Chemistry
A European
Journal

Chemistry Europe
European Chemical Societies Publishing

Accepted Article

Title: The Unexpected Base-pairing Behavior of Cyanuric Acid in RNA and Ribose versus Cyanuric Acid Induced Helicene Assembly of Nucleic Acids: Implications for the pre-RNA Paradigm.

Authors: Brooke Anderson, Kévin Fauche, Suneesh Karunakaran, Jayasudhan Yerabolu, Nicholas V Hud, and Ramanarayanan Krishnamurthy

To be cited as: *Chem. Eur. J.* 10.1002/chem.202004397

Link to VoR: https://doi.org/10.1002/chem.202004397

WILEY-VCH

10.1002/chem.202004397

WILEY-VCH

RESEARCH ARTICLE

The Unexpected Base-pairing Behavior of Cyanuric Acid in RNA and Ribose versus Cyanuric Acid Induced Helicene Assembly of Nucleic Acids: Implications for the pre-RNA Paradigm.

Brooke A. Anderson[a,c,‡], Kévin Fauché[a,c,‡], Suneesh C. Karunakaran[b,c], Jayasudhan R. Yerabolu[a,c], Nicholas V. Hud[b,c], Ramanarayanan Krishnamurthy*[a,c]

[a] Dr. B. A. Anderson, Dr. K. Fauché, Dr. J. R. Yerabolu, Prof. R. Krishnamurthy*
Department of Chemistry
The Scripps Research Institute
10550 North Torrey Pines Road, La Jolla, CA 92037, USA
E-mail: rkrishna@scripps.edu
[b] Dr. S. C. Karunakaran, Prof. N. V. Hud
School of Chemistry and Biochemistry
Parker H. Petit Institute for Bioengineering and Bioscience
Georgia Institute of Technology, 901 Atlantic Drive, Atlanta, GA 30332, USA
[c] NSF-NASA Center for Chemical Evolution, Atlanta, GA 30332, USA
[‡] These authors contributed equally to this work.

Supporting information for this article is given via a link at the end of the document.

Abstract: Cyanuric acid heterocycle (CA) forms supramolecular structures with adenine nucleobases/nucleosides and oligonucleotides leading to speculation that they can act as forerunners to RNA. Herein, we studied the assembly behavior of RNA containing CA and CA-ribose nucleoside. Contrary to previous reports, CA in RNA and as the ribonucleoside leads to destabilization of supramolecular assemblies, which led to reevaluation of the CA-adenine hexameric rosette structure. We propose an unprecedented non-covalent supramolecular helicene structure to account for the striking difference in behavior, which has implications for novel paradigms for reorganizing the structures of nucleic acids, synthesis of long helicenes, and pre-RNA world paradigms. The results caution against extrapolating the self-assembly behavior of individual heterocycles from the level of monomers to oligomers – since the base-paring properties of (non)canonical nucleobases are impacted by the type of oligomeric backbone to which they are attached.

Introduction

In a pre-RNA world scenario, there is interest to discover one-pot self-assembly of small heterocycles and their nucleosides into proto-biopolymers that could act as forerunners to RNA (Figure 1).[1] Prebiotically plausible triaminopyrimidine (TAP), cyanuric acid (CA) and barbituric acid (BA) with melamine (Mel) and their derivatives have been shown to spontaneously self-assemble into hexameric rosette structures[2] (Figure 1), providing opportunities for these noncovalent self-assemblies to form proto-oligonucleic acids.[3] Sleiman lab has shown that poly(A) can assemble into a higher-order hexad-structure in the presence of heterocyclic cyanuric acid (Figure 1).[4] Li et al showed cyanuric acid derivatives self-assemble with purine nucleobases and nucleotides (Figure 1).[5] More recently, Asanuma lab demonstrated hexaplex oligonucleotide constructs with D-threoninol backbone with cyanuric acid, triaminopyrimidine, and

