01 Mar 2013

Site-Specific Inter-Strand Cross-Links of DNA Duplexes

Miao Ye

Johan Guillame

Yu Liu

Ruojie Sha

et. al. For a complete list of authors, see https://scholarsmine.mst.edu/chem_facwork/2685

Recommended Citation

M. Ye et al., "Site-Specific Inter-Strand Cross-Links of DNA Duplexes," *Chemical Science*, vol. 4, no. 3, pp. 1319-1329, Royal Society of Chemistry, Mar 2013.
The definitive version is available at https://doi.org/10.1039/c2sc21775a

Chemical Science

RSC Publishing

EDGE ARTICLE

View Article Online
View Journal | View Issue

Site-specific inter-strand cross-links of DNA duplexes†

Cite this: *Chem. Sci.*, 2013, **4**, 1319

Miao Ye, Johan Guillaume, Yu Liu,‡ Ruojie Sha, Risheng Wang, Nadrian C. Seeman* and James W. Canary*

Received 17th October 2012
Accepted 19th December 2012

DOI: 10.1039/c2sc21775a

www.rsc.org/chemicalscience

We report the development of technology that allows inter-strand coupling across various positions within one turn of DNA. Four 2′-modified nucleotides were synthesized as protected phosphoramidites and incorporated into DNA oligonucleotides. The modified nucleotides contain either 5-atom or 16-atom linker components, with either amine or carboxylic acid functional groups at their termini, forming 10 or 32 atom (11 or 33 bond) linkages. Chemical coupling of the amine and carboxylate groups in designed strands resulted in the formation of an amide bond. Coupling efficiency as a function of trajectory distance between the individual linker components was examined. For those nucleotides capable of forming inter-strand cross-links (ICLs), coupling yields were found to depend on temperature, distance, and linker length, enabling several approaches that can control regioselective linkage. In the most favorable cases, the coupling yields are quantitative. Spectroscopic measurements of strands that were chemically cross-linked indicate that the global structure of the DNA duplex does not appear to be distorted from the B form after coupling. Thermal denaturing profiles of those strands were shifted to somewhat higher temperatures than those of their respective control duplexes. Thus, the robust amide ICLs formed by this approach are site-specific, do not destabilize the rest of the duplex, and only minimally perturb the secondary structure.

Introduction

The two strands of the DNA double helix can be separated readily because they are connected by hydrogen-bonded interactions, rather than by covalent bonds.[1] Nevertheless, there are a number of reasons to cross-link the two strands covalently. Examples include (1) to maintain 3D structures associated with functionalities,[2] (2) to enhance the stability of DNA nanostructures,[3,4] and (3) to examine features of DNA metabolism, such as repair.[5] As events closely related to cell death, cross-linking phenomena involved in cancer therapy have been studied intensely.[5,6] Consequently, the goal of this work is to connect specific nucleotides on opposite strands of nucleic acid double helices *via* organic moieties, which produce covalent bonds in a systematic regioselective fashion.

Cross-linking can be effected by an array of different chemistries. Inter-strand cross-links (ICLs) have been introduced into unmodified DNA by abasic sites,[7] by oxidants,[8] by chemical additives including drugs,[9] and by low pH.[10] ICLs can also form in modified DNA by reactions such as enzymatic restriction and ligation,[11] phosphate backbone linking,[12] disulfide bond formation,[13] *in situ* oxidation of furanyl modifications,[14] photocleavage induced abasic sites[15] and "click chemistry" such as the Cu(I) catalysed [3 + 2] azide–alkyne cycloaddition (CuAAC) reaction.[16,17] Most cross-linking strategies have involved linkages over short distances, such as between neighboring base pairs. ICLs over greater distances have also been reported.[16,18]

Although several groups have used different methods to introduce ICLs to fixed distances such as adjacent base pairs,[19,20] one base pair apart,[21] and two base pairs apart,[22] most current methods do not readily allow for the coupling of specific nucleotides in a nucleic acid structure, especially at varying distances. For example, psoralen is one of the most frequently used agents for cross-linking.[3,4,23] However, with psoralen, reactivity is dependent on the proper placement of pyrimidine nucleotides and it can be difficult to control the amount of cross-linking. Chemotherapeutic alkylating agents such as cisplatin[5,24] or nitrogen mustards[5] react regioselectively with the nucleophiles present in nucleic acids, but such complexes may be limited to studies that do not require homogenous or regiospecific cross-links.

In addition to biological processes, such modifications are certainly important in DNA nanotechnology,[3,4,25] where it becomes key to develop nucleic acid analogs with modifications that bring precise spatial control of chemical reactivity, or that allow them to assume functions that are not associated with a

Department of Chemistry, New York University, New York, NY 10003, USA. E-mail: ned.seeman@nyu.edu; james.canary@nyu.edu; Fax: +1 212 995 4367; Tel: +1 212 998 8422

† Electronic supplementary information (ESI) available: Synthesis of uridine phosphoramidite for U_N^{16}; Time course studies of N−1 to N−6 and N+1 to N+4 coupling reactions; MALDI-TOF data for coupled products of 32-atom linkers; Melting curves of uncoupled N−1 duplex, singly modified duplex with U_C or U_N modified nucleotides. See DOI: 10.1039/c2sc21775a

‡ Present address: Department of Chemistry, Northern Michigan University, Marquette, MI 49855, USA

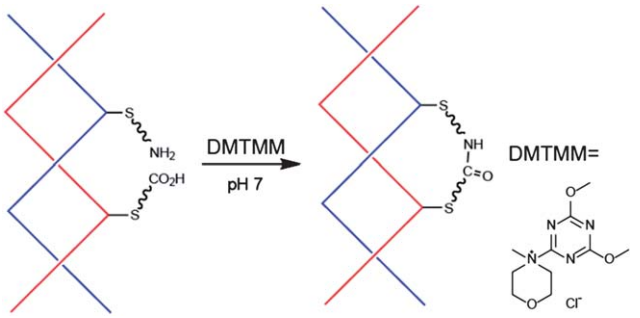


Fig. 1 Inter-strand cross-linking of DNA by amide bond formation, mediated by DMTMM. The major groove is not differentiated from the minor groove in this diagram, which represents coupling across either of the grooves schematically.

biological DNA context. Locking nucleic acid molecules (Fig. 1) into specific conformations would enable the formation of novel DNA motifs, structures and topologies, particularly through hierarchical approaches.

Previously, we reported the addition of 2′-alkylamine or 2′-alkylcarboxylate groups to nucleotides where the 2′-oxygen of a ribonucleotide had been replaced by a sulfur atom.[26–29] In these reports, we incorporated special nucleotides into DNA molecules and joined the derivatized nucleotides together by forming amide bonds in a reaction using the chemical coupling reagent DMTMM (4-(4,6-dimethoxy-1,3,5-triazin-2-yl)-4-methylmorpholinium chloride) in aqueous solution. We have also used this methodology to ligate DNA strands by 2′,2′ coupling.[27] Additionally, we reported the formation of novel DNA catenanes by forming an intra-strand cross-link across a full turn of double helix.[30] In this report, we have inserted special nucleotides containing these or similar modifications on opposite strands of the double helix. From Fig. 1, it is evident that this approach can join the two complementary strands of a double helical DNA molecule to each other by a covalent amide linkage. In principle, the linkage is capable of cross-linking strands in a simple double helix or in a more complex DNA motif. Here, we report a systematic study of amide-based inter-strand cross-linking in double helices at various distances with different linker lengths.

