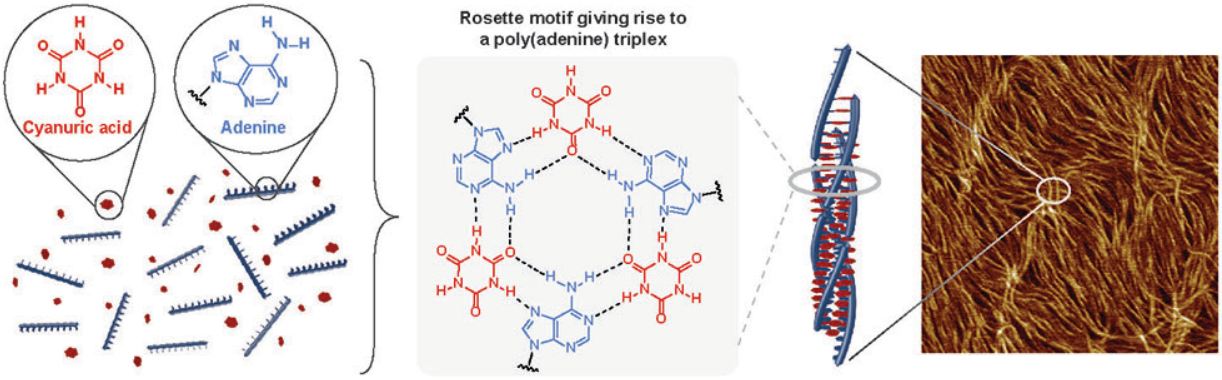
Reprogramming the assembly of unmodified DNA with a small molecule

Nicole Avakyan[1], Andrea A. Greschner[1, 2], Faisal Aldaye[1], Christopher J. Serpell[1, 3], Violeta Toader,[1] Anne Petitjean[4] and Hanadi F. Sleiman[1]*

Corresponding author: hanadi.sleiman@mcgill.ca

1. *Department of Chemistry and Centre for Self-assembled Chemical Structures, McGill University, 801 Sherbrooke Street West, Montreal, QC H3A 0B8, Canada*
2. *INRS: Centre Énergie Matériaux Télécommunications, 1650 boul. Lionel-Boulet, Varennes QC, J3X 1S2, Canada*
3. *School of Physical Sciences, Ingram Building, University of Kent, Canterbury, Kent, CT2 7NH, UK*
4. *Department of Chemistry, Queen's University, Chernoff Hall, 90 Bader Lane, Kingston ON, K7L 3N6, Canada*

The information storage and encoding ability of DNA arise from a remarkably simple 4-letter – A, T, G, C nucleobase code. Expanding this DNA 'alphabet' provides information about its function and evolution, and introduces new functionalities into nucleic acids and organisms. Previous efforts relied on the synthetically demanding incorporation of non-canonical bases into nucleosides. Here we report the discovery that a small molecule, cyanuric acid, with three thymine-like faces reprograms the assembly of unmodified poly(adenine) into stable, long and abundant fibers with a unique internal structure. Poly(A) DNA, RNA and PNA all form these assemblies. Our studies are consistent with the association of adenine and cyanuric acid units into a hexameric rosette, bringing together poly(A) triplexes with subsequent cooperative polymerization. Fundamentally, this study shows that small hydrogen-bonding molecules can be used to induce the assembly of nucleic acids in water, leading to new structures from inexpensive and readily available materials.



The standard DNA base-pairing alphabet gives rise to both the familiar double helix and a variety of alternate assembly motifs such as the guanine quadruplex, i-motif and triple helix[1,2]. DNA's stability, remarkable recognition properties and diversity in tertiary structure are not only relevant to biological function. Three decades of DNA nanotechnology research have revealed its power as a building block in developing nanomaterials[3]. Considering the wealth of intermolecular interactions exploited in supramolecular chemistry, the DNA base-pairing code is limited to just two hydrogen bonding motifs that harness only four letters of a vast potential alphabet. Efforts to expand this genetic 'alphabet' have been driven by several goals. First, information about the principles behind the selectivity, function and evolution of DNA can be obtained by replacing its components, bases, or backbone with artificial counterparts[4-6]. Second, selective incorporation of new properties relevant to the use of nucleic acids in materials science, synthetic biology, biotechnology and medicine can be achieved by using unnatural DNA analogs[7-11]. Finally, compared to the standard A-T/G-C duplex code, an expanded DNA alphabet grants access to increasingly complex DNA nanostructures, with potentially reduced defects[12,13].

The creation of artificial nucleic acid motifs has relied on synthetic incorporation of nucleobases with altered hydrogen-bonding patterns[14,15], or relying on hydrophobic interactions[16,17] and metal coordination[18-20]. Synthetic incorporation of unnatural bases which present more than one hydrogen-bonding 'face' in order to access higher-order DNA structures has been of great interest[21-25]. 'Janus'[26] monomers (complementary to two nucleobases simultaneously) were attached to PNA and associated into triplexes with the addition of two complementary DNA strands[27]. *Iso*-guanine strands formed a DNA pentaplex in the presence of large alkali ions[28,29]. In the studies cited above, changing the DNA self-assembly code required synthetically laborious attachment of new monomers onto nucleoside sugar moieties, their incorporation into DNA strands and assembly of these new strands into altered structures. Intercalation of crescent-shaped alkaloids such as coralyne has been shown to induce the formation of an antiparallel duplex from poly(dA) sequences,[30-34] and addition of π-extended molecules to guanine-rich sequences can chaperone the formation of quadruplex structures.[35]

Adenine contains two hydrogen-bonding (Watson-Crick and Hoogsteen) faces, each with a two-point donor-acceptor motif. In the presence of poly(thymine), poly(adenine) can form double and triple helices. On its own, poly(adenine) can form a parallel homoduplex in acidic conditions,[36,37] whereas an antiparallel homoduplex can be formed at neutral pH with the addition of small molecule intercalators such as coralyne[30,31]. In principle, we can reprogram poly(adenine) interactions to induce assembly of a higher-order structure if we combine adenine with a small symmetric molecule such as cyanuric acid, that contains three complementary thymine-like faces (Fig. 1a). Cyanuric acid (CA) and its analogs have been extensively used in supramolecular chemistry, most notably to form hydrogen-bonded tape and hexameric rosette architectures with melamine and its analogs[38-40]. Here, we show that simple addition of this small molecule can reprogram the assembly of unmodified poly(adenine) strands into a non-canonical motif. Moreover, CA mediates the cooperative growth of supramolecular polymers with very high aspect ratio. The preparation of CA-mediated poly(A) nanofibers is an extremely facile method to create new nucleic acid-based nanomaterials. Fundamentally, this illustrates that small hydrogen-bonded molecules can now be used, in water, to induce the assembly of nucleic acids into new forms. This orthogonal assembly mode will expand the 4-letter DNA alphabet and increase the range of

possible architectures in DNA nanotechnology. In addition, the structures described here can be generated in large quantities from biocompatible components, allowing prospective applications in tissue engineering and drug delivery.