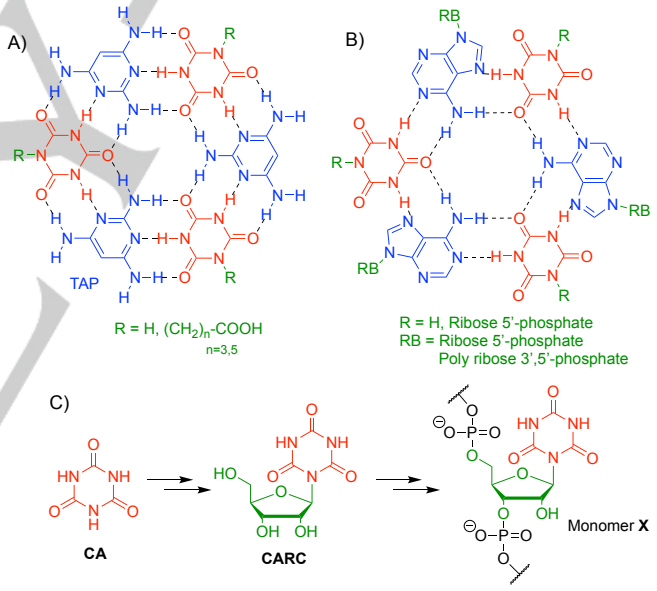
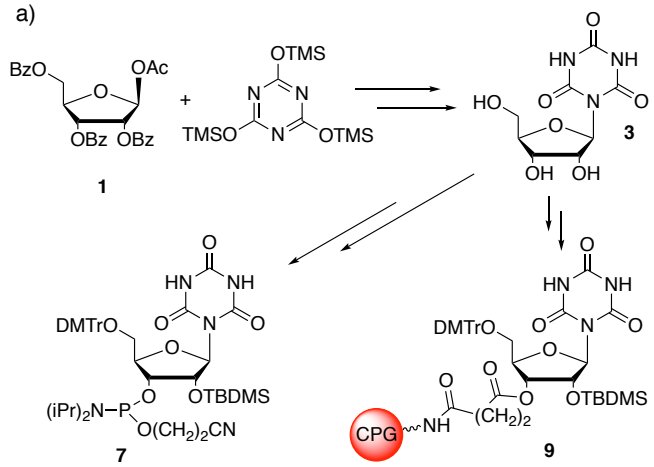


Figure 1. Cyanuric acid building block and its self-assemblies. Proposed hexad structure of cyanuric acid derivatives with (A) triaminopyrimidine (TAP)[2a, 2b] and (B) adenosine-5'-monophosphate (AMP)[5] and poly(A)[4]. (C) Structure of cyanuric acid CA, nucleoside CARC and the corresponding CARC-nucleotide (monomer X) as incorporated in the oligonucleotide.

aminopyrimidine.[6] Encouraged by these results, and drawing on the proposed pre-RNA paradigms,[1, 5] we initiated a 'top-down' approach at the oligomeric level to study the base-pairing and the self- assembly properties of cyanuric acid containing oligonucleotides. We synthesized the cyanuric acid β-ribofuranoside phosphoramidite **7** (Scheme 1), incorporated it into a series of oligoribonucleotides (ONs), and investigated their base-pairing and self-assembly properties (under a variety of concentrations, buffers, pH conditions) by

10.1002/chem.202004397

a)



b)

ON1 5′-r(A_8XA_7)
ON2 5′-r$(XA)_8$
ON3 5′-r(X_8A_8)
ON4 5′-FAM-r(X_{16})
ON5 5′-r($AX_{16}A$)

R1 5′-r(A_8UA_7)
R2 5′-r$(UA)_8$
R3 5′-r(U_8A_8)
R4 5′-r(U_{16})

R1c 3′-r(U_8AU_7)
R4c 3′-r(A_{16})
D1c 3′-d(A_{16})

Scheme 1. (a) Synthesis of 1-(β-D-ribofuranosyl)cyanuric acid phosphoramidite **7** and the CPG derivative **9**. (b) List of oligonucleotides used in this study. Cyanuric acid containing oligos **ON1**-**ON5** and the corresponding RNA and DNA oligos used in this study.

temperature-dependent UV-absorbance spectroscopy, circular dichroism (CD) spectroscopy, gel-electrophoresis and atomic force microscopy (AFM). Based on the unexpected results from the cyanuric acid containing oligoribonucleotides, we further investigated the potential of the cyanuric acid-ribose conjugate (CARC) to assemble with RNA and DNA oligonucleotides and compared the results of these investigations with those involving the free heterocycle cyanuric acid (CA). We observed an unexpected loss in propensity to form assemblies with RNA and DNA in going from the free CA- heterocycle to the CA-nucleoside (CARC) and CA-containing oligonucleotide. This unanticipated loss of propensity to assemble, both at the level of the CARC and the CA-oligoribonucleotide, forced a reevaluation of the hexad-models that have been proposed in the literature for the CA-heterocycle mediated self-assemblies (Figure 1).[4-5] Herein, based on molecular models (built from HGS Biochemistry kits), we suggest an alternative helicene model that is consistent with the observations reported here and by Sleiman and coworkers[4]. Further support for the helicene structure was obtained by Sherrill and coworkers, via *in silico* molecular dynamic simulations, the details of which are reported in the accompanying paper.[7] Finally, we place these unexpected observations in the context of a pre-RNA world paradigm and discuss its implications for pre-RNA—RNA transitions and reprogramming the nucleic acid structure by small molecules

Results and Discussion.

