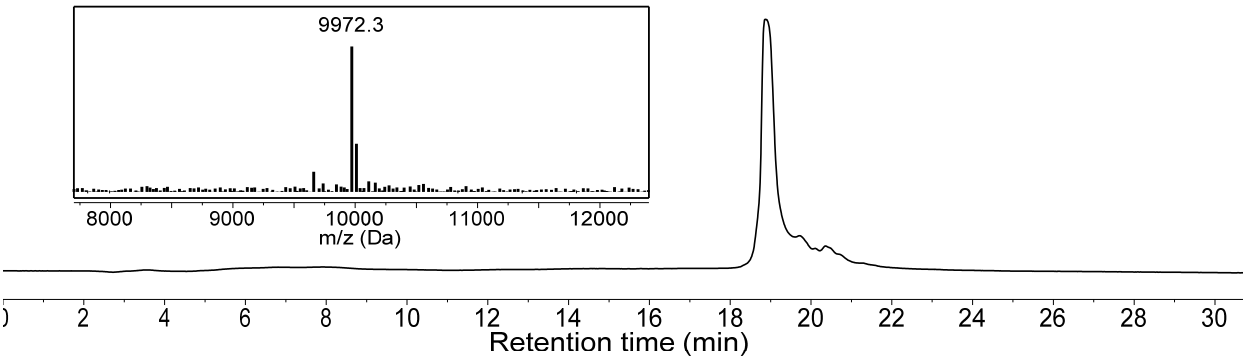


Figure S13. Analytical HPLC trace of 3′-**X**-5′-***T*$_3$**-3′-**Y**-5′ sequence (Table 1, entry 13). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 9972.3 (9976.6 theoretical)).

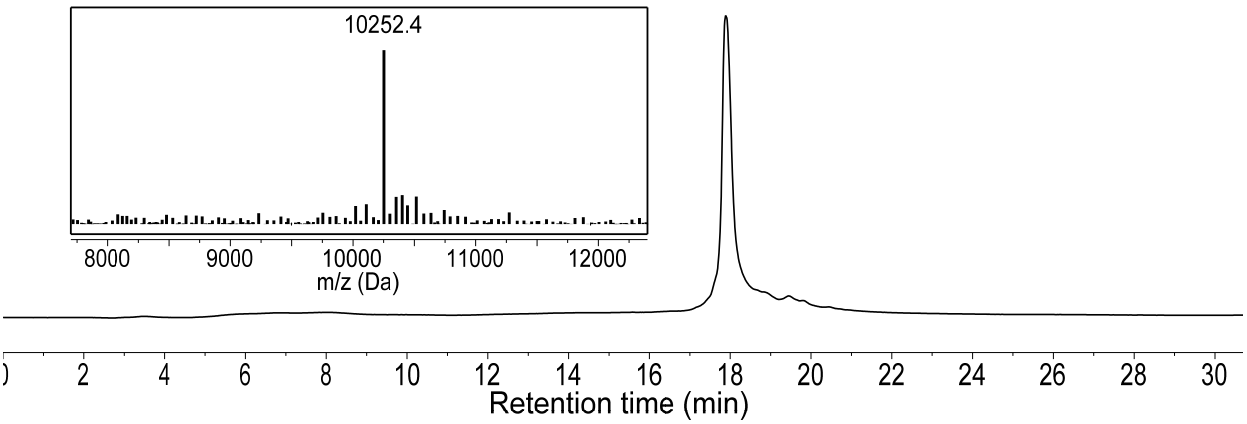


Figure S14. Analytical HPLC trace of 3′-**X′**-5′-***T*$_3$**-3′-**Y′**-5′ sequence (Table 1, entry 14). The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF spectra of the pure product (m/z = 10252.4 (10252.8 theoretical)).

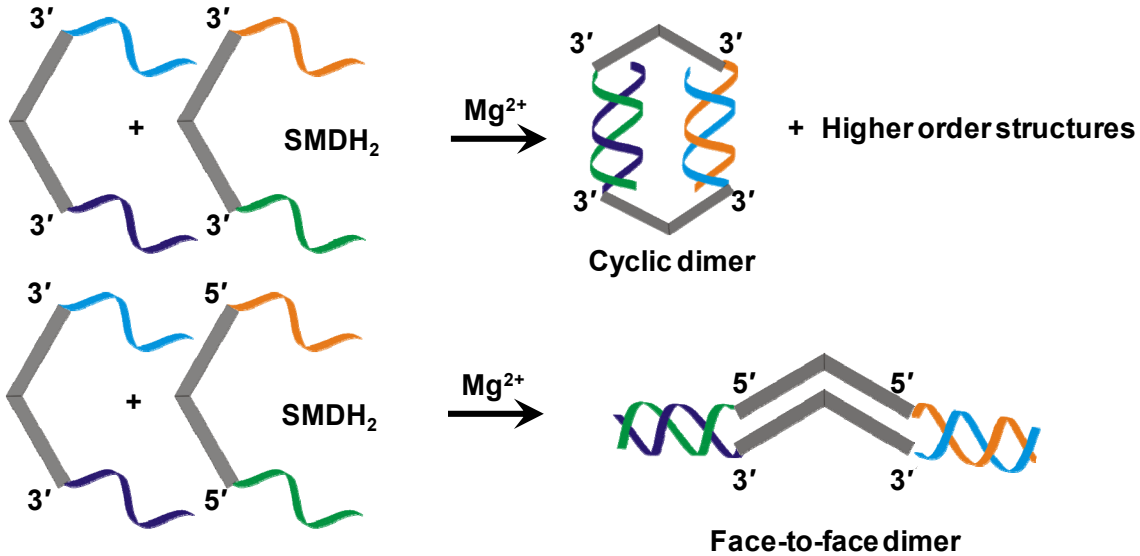


Figure S15. Schematic presentation of the self-assembly of cyclic versus face-to-face dimer structures controlled by the linkage direction of the organic core with the DNA duplex.

Section S3.

Hybridization procedure of the $SMDH_2$ assemblies (Table 2). Hybridization mixtures of the $SMDH_2$ building blocks and unmodified DNA were formed by combining equimolar amounts of the two complementary DNA species in TAMg buffer (40 mM Tris base, 20 mM acetic acid, 7.5 mM $MgCl_2 \cdot 6H_2O$) at 25 °C. The resulting mixtures were then annealed using either a normal or a slow cooling method.

1) **Normal cooling**. The mixture was heated to 90 °C in an Eppendorf® Thermomixer® R (Eppendorf, # 022670107) and kept there for 10 minutes. The instrument was turned off and the mixture was allowed to cool to room temperature over 2.5 h. Linear fitting the cooling profile of the thermomixer (Figure S16a) in the melting range of the $SMDH_2$ assemblies (62-40 °C) showed that the average cooling rate for this range is 0.34 °C/min.
2) **Slow cooling**. The mixture was placed in a Veriti® 96-Well Fast Thermal Cycler (Applied Biosystems®, # 4375305) and the instrument was programmed to cool from 60 °C to 25 °C over the course of 58 h at a rate of 0.01 °C/min.

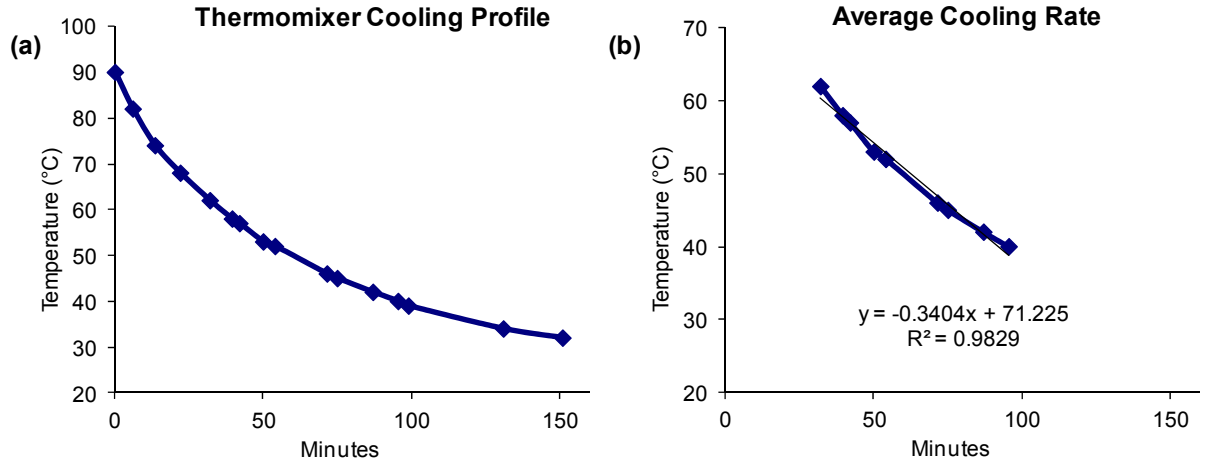


Figure S16. (a) Thermomixer cooling profile (90-25 °C). (b) The average thermomixer cooling rate (0.34 °C/min) in the melting range of the $SMDH_2$ assemblies (62-40 °C) was found by linear fitting the cooling profile.

PAGE-Gel experiments. Non-denaturing PAGE (6% acrylamide) experiments were carried out for DNA assemblies (5 µM total ss-DNA concentration). The polyacrylamide gel was prepared in 1X TAMg Buffer (40 mM Tris base, 20 mM acetic acid, 7.5 mM $MgCl_2 \cdot 6H_2O$). The PAGE was run at 4 °C for 2 h under a 200 V applied field).

Optical Melting Experiments. Hybridized mixtures (Table 2) were denatured by heating the samples from 30 °C to 75 °C at a rate of 1 °C/min while monitoring the UV absorbance at 260 nm at 0.1 °C intervals to observe the melting progress. Alpha curves of melting data were generated using the Varian Thermal Application software (version 3.0(182), Varian Australia Pty., Ltd.). The full-width-at-half-max (fwhm, Table 4) of the derivative was calculated by Gaussian fit in Origin 6.1 (v6.1052 (B232), OriginLab Corporation, Northampton, MA; see our previous work[S3] for details).