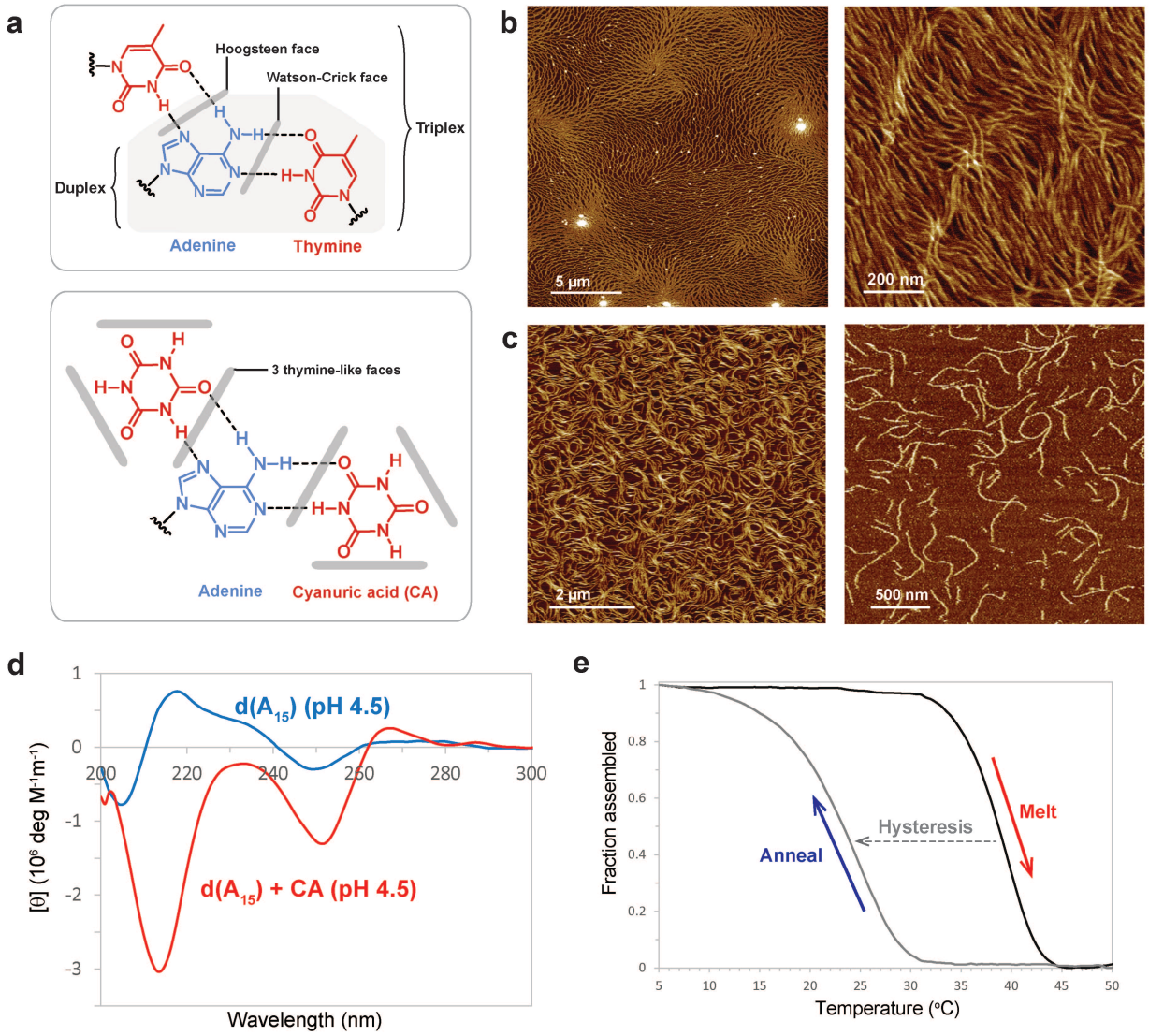


Figure 1. Cyanuric acid (CA) mediated assembly of d(A_n).) a) The Watson-Crick and Hoogsteen binding faces of adenine can participate in the formation duplex or triplex structures with thymine. CA, a small molecule which has 3 thymine-like faces, could form a non-canonical DNA assembly motif by interacting with adenine. Atomic force microscopy in air **(b)** and in solution **(c)** shows extensive fiber formation for d(A_{15}) in the presence of CA. **d)** CD spectra for d(A_{15}) the presence of CA (red) or alone (blue) in matching salt and pH conditions. **e)** Thermal denaturation of d(A_{15}) with CA followed by re-annealing shows a cooperative transition and hysteresis.

Evidence of CA-mediated assembly of $d(A)_{15}$

Cyanuric acid-mediated assembly of poly(A) DNA ($d(A_n)$) was first examined by atomic force microscopy (AFM) in air (Fig. 1b). Deposition of a mixture of $d(A_{15})$, excess CA and sodium chloride (NaCl) in water led to the formation of fibers of uniform height (2.5 ± 0.1 nm) on the mica surface upon drying (Supplementary Fig. S1). In contrast, no such features were observed for $d(A_{15})$ or CA alone deposited under the same pH and salt conditions (Supplementary Fig. S1). AFM imaging in fluid also revealed well-defined nanofibers microns in length (height 2.0 ± 0.1 nm), suggesting that CA-mediated assembly of $d(A_{15})$ is a solution-based phenomenon (Fig. 1c, Supplementary Fig. S2). Dynamic light scattering (DLS) experiments provided further evidence of fiber formation in solution (Supplementary Fig. S4). These findings indicate that CA, a small molecule with binding features complementary to a natural DNA base, can bring about formation of self-assembled nanofibers from simple, inexpensive components.

The structural nature of these fibers was studied by circular dichroism (CD) and UV-Vis spectroscopies. UV-Vis absorbance showed hypochromicity of the DNA absorbance peak at 260 nm for $d(A_{15})$ in the presence of excess CA and NaCl (pH 4.5) (Supplementary Fig. S5). In the same conditions, the CD spectrum of $d(A_{15})$ changed dramatically, with the appearance of strong negative bands at 212 and 252 nm and of weak positive bands at 267 and 285 nm (Fig. 1d). The CD spectrum is distinct from that of a B-DNA duplex and bears similarities to the spectrum of a poly(A)-poly(T)-poly(T) triplex, with a slight red-shift[41]. Control experiments show that CA addition has no such effect on DNA strands of random sequence (Supplementary Fig. S7). The same CD features characteristic of assembly were obtained in the presence of Mg^{2+} rather than Na^+ and at lower ion concentrations (250 µM Mg^{2+} *vs*. 20 mM Na^+) (Supplementary Fig. S8).

Thermal denaturation of the CA-mediated assembly was monitored by CD, UV-Vis and dynamic light scattering (DLS) (Fig. 1e and Supplementary Fig. S6). The curves showed a well-defined, cooperative transition with a sigmoidal shape akin to the thermal denaturation of a DNA double helix. This observation is a sign of uniform assembly that does not go through stable intermediate stages during the melting process. Scatter intensity for CA-mediated $d(A_{15})$ structures monitored by DLS similarly showed a sharp drop with increasing temperature, suggesting that large aggregates disassembled directly to their constituent components (Supplementary Fig. S6). This transition can also be qualitatively observed by AFM in solution (Supplementary Fig. S3). A kinetic barrier to assembly was evident in the hysteresis detected between the melting and annealing curves (Fig. 1e), an observation often associated with cooperative self-assembly processes[42].

Under highly acidic conditions (pH 3), poly(A) strands (both in their RNA and DNA forms) have been reported to self-associate into parallel homoduplexes as a result of protonation of the adenine units[36,37,43,44]. We ascertained that CA-mediated assembly of $d(A_{15})$ under the conditions described above (pH 4.5) is distinct from the pH-driven poly(A) duplex, based on major differences between the CD spectra of these species. CA-mediated assembly occurs in the 4 to 7 pH range and is thus favored when CA (pKa ~6.9) and adenine (pKaH ~3.6) are mostly uncharged (Supplementary Fig. S9). Overall, the CD and UV-Vis absorbance results reveal that $d(A_n)$ assembly is mediated by CA in a unique fashion, giving

rise to a structure with features distinct from the previously reported poly(A) duplexes that form in acidic conditions.