We began by synthesizing the cyanuric acid containing oligonucleotides. Starting from commercially available ribofuranose derivative **1** and following established literature protocols[8], phosphoramidite **7** (Scheme 1a) was synthesized in an overall yield of ~48% over 6 steps (Scheme S1). Controlled pore glass (CPG) **9** was synthesized and used in oligonucleotide synthesis in which a cyanuric acid is located at the 3′ terminus.[9] Development of successful synthetic (stepwise coupling yields of 95-97%), deprotection and purification/isolation protocols enabled the synthesis of cyanuric acid incorporated oligonucleotides **ON1** – **ON5** (Scheme 1b). The composition and purity of all modified ONs was ascertained by MALDI-MS analysis and reversed-phase HPLC (Table S1). We first evaluated the base-pairing properties of cyanuric acid containing oligonucleotides **ON1** – **ON3** by temperature-dependent UV- and CD-spectroscopy at various concentrations of 2-40 μM, under various buffer and salt (NaCl, $MgCl_2$) conditions and at three different pH values of 7.0, 6.0, and 4.5 (Tables S2 and S3). Acidic pHs were chosen in order to maintain the protonated version of monomer **X**, which has a pKa of around 6.5 (Figure S12). None of the temperature-dependent UV-scans, thermal- or CD-melts of **ON1** – **ON3** showed clear cut evidence for hydrogen bond mediated complex formation (Table S2, Figures 2 and S34-S57). For example, in the UV-scans there was no clear characteristic shoulder around 280-290 nm as was seen for r(A_{16})+CA combinations, which is attributed to supramolecular self-assemblies such as J-aggregates[4]. Thermal denaturation (UV-T_m) studies showed either duplex destabilization relative to unmodified reference RNA duplexes (for **ON1**+**R1c** versus **R1**+**R1c**, Table S2 and Figure S34) or undefined broad transitions (for self-complementary sequences **ON2** and **ON3**, Figure 2a, Tables S2 and S3, Figures S36-57). Determining accurate UV-T_m was difficult due to broadening of the denaturation curve (e.g. multiple transitions) and lack of clear upper and/or lower baselines (Figure 2A). The comparisons with the corresponding RNA standards **R2** and **R3** (Figures S36-S57) were again inconclusive. The sigmoidal UV-melt curve could not be easily distinguished between cyanuric acid having potential interaction with its dual-hydrogen bonding faces at the oligomeric level[10] and adenine-adenine(H^+) pairings that is ubiquitous in homo-adenine RNA oligonucleotides under acidic conditions.[11] Based on the previous study showing that CA-heterocycle aids the self-assembly of poly(A) oligos[4], we also investigated the behavior of **ON2** and **ON3** by temperature-dependent UV-spectroscopy in the presence of 15 mM cyanuric acid heterocycle (at pH 4.5 in the TAMg buffer). However, there was no difference in the temperature-dependent UV-curves between the presence and absence of cyanuric acid (Figures S58-S59), suggesting that the addition of cyanuric acid was not inducing supramolecular self- assembly in the self-complementary oligomers. CD-spectroscopic studies (Figure 2B), suggested that **ON2** forms a structure different from that of **ON3** (and also **R2** and **R3**). However, neither of the self-complementary sequences **ON2** (containing eight alternating cyanuric acid residues) and **ON3** (with eight consecutive cyanuric acid modifications) appeared to follow the previous observations of an overwhelmingly (and almost only) negative peak occurring at ~252 nm for d(A_{15}) or r(A_{15}) in the presence of heterocyclic cyanuric acid (Figures S36-S57).[4]

Furthermore, polyacrylamide gel electrophoresis (PAGE) studies of **ON2** and **ON3** under various pH conditions were conducted and compared with the PAGE behavior of rA_{16}+CA, which is known to form the supramolecular self-assemblies (Figures 2C and S81).[4] Both **ON2** and **ON3** showed a slower moving band at pH 4.5, which gradually decreased in intensity