Results and discussion

We prepared 2′-deoxy-2′-alkylmercaptophosphoramidite analogs of uridine-containing linker component groups of five atoms or sixteen atoms. The reagents were introduced into DNA strands using standard phosphoramidite chemistry[31] to yield DNA molecules containing these modified nucleotides (Fig. 2). The nomenclature of the compounds is such that the superscript refers to the length of the linker component from the 2′-S atom to the nitrogen atom or carboxylate carbon atom at the termini; the subscript refers to the nature of the reactive chemical groups at the end of the linker component. N represents an amine group and C represents a carboxylate group. The modified nucleotides with 5-atom linker components have aliphatic side chains that are terminated with an amine group, U^5_N, or a carboxylate group, U^5_C. Compounds with 16-atom

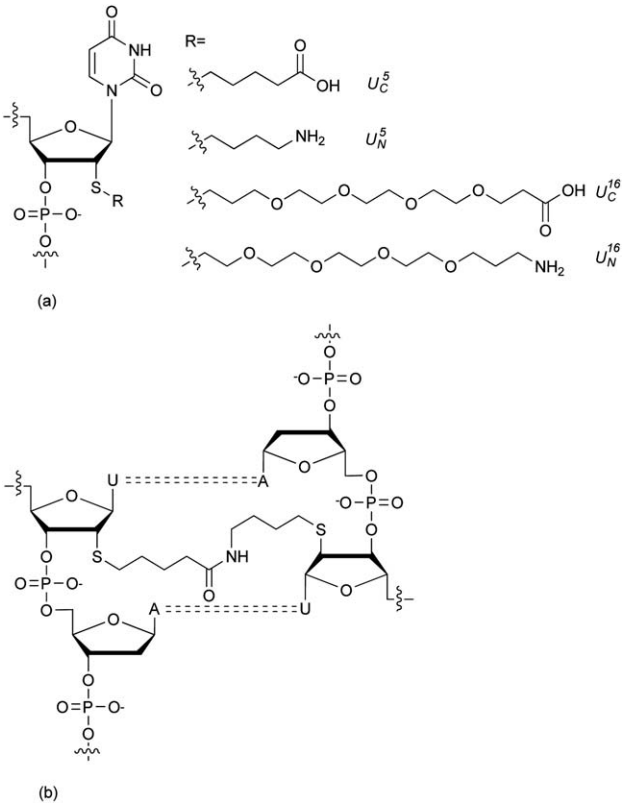


Fig. 2 (a) The chemical structures of modified nucleotides used to cross-link DNA strands. (b) A duplex crosslinked using U^5_N and U^5_C modified nucleotides separated by one base pair.

polyethylene glycol (PEG) linker components are designated U^{16}_N (amine terminus) or U^{16}_C (carboxylate). The syntheses of U^5_N, U^5_C, and U^{16}_C were reported previously.[26,30] The synthesis of U^{16}_N phosphoramidite followed a route similar to earlier compounds and is described in the ESI.† After incorporation into nucleic acid strands, the monomers were coupled *via* amide bond formation to produce ICLs. Two 5-atom linker components form a 10-atom linker (from 2′-S of one nucleotide to 2′-S of the other) and two 16-atom linker components form a 32-atom linker. From structural point of view, 10-atom linkers consist of 11 bonds from 2′-S to 2′-S and 32-atom linkers consist of 33 bonds.

As noted above, our goal is to develop and study methods to cross-link double helical DNA strands with high specificity. An ICL is one where a connection is made between one strand and its complement, as opposed to a connection within the same strand. The distance that is spanned by a specific cross-link is determined not only by the number of nucleotides that lie between the two modified nucleotides but also by the position of the second modified nucleotide relative to the first and the path that the connector must follow to reach it.

We choose the N′ position on the reference strand to place a modified nucleotide, then use its corresponding nucleotide N on the complementary strand as the starting position. (Fig. 3) The stereochemistry of a double stranded nucleic acid molecule is such that ICLs can be produced between N′ and nucleotides in either a positive (N+*i*) direction or a negative (N−*i*) direction

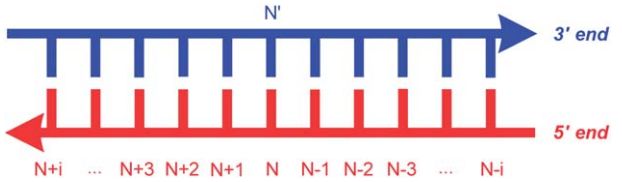


Fig. 3 Nomenclature for two-nucleotide inter-strand cross-links (ICLs). Cross-links are formed between N′ and N ± *i*. To determine the location of a cross-link in relation to a nucleotide N, both strands in the duplex are read from the 5′ end toward the 3′ end. Cross-links formed between N′ and a nucleotide that is before N are assigned negative values and those after N are assigned positive values.

from the starting nucleotide N. Our notation uses the standard convention that strands progress from the 5′ end to the 3′ end. From N, any nucleotide positioned towards its 5′ end is assigned a negative value; these sites are labelled as N−1, N−2, N−3, and so on. The nucleotides that are positioned towards the 3′ end of N are assigned a positive value, such as N+1, N+2, *etc.*

In this study, U^5_N nucleotides are at a fixed position on the reference strand (N′), U^5_C nucleotides are located on complementary strands with systematic variation in separation in both the 5′ and 3′ directions (N ± *i*). Similarly, U^{16}_C nucleotides are fixed in one position on the reference strand, and U^{16}_N nucleotides are located in the 5′ direction of the complementary strands. To simplify our nomenclature and maintain consistency, we also use "N+*i*" and "N−*i*" to denote reactions and products of certain linkages. For example, N−1 can mean both the reaction between N and N−1 or the cross-linked product strand after the reaction. The designation "*i*" is a number varying from 1 to 7, denoting the separation in terms of number of nucleotides. Similar nomenclature such as N−6(L), N−7(L) and N−8(L) are given to reactions and products involving the longer linkers.

After DNA strands containing modified nucleotides were synthesized, they were purified by reverse phase HPLC and analyzed by MALDI-TOF mass spectrometry. The sequences and mass data for the oligonucleotides used in this study are listed in Table 1. The sequences were chosen to vary nucleotides evenly so as to avoid local distortions caused by the accumulation of purines or pyrimidines. Modified nucleotides were positioned near the centers of the duplexes to minimize potential impact from thermal motion at the ends. The two complementary strands were of differing length (18 and 16 nucleotides each) so that they could be monitored separately by denaturing PAGE. To form ICLs, complementary strands containing amine- and carboxylic acid-modified nucleotides were annealed to form double helices first, and were then coupled to form amide bonds, as shown in Fig. 1. The cross-linked products were purified using gel electrophoresis and then analyzed by MALDI-TOF (Tables 2 and 3).

We conducted N−*i* cross-linking reactions at both 4 °C and 20 °C to explore the temperature dependence of reactions in this range. Fig. 4 shows the denaturing PAGE gel of coupled DNA nucleotides from N−1 to N−7 after reaction for 48 hours.