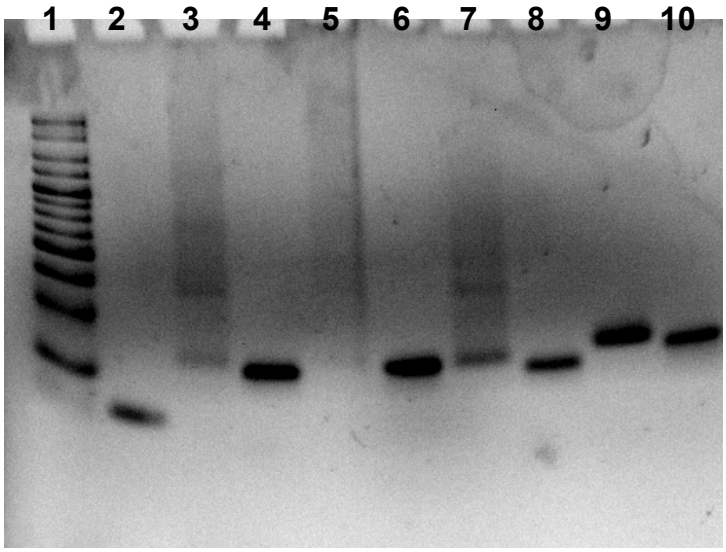


Figure S17. From left to right: *lane 1* = HL5 DNA ladder, *lane 2* = 15-bp (**X**:**X′**), *lane 3* = cyclic-[5′-***C***-3′]:[5′-***C***-3′]′, *lane 4* = control-[5′-***C***-3′], *lane 5* = cyclic-[5′-***C***-5′]:[5′-***C***-5′]′, *lane 6* = control-[5′-***C***-5′], *lane 7* = cyclic-[3′-***C***-3′]:[3′-***C***-3′]′, *lane 8* = control-[3′-***C***-3′], *lane 9* = cyclic-[5′-$\boldsymbol{T_3CT_3}$-3′]:[5′-$\boldsymbol{T_3CT_3}$-3′]′, and *lane 10* = control-[5′-$\boldsymbol{T_3CT_3}$-3′].

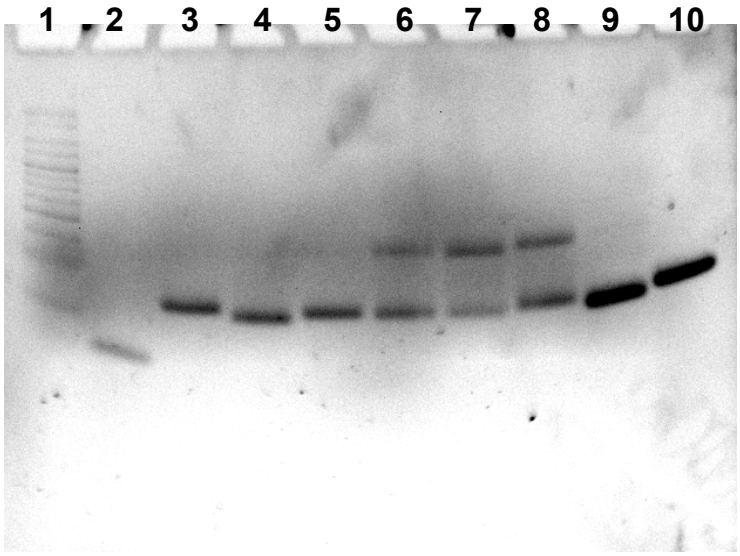


Figure S18. From left to right: *lane 1* = HL5 DNA ladder, *lane 2* = 15-bp (**X**:**X′**), *lane 3* = cyclic-[5′-$\boldsymbol{T_6}$-3′]:[5′-$\boldsymbol{T_6}$-3′]′, *lane 4* = cyclic-[5′-$\boldsymbol{T_3}$-3′]:[5′-$\boldsymbol{T_3}$-3′]′, *lane 5* = cyclic-[5′-$\boldsymbol{T_6}$-5′]:[5′-$\boldsymbol{T_6}$-5′]′, *lane 6* = cyclic-[5′-***C***-3′]:[5′-$\boldsymbol{T_6}$-3′]′, *lane 7* = cyclic-[5′-***C***-5′]:[5′-$\boldsymbol{T_6}$-5′]′, *lane 8* = cyclic-[5′-***C***-3′]:[5′-$\boldsymbol{T_3CT_3}$-3′]′, *lane 9* = ff-[3′-***C***-3′]:[5′-***C***-5′]′, and *lane 10* = ff-[3′-***C***-3′]:[5′-$\boldsymbol{T_6}$-5′]′.

Computational details

Section S4. Parameterization of the organic core

Before molecular dynamics (MD) simulations can be carried out on the $SMDH_2$ molecules, parameters necessary to represent the organic core must be obtained. Amber force field parameters for the organic core (Figure S19) were calculated as described previously.[S3] GAMESS[S4] was used to optimize and calculate the electrostatic potentials at a set of grid points. Restrained electrostatic potential (RESP) charges of organic core atoms were then calculated using the RESP program previously described (see Figure S19 and Table S1).[S3, S5]

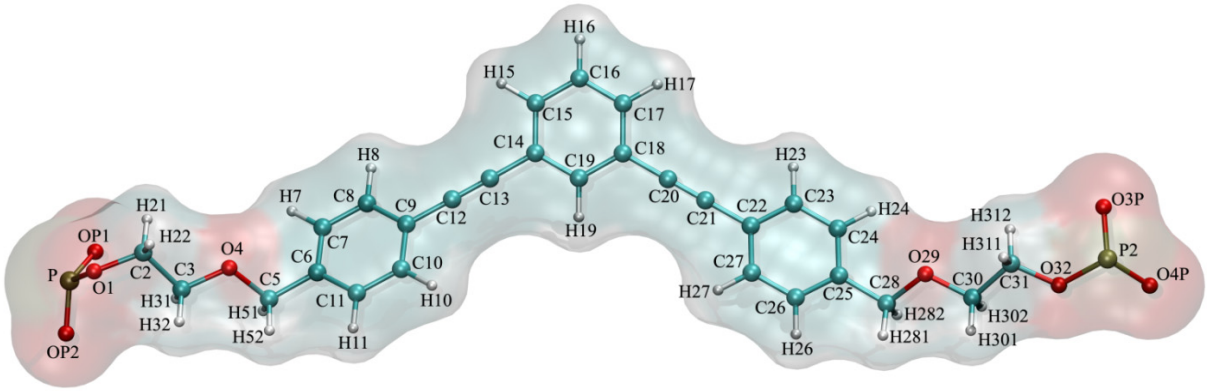


Figure S19. Atom naming in organic core (LN3 = ***C*** = **3**). See Table S1 for atom types and charges.

Table S1. Atom types and charges for the organic core (LNX, X = 1,2,3).

		Charges		
Atom Name	**Atom Type**	**LN1**	**LN2**	**LN3**
P	P	----------	1.166200	1.166200
O1P	O2	----------	-0.776000	-0.776000
O2P	O2	----------	-0.776000	-0.776000
O1	os	-0.498900	-0.498900	-0.498900
C2	c3	-0.084795	-0.083195	-0.083195
H21	h1	0.094012	0.094012	0.094012
H22	h1	0.094012	0.094012	0.094012
C3	c3	0.257907	0.257907	0.257907
H31	h1	0.038700	0.038700	0.038700
H32	h1	0.038700	0.038700	0.038700
O4	os	-0.386434	-0.386434	-0.386434
C5	c3	-0.005304	-0.005304	-0.005304
H51	h1	0.053131	0.053131	0.053131
H52	h1	0.053131	0.053131	0.053131
C6	ca	0.194704	0.194704	0.194704
C7	ca	-0.224620	-0.224620	-0.224620
H7	ha	0.136694	0.136694	0.136694
C8	ca	-0.207087	-0.207087	-0.207087
H8	ha	0.155649	0.155649	0.155649
C9	ca	0.200738	0.200738	0.200738
C10	ca	-0.159992	-0.159992	-0.159992
H10	ha	0.127975	0.127975	0.127975
C11	ca	-0.124824	-0.124824	-0.124824
H11	ha	0.093719	0.093719	0.093719
C12	cg	-0.168655	-0.168655	-0.168655
C13	ch	-0.089869	-0.089869	-0.089869
C14	ca	0.146305	0.146305	0.146305
C15	ca	-0.137915	-0.137915	-0.137915
H15	ha	0.129092	0.129092	0.129092
C16	ca	-0.123769	-0.123769	-0.123769
H16	ha	0.123189	0.123189	0.123189
C17	ca	-0.193351	-0.193351	-0.193351
H17	ha	0.143317	0.143317	0.143317

C18	ca	0.268921	0.268921	0.268921
C19	ca	-0.232822	-0.232822	-0.232822
H19	ha	0.145948	0.145948	0.145948
C20	cg	-0.220983	-0.220983	-0.220983
C21	ch	-0.054183	-0.054183	-0.054183
C22	ca	0.106880	0.106880	0.106880
C23	ca	-0.090752	-0.090752	-0.090752
H23	ha	0.103069	0.103069	0.103069
C24	ca	-0.187479	-0.187479	-0.187479
H24	ha	0.120245	0.120245	0.120245
C25	ca	0.113914	0.113914	0.113914
C26	ca	-0.227756	-0.227756	-0.227756
H26	ha	0.126157	0.126157	0.126157
C27	ca	-0.112719	-0.112719	-0.112719
H27	ha	0.123411	0.123411	0.123411
C28	c3	0.328443	0.328443	0.328443
H281	h1	-0.028416	-0.028416	-0.028416
H282	h1	-0.028416	-0.028416	-0.028416
O29	os	-0.495623	-0.495623	-0.495623
C30	c3	0.253631	0.253631	0.253631
H301	h1	0.035104	0.035104	0.035104
H302	h1	0.035104	0.035104	0.035104
C31	c3	-0.000810	-0.000810	-0.000810
H311	h1	0.076236	0.076236	0.076236
H312	h1	0.076236	0.076236	0.076236
O32	os	-0.524600	-0.524600	-0.523000
P2	P	----------	----------	1.166200
O3P	O2	----------	----------	-0.776000
O4P	O2	----------	----------	-0.776000

Section S5

Preparation of the DNA systems for MD Simulations. All the model DNA systems (with sequences described in Table 1) were created in B-form conformations using the Nucgen module of AMBER 9.[S6] Amber99 force field[S7, S8] with revised χ,[S9] and α/γ[S10] torsional parameter sets were used to define the DNA parameters. Systems in Table 2 (entries 1-7) were then prepared where organic cores were attached at either 3′- or 5′-ends of the DNA (Figure 6).