10.1002/chem.202004397

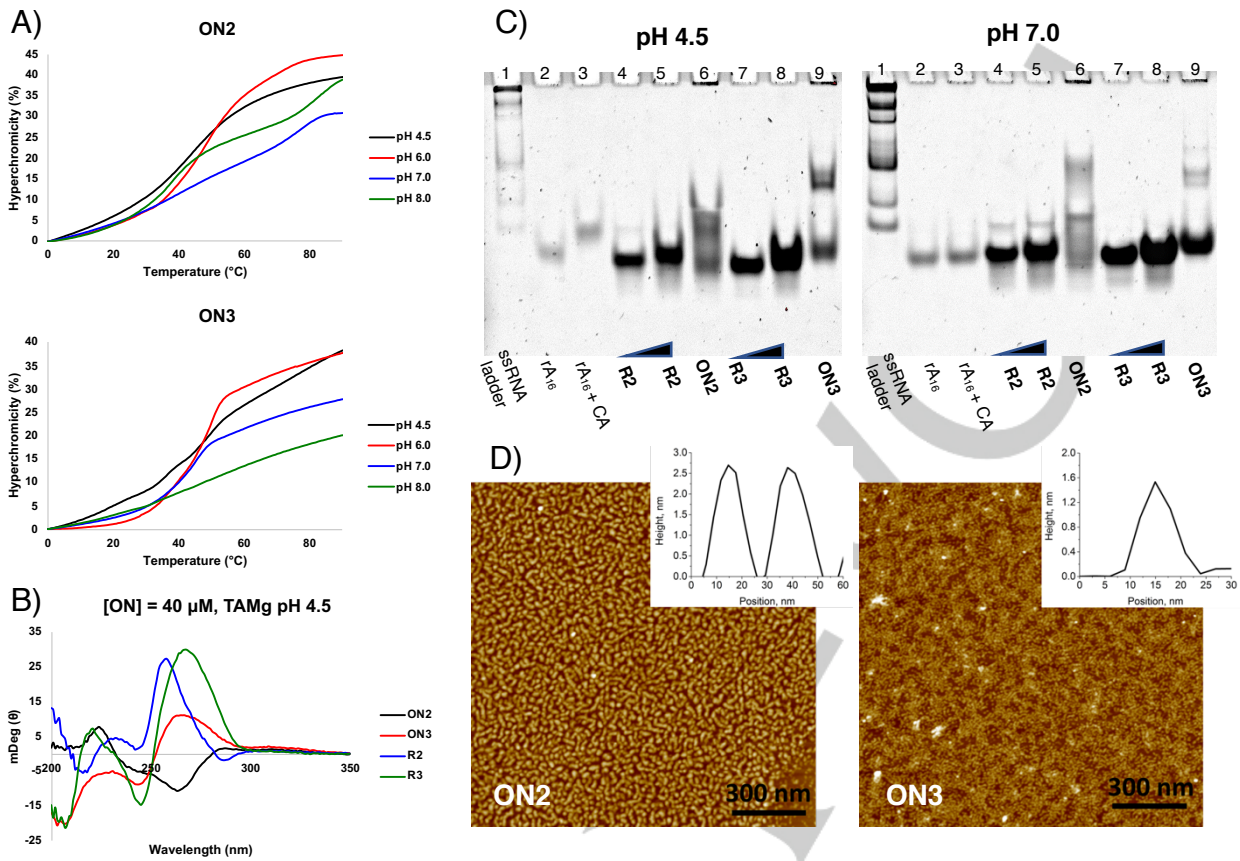


Figure 2. Base-pairing behavior of cyanuric acid containing oligos **ON1**, **ON2** and **ON3** sequences in TAMg buffer ([tris] = 40 mM, [$MgCl_2$] = 7.6 mM). A) thermal denaturation curves of **ON2** (top) and **ON3** (bottom) at different pH (4.5, 6.0, 7.0, or 8.0), heating curve, [ON] = 40 µM; B) comparison of the CD spectra of **ON2**, **ON3**, **R2** and **R3** at 0°C at pH 4.5, [ON] = 40 µM. ; C) PAGE experiments at pH 4.5 (left) and 7.0 (right) with [ON] = 50 µM. Lane 1: low range ssRNA ladder (50, 80, 150, 300, 500, 1000-mer); lane 2: single-stranded rA_{16}; lane 3: rA_{16} + 15 mM CA; lane 4: 0.25 nmol **R2**; lane 5: 0.50 nmol **R2**; lane 6: **ON2**; lane 7: 0.25 nmol **R3**; lane 8: 0.50 nmol **R3**; lane 9: **ON3**; D) AFM images of **ON2** (left) and **ON3** (right). For methods and conditions of measurements see SI.