Fig. 4A shows the denaturing PAGE gel of the coupling reactions performed at 20 °C. Fig. 4B shows the denaturing

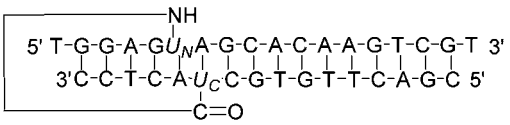
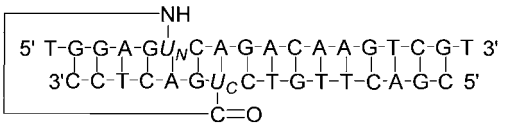
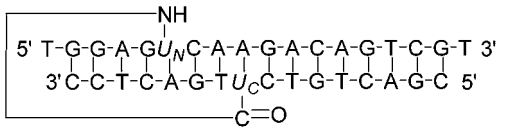
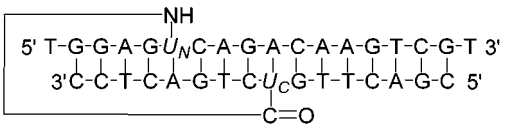
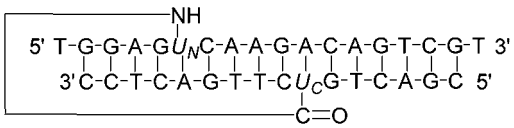
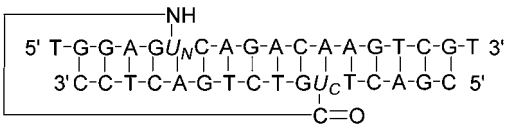
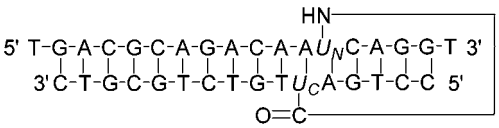
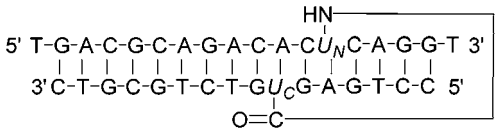
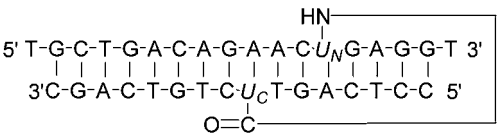
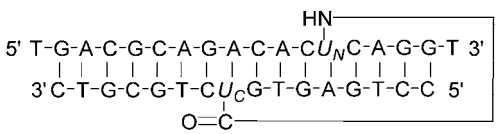
Table 1 Modified oligonucleotides used and their analysis by MALDI-TOF MS

		Mass	
ODNs	Sequences	Calculated	Found
N1	5′-TGGAG U^5_NCAAGACAGTCGT-3′	5652	5650
C1	5′-CGACTGTCU^5_CTGACTCC-3′	4925	4925
C2	5′-CGACTGU^5_CCTTGACTCC-3′	4925	4925
C3	5′-CGACU^5_CGTCTTGACTCC-3′	4925	4926
N2	5′-TGGAG U^5_NCAGACAAGTCGT-3′	5652	5652
C4	5′-CGACTTGTCU^5_CGACTCC-3′	4925	4925
C5	5′-CGACTTGU^5_CCTGACTCC-3′	4925	4925
C6	5′-CGACTU^5_CGTCTGACTCC-3′	4925	4925
N3	5′-TGGAGU^5_NAGCACAAGTCGT-3′	5652	5652
C7	5′-CGACTTGTGCU^5_CACTCC-3′	4925	4925
N4	5′-TGCTGACAGAACU^5_NGAGGT-3′	5652	5651
C8	5′-CCTCAGTU^5_CCTGTCAGC-3′	4925	4926
C9	5′-CCTCAGTTCU^5_CGTCAGC-3′	4925	4925
C10	5′-CCTCAGTTCTGU^5_CCAGC-3′	4925	4926
N5	5′-TGACGCAGACACU^5_NCAGGT-3′	5597	5597
C11	5′-CCTGAGU^5_CGTCTGCGTC-3′	4981	4982
C12	5′-CCTGAGTGU^5_CCTGCGTC-3′	4981	4981
C13	5′-CCTGAGTGTCU^5_CGCGTC-3′	4981	4980
N6	5′-TGACGCAGACAAU^5_NCAGGT-3′	5621	5621
C14	5′-CCTGAU^5_CTGTCTGCGTC-3′	4956	4957
C15	5′-TCGAU^{16}_CAGTGCAAACTCCT-3′	5736	5738
N7	5′-GGAGU^{16}_NTTGCACTATCG-3′	5161	5162
N8	5′-GGAGTU^{16}_NTGCACTATCG-3′	5161	5162
N9	5′-GGAGTTU^{16}_NGCACTATCG-3′	5161	5163

PAGE gel of the coupling reactions at 4 °C. In both panels A and B, lane 1 contains negative controls with non-complementary U_N and U_C strands treated under conditions identical to those for other reaction groups.

The U_N and U_C strands migrate as expected for 18 and 16 nucleotides respectively. Upon coupling, a new upper band appeared. The new band migrates slightly more slowly than the combined mass of 34 nucleotides, which might be attributable to the presence of intra-strand linkers producing a branched structure. Lower bands could still be observed around 18 and 16 nucleotides, representing uncoupled strands. The 20 °C reactions and 4 °C reactions display similar trends. For 20 °C reactions, N−1, N−2, N−3 and N−5 are apparently completely coupled, with one starting material band being undetectable. Cross-linking with N−4 and N−6 gave lower but still high yields. However, the N−7 reaction showed a sharp decrease in yield compared to the others. The coupling reaction yield of N−7 at 20 °C was barely observable, while it was not detected at 4 °C. Reaction yield differences were expected at different reaction temperatures.[27] The consistency of trends suggested that the linking reactivity is caused by the structure of the duplex itself rather than thermal fluctuations related to the temperature change. Once fixed on the duplex, U_N and U_C coupling reactions have high effective concentrations of the two reactants if they are within reach (N−1 to N−6). Once the two arms of linker components are out of reach (N−7), reactions can hardly occur, owing to the rigidity of the DNA duplex, even though the effective concentration is still high. Trace amounts of product can still be observed for the N−7 reaction at 20 °C,

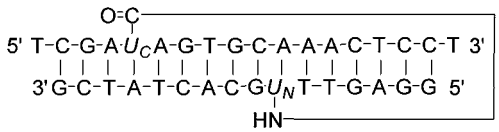
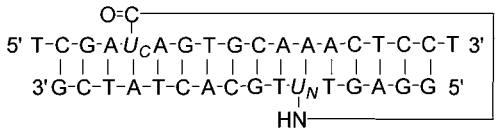
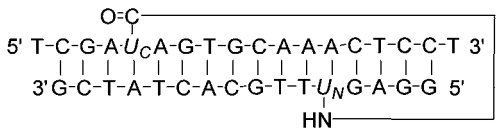
Table 2 Five-atom linker components coupling reaction products and their analysis by MALDI-TOF MS

Coupled Products	Reactant ODNs	Separation (in nucleotides)	Structures	Mass Calculated	Mass Found
N−1	N3 + C7	1		10 560	10 560
N−2	N2 + C4	2		10 560	10 560
N−3	N1 + C1	3		10 560	10 559
N−4	N2 + C5	4		10 560	10 560
N−5	N1 + C2	5		10 560	10 559
N−6	N2 + C6	6		10560	10 559
N+1	N6 + C14	1		10 560	10 560
N+2	N5 + C11	2		10 560	10 559
N+3	N4 + C8	3		10 560	10 558
N+4	N5 + C12	4		10 560	10 558

which may be attributable to thermal fluctuations. However, the amount of product is minimal, relative to the other reactions.

Fig. 5 contains a denaturing PAGE gel of coupled DNA nucleotides from N+1 to N+7 after reaction for 48 hours. Fig. 5A contains the denaturing PAGE gel of coupling reactions occurring at 20 °C. Fig. 5B contains the denaturing PAGE gel of coupling reactions occurring at 4 °C. In both A and B, lane 1 contains control groups with non-complementary U_N and U_C strands treated under conditions identical to the other reaction groups. Similar to the N−*i* reactions, U_N and U_C N+*i* strands migrate around 18 and 16 nucleotides on the gel respectively, while the product band migrates between 40 to 50 nucleotides. Reactions at 20 °C and 4 °C displayed similar trends. For 20 °C reactions, N+1, N+2 and N+3 were close to complete, with one

Table 3 16-atom linker components coupling reaction products and their analysis by MALDI-TOF MS

Coupled products	Reactant ODNs	Separation (in nucleotides)	Structures	Mass	
				Calculated	Found
N−6(L)	N9 + C15	6		10 879	10 884
N−7(L)	N8 + C15	7		10 879	10 882
N−8(L)	N7 + C15	8		10 879	10 885