Data Analysis.

Solvent-Accessible Surface Area (SASA) analysis. As the organic cores are hydrophobic in nature, they will try to reduce their SASA in aqueous environments. The time evolution of this process was investigated using the AREAIMOL program from ccp4 ver. 6.2.0[S11] on all the snapshots extracted from the trajectories at intervals of 50 ps (Tables S2-S4). Only the benzyl and acetylenes moieties of organic cores were included in these calculations.

Root mean square (RMS) deviation analysis. RMS deviation analysis with respect to B-form conformation was performed on the DNA duplexes to observe DNA distortions as a function of time using the Ptraj module of AMBER 11[S12] (Tables S2-S4 and Figures S25-S31).

Time evolution of total hydrogen-bonds. Given the dynamic nature of the DNA hybrids explored in this paper, the total number of hydrogen-bonds formed will change with time. In particular, the base-pairing patterns will change when the terminal base pairs are distorted because of fraying effects. As such, we calculated the total numbers of hydrogen bonds as a function time using the Ptraj module of AMBER 11[S12] (Table S2-4). Snapshots extracted from the trajectories at intervals of 50 ps were used in these calculations, where a hydrogen-bond was assumed to form when the distance between the donor and acceptor atoms was less than 3.5 Å and the angle forming the hydrogen bond was greater than 120.0°.

Section S6. Unrestrained MD simulations.

*cyclic-[5′-**C**-3′]:[5′-**C**-3′]′*. Organic cores were attached to 3′- and 5′- ends of DNA sequences to create the assembly shown in Table 2, entry 1. Cyclic dimer, tetramer, and hexamer nanostructures (Figures 6b and 7a,b) were prepared to investigate the structural properties of the dimer as well as the higher order structures that were observed in PAGE analysis.

*cyclic-[5′-**C**-5′]:[5′-**C**-5′]′ and cyclic-[3′-**C**-3′]:[3′-**C**-3′]′*. The organic cores were attached to either 5′- or 3′-ends of DNA sequences to create the assemblies shown in Table 2, entries 2 and 3, respectively. Cyclic dimer nanostructures for these systems were then created (Figures 6c and 6a).

*cyclic-[5′-**T**$_3$**C**T$_3$-3′]:[5′-**T**$_3$**C**T$_3$-3′]′*. T_3 spacers were inserted between the organic cores and DNA sequences to create the assembly shown in Table 2, entry 4, which was then used to prepare the cyclic dimer nanostructure shown in Figure 6d.

*cyclic-[5′-**T**$_6$-3′]:[5′-**T**$_6$-3′]′*. Rather using organic linkers, T_6 spacers were inserted between DNA sequences to create the assembly shown in Table 2, entry 5, which then was then used to prepare the cyclic dimer nanostructure shown in Figure 6e. Compared to the organic linkers, T_6 spacers are more flexible and hydrophilic and do not produce any higher-order structures.

control-[5′-C-3′]. The dimer and higher-order structures described above (Table 2, entries 1-10) are cyclic nanostructures. A model DNA system with only one organic core (Table 2, entry 13) was prepared that represents single core 3′-***C***-5′ system (Figure 5). This system was specifically prepared to evaluate any potential stacking preference of the hydrophobic organic cores when sandwiched between two DNA duplexes and while holding back any potential strains created by connecting the organic cores to both DNA duplexes, such as the case observed for cyclic dimer nanostructures (see Movie S1).

All unrestrained MD simulations were run in Generalized Born (GB)-implicit solvent model (GB^{HCT})[S13, S14] with 0.3 M salt concentrations. Systems were initially minimized in two steps as previously described.[S3] Chirality restraints were imposed on the DNAs during the minimization. After minimization, temperature was gradually increased to 300 K from 0 K. Again, chirality restraints were imposed on DNA residues to keep them in regular DNA orientations during the temperature increase.

The PMEMD module of the AMBER[S12] simulation package was used in both equilibration and production runs. At each equilibration step, temperature was increased by 50 K every 200 ps with a 2 fs time step. A long-range cutoff of 20 Å was used. Trajectories were written every 1000 steps. In the final equilibration step, the temperature was kept at 300 K.

The production runs of dimer, tetramer, hexamer, and single core control-[5′-***C***-3′] systems were similar to the equilibration steps described above except that no chirality restraints were imposed on the systems. Temperature was kept at 300 K, and trajectories were written every 5000 steps. A time step of 1 fs was used in all production runs. Four independent MD simulations were run on dimer systems, with over 90 ns each. Two independent MD simulations were run on a single core control-[5′-***C***-3′] system, with over 85 ns each. MD simulations that are 400 ns and 243 ns long were run on tetramer and hexamer systems, respectively.

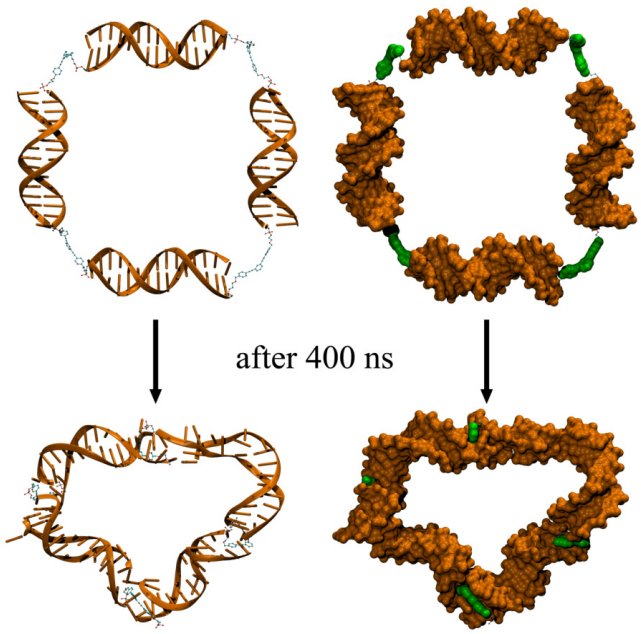


Figure S20. Initial (top) and final (below) cyclic tetramer structures of [5′-***C***-3′]:[5′-***C***-3′]′after 400 ns in cartoon (left) and molecular surface (right) representations. The organic cores are represented in green color (lower right) and are stacked between neighboring DNA duplexes. The distortions of the DNA duplexes here are not as severe as those in the cyclic dimers (see Figure 7 and Table 3).

Table S2. Details of the RMSD, SASA, and H-bond analyses. Cyclic dimer, tetramer, and hexamer (30, 60, and 90 bp, respectively) were examined for [5′-***C***-3′]:[5′-***C***-3′]′. Cyclic dimers were examined for [3′-***C***-3′]:[3′-***C***-3′]′ and [5′-***C***-5′]:[5′-***C***-5′]′. The first 20 ns of MD time were excluded from the calculations of the dimer systems and the first 50 ns of MD time were excluded from the remaining calculations.

Case	Duplex RMSD	Duplex RMSD (middle 9 bp)	SASA (Å^2) (per core)	# of H-bond	# of H-bond (new)	% of H-bonds lost	% of H-bonds lost[a]
hexamer [5′-***C***-3′]:[5′-***C***-3′]′	6.63 ± 1.42	3.66 ± 0.36	122.21	163.40	191.41	% 17	% 3
tetramer [5′-***C***-3′]:[5′-***C***-3′]′	6.50 ± 1.14	3.89 ± 0.64	133.79	102.95	127.98	% 22	% 3
dimer [5′-***C***-3′]:[5′-***C***-3′]′	7.42 ± 0.54	3.92 ± 0.49	162.21 ± 31.62	52 ± 2	60 ± 1	% 21	% 9
MD # 1	8.29	4.55	180.89	50.92	57.95	% 23	% 12
MD # 2	7.42	3.65	150.71	52.96	59.16	% 20	% 10
MD # 3	7.16	3.28	116.83	54.73	61.95	% 17	% 6
MD # 4	6.82	4.21	200.39	50.41	60.27	% 24	% 9
dimer [3′-***C***-3′]:[3′-***C***-3′]′	7.68 ± 1.26	4.49 ± 1.45	165.24 ± 18.94	51 ± 3	59 ± 2	% 24	% 11
MD # 1	6.64	3.66	163.56	51.63	60.55	% 22	% 8
MD # 2	9.75	7.00	138.59	45.71	55.32	% 31	% 16
MD # 3	6.71	3.52	166.75	51.54	60.41	% 22	% 8
MD # 4	7.60	3.79	192.06	53.33	59.51	% 19	% 10
dimer [5′-***C***-5′]:[5′-***C***-5′]′	7.92 ± 1.78	3.79 ± 0.61	128.7 ± 21.52	49 ± 2	58 ± 2	% 26	% 12
MD # 1	5.88	3.29	140.37	45.69	57.93	% 31	% 12
MD # 2	7.98	3.59	107.46	50.66	61.06	% 23	% 7
MD # 3	7.10	3.45	108.80	48.41	56.59	% 27	% 14
MD # 4	10.73	4.82	158.17	50.70	57.44	% 23	% 13
dimer [5′-***T*$_3$*CT*$_3$**-3′]:[5′-***T*$_3$*CT*$_3$**-3′]′	5.87 ± 0.46	3.57 ± 0.11	136.75 ± 28.53	60 ± 3	68 ± 1	% 8	- % 3
MD # 1	6.59	3.63	118.26	61.28	67.88	% 7	- % 3
MD # 2	5.91	3.49	123.83	63.75	68.23	% 3	- % 3
MD # 3	5.61	3.44	186.02	60.52	68.82	% 8	- % 4
MD # 4	5.37	3.72	118.90	56.15	66.83	% 5	- % 1
dimer [5′-***T*$_6$**-3′]:[5′-***T*$_6$**-3′]′	6.58 ± 0.41	3.86 ± 0.15	-	57 ± 3	68 ± 3	% 13	- % 4
MD # 1	6.91	3.73	-	57.45	66.75	% 13	- % 1
MD # 2	6.67	4.09	-	58.85	69.26	% 11	- % 5
MD # 3	5.89	3.71	-	60.11	71.30	% 9	- % 8
MD # 4	6.86	3.89	-	53.17	63.81	% 19	% 3

[a] Negative value means percentage hydrogen bond gained.