with increasing pH (Figure 2C), but with one important difference. The band for single-strand **ON2** and **ON3** was always present at lower pH as opposed to the rA_{16}+CA lane where the band observed for rA_{16} alone completely disappeared. Moreover, while dose-response gels of rA_{16}+CA showed behavior consistent with complex formation[4], the relative intensity for all bands of **ON2** and **ON3** remained constant with increasing concentrations (Figure S82). This observation indicated the absence of any conversion of the self-complementary single-strands to higher-complexes. Oligomer **ON2** (analyzed under similar conditions as above, except without the use of CA) resulted in the formation of a large streaking/smeared band with increasing ON concentrations at all pHs (Figures 2C and S81). For **ON3**, a distinct separation of two bands was observed at all pH conditions studied, and the band intensity fades as pH increases (Figures 2C and S81). At higher concentrations, both **ON2** and **ON3** additionally exhibit some immobile 'residue' in the gel loading wells, suggestive of either high molecular weight aggregates or precipitation. At this point, given the results from the temperature-dependent UV, CD and gel studies with self-complementary sequences **ON2** and **ON3** we are forced to conclude that – although there may be some level of self-aggregation – there is no evidence supporting formation of higher-order assemblies by these oligonucleotides as is observed for the free cyanuric acid interactions with poly(dA) or poly(rA)[4]. This conclusion was further supported by solution AFM-studies on **ON2** and **ON3** under the TAMg buffer at pH 4.5. While some indiscriminate aggregates were observed, no well-defined structures (such as long fibers) were seen (Figure 2D) in contrast to observations of Sleiman and coworkers in their studies[4] of poly(rA) or poly(dA) with CA-heterocycle. The contrasting pH dependent UV-T_m behavior between **ON2** and **ON3** (Figure 2A) seemed to suggest that a consecutive string of cyanuric acid residues in the oligonucleotide may be more apt to forming ordered structures – an expectation that would be in line with previous observations of Sleiman and coworkers[4], where they showed that free cyanuric acid heterocycle is able to base pair with d(A_{15}) and r(A_{15}). Therefore, we synthesized 6-carboxyfluorescein (FAM) labelled cyanuric acid modified RNA hexadecamer **ON4** to explore complex formation with the corresponding complementary sequence such as **R4c**, r(A_{16}).[6] We also synthesized the **ON5** which contains sixteen cyanuric acid without the FAM-label (but with adenine residues at both ends). At pH 4.5, **ON4** + ^{dabcyl}**R4c** showed a sharp, sigmoidal transition with no observable hysteresis upon cooling (Figure 3A and Table S2). At first glance, this seemed promising for an interaction between **ON4:** ^{dabcyl}**R4c**. However, the thermal stability was similar to that of the unmodified reference duplex **R4**:**R4c** (Figure 3A).