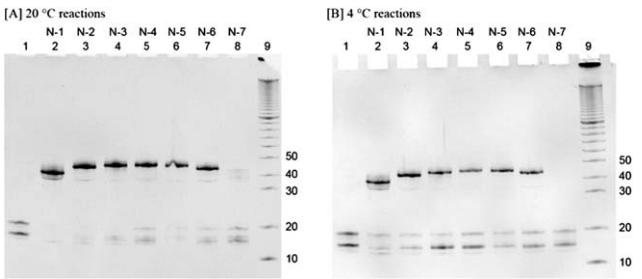


Fig. 4 Denaturing PAGE gels of coupled DNA nucleotides from N−1 to N−7. U_N and U_C ODNs contain 18 and 16 nucleotides, respectively. Upon coupling, a new upper band appeared. In both A and B, lane 1 contains control groups with non-complementary U_N and U_C strands treated under the same conditions as other reaction groups. Lanes 2 to 9 represent, respectively, N−1 to N−7 coupling reactions; lane 2, N−1; lane 3, N−2; lane 4, N−3; lane 5, N−4; lane 6, N−5; lane 7, N−6; lane 8, N−7; lane 9, 10 bp marker. (A) Denaturing PAGE gel of coupling reactions occurring at 20 °C. (B) Denaturing PAGE gel of coupling reactions occurring at 4 °C.

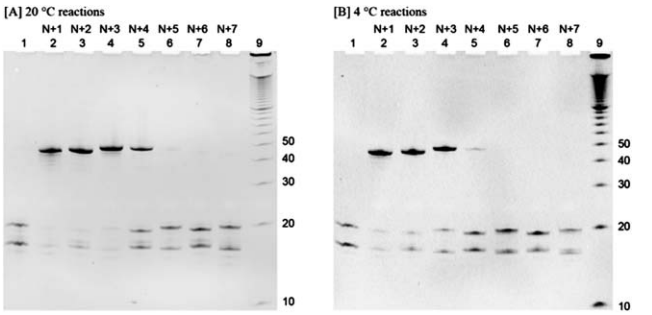


Fig. 5 Denaturing PAGE gels demonstrating the coupling reaction of DNA nucleotides from N+1 to N+7. U_N and U_C ODNs contain 18 and 16 nucleotides respectively. When coupled, a new upper band appeared. In both A and B, lane 1 contains control groups with non-complementary U_N and U_C strands treated under the same conditions as other reaction groups. Lanes 2 to 9 represent N+1 to N+7 coupling reactions; lane 2, N+1; lane 3, N+2; lane 4, N+3; lane 5, N+4; lane 6, N+5; lane 7, N+6; lane 8, N+7; lane 9, 10 bp marker. (A) Denaturing PAGE gel of coupling reactions occurring at 20 °C. (B) Denaturing PAGE gel of coupling reactions occurring at 4 °C.

starting material band being undetectable. The N+4 reaction had a lower yield but was still clearly visible. Traces of the upper band could be seen from N+5 to N+7 with a sharp reaction yield drop from N+4 to N+5. For 4 °C reactions, the same trends were retained, with a lower yield for each reaction. It is noticeable that the N+4 reaction yield dropped sharply compared to the 20 °C reactions. The N+5 to N+7 products were undetectable on denaturing gels. These observations suggest that N+4 is the limit for cross-linking reactions along the N+*i* direction, and that thermal fluctuations contribute to the reaction yield difference.

From cross-linking reactions in both the 3′- and 5′-directions, mismatched sequences resulted in undetectable product formation under identical concentration and coupling conditions. This finding illustrates the effect of templation, which is a consequence of complementary strands forming a well-defined and predictable structure.[27,32] However, not all templated reactants resulted in successful linkage formation. Certain limits of the cross-linking reactions were also observed. The limit of cross-linking for N−*i* (5′ direction of complementary strands) is a separation of 7 nucleotides, and the limit for N+*i* (3′ direction of complementary strands) is 4 nucleotides. Surprisingly, the different distance requirement displayed in the 3′ *vs.* 5′ directions is not consistent with the structural requirement for N−*i* linkages to cross the major groove of DNA duplex and for N+*i* linkages to cross the minor groove. Since the width of major groove is wider (~22 Å) than the minor groove (~12 Å), we would naïvely expect that the linking of N−*i* should be more difficult than that of N+*i*, which contradicts our observations (7 nucleotides *versus* 4). We suspect that not only the distance but also the linker trajectory plays a key role in successful cross-linking. To challenge this hypothesis, we examined computerized DNA models, which will be discussed below in detail.

Since we postulated that the inability of short linkers to span the stiff DNA duplex led to the N−7 limit of N−*i* reactions, we

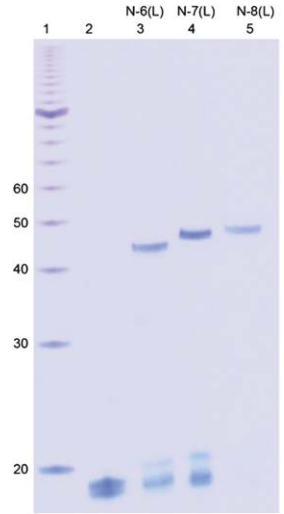


Fig. 6 Denaturing PAGE gel showing inter-strand coupling for 16-atom linker components at 20 °C. Lane 1: 10 bp DNA marker, lane 2: uncoupled strands, lane 3: N−6(L), lane 4: N−7(L), lane 5: N−8(L).

decided to see whether a longer linker could enable us to cross this boundary and cross-link at greater separations. We synthesized nucleotides with 16 atom linker components to go from N−6 to N−8 positions, covering the range of boundaries set by short linkers. The result of this coupling is shown in Fig. 6. Cross-coupled products can be seen clearly from the denaturing gel, which demonstrates that the boundary restraining the cross-linking of short linkers can be overcome by extending linker length.

To understand the rate of inter-strand linkage formation, we studied N+*i* and N−*i* coupling reactions to form 10-atom linkers at both 20 °C and 4 °C. (Fig. S1 and S2†). A crude calibration was conducted using a time course study to give a semi-quantitative estimate of reaction yields. Generally, these ICLs formed efficiently at 20 °C: around 50% of the final products were observed after 4 hours of reaction, while reactions at 4 °C generally have lower yields and reaction rates. The reaction yields correlated with distances in terms of base pairs separating two modified bases. Further separated modified base pairs gave lower reaction yields and slower reaction rates.

The effects of all inter-strand linkages on DNA duplex structure were examined by comparing the melting temperatures of the coupling reaction products with their respective unmodified control duplexes. Melting temperatures (T_m) are summarized in Table 4. For both N−*i* and N+*i* products, linked products generally have wider melting transitions. Products linked between proximal nucleotides tend to have wider melting transitions and higher T_m values than more distantly linked products. (Fig. 7)

The data listed in Table 4 show that inter-strand coupling of the amino and carboxyl linker groups led to higher stability of the duplex form of the products. The degree of T_m rise corresponds to the direction and location of cross-links. The T_m of the N−1 product was not reached under experimental conditions (up to 93 °C). The N−2 product gave a 17.8 °C increase. The difference (ΔT_m) in T_m of the coupled product *vs.* the control decreased for linkers that are stretched further apart in the same direction. When linkers are separated by 3 and 4 base pairs, the ΔT_m increases dropped to 15.0 °C (N−3) and 6.1 °C (N−4). ΔT_m was further reduced for 5 and 6 base separations, (4.7 °C for N−5 and 4.6 °C for N−6). Similar trends were also observed for N+*i* products, with less pronounced differences in ΔT_m. Separation by one base pair gave an 18.7 °C (N+1) of ΔT_m while two and three base pair separations give increases of 17.7 °C (N+2) and 11.9 °C (N+3) respectively.

Overall, the melting profiles of the strands that were chemically cross-linked were shifted to somewhat higher temperatures than that of the control, indicating stabilization effects.[11,33] The extent of increase is related to the separations between linkers.