Investigation of fiber structure

The ability of a small molecule to mediate hydrogen-bonded interactions in water may seem counterintuitive, but is not unprecedented[45]. Hud et al. have demonstrated self-assembly of CA with hydrophilic derivatives of 2,4,6-triaminopyrimidine (TAP) in water, including a monomeric TAP-ribose derivative[46]. These mixtures produced hexameric rosettes that stacked into long fibers with a similar diameter to the d(A_n):CA structure described here. Comparably to our studies, Hud *et al.* observed the absence of intermediates in the assembly process and ascribed this property to the large hydrophobic surface of the rosette (> 1 nm^2) that imposes a high energetic cost (+27 kcal/mol) for its exposure to the aqueous solvent in intermediate structures. In the d(A_n):CA system, the adenine units are tethered by a phosphodiester backbone. In addition, d(A_n) single strands are significantly preorganized[36], which is expected to reduce the entropic cost of assembly.

To verify that monomeric adenine is capable of forming rosette structures with CA, we prepared deoxy-tetraisopropyldisiloxane-adenosine (TIPDS-dA) and hexyl-CA (Hex-CA) derivatives and tested their assembly in toluene (Fig. 2a)[47]. Vapor pressure osmometry (VPO) studies show that in the 60-120 mM concentration range, these monomers assemble into aggregates with a mass consistent with a hexameric rosette ($M_{W\ Rosette}$ = 2 121 g/mol, $M_{W\ VPO}$ = 2050 ± 150 g/mol) from a 1:1 mixture. Furthermore, diffusion ordered NMR spectroscopy (DOSY) of an analogous mixture showed the formation of a single slow diffusing species, rather than a mixture of oligomers with different diffusion coefficients. Using the diffusion coefficient of monomeric TIPDS-dA and its molecular weight as references, a molecular weight range for the species in the mixture can be estimated and is consistent with a hexameric rosette[48].

Having demonstrated that modified adenosine and CA monomers are capable of forming hexameric rosettes in organic medium, we turned to the native poly(A):CA assembly in aqueous solution. In order to examine CA:A stoichiometry in the CA:d(A_n) structure, we carried out equilibrium dialysis experiments[49,50]. With d(A_{15}) in one chamber of the dialysis unit, and CA in a range of concentrations in the other (both in TAMg pH 4.5) the system was allowed to equilibrate with gentle agitation at 4°C for 24 hrs. d(A_n) remained sequestered by the dialysis membrane in one chamber where it assembled with CA that freely diffused through the membrane from the other side. Once equilibrium was reached, solutions from both chambers were analyzed by high performance liquid chromatography (HPLC) to determine the concentrations of 'free' CA and CA 'bound' through participation in assembly ($[CA]_{Free}$ and $[CA]_{Bound}$). These values were used to determine the CA:A stoichiometry and to build a binding curve (Fig. 2b). The curve revealed CA:A saturation at 1, consistent with a 1:1 CA:A ratio in the structure. A complication of the equilibrium dialysis experiment is that CA binding to d(A_{15}) is coupled to a structural change of the single-stranded DNA into a new assembly, as well as its elongation into fibers[33]. This experiment is different from binding studies of small molecules with double stranded DNA, in which the duplex structure is relatively maintained throughout the dialysis experiment[51].

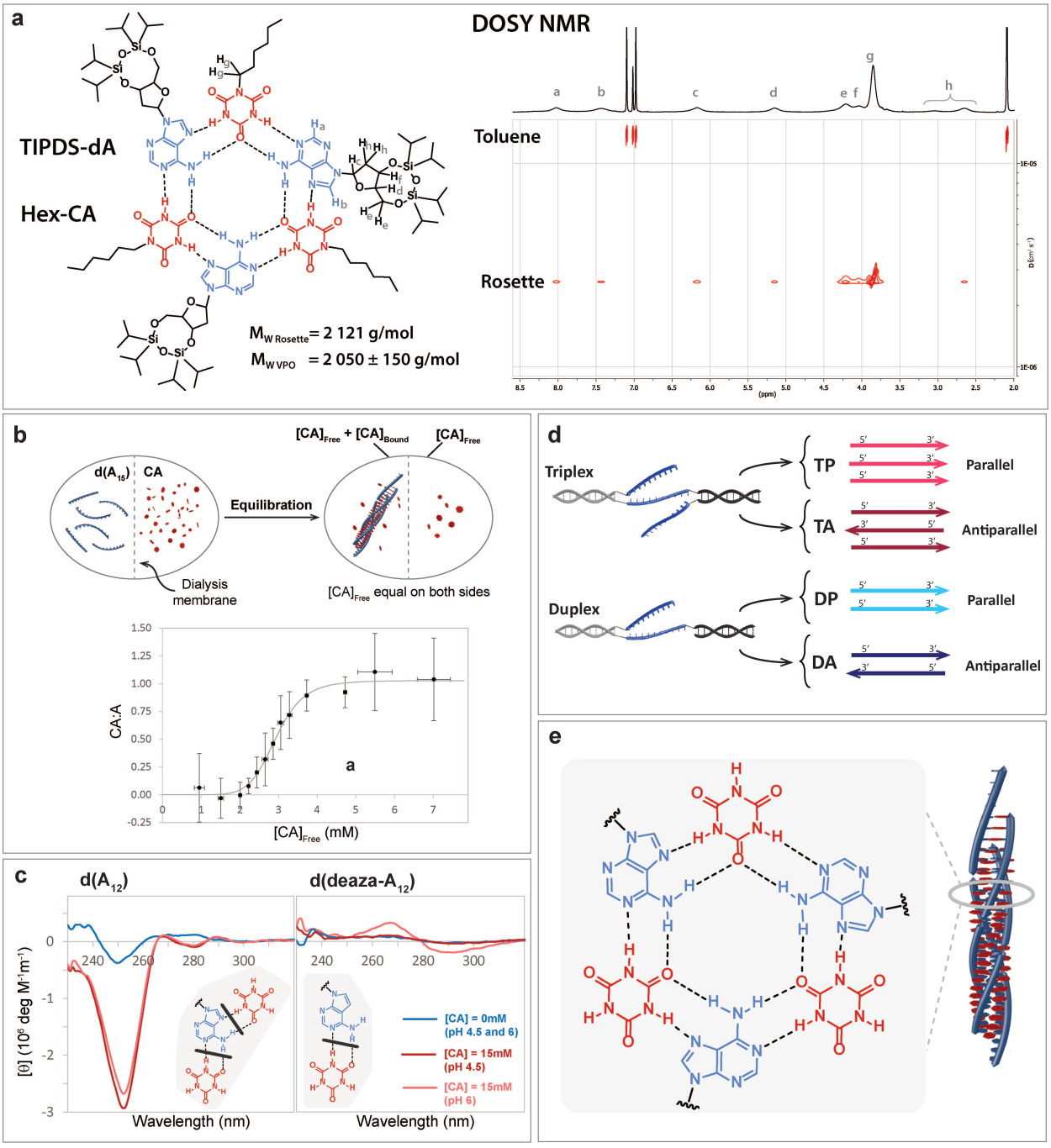


Figure 2. Proposed hexameric rosette structure. a) VPO and DOSY NMR studies of a 1:1 mixture of TIPDS-dA and Hex-CA in toluene suggest that these monomers form a single structure in solution with a mass consistent with a hexameric rosette. **b)** Equilibrium dialysis studies show that the CA:A stoichiometry within the CA mediated d(A_n) assembly is 1:1. **c)** d(deaza-A_{12}) (right) vs d(A_{12}) (left) in the presence of CA by CD. The adenine analog where only the Watson-Crick face is available for interaction with CA does not show assembly. **d)** Pre-organized test system to study the influence of strand orientation and molecularity on assembly. **e)** Proposed hexameric rosette structure leads to d(A_n) triplex formation.