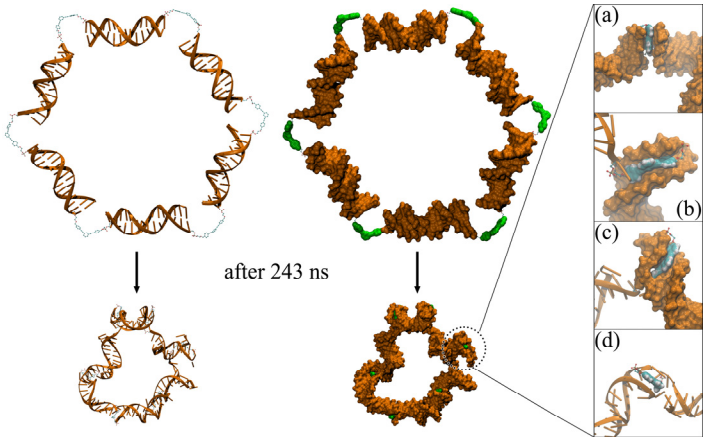


Figure S21. Initial (top) and final (below) cyclic hexamer structures of [5′-***C***-3′]:[5′-***C***-3′]′ after 243 ns in cartoon (left) and molecular surface (right) representations. The organic cores are represented in green color (lower middle) and are stacked in different forms: (a) direct stacking between neighboring DNA duplexes, (b) perfect stacking in the minor groove, (c) distorted stacking in the minor groove, and (d) stacking within distorted terminal base pairs of neighboring DNA duplexes. The distortions of the DNA duplexes here are not as severe as those in cyclic dimers (see Figure 7 and Table 3).

Table S3. SASA analysis of restrained MD simulations for cyclic-dimer $SMDH_2$ systems with varying DNA lengths. The values shown in parentheses are the lowest SASA results extracted from simulated annealing MD simulations for structures having final restraint energies and duplex rmsd values less than 10 kcal/mol and 10 Å, respectively (Figure S24 and Table S4). The errors reported are the standard deviations.

# of base pair	cyclic-[3′-***C***-3′]:[3′-***C***-3′]′ (Å^2)	cyclic-[5′-***C***-3′]:[5′-***C***-3′]′ (Å^2)	cyclic-[5′-***C***-5′]:[5′-***C***-5′]′ (Å^2)
11	489.8 ± 38.3 (302.28)	550.1 ± 44.1 (276.16)	517.5 ± 31.0 (435.76)
13	494.3 ± 22.9 (254.21)	478.5 ± 28.2 (433.95)	479.1 ± 35.2 (383.84)
15	462.2 ± 25.4 (273.46)	497.6 ± 22.4 (390.27)	677.0 ± 53.2 (425.87)
17	500.6 ± 23.4 (237.64)	624.1 ± 47.4 (390.27)	610.2 ± 34.9 (425.87)
19	549.2 ± 84.2 (375.99)	627.8 ± 42.9 (412.52)	700.3 ± 57.3 (491.68)
21	528.0 ± 36.8 (404.17)	534.0 ± 42.9 (272.39)	570.0 ± 63.6 (474.33)
23	482.5 ± 24.0 (287.07)	576.2 ± 78.9 (509.27)	542.5 ± 47.0 (518.07)

Section S7. Implicit- versus explicit-solvent models in restrained MD simulations.

In implicit-solvent MD simulations, GB-implicit solvent models were used to mimic solvation effects. While these implicit-solvent models are useful in the simulation of big systems, they are less realistic because empirical parameters were used to calculate the solvation free energies. Thus, they cause DNA molecules to be more flexible than found in explicit-solvent MD studies: the DNA backbone can be distorted and terminal base pairs can be frayed in a non-physical manner. The backbone distortions can be attributed to the stress induced by the hydrophobic cores on the DNA duplex arms while they try to minimize their SASA (see Methods in main article for details). To observe the stacking effects of hydrophobic cores without distorting the DNA, Watson-Crick base pairing, torsional, and chirality restraints were imposed on the assemblies investigated herein to keep them in the native B-form DNA conformations. A similar procedure described for unrestrained MD simulation was followed in these MD simulations. Each dimer DNA-hybrid system was run for over 90 ns. In the DNA-hybrid systems with one organic core, only Watson-Crick base pairing and torsional restraints were used to run over 50 ns MD simulations.

DNA linked to one $SMDH_2$ molecule. Because the PAGE results for 3′- and 5′- linked nanostructures showed differences in dimer formation, we decided to focus on two simple model systems. To understand if an organic core attached at either 3′- or 5′- ends of DNA could result in different stacking properties upon assembly, we prepared two model systems where only one core was attached at either 3′- or 5′-ends of 11 bp DNA duplex (Figure S22b,c). Moreover, explicit-solvent MD simulations were utilized on each final conformation to investigate the structural stabilities of each system. Each system was neutralized with Na^+/Cl^- ions[S15] and solvated with TIP3P water molecules[S16] in a truncated octahedral box. Each final system included 7057 H_2O molecules, 31 Na^+ ions, and 10 Cl^- ions. Explicit-solvent MD simulations included minimization, equilibration, and product run sequentially as previously described.[S17] During the production runs, no restraints were imposed on DNA structures. See Figure S22 for results.

Cyclic-dimer $SMDH_2$ systems with varying DNA lengths. In our previous work,[S3] we showed that the hydrophobic cores prefer to stack either on terminal base pairs or within deoxythymidine spacers if available. Otherwise, the self-assembly of dimer nanostructures would not happen due to the hydrophobic cores not being able to reduce their SASAs without distorting the DNA backbone. The effect of DNA length on the SASA of the organic cores in cyclic $SMDH_2$ nanostructures was, therefore, of interest in this study. Consequently, we prepared dimer cyclic-[5′-**C**-3′]:[5′-**C**-3′]′, cyclic-[5′-**C**-3′]:[5′-**C**-3′]′, and cyclic-[5′-**C**-3′]:[5′-**C**-3′]′ nanostructures with varying DNA lengths (11, 13, 15, 17, 19, 21, and 23 base pairs), which totaled 21 systems. The correlation between the lengths of the DNA duplexes and the SASAs of the organic cores in dimer nanostructures was investigated. Different sequence lengths were created using 15 bp DNA sequences described in Table 1 as building blocks.

Section S8. Simulated-annealing MD simulations.

As discussed in Section S7, in the MD simulations of cyclic dimer $SMDH_2$ systems with varying lengths, Watson-Crick, torsional, and chirality restraints were employed to keep the DNA duplexes in B-form. Even though this is a useful method to probe only the effects of hydrophobicity of the organic cores, it is possible that the structures might get stuck in local minimum states. As a result, simulated-annealing MD simulations were carried out for all the 21 dimer $SMDH_2$ DNA-hybrid systems (Table S4 and Figures S24). The final structures of the restrained MD simulations of dimer $SMDH_2$ DNA-hybrid systems were used as initial structures in the simulated-

annealing MD simulations where Watson-Crick base pairing, torsional, and chirality restraints were imposed on DNA duplexes again to keep them in B-form conformations during the MD simulations. Temperature was increased from 0 to 3000 K in 5000 steps, while it was cooled down to 100 K gradually within the next 95000 steps with a 2 fs time step. For each system, 301 simulated-annealing MD simulations were run continuously where the starting structure for each run was taken from the final structure of the previous run.

As shown in Figure 8, the total SASA values for the organic cores in the [5′-***C***-3′]:[5′-***C***-3′]′ (red) cyclic dimers (black) are in between those for the [3′-***C***-3′]:[3′-***C***-3′]′ and [5′-***C***-5′]:[5′-***C***-5′]′ cyclic dimers (green); the two exceptions are for the dimers with 11 and 21 bp. Computationally, as the length of DNA becomes too long, as in the case for the DNA duplexes with 21 bp, it is harder to completely sample the phase space for the dimer even with simulated-annealing MD simulations. In addition, as the length of the DNA duplexes increases, the flexibility of the whole $SMDH_2$ system increases and it is possible that the 3′-linked organic cores in some cyclic-[5′-***C***-3′]:[5′-***C***-3′]′ system can be inserted into the minor grooves of the DNA duplexes (Figure S24, 21 bp). Finally, the off-trend points at 11 and 21 bp could also be attributed to the 10 bp periodicity in B-form DNA structures. Yet, the properties of these complex $SMDH_2$ dimers, such as hydrophobicity of organic cores, linkage type, and flexibility of DNA duplexes with respect to sequence length, are so entangled with each other that it is not simple to elucidate the effects of these individual terms simply based on the SASA of the organic cores.