We also compared the thermal denaturation profiles of the coupled strands with the profiles of the uncoupled strands and control duplexes for N−1 product. The T_m of the uncoupled duplex containing two modified nucleotides is 11 °C lower than control duplex, which is more than 29 °C lower than linked products (Fig. S5†). This result agrees with our earlier studies of intra-strand linkages in which uncoupled, modified sequences were less stable, while coupled linkages showed increased thermal stability.[28]

To understand the global structure of the cross-linked products compared to unmodified control duplexes, circular

Table 4 T_m values of coupling reaction products and respective control duplexes

(a) N−*i* products and controls

Products	T_m (°C)	Control duplexes T_m (°C)	ΔT_m[a]
N−1	>93.0	63.9[b]	>29.1
N−2	81.6	63.8[b]	17.8
N−3	79.6	64.6[b]	15.0
N−4	69.9	63.8[b]	6.1
N−5	69.3	64.6[b]	4.7
N−6	68.4	63.8[b]	4.6

(b) N+*i* products and controls

Products	T_m (°C)	Control duplexes T_m (°C)	ΔT_m[a]
N+1	83.6	64.9[b]	18.7
N+2	84.6	66.9[b]	17.7
N+3	75.6	63.7[b]	11.9

[a] ΔT_m is the T_m difference between coupled reaction products and corresponding control duplexes. [b] Control duplexes for each product: N−1-control for N−1; N−3,5-control for N−3 and N−5; N−2,4,6-control for N−2, N−4 and N−6; N+1-control for N+1; N+2-control for N+2; N+3-control for N+3.

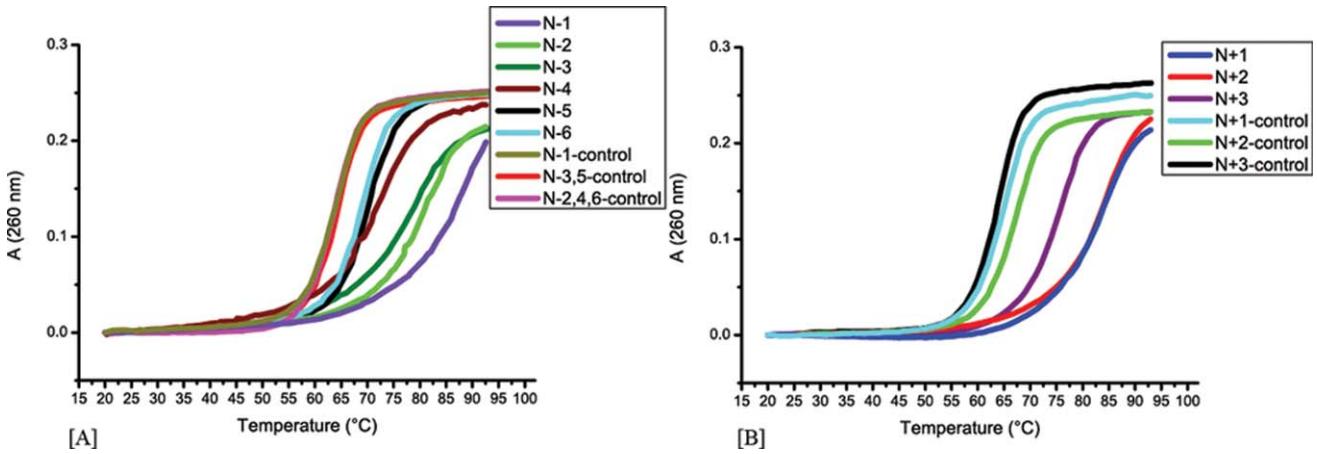


Fig. 7 Melting curves for coupling reaction products and their respective control duplexes. Control duplexes share the sequence of respective products with modified nucleotides replaced by T. (Table 2) Three sets of control duplexes are listed because sequences of products vary. Control duplexes for each product: N−1-control for N−1; N−3,5-control for N−3 and N−5; N−2,4,6-control for N−2, N−4 and N−6; N+1-control for N+1; N+2-control for N+2; N+3-control for N+3. (A) Reaction N−1 to N−6 products and their respective control duplexes. N−1-control, N−3,5-control, N−2,4,6-control largely overlap on the left. N−2,4,6-control may not be seen clearly. (B) Reaction N+1 to N+3 products and their respective control duplexes. $A = (A(T) - A(T_0))/A(T_0)$, where $A(T_0)$ is the absorbance at 20 °C.

dichroism (CD) spectra were obtained. In Fig. 8, the overall CD curves of product strands are similar to their respective control duplexes, consistent with an overall B-form structure. All coupled strands show a less intense negative band around 250 nm and a similar or more intense positive band around 270–280 nm. The latter band is generally shifted toward shorter wavelength compared to their respective control duplexes. Similar effects were observed for N+*i* products, for which CD data are shown in Fig. 9. However, the changes in spectra of coupled products compared to control duplexes are less pronounced. The CD spectra of N+1 and its corresponding control duplex are nearly super-imposable. Therefore, the overall B-form conformation of DNA duplex is retained with only minor perturbation.

The CD spectrum of the N−1 product (Fig. 8a) looks different from others when compared to its control. In the range between 240 nm to 260 nm, there is an apparent inversion of the curve, which is likely owing to a perturbation caused by the linker.

As mentioned above, despite the fact that the major groove is around 10 Å wider than the minor groove in B-form DNA, our experiments clearly show that 10-atom linkers reach across the major groove more easily. To explain the difference in linking abilities, we examined the geometries with molecular models. Since CD experiments had shown that the DNA duplex adopted an overall B-form conformation, it is reasonable to build models based on a B-form structure, although the local structure of the modified nucleosides is likely to be 3′ endo (corresponding to A-form), because of the derivitization of their 2′ carbons.

In considering the required reach between linker components in real space, the trajectory of the linker between two components must be taken into account. Fig. 10 shows that in the B-form structure, N−*i* linkers are positioned by the orientations of the sugars: The vectors from the 2′-ribosyl carbon attachments are directed towards each other. Meanwhile, pairs of N+*i* linkers must cross the minor groove such that all linker components face away from one another. (Fig. 11) Thus, N−*i* linker components easily reach across the major groove of B-form DNA duplex, while N+*i* linker components need to follow a contour across the minor groove by wrapping around the phosphodiester backbone. We suggest that it is the cross-linker

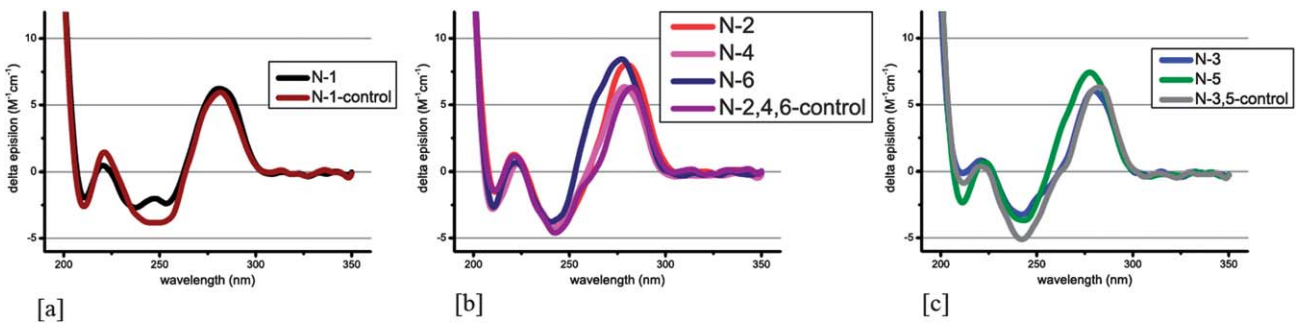


Fig. 8 CD spectra of N−1 to N−6 reaction products and unmodified control duplexes. (a) Overlay of N−1 product and control duplex N−1-control. (b) Overlay of N−3, N−5 products and control duplex N−2,4,6-control. (c) Overlay of N−2, N−4, and N−6 products and control duplex N−3,5-control.