10.1002/chem.202004397

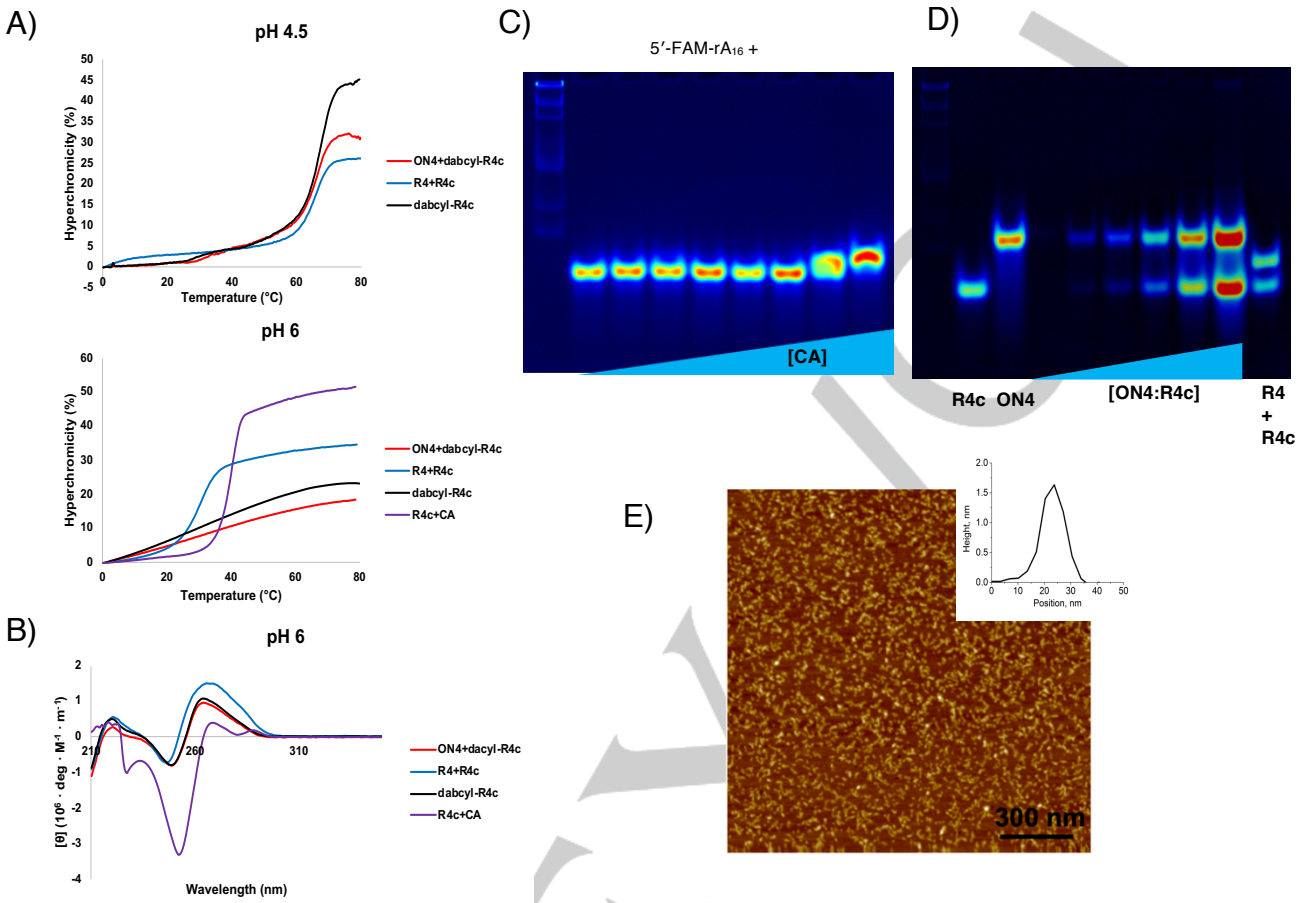


Figure 3. Base-pairing behavior of cyanuric acid containing oligo ON4 compared with its RNA parent sequences in acetate or phosphate buffer ([AcONa] or [NaH_2PO_4/Na_2HPO_4] = 10 mM, [NaCl] = 150 mM, [EDTA] = 0.1 mM, pH = 4.5 or 6, respectively). A). Thermal denaturation curves of **ON4** + ^{dabcyl}**R4c** vs **R4** + **R4c** vs ^{dabcyl}**R4c** alone at pH 4.5 (top) or of **ON4** + ^{dabcyl}**R4c** vs **R4** + **R4c** vs ^{dabcyl}**R4c** vs **R4c** + 15 mM CA at pH 6 (bottom), cooling curves, [ON] = 2–4 μM. B). CD of **ON4** + ^{dabcyl}**R4c** vs **R4** + **R4c** vs ^{dabcyl}**R4c** vs **R4c** + 15 mM CA at 0°C at pH 6, [ON] = 2–4 μM. C) PAGE of 5′-FAM-rA_{16} (0.25 nmol, 50 μM) + CA (0, 1, 2, 3, 4, 5, 10, and 15 mM); Lane 1 is a low range RNA ladder (50, 80, 150, 300, 500, 1000-mer). Running buffer is TAMg pH 4.5. D) Concentration dependent PAGE analysis of **ON4** and 5′-FAM-rA_{16}: Lane 1 is a low range RNA ladder (50, 80, 150, 300, 500, 1000-mer), lane 2 and 3 are 5′-FAM-rA_{16} alone and **ON4** alone, respectively (0.50 nmol, 50 μM), lanes 4 – 9 are the dose response (1, 5, 10, 25, 50, 100 μM), lane 10 is 5′-FAM-rU_{16} + 5′-FAM-rA_{16} (0.5 nmol, 50 μM). Running buffer is TAMg pH 4.5. E) AFM image of **ON4** + **R4c** at pH 4.5. For methods and conditions of measurements see SI. See also Figures S60-S70. ^{dabcyl}**R4c** = 5'-dabcyl labeled R4c.