The error bars associated with the data points (despite multiple repetitions) are likely the result of these complications, and also arise from precipitation of growing assemblies and adhesion of mixture components to surfaces of the dialysis chamber.

Furthermore, we have confirmed that both the Watson-Crick and Hoogsteen faces of adenine are necessary for CA-mediated d(A_n) assembly. We synthesized a DNA oligonucleotide (d(deaza-A_{12})) composed of the adenine analog 7-deazaadenine, which lacks the hydrogen bonding N^7 atom, thus eliminating the Hoogsteen binding face of adenine. The d(deaza-A_{12}) strand incubated with excess CA (TAMg pH 4.5 and 6) did not display the change in CD signal characteristic with CA-mediated assembly. At pH 4.5 (below the 5.3 pKa of 7-deazaadenine) the spectrum remained unchanged, whereas at pH 6, it showed features reminiscent of Z-DNA (Fig. 2c) instead[41]. In both pH conditions, the spectral features are distinct from those observed CA-mediated assembly of d(A_n). This suggests that both hydrogen bonding faces of adenine are required for CA-mediated assembly.

To provide further evidence for this poly d(A_n):CA structure and examine possible strand orientations, a preorganized, intramolecular test system that is incapable of fiber elongation was designed (Fig. 2d, see Supplementary information Section Ve for details). In it, three d(A_{10}) regions are spatially tethered into position using two double-stranded clipping duplexes. These clipping strands are linked to each of the poly-adenine regions via a hexane-diol molecule to alleviate any potential structural strain that might occur during the self-assembly process. We used the two DNA duplexes to template four distinct d(A_n) environments (Fig. 2d): (i) 3 parallel d(A_{10}) strands (TP) ; (ii) 3 antiparallel (TA); (iii) two parallel (DP) and (iv) two antiparallel strands (DA). d($CA_{10}C$) was used as a non-preorganized control strand, termed "free d(A_{10})". CD spectra obtained after annealing the extremities of the DNA duplexes and adding CA showed both features for B-duplex DNA and the d(A_n):CA distinct peaks. Thermal denaturation was studied by monitoring the CD and UV-Vis signals at 252 nm, a wavelength that tracks both the d(A_n):CA and duplex environments. Two distinct melting transitions were noted, the first for the d(A_n):CA structure, and the second for the DNA duplexes at the extremities of the structure. Interestingly, the parallel triplex structure TP showed a T_M value that is closest to that of the CA-mediated assembly of "free d(A_{10})", while the antiparallel duplex DA construct showed the lowest T_M value (Supplementary Fig. S15). These observations agree with the proposed triplex-rosette structure, and importantly, suggest that there is no need to invoke a structure of molecularity higher than a triplex to explain the thermal denaturation results for free d(A_{10}).

Based on the VPO and DOSY studies with modified adenosine and CA units in toluene, the 1:1 CA:A ratio in the CA-mediated structure of poly(A) in aqueous solution determined by equilibrium dialysis, the participation of both the Watson-Crick and Hoogsteen faces of adenine in CA-mediated assembly, AFM height measurements as well as a model intramolecular system, we propose that CA-mediated d(A_n) assemblies are brought about by the formation of hexameric CA:A rosettes which lead to the growth of d(A_n) triplex fibers (Fig. 2e).

Mechanism of CA-mediated d(A_n) assembly

The material generated here not only has a unique internal structure, its robust elongation into micron-long polymers encourages its investigation for a number of biological and materials applications. The long-range organization of short oligonucleotides into fibers can be explained by end-to-end π-stacking of finite structures, or by a phase shift between the participating DNA strands during the growth process (Fig. 3a). The term 'phase shift' refers to staggered oligonucleotide assembly, producing overhangs that are further capable of bringing small constructs together longitudinally into fibers.

In order to examine the manner of CA-mediated fiber growth, we synthesized strand d(A_{20})-2F, labelled with a Cy3 dye at its 5'-end, and with a Cy5 dye in the middle of the strand (after the 10th adenosine unit) (Fig. 3b). Depending on whether fiber growth proceeds via blunt-ended stacking or staggered assembly, different fluorescence resonance energy transfer (FRET) behaviors are predicted for the dye pair. Due to increased proximity between donor and acceptor molecules in the case of staggered assembly, FRET efficiency (E_T) is expected to be higher than in the case of blunt-ended stacking, where identical molecules will be brought next to each other and the Cy3 and Cy5 dyes are separated by 10 bases. To set benchmarks for the two possible cases, we prepared fluorescently labeled DNA duplexes of mixed sequence as control structures, organizing the donor and acceptor molecules in distinct "blunt" and "staggered" arrangements (Fig. 3b and Supplementary Scheme S2). Comparison of the E_T values determined for d(A_{20})-2F incubated with 10 mM CA (TAMg pH 6) and the control duplexes (measured at the same pH, with or without added CA) suggests that CA-mediated fiber growth likely occurs through staggered rather than blunt assembly, as E_T for d(A_{20})-2F is significantly higher than that of the "blunt" duplex, and is similar to the E_T in the "staggered" duplex.

The process of CA-mediated d(A_n) assembly is a supramolecular polymerization. It may thus occur via two distinct mechanisms: isodesmic (step-growth), or cooperative (nucleation-growth) polymerization[42]. To investigate the mechanism of polymerization, we monitored the assembly outcome in relation to the concentrations of both d(A_n) and CA (pH 4.5). Interestingly, we observed that assembly only occurred above a critical CA concentration $[CA]_{cr}$. As an example, Figure 3c shows the CD spectra obtained at 40 μM d(A_{15}) in a range of CA concentrations. From these traces, the fraction of assembled d(A_{15}) was calculated and plotted *vs*. CA concentration for each d(A_{15}) concentration (Fig. 3d, Supplementary Fig. S17). In the range of 25-75 μM d(A_{15}), we noted non-sigmoidal curves with $[CA]_{cr}$ = 3 mM (corresponding to a CA:A ratio >3:1), below which no assembly was detected. This behavior suggests a cooperative nucleation-growth polymerization mechanism, whereby elongation of fiber nuclei to the observed higher-order structure only occurs above $[CA]_{cr}$ (Fig. 3e). Above this concentration, the intensity of the characteristic CD peak at 252 nm grew with increasing CA concentration, and UV-Vis showed a more pronounced hypochromic effect (Fig. 3c, Supplementary Fig. S18 and S19). Native polyacrylamide gel electrophoresis (PAGE) also showed the onset of d(A_{15}) assembly above $[CA]_{cr}$ (Supplementary Fig. S22). Hud et al. showed that dissociation of coralyne-induced poly(dA) duplexes into single-strands upon dilution occurs with a sharp transition as a function of concentration, indicating a similar cooperative assembly onset at a critical concentration.[33]