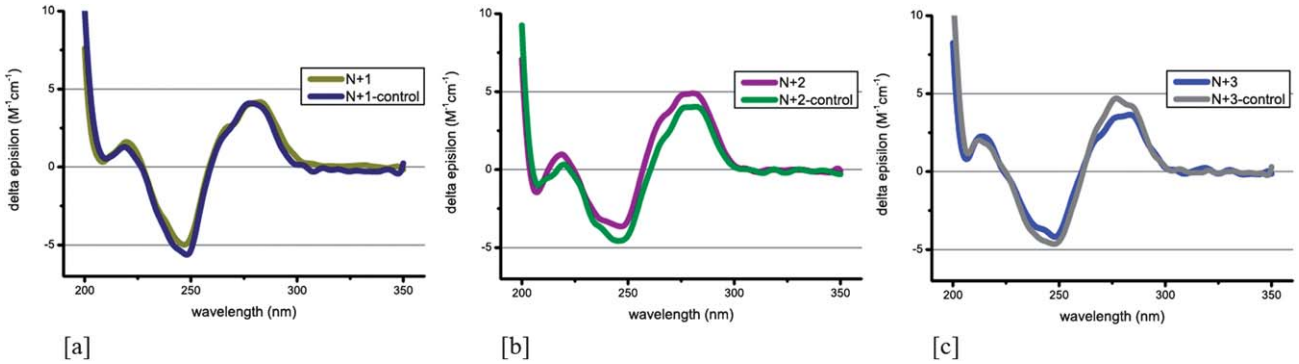


Fig. 9 CD spectra of the N+1 to N+3 reaction products and unmodified control duplexes. (a) Overlay of the N+1 product and control duplex. (b) Overlay of the N+2 product and control duplex. (c) Overlay of the N+3 product and control duplex.

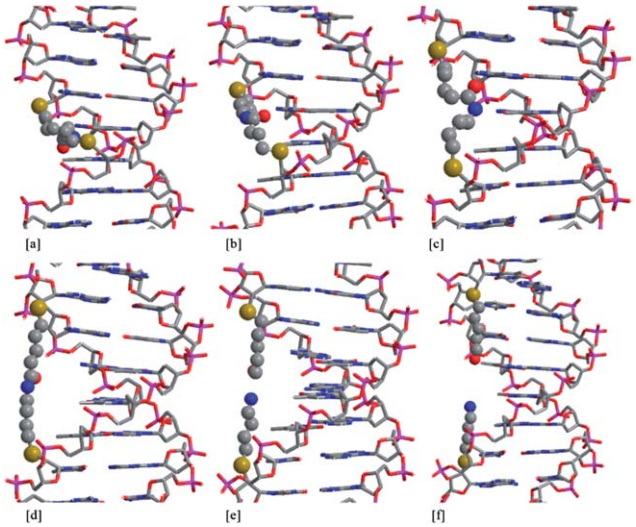


Fig. 10 Molecular models of inter-strand linkages in N−*i* products across the major groove. (a) N−2 product. (b) N−3 product. (c) N−4 product. (d) N−5 product. (e) N−6 duplex, the distance between the nitrogen atom of U_N linker component and carboxyl carbon of U_C linker component is 2.8 Å. (f) N−7 duplex, the distance between the nitrogen atom of U_N linker component and carboxyl carbon of U_C linker component is 7.5 Å.

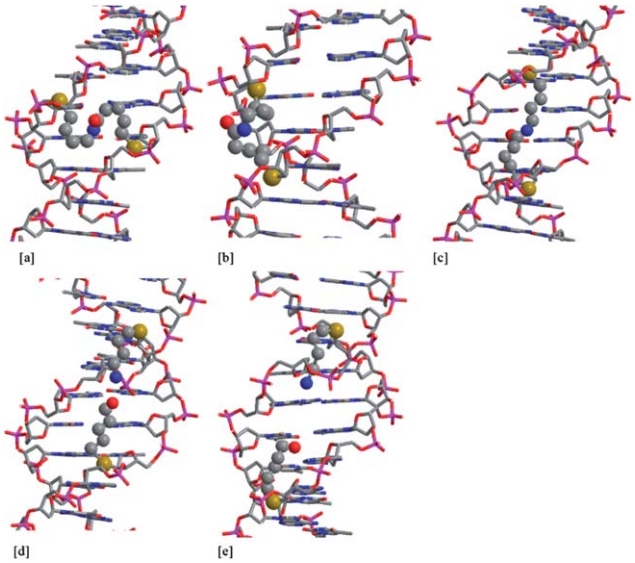


Fig. 11 Molecular models of inter-strand linkages in N+*i* products across the minor groove. (a) N+1 product. (b) N+2 product. (c) N+3 product. (d) N+4 duplex, the distance between the nitrogen atom of U_N linker component and carboxyl carbon of U_C linker component is 3.8 Å. (e) N+5 duplex, the distance between the nitrogen atom of U_N linker component and carboxyl carbon of U_C linker component is 6.8 Å.

molecular trajectory that limits linking reactions in this direction, thus accounting for the difference.

To be more specific, as we can see from models in Fig. 10 and 11, N−1 to N−5 and N+1 to N+3 linkages can form readily; their conformations have been estimated by the methods described. However, distances between two functional groups increase as the number of base pairs separating them increases. For duplexes in which linkers cannot form straightforwardly, distances between the nitrogen atom of the amine groups and the carbon atom of the carbonyl groups of their respective linker components were measured.

The distances represented by duplexes N−7 and N+5 represent the extent to which functional groups cannot reach each other in space. At these locations, distances are 7.5 Å for N−7 (Fig. 10f) and 6.8 Å for N+5 (Fig. 11e), which explains the trace quantities of products observable. In the case of N−6 (Fig. 10e), the distance is around 2.8 Å. Given that the C–N bond length of an amide is around 1.33 Å, the 1.5 Å gap between those two groups must be compensated by thermal fluctuations of the DNA duplex, so as to form the product.

Similarly, in the case of N+4 (Fig. 11d), the closest distance between these two atoms is around 3.8 Å. To make an amide bond, a 2.5 Å distance gap needs to be overcome by thermal fluctuations. Compared to the case of N−6 (Fig. 10e), fewer base pairs (4 *vs.* 6) separating two modified nucleotides may contribute to a reduced flexibility, thereby resulting in a much lower reaction yield.

In the case of N−1, if the duplex conformation is maintained undistorted, the span between linked nucleotides is prevented by van der Waals violations with neighboring base pairs. (Fig. 12a) In an attempt to make a more sensible model to avoid this severe steric effect, a bent conformation was examined. (Fig. 12b) The bent model has a much lower energy, and

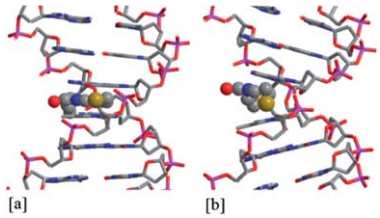


Fig. 12 Proposed structures of N−1 product (a) N−1 product unbent conformations, the linker goes through the space between two adjacent base pairs. (b) N−1 product bent conformation, adjacent base pairs neighboring linking sites are tilted to accommodate the linker.

accommodates linkers without violating van der Waals radii. Since product N−1 shows several unusual properties in comparison with the other products, including very high T_m, faster mobility, and a more pronounced CD deviation from the control, we postulate that all of these properties are somehow related to a distorted conformation. However, further experimental data will be needed to obtain a more precise structure of the product.