These observations raised questions about the actual nature of the complex formed and whether poly(rA) prefers to form homo-duplex even in the presence of its complementary pairing strand r(U_{16}), in view of the known homo-duplex behavior of poly(rA) under acidic pH.[11] Confirming our suspicion, **R4c** (rA_{16}) alone under the same conditions (pH 4.5) showed the same homo-duplex formation suggesting that **R4c** is not interacting with either **ON4** or **R4** (Figure 3A). Additional corroboration was provided by (a) repeating the UV- and CD-melt experiments with **ON4** + ^{dabcyl}**R4c** at pH 6 (a condition where **R4c** does not undergo self-pairing), which showed no complex formation and (b) comparing it with **R4c** + 15 mM cyanuric acid (CA) heterocycle at pH 6 which clearly showed a complex formation consistent with previous observations (Figure 3B). [4-5] To check if poly(dA) would act differently,[4] we investigated d(A_{16}) in place of r(A_{16}), but again there was no interaction between **ON4** and **D1c** as indicated by UV- and CD-spectroscopy (Figures S61 and S67). We also mixed the unlabeled version **ON5** with r(A_{16}) and observed identical results indicating that the FAM-label is not the cause of these observations (Figure S71). Gel-electrophoresis (Figures 3C versus 3D) was used to further confirm the behavior of the FAM-labeled **ON4** + r(A_{16}) and was compared separately with the gel-electrophoresis behavior of the free CA-heterocycle + r(A_{16}). First, the CA-heterocycle concentration dependence of 5′-FAM-r(A_{16}) from single-stranded to assembly was monitored (Figure 3C). Full assembly was realized at 15 mM CA, as no more single-stranded 5′-FAM-rA_{16} was present at this concentration (Figures 3C and S83). When **ON4** (5′-FAM-rX_{16}) + complementary rA_{16} (labeled with FAM on either the 5′-end, 3′-end, or unlabeled) was loaded in TAMg pH 4.5 buffer, no slower migrating bands (relative to single-stranded oligonucleotides) were observed (Figures 3D and S83). Although thermal stability experiments of **ON4** mixed with rA_{16} showed sigmoidal transitions at pH 4.5, neither of the single-strands disappeared from the lanes to form a slower moving complex in PAGE.