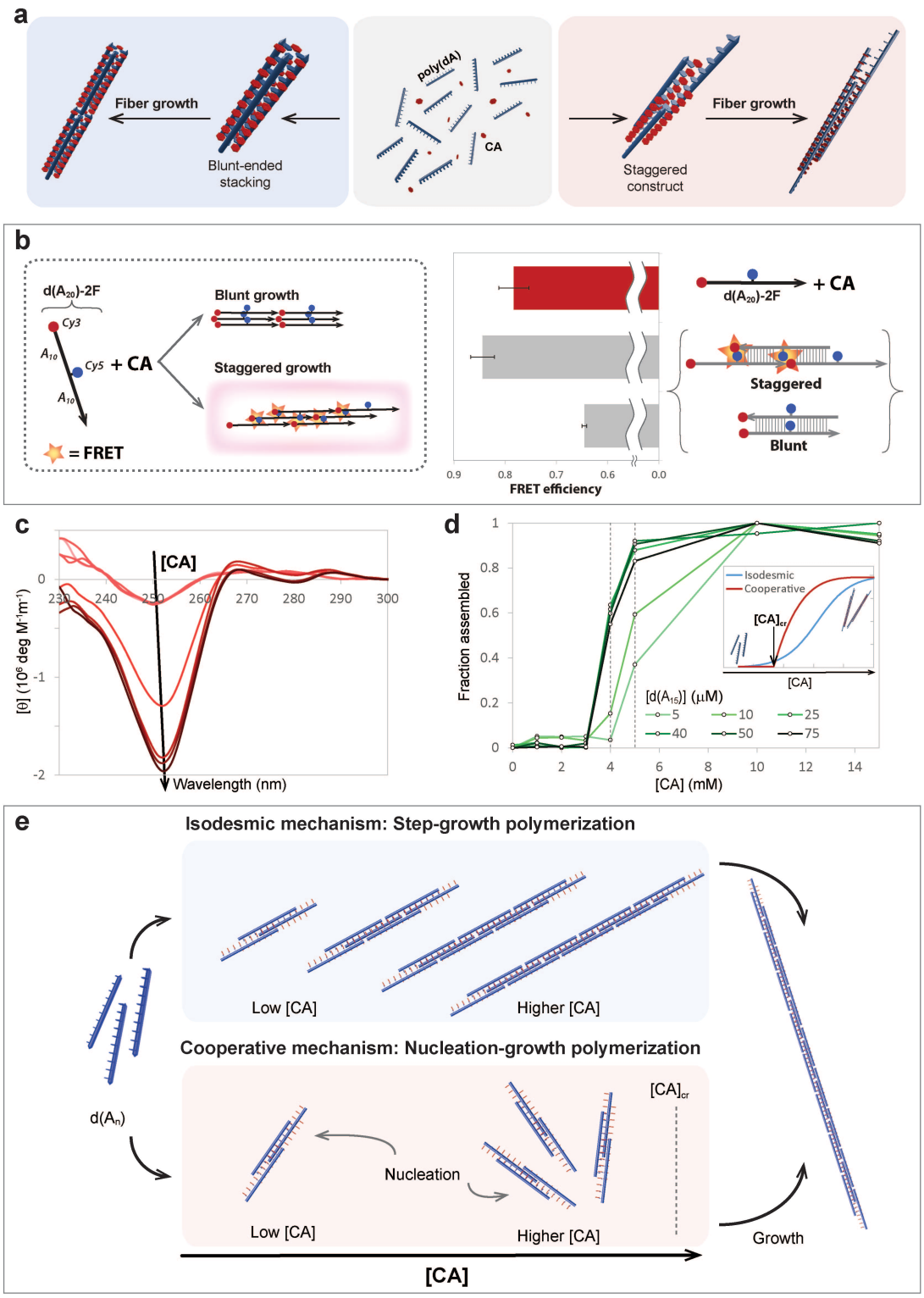


Figure 3. Mechanism of supramolecular polymerization. a) Fiber growth can occur through π-stacking of blunt-ended constructs or through a phase shifting mechanism. **b)** (LEFT) The FRET behavior of dually labelled strand d(A_{20})-2F was used to determine whether CA-mediated fiber growth occurs via blunt-

ended stacking or staggered assembly. (RIGHT) FRET efficiency comparison between the CA-mediated assembly of d(A_{20})-2F and control duplexes with "staggered" and "blunt" spatial organization of dyes. **c)** The intensity of the characteristic CD peak at 252 nm increases with CA concentration (data set shown for [d(A_{15})]=40 μM). **d)** Fraction of d(A_{15}) assembled as a function of CA concentration in a range of d(A_{15}) concentrations. The shape of the curves indicates that the supramolecular polymerization giving rise to CA-mediated poly(A) fibers proceeds by a cooperative mechanism (inset shows theoretical curves for isodesmic (blue) *vs*. cooperative (red) polymerization). **e)** Representation of isodesmic *vs.* cooperative supramolecular polymerization mechanisms.

Thermal denaturation of d(A_{15}) assemblies prepared in a range of CA concentrations (0 to 18 mM CA) was monitored by CD spectroscopy at 252 nm (Fig. 4a). Above the CA concentration threshold, the melting temperature (T_M) values determined from these curves increased with increasing CA concentration, with broadening of the melting transition. This structure stabilization effect occurring above $[CA]_{cr}$, and suggests that CA molecules may be participating in additional stabilizing interactions with the fibers, possibly through non-specific binding to their exterior. d(A_n) length also had an effect on the kinetic and thermodynamic properties of the assembly. Below d(A_{15}), the structures displayed monophasic transitions, however with increasing d(A_n) length, multistep melting transitions emerged (Fig. 4b). With the assembly of short strands, in order to ensure polymerization, strands in a triplex unit are likely shifted with respect to each other rather uniformly, creating a population of fibers that melt at the same temperature. With longer strands, there are now more possibilities for the length of the region of overlap between the strands that will still allow for fiber elongation. This will then create a distribution where fibers, or regions therein, melt at different temperatures such that multistep melting transitions are observed. An additional possibility in the mechanism of assembly of longer strands is the possibility of flexible strand ends bending back on the strand to form intramolecular CA-mediated associations.

These features grant access to simple assembly of fibers with a variety of definable melting temperatures. Importantly, the longer d(A_n) strands with multiphasic melting transitions had the same spectral CD features as the shorter, monophasic d(A_n) strands. This suggests the formation of a single class of structures for all d(A_n) lengths (as opposed to distinct structures with individual T_Ms). In all cases, the assembly kinetics were slow, but accelerated with increasing CA concentration and d(A_n) length.

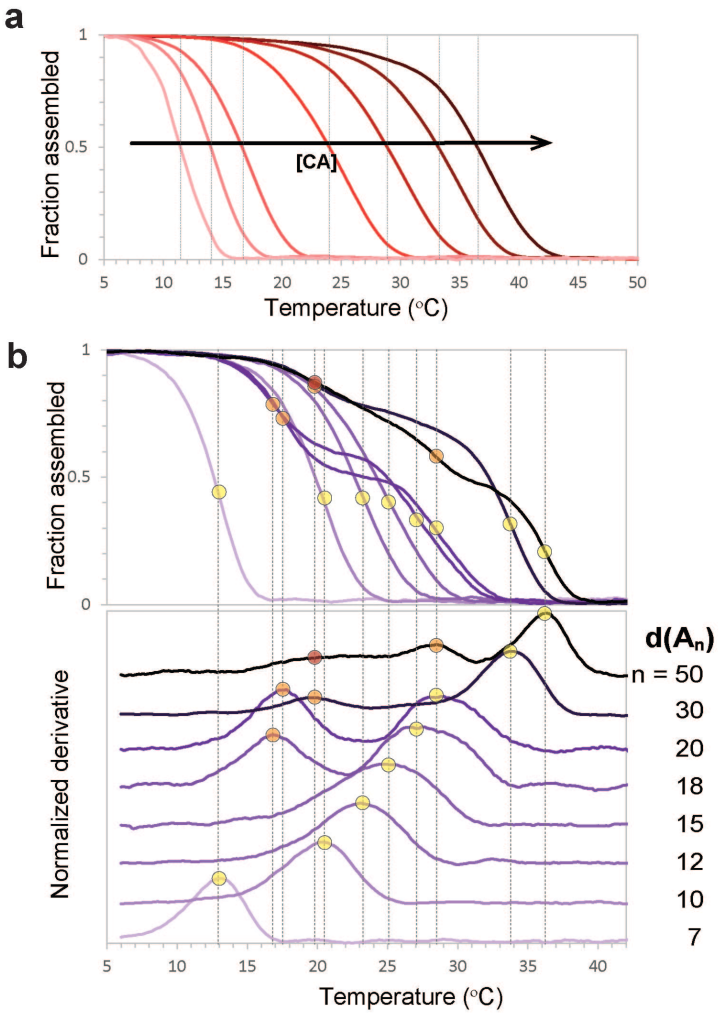


Figure 4. Thermal denaturation studies monitored by CD at 252 nm. a) Changes in T_M for d(A_{15}) as a function of CA concentration. **b)** d(A_n) T_M increases with growing strand length. For n>15, multiphasic melting transitions appear.