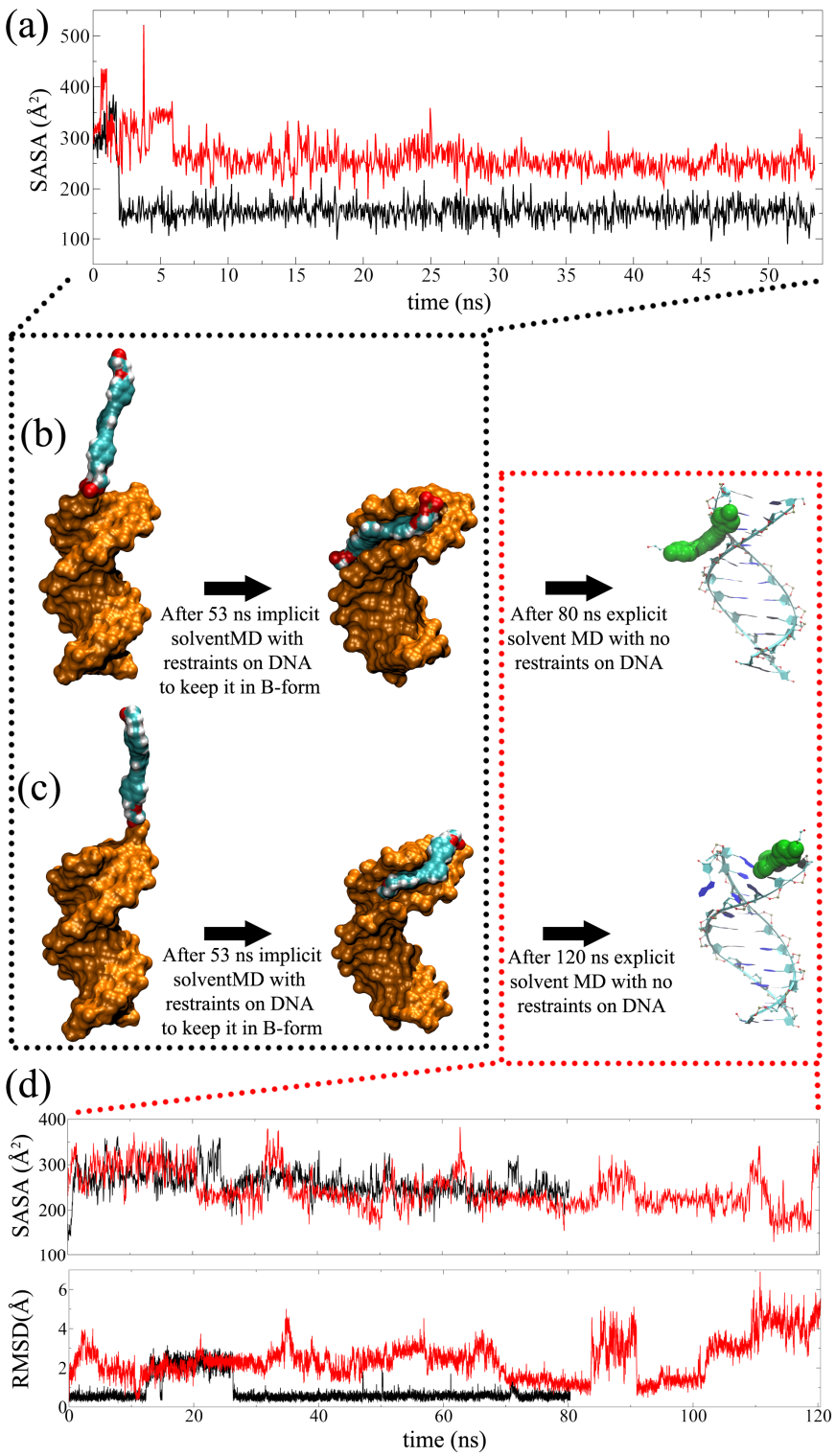


Figure S22. Computational results for 3′- and 5′-linked single core-DNA systems. In (a), SASA values of hydrophobic cores are shown with respect to time for 3′- and 5′-linked core-DNA systems in black and red, respectively, as extracted from implicit-solvent MD simulations. Initial and final structures of 3′- and 5′-linked core-DNA systems are shown in (b) and (c), respectively. The final structures (b and c) of implicit-solvent MD simulations (highlighted within dashed black rectangle) were then solvated with water and Na^+/Cl^- ions to investigate the structural stabilities of these conformations in explicit solvents. In (d), SASA values of the hydrophobic cores and the rmsd of terminal DNA base pairs are shown with respect to time for 3′- and 5′-linked core-DNA systems in black and red, respectively, as extracted from explicit-solvent MD simulations. The 3′-linked core-DNA system appears to be stable while the 5′-linked core-DNA system appears to be fluxional. The final structures of 3′- and 5′-linked core-DNA systems in explicit-solvent MD simulations are highlighted within dashed red rectangle for (b) and (c). Interestingly, the terminal basepair of the 5′-linked core-DNA system is fully lost. See main text for details.

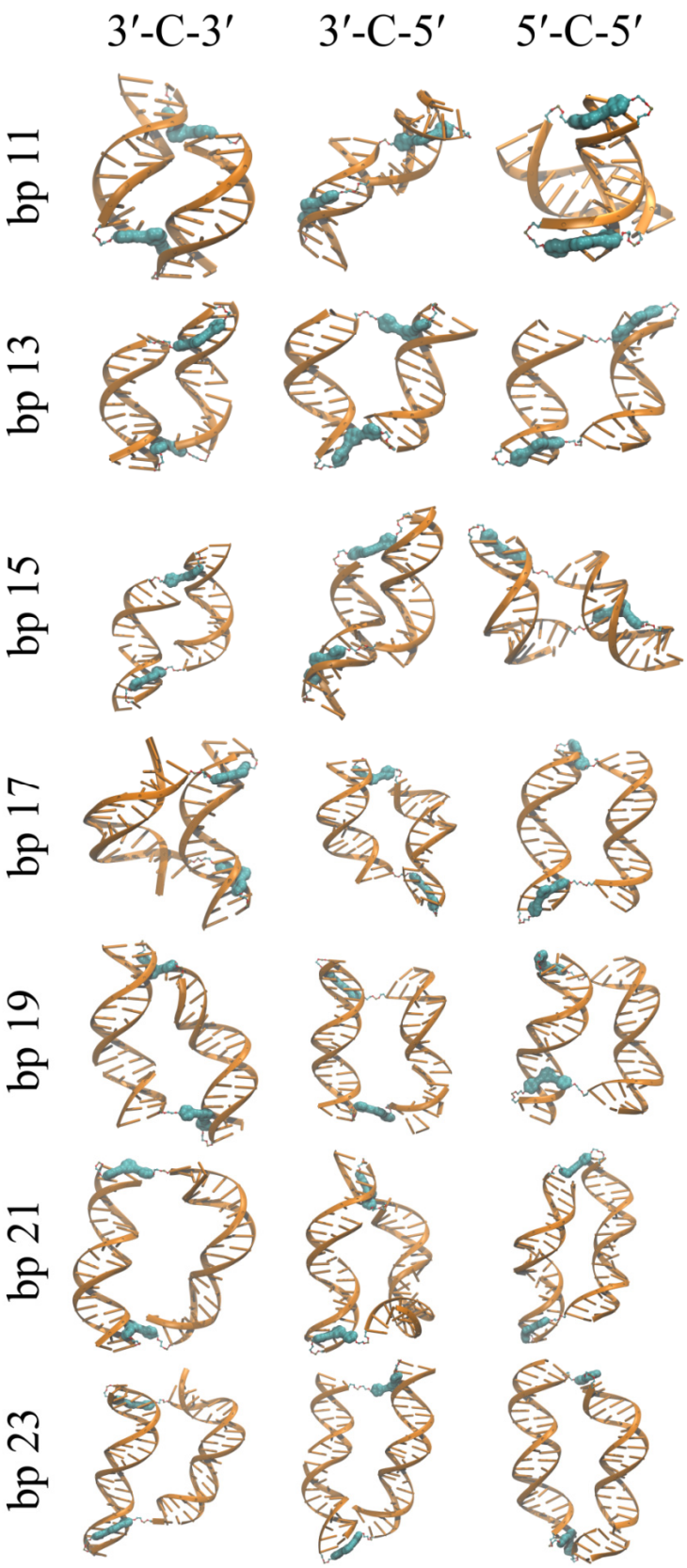


Figure S23. Cyclic-dimer $SMDH_2$ systems with lowest SASA values extracted from simulated-annealing MD simulations. Structures have final restraint energies and duplex rmsd values less than 10 kcal/mol and 10 Å, respectively.

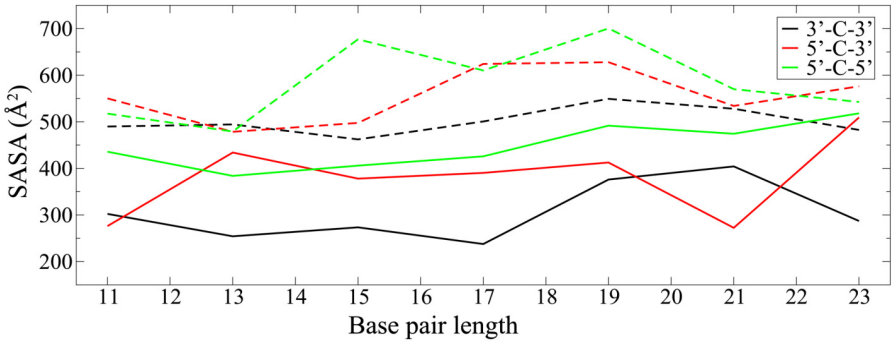


Figure S24. Plots of SASAs of the hydrophobic cores as a function of the DNA duplex length in cyclic-dimer DNA-hybrid systems; data were extracted from restrained MD simulations (Table 4). The black, red, and green profiles represent the results for cyclic dimers in [3′-***C***-3′]:[3′-***C***-3′]′, [5′-***C***-3′]:[5′-***C***-3′]′, and [5′-***C***-5′]:[5′-***C***-5′]′, respectively. The dashed lines are the results extracted from restrained MD simulations while the solid lines represent the lowest SASA results for the simulated-annealing MD simulations with final restraint energies and duplex rmsd values less than 10 kcal/mol and 10 Å, respectively. Note that the SASA values of cyclic-dimers in [3′-***C***-3′]:[3′-***C***-3′]′ systems are in general lower than the other two cases.