Conclusions

In summary, we have developed convenient tools to cross-link selected positions on the opposite strands of a DNA double helix. We have thoroughly examined the linking ability of ten-atom linkers. We established the linking constraints of the ten-atom linkers to be N−1 to N−6, and N+1 to N+4. Reactions at 20 °C and 4 °C both display the same trends, showing that the linking selectivity of these reactions is largely temperature independent in this range. Reaction yields and rates are generally related to the number of nucleotides separating two linking positions. The shortest separations provide quantitative yields. Longer linkers generally have lower reaction yields and lower reaction rates. To explain the difference between linking yields in the two directions, we have examined structural models. Structures based on B-form DNA demonstrate that the orientations of linkers and the resulting contour lengths of the linker molecular trajectories between them largely determine reaction yields. However, if linker lengths are extended, the set of constraints originating in the DNA duplex structure applied to 10-atom linkers can easily be overcome by using 32-atom linkers, suggesting the possibility of expanding our system to realize cross-linking selectively over larger distances if desired.

We employed several approaches to characterize cross-linked products. MALDI-TOF mass spectrometry and denaturing gel electrophoresis verified successful links between strands. Thermal denaturation showed that the existence of linkers increases the melting temperatures of duplexes. While a general stabilizing effect was discovered, the extent depends heavily on the separation of the linkers. Short linkers can greatly increase T_m, whereas longer linkers contribute much smaller changes, as would be expected from the extent of restraint the linkers impose on the double helix. CD spectra demonstrated that the overall character of B-form DNA is maintained despite the fact that the uridine nucleotides containing the linkers are ribonucleotides and likely to prefer the 3′-endo, A-form conformation.

The N−1 product revealed unusual properties, such as a very high ΔT_m, disturbance of the CD spectrum and faster mobility on denaturing gels. We suggest that there may be a stronger structural perturbation at play.

The cross-linking procedure that we have introduced here should find wide applications in nucleic acid chemistry, biotechnology and nanotechnology. Readily accessible modified bases can be introduced using phosphoramidite-based oligonucleotide synthesis[31] to produce strands ready for cross-linking without further functionalization. It should be straightforward to insert a fragment containing cross-linked or cross-linkable synthetic fragments into a larger piece of DNA or other nucleic acids, such as RNA or LNA. Such fragments will prove resistant to strand separation and to strand displacement through the processes of strand invasion and branch migration, thereby facilitating DNA-based hierarchical assembly. Likewise, the type of strand exchange needed for PX-DNA formation[34] is likely to be impaired. Both DNA nanotechnology and DNA-based computation will be able to benefit from molecules that are resistant to strand invasion, branch migration and separation. Furthermore, it is likely that analysis of those biological processes that entail these phenomena will be facilitated by the developments reported here. In addition, the possible roles of strand separation and empty grooves, perhaps currently unsuspected in a variety of biological processes, can be explored and challenged using this system.

Materials and Methods

General procedure for MALDI sample preparation

Amicon Ultra-0.5 centrifugal filter devices with nominal molecular weight limit (NMWL) 3000 Da were used to desalt samples. The matrix solution consisted of 2 : 1 v/v ratio of 2′,4′,6′-trihydroxyacetophenone (THAP) (35 mg, 0.2 mmol in 1 mL of 50% MeOH–H_2O) to ammonium citrate solution (0.1 M). Samples and standards (*e.g.*, unmodified T_{18}) were prepared in vials, to which the matrix solution was added. Each sample (1 μL) was mixed with matrix solution (3 μL) using vortex mixing followed by centrifugation (5 s). A total volume of 4 μL of the mixture was then added to wells on the target.

General procedure for amide bond formation on DNA

A U^5_N-containing strand (100 pmol, DNA oligomer modified with a 5-atom amine linker component) and a U^5_C-containing strand (100 pmol, DNA oligomer modified with a 5-atom carboxylic acid linker component) were dissolved in 100 μL of MOPS buffer (100 mM MOPS, 1 M NaCl, pH 7.0). The solution was annealed overnight. A portion of 100 mM DMTMM reagent (100 μL) in MOPS buffer was added to the DNA mixture so that the final duplex solution became 500 nM and the final DMTMM concentration was 50 mM. The reaction was allowed to proceed for 48 hours in the dark.

General procedure for time-dependent amide formation on DNA

A U_N^5-containing strand (500 pmol) and a U_C^5-containing strand (500 pmol) were dissolved in 500 μL of MOPS buffer (100 mM MOPS, 1 M NaCl, pH 7.0) and annealed overnight. A portion (500 μL) of 100 mM DMTMM solution was added to start the reaction. An aliquot of the reaction mixture (200 μL) was taken at the designated time and then quenched by 300 μL of 1 M ammonium acetate solution. The quenched solutions were then desalted and analyzed by denaturing PAGE. The same procedure was performed at both 4 °C and 20 °C.

General procedure to prepare samples for thermal denaturing studies

Each complementary strand (2 nmol) was dissolved in 500 μL of sodium cacodylate buffer (40 mM sodium cacodylate, 100 mM NaCl). The solution containing DNA strands was pre-annealed. The melting curve was obtained at 260 nm using a Varian Cary 100 Bio UV-Vis Spectrophotometer. The temperature was changed at the rate of 0.2 °C min^{-1} at a data interval of 0.5 °C. Each sample was heated from 20.0 °C to 93.0 °C, then cooled back down to 20.0 °C. This temperature was maintained for 5 min at both 20.0 °C and 93.0 °C. Two such runs were repeated for each sample.

CD experiments

The CD spectra were recorded from 350 nm to 200 nm on an AVIV Model 202SF Circular Dichroism Spectropolarimeter at ambient temperature in 0.2 cm cuvettes. Two complementary strands (2.8 nmol each) were pre-annealed and then dissolved in buffer (10 mM sodium phosphate, 150 mM sodium chloride, and 0.1 mM EDTA) to a final volume of 700 μL. The duplex solution was filtered through a 0.2 μm syringe filter (Millex) before recording CD spectra. At least three scans were collected for each sample and a buffer baseline was subtracted from the average of these scans to yield the CD plots.

Denaturing polyacrylamide gel electrophoresis (PAGE) analysis and purification

The coupling products were analyzed and purified on 20% polyacrylamide gels. The running buffer contained 89 mM Tris–HCl, pH 8.0, 89 mM boric acid, and 2 mM EDTA. The tracking dye contained buffer, 50% glycerol, and a trace amount of bromophenol blue and xylene cyanol FF. The gel was stained with ethidium bromide.

For analysis, gels were imaged on a Kodak Image Station Gel Logic 200. Semi-quantitative estimation of yields was calculated by measuring the band intensity using Kodak Molecular Imaging software.

For purification, product bands were cut out and eluted in a solution containing 500 mM ammonium acetate, 10 mM magnesium acetate, and 1 mM EDTA. Isolated yields were around 50% to 60%, estimated by OD_{260}.

General procedure for DNA modelling

The DNA duplex structures were prepared using coordinates derived from X-ray fiber diffraction data.[35] An 18 base-pair double-stranded B-DNA molecule was introduced. Linker components and linkers were built using Maestro (Maestro, version 9.2, Schrödinger, LLC). For unlinked duplexes, conformations with the shortest distances between amine and carboxyl were found and measured. For linked duplexes, conformational searches were conducted using the Monte Carlo (MCMM) method as implemented in MacroModel (MacroModel, version 9.9, Schrödinger, LLC). All atoms and all rotatable single bonds of linkers were included in the conformational search. All base pairs were maintained stationary; modified and unmodified deoxyribose and adjacent phosphate diester linkages were allowed to move only when at linking sites or in close proximity to them. The energy minimizations were carried out using the Powell–Reeves conjugate gradient (PRCG) method and the AMBER* force field, as implemented in MacroModel. The conformational searches were done for an aqueous solution using a constant dielectric electrostatic treatment with default settings. To bend cross-linked product N−1 from the starting duplex, the angle between two adjacent bases sandwiching the linker was estimated from a 4 base pair fragment of the original starting duplex. Following this, a conformational search was performed with the bent form.