Probing the versatility of CA-mediated assembly

With its numerous biological functions and diverse self-assembly motifs, RNA is emerging as another exciting nucleic acid building block[52-54]. In addition, messenger RNA (mRNA) in eukaryotic cells has a 3'-poly(A) tail composed of ~200-300 adenosine residues[55,56]. This tail is essential for mRNA translation and plays an important role in its stability and maturation. We studied CA interaction with A_{15} RNA (r(A_{15})) (TAMg pH 4.5) by CD. In these conditions, the CD spectrum for r(A_{15}) acquired features very similar to the spectrum characteristic of d(A_{15}) assembly with CA and has similar thermal denaturation properties. In the absence of CA at this pH, RNA forms a parallel homoduplex[36,43], as evidenced by CD. AFM imaging in solution for r(A_{15}) with CA showed extensive formation of fibers (Fig. 5a), as in the case of d(A_{15}) under similar conditions. As polyadenylation is an important step in the

production of mature messenger RNA, CA-mediated poly(A) assembly can potentially find applications in probing or modulating protein expression.

Peptide nucleic acid (PNA), is an uncharged, achiral nucleic acid analog. We monitored the assembly of PNA A_7 ($p(A_7)$) with CA by UV-Vis spectroscopy. Upon addition of CA to $p(A_7)$ (in water or TAMg pH 4.5), a hypochromic effect and slight red-shift were observed for the nucleobase signal at 260 nm. Short fibers mostly around 100 nm in length and 2.0 ± 0.1 nm in height were observed by AFM (Fig.5b). On the whole, the assembly of poly(A) DNA, RNA and PNA can be robustly achieved with the simple external addition of CA to the system, and produces long-range fibers of uniform, well-defined structure.

To probe whether the fibers can act as templates for the organization of other materials, we tested the CA-mediated assembly of 5'-biotinylated $d(A_{15})$ and unmodified $d(A_{15})$ mixtures (TAMg pH 6) at different molar ratios. Long-range fibers were formed, as evidenced by AFM in solution for all mixtures. Addition of streptavidin (STV) visibly altered the original fibers, resulting in elongated features of greater height (Fig. 5c). At lower streptavidin concentrations and smaller biotinylated:unmodified $d(A_{15})$ ratios, individual STV molecules could be clearly observed on the fibers (Supplementary Fig. S25). With a large STV excess, the fibers appeared completely covered with the protein, and these fully substituted structures tend to shorten upon further standing. This experiment demonstrated that the fibers can be conjugated with streptavidin, allowing the CA-mediated organization of other biotinylated functional materials.

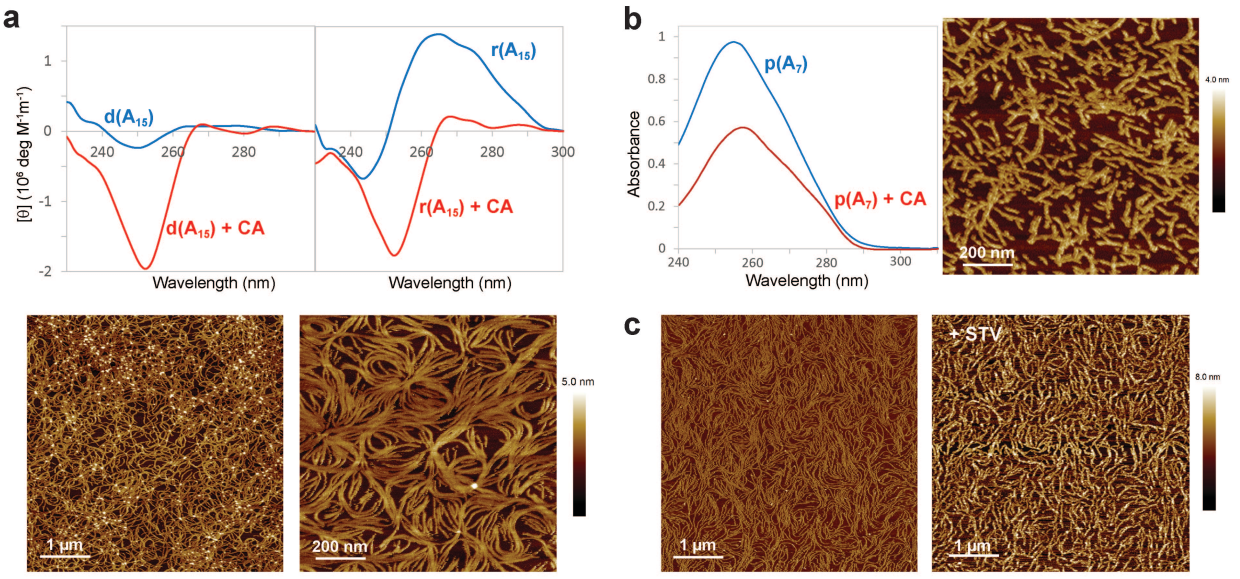


Figure 5. CA-mediated assembly of other oligonucleotides (RNA, PNA) and fiber functionalization. a) $r(A_{15})$ also assembles with CA as evidenced by the CD spectrum obtained and AFM imaging in solution. **b)** $p(A_7)$ shows hypochromicity in the presence of CA and forms short fibers observed by AFM in solution. **c)** CA-mediated fibers prepared with biotinylated $d(A_{15})$ can be decorated with the addition of streptavidin.

Conclusions

We have shown that a small molecule, cyanuric acid (CA), with three thymine-like faces, can induce the assembly of unmodified poly(adenine) strands into a new non-canonical motif. This results in the cooperative growth of micron-long supramolecular fibers. This assembly is solution-based and the resulting fibers are consistent and abundant, as determined by atomic force microscopy (in air and in solution) and dynamic light scattering. Circular dichroism spectroscopy reveals spectral features distinct from both B-DNA and previously reported poly(A) duplexes. A minimum threshold concentration of cyanuric acid, $[CA]_{cr}$, is required to initiate long-range assembly, and the supramolecular polymerization is consistent with a nucleation-growth mechanism. Investigations into the structure of the fibers support a proposed hexameric rosette hydrogen-bonded array as the underlying assembly motif. AFM height measurements, assembly experiments with an adenine analog and equilibrium dialysis are consistent with the formation of this motif, which brings together poly(A) strands into a triplex. This phenomenon is observed for DNA, RNA, and PNA backbones.