Table S4. Detailed simulated-annealing results of the lowest SASA structures with final restraint energies less than 10 kcal/mol and duplex rmsd values less than 10 Å.

Base pair Length	**System type**	**SASA (Å²)**	**Final Restraint Energy (kcal/mol)**	**Duplex 1 RMSD (Å)**	**Duplex 2 RMSD (Å)**
bp_11	[3′-***C***-3′]:[3′-***C***-3′]′	302.28	5.1	1.8	3.5
	[5′-***C***-3′]:[5′-***C***-3′]′	276.16	3.5	1.3	4.1
	[5′-***C***-5′]:[5′-***C***-5′]′	435.76	4.0	3.1	4.5
bp_13	[3′-***C***-3′]:[3′-***C***-3′]′	254.21	5.9	2.9	2.3
	[5′-***C***-3′]:[5′-***C***-3′]′	433.95	7.1	2.5	3.8
	[5′-***C***-5′]:[5′-***C***-5′]′	383.84	5.3	4.2	3.0
bp_15	[3′-***C***-3′]:[3′-***C***-3′]′	273.46	3.9	5.1	3.6
	[5′-***C***-3′]:[5′-***C***-3′]′	378.00	7.7	3.1	3.5
	[5′-***C***-5′]:[5′-***C***-5′]′	405.86	8.1	7.2	6.5
bp_17	[3′-***C***-3′]:[3′-***C***-3′]′	237.64	8.0	4.0	9.9
	[5′-***C***-3′]:[5′-***C***-3′]′	390.27	8.4	6.5	6.1
	[5′-***C***-5′]:[5′-***C***-5′]′	425.87	5.7	3.2	4.1
bp_19	[3′-***C***-3′]:[3′-***C***-3′]′	375.99	7.7	6.2	5.7
	[5′-***C***-3′]:[5′-***C***-3′]′	412.52	7.8	3.2	6.0
	[5′-***C***-5′]:[5′-***C***-5′]′	491.68	6.7	6.0	5.3
bp_21	[3′-***C***-3′]:[3′-***C***-3′]′	404.17	6.2	4.8	7.8
	[5′-***C***-3′]:[5′-***C***-3′]′	272.39	4.7	4.7	8.4
	[5′-***C***-5′]:[5′-***C***-5′]′	474.33	7.7	5.5	4.4
bp_23	[3′-***C***-3′]:[3′-***C***-3′]′	287.07	8.1	5.1	9.3
	[5′-***C***-3′]:[5′-***C***-3′]′	509.27	9.3	4.6	4.7
	[5′-***C***-5′]:[5′-***C***-5′]′	518.07	8.7	6.1	6.6

Table S5. Stability analysis of the structures shown in Figure S23. MD simulations, at >125 ns each, were run on each system with no restraints imposed on DNA molecules. The RMSD values, with respect to the B-form of each duplex in each system, are mostly >5 Å. This is due to the shortcomings of implicit-solvent models developed for nucleic acids. See main text for details.

Base pair Length	System type	SASA (Å^2)	# of H-bond	% of H-bonds lost	Duplex 1 RMSD (Å)	Duplex 2 RMSD (Å)
bp_11	[3′-***C***-3′]:[3′-***C***-3′]′	300.2 ± 22.7	40 ± 3	0.20	5.4 ± 0.7	6.1 ± 0.3
	[5′-***C***-3′]:[5′-***C***-3′]′	287.8 ± 37.2	45 ± 2	0.10	5.4 ± 0.7	6.1 ± 0.3
	[5′-***C***-5′]:[5′-***C***-5′]′	361.6 ± 50.7	44 ± 2	0.12	4.6 ± 0.5	8.2 ± 0.6
bp_13	[3′-***C***-3′]:[3′-***C***-3′]′	306.8 ± 25.8	52 ± 3	0.10	6.2 ± 0.6	5.5 ± 0.6
	[5′-***C***-3′]:[5′-***C***-3′]′	265.6 ± 66.2	45 ± 3	0.22	5.0 ± 0.7	7.0 ± 0.8
	[5′-***C***-5′]:[5′-***C***-5′]′	232.7 ± 49.6	50 ± 2	0.14	5.7 ± 0.6	5.6 ± 0.7
bp_15	[3′-***C***-3′]:[3′-***C***-3′]′	307.4 ± 61.6	55 ± 3	0.17	6.0 ± 0.8	5.8 ± 0.7
	[5′-***C***-3′]:[5′-***C***-3′]′	439.1 ± 50.4	52 ± 2	0.21	4.6 ± 0.7	9.3 ± 0.7
	[5′-***C***-5′]:[5′-***C***-5′]′	334.9 ± 58.2	51 ± 4	0.23	10.8 ± 0.7	7.7 ± 0.5
bp_17	[3′-***C***-3′]:[3′-***C***-3′]′	412.1 ± 55.9	62 ± 2	0.16	6.2 ± 0.6	7.4 ± 0.7
	[5′-***C***-3′]:[5′-***C***-3′]′	360.8 ± 35.2	61 ± 3	0.18	6.5 ± 0.7	7.2 ± 0.9
	[5′-***C***-5′]:[5′-***C***-5′]′	250.7 ± 41.1	47 ± 4	0.36	9.3 ± 0.5	37.4 ± 2.1
bp_19	[3′-***C***-3′]:[3′-***C***-3′]′	388.0 ± 61.2	68 ± 4	0.21	6.3 ± 1.1	6.3 ± 0.9
	[5′-***C***-3′]:[5′-***C***-3′]′	266.9 ± 39.9	73 ± 3	0.15	5.6 ± 0.7	11.8 ± 0.7
	[5′-***C***-5′]:[5′-***C***-5′]′	229.6 ± 26.0	67 ± 4	0.22	11.9 ± 1.0	12.4 ± 0.6
bp_21	[3′-***C***-3′]:[3′-***C***-3′]′	238.6 ± 42.1	69 ± 4	0.27	8.7 ± 0.8	22.7 ± 2.6
	[5′-***C***-3′]:[5′-***C***-3′]′	271.9 ± 31.7	70 ± 3	0.26	7.5 ± 1.0	11.0 ± 1.0
	[5′-***C***-5′]:[5′-***C***-5′]′	217.2 ± 50.1	70 ± 4	0.26	13.4 ± 0.8	12.6 ± 0.9
bp_23	[3′-***C***-3′]:[3′-***C***-3′]′	248.7 ± 35.1	74 ± 4	0.27	6.6 ± 0.9	27.1 ± 1.1
	[5′-***C***-3′]:[5′-***C***-3′]′	182.1 ± 51.0	70 ± 3	0.31	7.2 ± 0.8	12.3 ± 0.5
	[5′-***C***-5′]:[5′-***C***-5′]′	203.3 ± 49.6	76 ± 4	0.25	8.9 ± 1.0	17.0 ± 2.1

Section S9. Unrestrained MD simulation results.

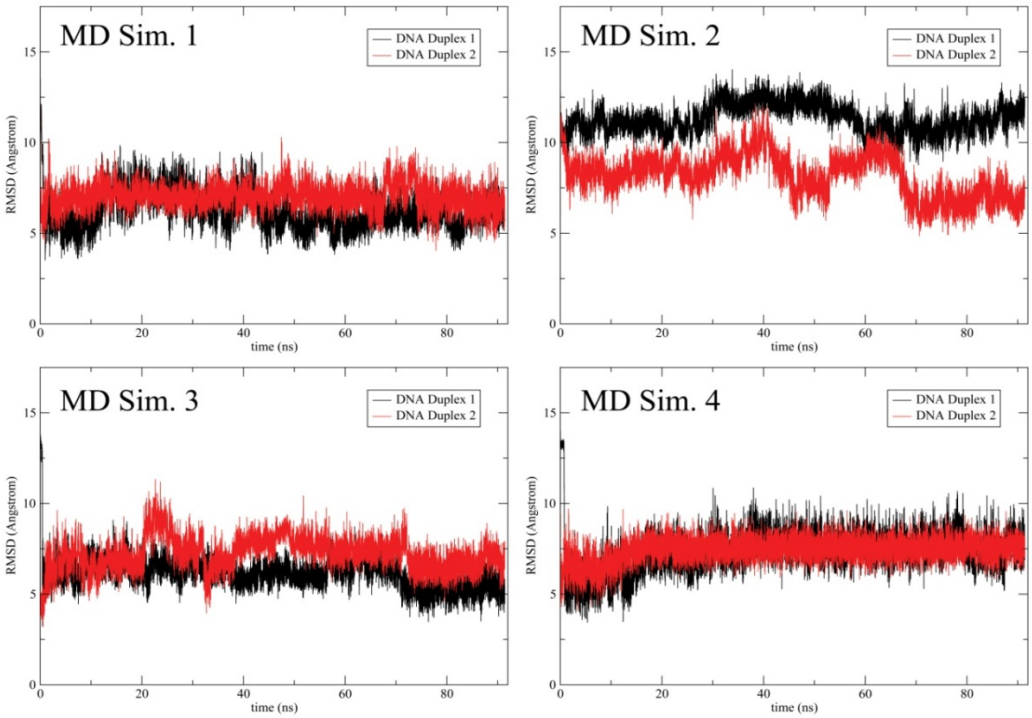


Figure S25. MD simulation results for cyclic dimer-[3′-***C***-3′]:[3′-***C***-3′]′ (30 bp): the RMSD of heavy atoms in both duplexes were plotted as functions of time.