Synthesis of oligonucleotides

Oligonucleotide strands were prepared from synthetic or commercial phosphoramidite reagents on an Applied Biosystem Model 394 automated synthesizer using standard phosphoramidite techniques.[31] Synthetic amidites were prepared as described previously,[26,30] except for U_N^{16} which was prepared as shown in Scheme S1 in the ESI.†

Funding

We (JWC and NCS) are grateful for funding to the National Science Foundation (CTS-0608889) and the Office of Naval Research (N000140911118). This work has also been supported by grants to NCS from the National Institute of General Medical Sciences (GM-29544) the National Science Foundation (CCF-1117210), the Office of Naval Research (N00014-11-1-0729) the Army Research Office (W911NF-11-1-0024), grant DE-SC0007991 from the Department of Energy. This work was supported partially by the MRSEC Program of the National Science Foundation under Award Number DMR-0820341. We thank the National Science Foundation for instrument grants CHE-0234863, CHE-0116222, CHE-0958457, and DUE-0126958, and the National Institutes of Health for construction grant C06 RR-16572-01.

Acknowledgements

We thank Liang Ding for contributions during early phases of this work.

Notes and references

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
2 (*a*) S. T. Sigurdsson, Site-Specific Sulfhydryl Groups for Study of RNA Conformation *via* Disulfide Cross-Linking, *Methods in Enzymology*, Academic Press, 2000, vol. 318, pp. 165–166; (*b*) T. K. Stage-Zimmermann and O. C. Uhlenbeck, *Nat. Struct. Biol.*, 2001, **8**, 863.
3 N. C. Seeman, *Nature*, 2003, **421**, 427–431.
4 A. Rajendran, M. Endo, Y. Katsuda, K. Hidaka and H. Sugiyama, *J. Am. Chem. Soc.*, 2011, **133**, 14488–14491.
5 D. M. Noll, T. M. Mason and P. S. Miller, *Chem. Rev.*, 2006, **106**, 277–301.
6 P. A. Muniandy, J. Liu, A. Majumdar, S. T. Liu and M. M. Seidman, *Crit. Rev. Biochem. Mol. Biol.*, 2010, **45**, 23–49.
7 S. Dutta, G. Chowdhury and K. S. Gates, *J. Am. Chem. Soc.*, 2007, **129**, 1852–1853.
8 N. B. Edfeldt, E. A. Harwood, S. T. Sigurdsson, P. B. Hopkins and B. R. Reid, *Nucleic Acids Res.*, 2004, **32**, 2795–2801.
9 (*a*) E. Champeil, M. M. Paz, S. Ladwa, C. C. Clement, A. Zatorski and M. Tomasz, *J. Am. Chem. Soc.*, 2008, **130**, 9556–9565; (*b*) H. Huang, H. Y. Kim, I. D. Kozekov, Y. J. Cho, H. Wang, A. Kozekova, T. M. Harris, C. J. Rizzo and M. P. Stone, *J. Am. Chem. Soc.*, 2009, **131**, 8416–8424.
10 E. Freese and M. Cashel, *Biochim. Biophys. Acta*, 1964, **91**, 67–77.
11 T. J. Santangelo and J. W. Roberts, Formation of Long DNA Templates Containing Site-Specific Alkane–Disulfide DNA Interstrand Cross-Links for Use in Transcription Reactions, *Methods in Enzymology*, ed. S. L. Adhya and S. Garges, Academic Press, 2003, vol. 371, pp. 120–132.
12 M. Endo and T. Majima, *J. Am. Chem. Soc.*, 2003, **125**, 13654–13655.
13 S. Alefelder and S. T. Sigurdsson, *Bioorg. Med. Chem.*, 2000, **8**, 269–273.
14 M. Op de Beeck and A. Madder, *J. Am. Chem. Soc.*, 2011, **133**, 796–807.
15 J. T. Sczepanski, C. N. Hiemstra and M. M. Greenberg, *Bioorg. Med. Chem.*, 2011, **19**, 5788–5793.
16 H. Xiong and F. Seela, *J. Org. Chem.*, 2011, **76**, 5584–5597.
17 (*a*) A. El-Sagheer and T. Brown, *Chem. Soc. Rev.*, 2010, **39**, 1388–1405; (*b*) R. Kumar, A. El-Sagheer, J. Tumpane, P. Lincoln, L. M. Wilhelmsson and T. Brown, *J. Am. Chem. Soc.*, 2007, **129**, 6859–6864.
18 C. S. Andersen, M. M. Knudsen, R. Chhabra, Y. Liu, H. Yan and K. V. Gothelf, *Bioconjugate Chem.*, 2009, **20**, 1538–1546.
19 S. A. Wolfe and G. L. Verdine, *J. Am. Chem. Soc.*, 1993, **115**, 12585–12586.
20 D. M. Noll, M. W. Silva, A. M. Noronha, C. J. Wilds, O. M. Colvin, M. P. Gamcsik and P. S. Miller, *Biochemistry*, 2005, **44**, 6764–6775.
21 T. Angelov, A. Guainazzi and O. D. Schärer, *Org. Lett.*, 2009, **11**, 661–664.
22 S. A. Wolfe, A. E. Ferentz, V. Grantcharova, M. E. A. Churchill and G. L. Verdine, *Chem. Biol.*, 1995, **2**, 213–221.
23 G. D. Cimino, H. B. Gamper, S. T. Isaacs and J. E. Hearst, *Annu. Rev. Biochem.*, 1985, **54**, 1151.
24 E. R. Jamieson and S. J. Lippard, *Chem. Rev.*, 1999, **99**, 2467–2498.
25 N. C. Seeman, *Methods Mol. Biol.*, 2005, **303**, 427–431.
26 L. Zhu, P. S. Lukeman, J. W. Canary and N. C. Seeman, *J. Am. Chem. Soc.*, 2003, **125**, 10178–10179.
27 Y. Liu, R. Sha, R. Wang, L. Ding, J. W. Canary and N. C. Seeman, *Tetrahedron*, 2008, **64**, 8417–8422.
28 Y. Liu, R. Wang, L. Ding, R. Sha, P. S. Lukeman, J. W. Canary and N. C. Seeman, *ChemBioChem*, 2008, **9**, 1641–1648.
29 Y. Liu, R. Wang, L. Ding, R. Sha, N. C. Seeman and J. W. Canary, *Chem. Sci.*, 2012, **3**, 1930–1937.
30 Y. Liu, A. Kuzuya, R. Sha, J. Guillaume, R. Wang, J. W. Canary and N. C. Seeman, *J. Am. Chem. Soc.*, 2008, **130**, 10882–10883.
31 M. H. Caruthers, A. D. Barone, S. L. Beaucage, D. R. Dodds, E. F. Fisher, L. J. McBride, M. Matteucci, Z. Stabinsky and J. Tang, Chemical Synthesis of Deoxyoligonucleotides by the Phosphoramidite Method, *Methods in Enzymology*, ed. R. Wu and L. Grossman, Academic Press, 1987, vol. 154, pp. 287–313.
32 T. M. Snyder, B. N. Tse and D. R. Liu, *J. Am. Chem. Soc.*, 2008, **130**, 1392–1401.
33 A. S. Fridman, E. N. Galyuk, V. I. Vorobiev, A. N. Skvortsov and D. Y. Lando, *J. Biomol. Struct. Dyn.*, 2008, **26**, 175–186.
34 X. Wang, X. Zhang, C. Mao and N. C. Seeman, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 12547–12552.
35 S. Arnott, D. W. L. Hukins and S. D. Dover, *Biochem. Biophys. Res. Commun.*, 1972, **48**, 1392.