We believe that this study has important consequences in a number of areas. First, this is an inexpensive and extremely facile method to create new nucleic acid-based nanomaterials. The poly(A):CA material described here can be generated in large quantities. CA is an inexpensive and relatively non-toxic molecule widely used to stabilize chlorination in swimming pools. These structures are thus made from biocompatible units; they are stable over a range of pH values and their thermal denaturation temperatures can be finely controlled. The CA-mediated poly(A) fibers may have interesting applications in tissue engineering and drug delivery. Second, we have shown that this new motif can coexist with regular DNA Watson-Crick base-pairing. This method then introduces a new orthogonal interaction to expand the 4-letter DNA alphabet. Along with intercalator-induced poly(A) assembly (such as coralyne), it is expected to expand the repertoire and complexity of DNA nanostructures, with reduced 'errors' or defects, as compared to the use of a simple A-T, G-C duplex code. Third, the guest-induced assembly of DNA strands by a small molecule with more than one hydrogen bond face may be relevant in chemical biology. Long strands of poly(A) are added to messenger RNA before its translation to a protein and poly(purine) stretches in general are biologically important. Our study raises the possibility that cyanuric acid (or derivatives) can be used to probe and potentially modulate protein translation. There are a number of biologically relevant bifacial "Janus" molecules, such as 8-oxo-guanine (G/T) or folic acid (G/C), among many others, and they may also be able to form higher-order structures with nucleic acid strands. To our knowledge, the supramolecular behavior of small hydrogen bonding molecules with nucleic acids has not been examined in detail. Finally, the fact that a small molecule such as cyanuric acid can organize poly(A) strands into long-range polymers in aqueous media may have implications in prebiotic chemistry, specifically in answering the question of how individual nucleobases originally assembled into a nucleic acid polymeric backbone.

Acknowledgements

The authors acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs program, the Canada Foundation for Innovation (CFI), the Centre for Self-Assembled Chemical Structures (CSACS) and the Canadian Institute for Advanced Research (CIFAR) for financial support. N.A. and A.G. thank NSERC and FQRNT for doctoral fellowships. H.F.S. is a Cottrell Scholar of the Research Corporation.

Author contributions

H.F.S and F.A. conceived the study. N.A., F.A. and H.F.S designed the experiments. N.A. performed the experimental studies with assistance from A.A.G. (DNA synthesis), C.J.S. (adenosine/CA crystal growth and X-ray analysis), A.P. (vapor pressure osmometry) and V.T. (synthesis of TIPDS-dA and hex-CA). N.A. and H.F.S analyzed data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Methods

Preparation of CA-mediated poly(A) assembly and analysis by CD and UV-Vis spectroscopies. Poly(A) (concentrations ranging from 5 μM to 75 μM) and CA (concentrations ranging from 0 to 18 mM) were thermally annealed (50 °C for 10 minutes, followed by cooling to 5 °C at a rate of 0.5 °C/min) in 200 mM NaCl (unbuffered) or in 1xTAMg at pH 4.5 (buffer) and incubated at 5 °C overnight prior to any measurements. CD and UV-Vis spectra were obtained (200 nm to 350 nm range, 100 nm/min scan rate, 1 nm bandwidth, 2 accumulations) at 5°C. Thermal denaturation (1 °C/min rate) was monitored by CD and UV-Vis absorbance at 252 nm and 0.5 °C intervals.

Atomic force microscopy (AFM) imaging in air and in solution. Imaging in air was carried out under ambient conditions in air using a Multimode scanning probe microscope and Nanoscope IIIa controller (Digital Instruments, Santa Barbara, CA). Topography and phase contrast were simultaneously acquired in tapping mode with silicon probes (AC160TS from Olympus, nominal spring constant 42 N/m, resonant frequency of 300 kHz and tip radius < 10 nm). Samples were prepared as described above and diluted as necessary for deposition of a 5 μL drop containing ~0.1 nmol of poly(A). Upon deposition on freshly cleaved mica, samples were incubated at 5 °C for 4-6 hours, followed by drying under vacuum overnight. Imaging in solution was carried out under ambient conditions or with temperature maintained at 12 °C using a MultiMode 8 microscope with a Nanoscope V controller (Bruker) in ScanAsyst mode. Silicon nitride levers with nominal spring constant of 0.7 N/m, resonant frequency of 150 kHz and tip radius <

20 nm were used (ScanAsyst Fluid). Samples were prepared as described above and contained 5 to 10 µM poly(A). A 5 µL drop was deposited on a freshly cleaved mica surface (or a surface pre-treated with $NiCl_2$ solution to promote sample adhesion) and incubated for 2 minutes followed by addition of buffer appropriate to the sample to the closed MTFML (Bruker, Santa Barbara, CA) fluid cell and imaging.

Equilibrium dialysis. Equilibrium dialysis was used to establish the CA:A stoichiometry in the CA-mediated assembly of poly(A). Equal volumes (25 µL) of $d(A_{15})$ (40 µM) and CA (2, 3, 4, 4.5, 5, 5.5, 6, 6.5, 7, 8, 10, 12 and 15 mM) were allowed to equilibrate at 4 °C for 24 hours with gentle agitation, separated by a dialysis membrane (5000 Dalton molecular weight cut-off). Both solutions were then analyzed by RP-HPLC to assess the equilibrium concentration of CA on either side of the membrane. The free concentration of CA, $[CA]_{free}$, and the ratio of CA bound per A, CA:A, were calculated and plotted to obtain a binding curve. The analysis was performed in triplicate 4 times and all results were compiled.

References

1 Watson, J. D. & Crick, F. H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738, (1953).

2 Choi, J. & Majima, T. Conformational changes of non-B DNA. *Chem Soc Rev* **40**, 5893-5909, (2011).

3 Aldaye, F. A., Palmer, A. L. & Sleiman, H. F. Assembling materials with DNA as the guide. *Science* **321**, 1795-1799, (2008).

4 Egholm, M., Buchardt, O., Nielsen, P. E. & Berg, R. H. Peptide Nucleic-Acids (PNA) - Oligonucleotide Analogs with an Achiral Peptide Backbone. *J. Am. Chem. Soc.* **114**, 1895-1897, (1992).

5 Egholm, M. *et al.* PNA hybridizes to complementary oligonucleotides obeying the Watson-Crick hydrogen-bonding rules. *Nature* **365**, 566-568, (1993).

6 Chen, M. C. *et al.* Spontaneous prebiotic formation of a beta-ribofuranoside that self-assembles with a complementary heterocycle. *J. Am. Chem. Soc.* **136**, 5640-5646, (2014).

7 Malyshev, D. A. *et al.* A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509**, 385-388, (2014).

8 Yang, Z., Chen, F., Chamberlin, S. G. & Benner, S. A. Expanded genetic alphabets in the polymerase chain reaction. *Angew. Chem. Int. Ed. Engl.* **49**, 177-180, (2010).

9 Teo, Y. N. & Kool, E. T. DNA-Multichromophore Systems. *Chem. Rev.* **112**, 4221-4245, (2012).

10 Winnacker, M. & Kool, E. T. Artificial genetic sets composed of size-expanded base pairs. *Angew. Chem. Int. Ed. Engl.* **52**, 12498-12508, (2013).

11 Sessler, J. L. & Jayawickramarajah, J. Functionalized base-pairs: versatile scaffolds for self-assembly. *Chem. Commun.*, 1939-1949, (2005).

12 Woo, S. & Rothemund, P. W. K. Programmable molecular recognition based on the geometry of DNA nanostructures. *Nat. Chem.* **3**, 620-627, (2011).

13 Yang, H. *et al.* Metal-nucleic acid cages. *Nat. Chem.* **1**, 390-396, (2009).

14 Piccirilli, J. A., Krauch, T., Moroney, S. E. & Benner, S. A. Enzymatic Incorporation of a New Base Pair into DNA and Rna Extends the Genetic Alphabet. *Nature* **343**, 33-37, (1990).

15 Kimoto, M., Kawai, R., Mitsui, T., Yokoyama, S. & Hirao, I. An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res.* **37**, e14, (2009).