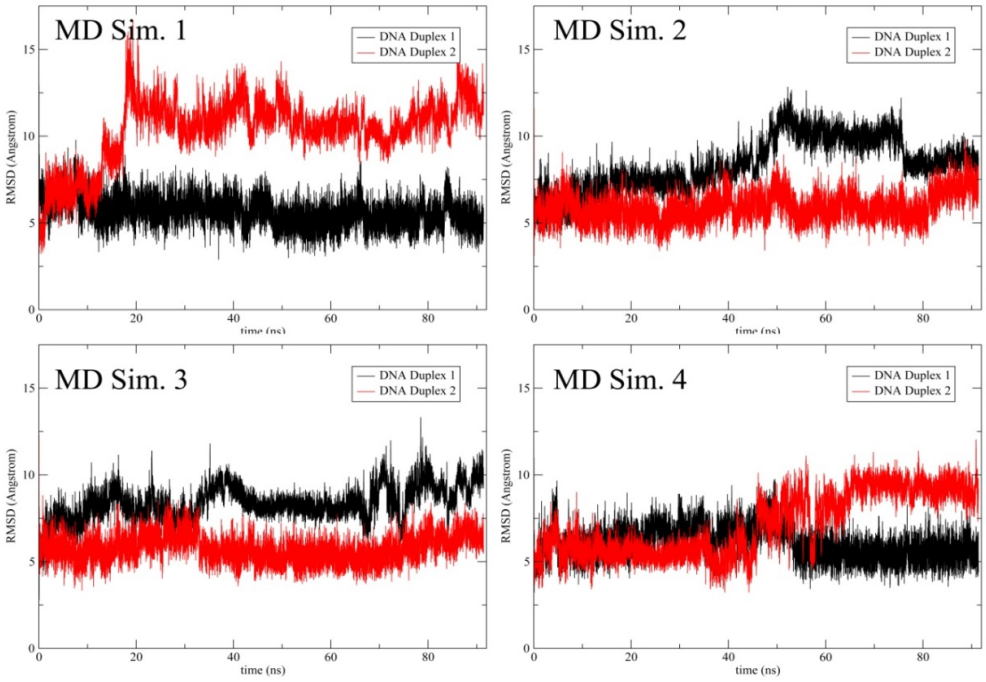


Figure S26. MD simulation results for cyclic dimer-[5′-***C***-3′]:[5′-***C***-3′]′ (30 bp): the RMSD of heavy atoms in both duplexes were plotted as functions of time.

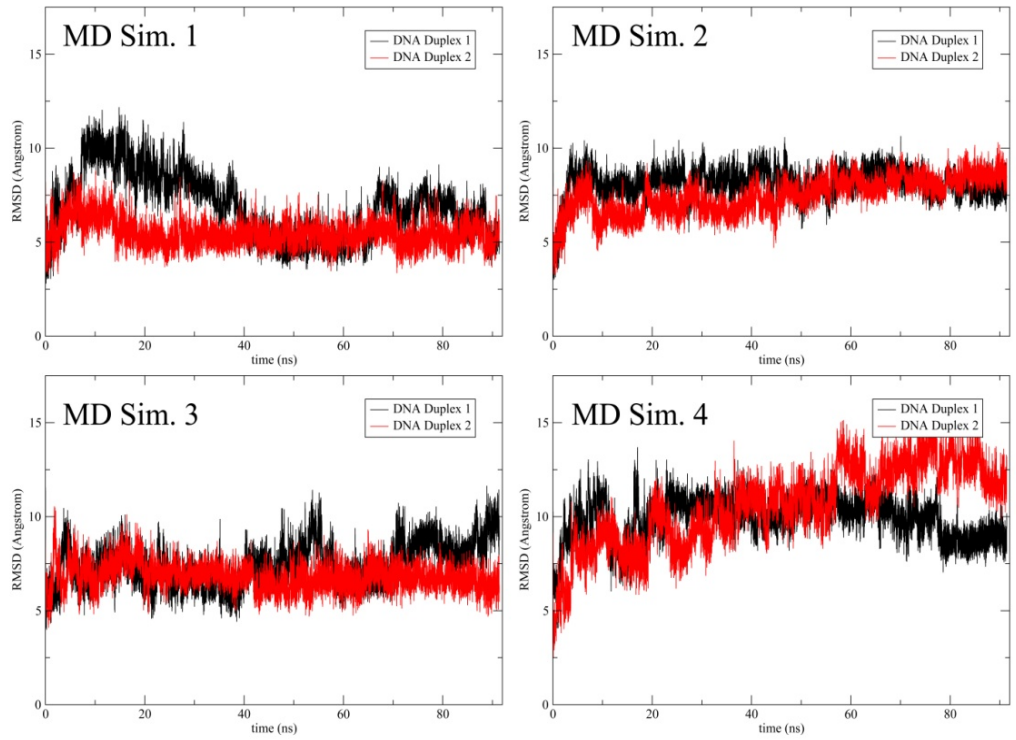


Figure S27. MD simulation results for cyclic dimer-[5′-***C***-5′]:[5′-***C***-5′]′ (30 bp): the RMSD of heavy atoms in both duplexes were plotted as functions of time.

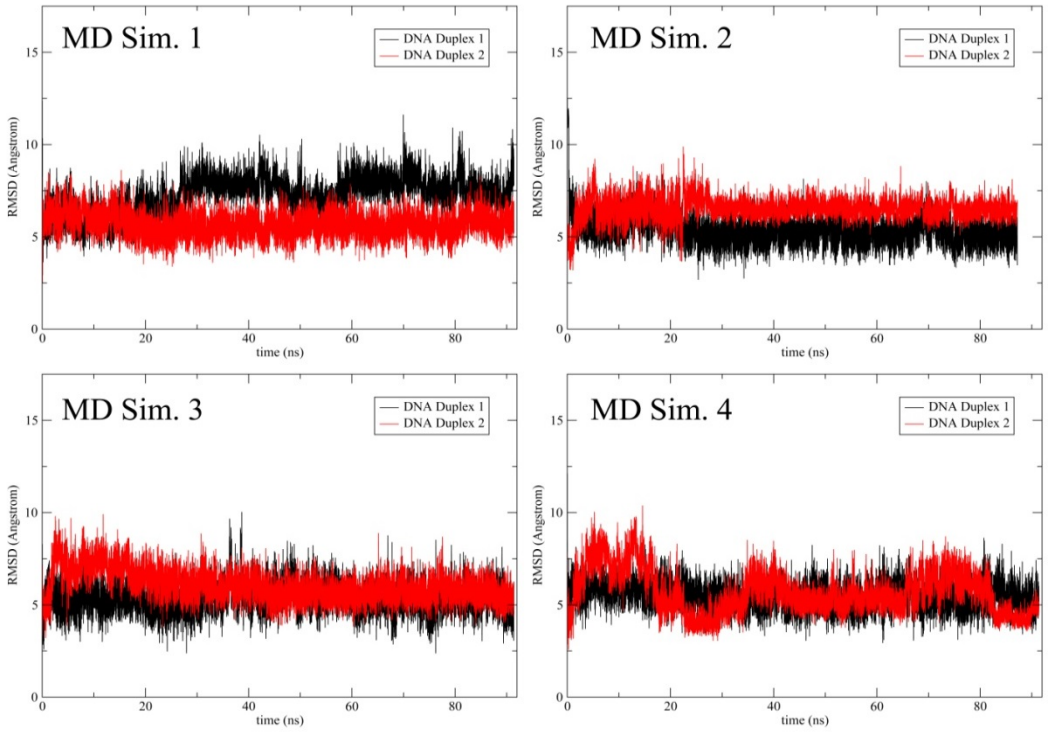


Figure S28. MD simulation results for cyclic dimer-[5′-***T₃CT₃***-3′]:[5′-***T₃CT₃***-3′]′ (30 bp): the RMSD of heavy atoms in both duplexes were plotted as functions of time.

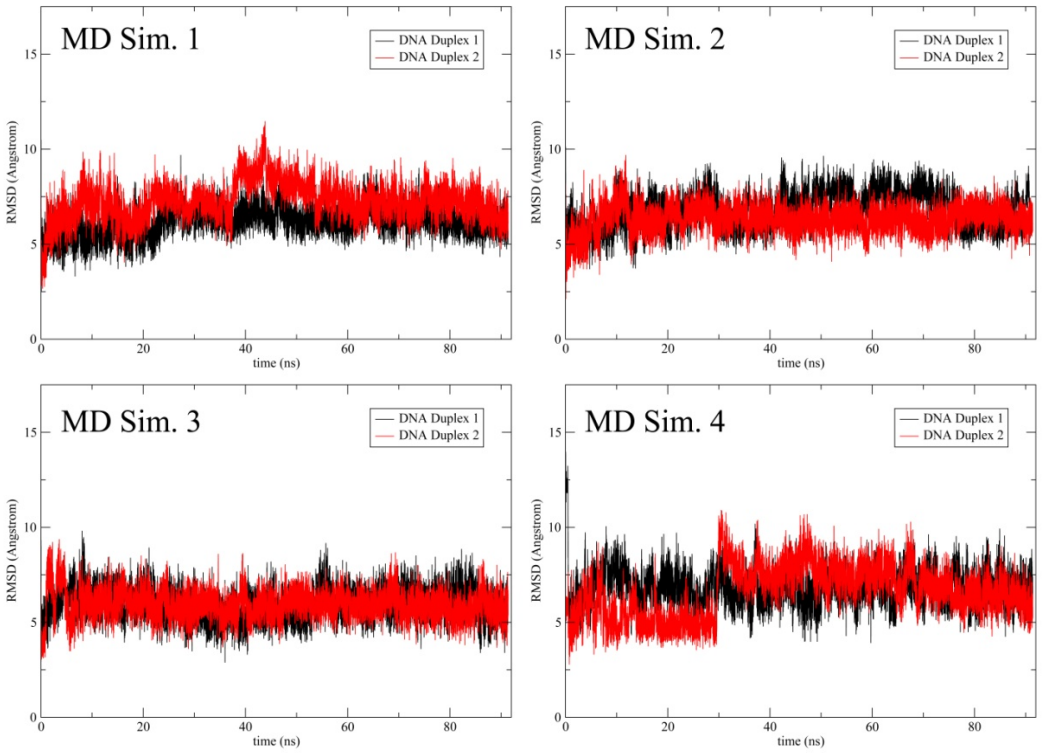


Figure S29. MD simulation results for cyclic dimer-[5′-***T*$_6$-3′]:[5′-***T*$_6$-3′]′ (30 bp): the RMSD of heavy atoms in both duplexes were plotted as functions of time.