16 Ogawa, A. K. *et al.* Efforts toward the expansion of the genetic alphabet: Information storage and replication with unnatural hydrophobic base pairs. *J. Am. Chem. Soc.* **122**, 3274-3287, (2000).
17 Wu, Y. Q. *et al.* Efforts toward expansion of the genetic alphabet: Optimization of interbase hydrophobic interactions. *J. Am. Chem. Soc.* **122**, 7621-7632, (2000).
18 Meggers, E., Holland, P. L., Tolman, W. B., Romesberg, F. E. & Schultz, P. G. A novel copper-mediated DNA base pair. *J. Am. Chem. Soc.* **122**, 10714-10715, (2000).
19 Shionoya, M. & Tanaka, K. Synthetic incorporation of metal complexes into nucleic acids and peptides directed toward functionalized molecules. *Bull. Chem. Soc. Jpn.* **73**, 1945-1954, (2000).
20 Martin-Ortiz, M., Gomez-Gallego, M., Ramirez de Arellano, C. & Sierra, M. A. The selective synthesis of metallanucleosides and metallanucleotides: a new tool for the functionalization of nucleic acids. *Chem. Eur. J.* **18**, 12603-12608, (2012).
21 Chen, H., Meena & McLaughlin, L. W. A Janus-Wedge DNA Triplex with A-W1-T and G-W2-C Base Triplets. *J. Am. Chem. Soc.* **130**, 13190-13191, (2008).
22 Shin, D. & Tor, Y. Bifacial nucleoside as a surrogate for both T and A in duplex DNA. *J. Am. Chem. Soc.* **133**, 6926-6929, (2011).
23 Yang, H. Z., Pan, M. Y., Jiang, D. W. & He, Y. Synthesis of Janus type nucleoside analogues and their preliminary bioactivity. *Org. Biomol. Chem.* **9**, 1516-1522, (2011).
24 Pan, M.-Y. *et al.* Janus-type AT nucleosides: synthesis, solid and solution state structures. *Org. Biomol. Chem.* **9**, 5692-5702, (2011).
25 Largy, E., Liu, W., Hasan, A. & Perrin, D. M. Base-pairing behavior of a carbocyclic Janus-AT nucleoside analogue capable of recognizing A and T within a DNA duplex. *ChemBioChem* **14**, 2199-2208, (2013).
26 Marsh, A., Silvestri, M. & Lehn, J. M. Self-complementary hydrogen bonding heterocycles designed for the enforced self-assembly into supramolecular macrocycles. *Chem. Commun.*, 1527-1528, (1996).
27 Zeng, Y., Pratumyot, Y., Piao, X. & Bong, D. Discrete assembly of synthetic peptide-DNA triplex structures from polyvalent melamine-thymine bifacial recognition. *J. Am. Chem. Soc.* **134**, 832-835, (2012).
28 Chaput, J. C. & Switzer, C. A DNA pentaplex incorporating nucleobase quintets. *PNAS* **96**, 10614-10619, (1999).
29 Jiang, D. & Seela, F. Oligonucleotide duplexes and multistrand assemblies with 8-aza-2'-deoxyisoguanosine: a fluorescent isoG(d) shape mimic expanding the genetic alphabet and forming ionophores. *J. Am. Chem. Soc.* **132**, 4016-4024, (2010).
30 Polak, M. & Hud, N. V. Complete disproportionation of duplex poly(dT)·poly(dA) into triplex poly(dT)·poly(dA)·poly(dT) and poly(dA) by coralyne. *Nucleic Acids Res.* **30**, 983-992, (2002).
31 Persil, Ö., Santai, C. T., Jain, S. S. & Hud, N. V. Assembly of an antiparallel homo-adenine DNA duplex by small-molecule binding. *J. Am. Chem. Soc.* **126**, 8644-8645, (2004).
32 Jain, S. S., Polak, M. & Hud, N. V. Controlling nucleic acid secondary structure by intercalation: effects of DNA strand length on coralyne‐driven duplex disproportionation. *Nucleic Acids Res.* **31**, 4608-4615, (2003).
33 Cetinkol, O. P. & Hud, N. V. Molecular recognition of poly(A) by small ligands: an alternative method of analysis reveals nanomolar, cooperative and shape-selective binding. *Nucleic Acids Res.* **37**, 611-621, (2009).
34 Park, K. S. & Park, H. G. A DNA-templated silver nanocluster probe for label-free, turn-on fluorescence-based screening of homo-adenine binding molecules. *Biosens. Bioelectron.* **64**, 618-624, (2015).

35 Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discovery* **10**, 261-275, (2011).
36 Rich, A., Davies, D. R., Crick, F. H. & Watson, J. D. The molecular structure of polyadenylic acid. *J Mol Biol* **3**, 71-86, (1961).
37 Chakraborty, S., Sharma, S., Maiti, P. K. & Krishnan, Y. The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res.* **37**, 2810-2817, (2009).
38 Zerkowski, J. A., Seto, C. T. & Whitesides, G. M. Solid-state structures of rosette and crinkled tape motifs derived from the cyanuric acid melamine lattice. *J. Am. Chem. Soc.* **114**, 5473-5475, (1992).
39 Rakotondradany, F. *et al.* Hydrogen-bond self-assembly of DNA-analogues into hexameric rosettes. *Chem. Commun.*, 5441-5443, (2005).
40 Roy, B., Bairi, P. & Nandi, A. K. Supramolecular assembly of melamine and its derivatives: nanostructures to functional materials. *RSC. Adv.* **4**, 1708-1734, (2014).
41 Berova, N., Nakanishi, K. & Woody, R. *Circular Dichroism: Principles and Applications*. (Wiley-VCH, New York, 2000).
42 De Greef, T. F. *et al.* Supramolecular polymerization. *Chem Rev* **109**, 5687-5754, (2009).
43 Safaee, N. *et al.* Structure of the parallel duplex of poly(A) RNA: evaluation of a 50 year-old prediction. *Angew. Chem. Int. Ed. Engl.* **52**, 10370-10373, (2013).
44 Krishnan, Y. & Simmel, F. C. Nucleic acid based molecular devices. *Angew. Chem. Int. Ed. Engl.* **50**, 3124-3156, (2011).
45 Ma, M. & Bong, D. Determinants of cyanuric acid and melamine assembly in water. *Langmuir* **27**, 8841-8853, (2011).
46 Cafferty, B. J. *et al.* Efficient self-assembly in water of long noncovalent polymers by nucleobase analogues. *J. Am. Chem. Soc.* **135**, 2447-2450, (2013).
47 Whitesides, G. M. *et al.* Noncovalent Synthesis: Using Physical-Organic Chemistry To Make Aggregates. *Acc. Chem. Res.* **28**, 37-44, (1995).
48 Timmerman, P. *et al.* NMR diffusion spectroscopy for the characterization of multicomponent hydrogen-bonded assemblies in solution. *J. Chem. Soc., Perkin Trans. 2*, 2077-2089, (2000).
49 Davies, R. J. & Davidson, N. Base pairing equilibria between polynucleotides and complementary monomers. *Biopolymers* **10**, 1455-1479, (1971).
50 Davies, R. J. Complexes of poly(adenylic acid) with complementary monomers. *Eur. J. Biochem.* **61**, 225-236, (1976).
51 Pyle, A. M. *et al.* Mixed-ligand complexes of ruthenium(II): factors governing binding to DNA. *J. Am. Chem. Soc.* **111**, 3051-3058, (1989).
52 Chworos, A. *et al.* Building programmable jigsaw puzzles with RNA. *Science* **306**, 2068-2072, (2004).
53 Hao, C. *et al.* Construction of RNA nanocages by re-engineering the packaging RNA of Phi29 bacteriophage. *Nat. Commun.* **5**, (2014).
54 Geary, C., Rothemund, P. W. K. & Andersen, E. S. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804, (2014).
55 Weill, L., Belloc, E., Bava, F.-A. & Mendez, R. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nat. Struct. Mol. Biol.* **19**, 577-585, (2012).
56 Di Giammartino, D. C., Nishida, K. & Manley, J. L. Mechanisms and consequences of alternative polyadenylation. *Mol. Cell* **43**, 853-866, (2011).