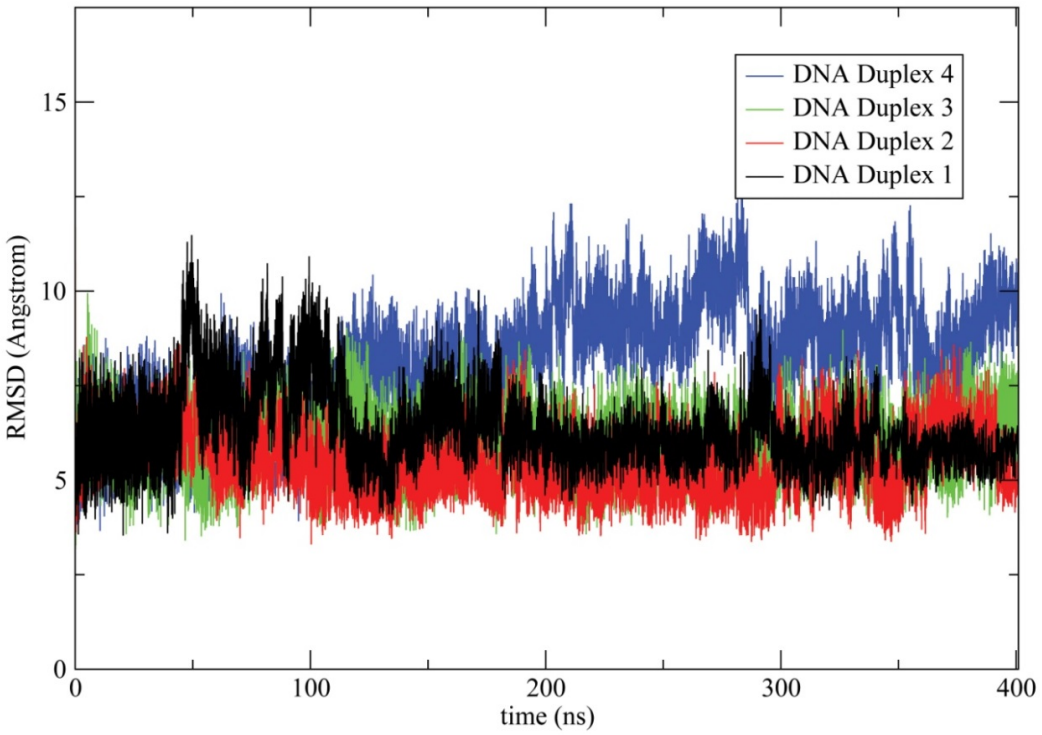


Figure S30. MD simulation results for cyclic tetramer-[5′-***C*-3′]:[5′-***C*-3′]′ (60 bp): the RMSD of heavy atoms in all four duplexes were plotted as functions of time.

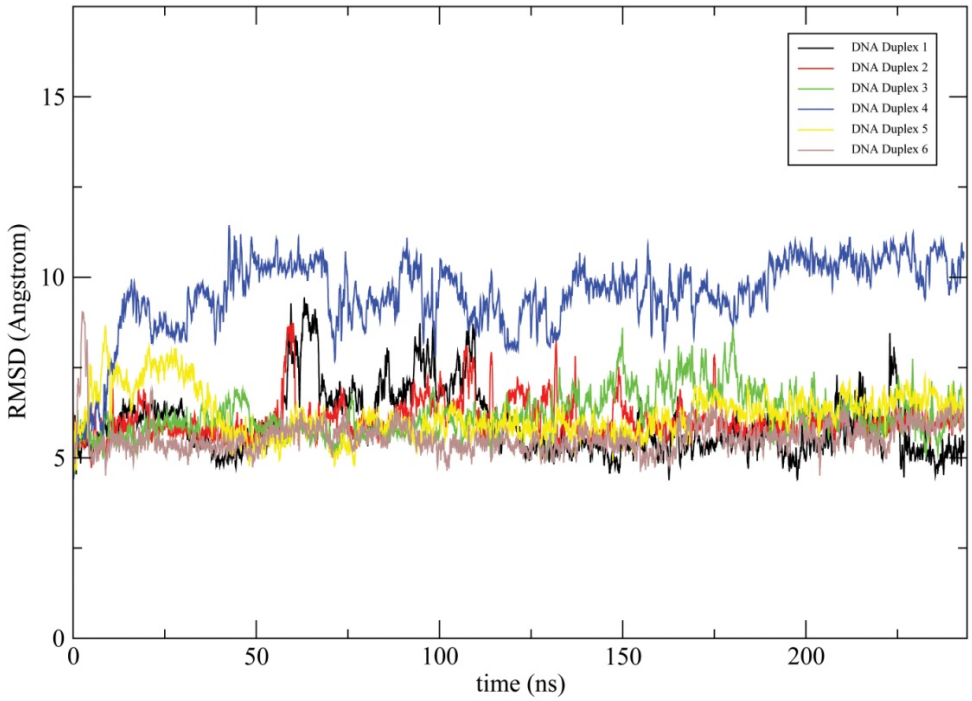


Figure S31. MD simulation results for cyclic hexamer-[5′-***C***-3′]:[5′-***C***-3′]′ (90 bp) the RMSD of heavy atoms in all four duplexes were plotted as functions of time.

Section S10. Movie description

Movie S1. MD simulation trajectory (first 12 ns) of the control-[5′-C-3′] system with respect to time (Figure 3b). Simulation time is shown in the movie to emphasize the structural change happening in this system. Note that the initial structure of this system is linear and the organic core is exposed to water. As time progresses, the core inserts itself into the minor groove of the left DNA duplex. Later on, the core orients itself in such a way that it is sandwiched between the minor groove of the left DNA duplex and the terminal base pair of the right DNA duplex. In such manners, the SASA of the core is minimized.

Section S11. Full author list for references 50, 51, and 52.

50) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S. R.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A. *AMBER 9* **2006**, University of California, San Francisco.

(51) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Crowley, M.; Walker, R. C.; Zhang, W.; Merz, K. M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Kollman, P. A. *AMBER 10* **2008**, University of California, San Francisco.

(52) Case, D. A.; Darden, T. A.; Cheatham III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 11* **2010**, University of California, San Francisco.

Section S12. References.

(S1) Jessen, C. H.; Pedersen, E. B. *Helv. Chim. Acta* **2004**, *87*, 2465.
(S2) Eryazici, I.; Prytkova, T. R.; Schatz, G. C.; Nguyen, S. T. *J. Am. Chem. Soc.* **2010**, *132*, 17068.
(S3) Eryazici, I.; Yildirim, I.; Schatz, G. C.; Nguyen, S. T. *J. Am. Chem. Soc.* **2012**, *134*, 7450.
(S4) Shao, Y.; Molnar, L. F.; Jung, Y.; Kussmann, J.; Ochsenfeld, C.; Brown, S. T.; Gilbert, A. T. B.; Slipchenko, L. V.; Levchenko, S. V.; O'Neill, D. P.; DiStasio, R. A.; Lochan, R. C.; Wang, T.; Beran, G. J. O.; Besley, N. A.; Herbert, J. M.; Lin, C. Y.; Van Voorhis, T.; Chien, S. H.; Sodt, A.; Steele, R. P.; Rassolov, V. A.; Maslen, P. E.; Korambath, P. P.; Adamson, R. D.; Austin, B.; Baker, J.; Byrd, E. F. C.; Dachsel, H.; Doerksen, R. J.; Dreuw, A.; Dunietz, B. D.; Dutoi, A. D.; Furlani, T. R.; Gwaltney, S. R.; Heyden, A.; Hirata, S.; Hsu, C. P.; Kedziora, G.; Khalliulin, R. Z.; Klunzinger, P.; Lee, A. M.; Lee, M. S.; Liang, W.; Lotan, I.; Nair, N.; Peters, B.; Proynov, E. I.; Pieniazek, P. A.; Rhee, Y. M.; Ritchie, J.; Rosta, E.; Sherrill, C. D.; Simmonett, A. C.; Subotnik, J. E.; Woodcock, H. L.; Zhang, W.; Bell, A. T.; Chakraborty, A. K.; Chipman, D. M.; Keil, F. J.; Warshel, A.; Hehre, W. J.; Schaefer, H. F.; Kong, J.; Krylov, A. I.; Gill, P. M. W.; Head-Gordon, M. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3172.
(S5) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269.
(S6) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Pearlman, D. A.; Crowley, M.; Walker, R. C.; Zhang, W.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Wong, K. F.; Paesani, F.; Wu, X.; Brozell, S.; Tsui, V.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Mathews, D. H.; Schafmeister, C.; Ross, W. S.; Kollman, P. A.; San Francisco, CA: University of California San Francisco, 2006.
(S7) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(S8) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049.
(S9) Yildirim, I.; Stern, H. A.; Kennedy, S. D.; Tubbs, J. D.; Turner, D. H. *J. Chem. Theory Comput.* **2010**, *6*, 1520.
(S10) Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E.; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817.
(S11) 4, C. C. P. N. *Acta Crystallogr., Sect. D* **1994**, *50*, 760.
(S12) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A.; San Francisco, CA: University of California San Francisco, 2010.
(S13) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122.
(S14) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(S15) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020.
(S16) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(S17) Yildirim, I.; Park, H.; Disney, M. D.; Schatz, G. C. *J. Am. Chem. Soc.* **2013**, *135*, 3528.