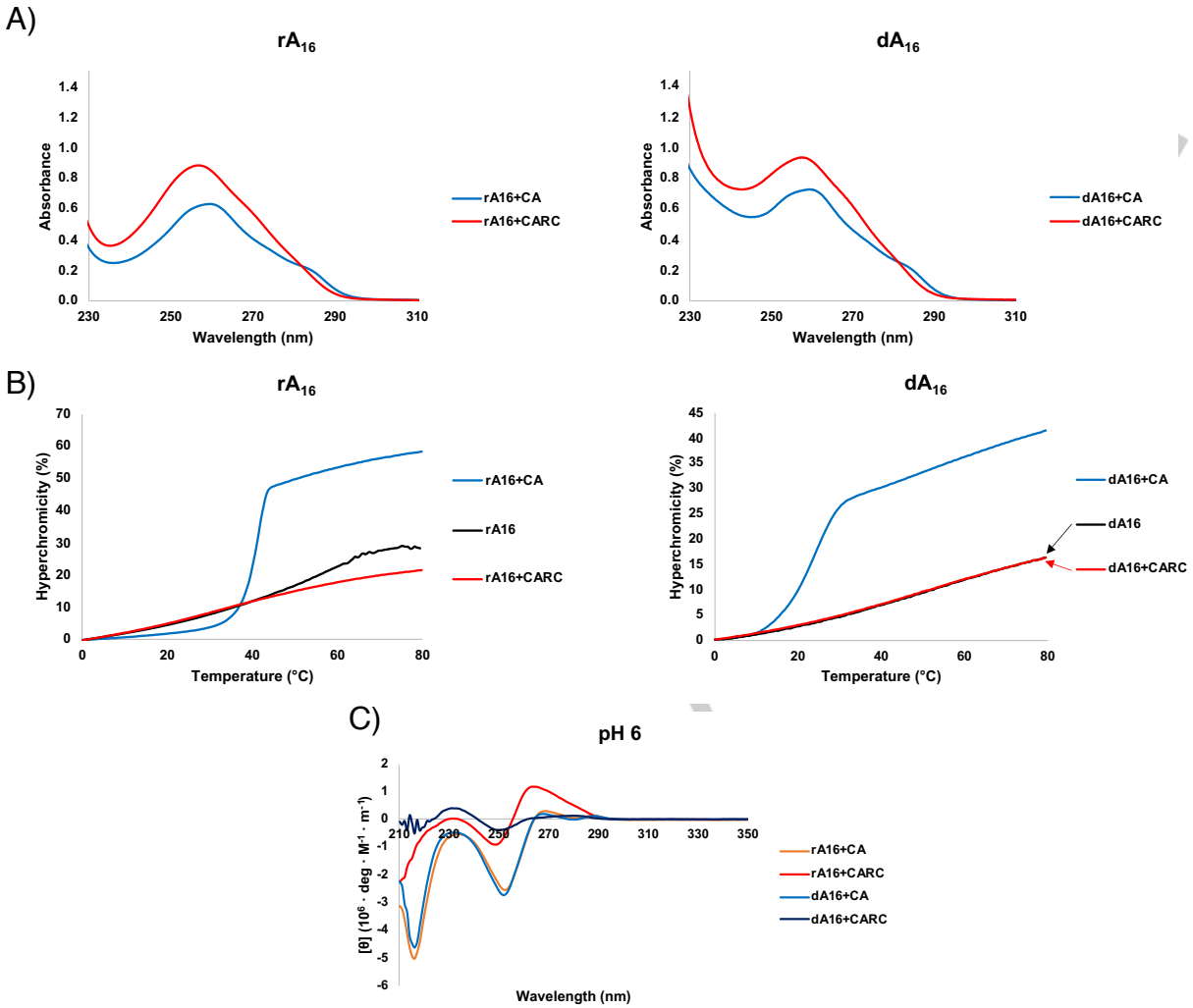


Figure 4. Comparing the base-pairing behavior of CARC versus cyanuric acid with r(A_{16}) and d(A_{16}) oligonucleotides in 200 mM NaCl at pH 6, [ON] = 40 μM. A) absorbance spectra at 0°C for rA_{16} + CA vs rA_{16} + CARC (left) or dA_{16} + CA vs dA_{16} + CARC (right). B) thermal denaturation curves of rA_{16} alone vs rA_{16} + CA vs rA_{16} + CARC (left) and dA_{16} alone vs dA_{16} + CA vs dA_{16} + CARC (right), cooling curves. C) comparison of the CD spectra of rA_{16} + CA vs rA_{16} + CARC vs dA_{16} + CA vs dA_{16} + CARC at 0°C. For methods and conditions of measurements see SI.

This result confirmed that the transitions observed in thermal stability studies at acidic pHs are from the single-stranded r(A_{16}) forming a homoduplex that is most stable at pH 4.5.[11] Solution AFM studies, once again, confirmed the absence of any ordered structures (Figure 3E) when compared to those of free CA-heterocycle + r(A_{15}) which is known to form supramolecular assemblies[4].

Thus, cyanuric acid when incorporated in an RNA backbone, irrespective of sequence combinations, appears to lose its capacity for hydrogen-bond mediated base-pairing with poly(rA). While the compromising impact of single cyanuric acid insertions on thermal stability of duplexes would be in line with our own observations (**ON1**+**R1c** versus **R1**+**R1c**) and others[12], the complete loss of base-pairing in **ON2** – **ON5** stands in stark contrast to ability of free CA to form supramolecular assemblies with various monomers[2b],[5] and oligomers of adenosine-5'-monophosphate[4]. Moreover, strategically placed single substitution in a PNA backbone has shown an increase in the thermal stability of PNA_2:DNA triplexes.[10] And, "trivalent cyanuric acid" derivatives form robust assemblies in water with melamine.[13] We wondered if this loss of base-pairing is because of the constraints imposed on CA in an oligomeric RNA-backbone, and questioned whether the monomeric ribose-nucleoside of CA (CARC, **3**) would behave like the free CA-heterocycle or the CA-oligonucleotide in its ability to form hydrogen-bond mediated supramolecular assemblies. In other words, how would a ribose-substitution on CA (leading to CARC) affect the ability of cyanuric acid to mediate supramolecular self-assemblies? According to the hexad-model that has been proposed[4] (Figure 1) and taken further[5], the substitution on the *N*1-position of cyanuric acid is not expected to have any detrimental effect, since -at first glance- it is on the periphery of the hexad/rosette structure.

We therefore used the techniques described above to investigate the potential for CARC (**3**) to assemble with r(A_{16}) and d(A_{16}) and compared the results of these investigations with those of the free heterocycle-CA (Figures S72-S80). Accordingly, we chose 15 mM of CARC in 200 mM NaCl at pH 6 as the standard condition for comparison since (a) it is the concentration of free

10.1002/chem.202004397

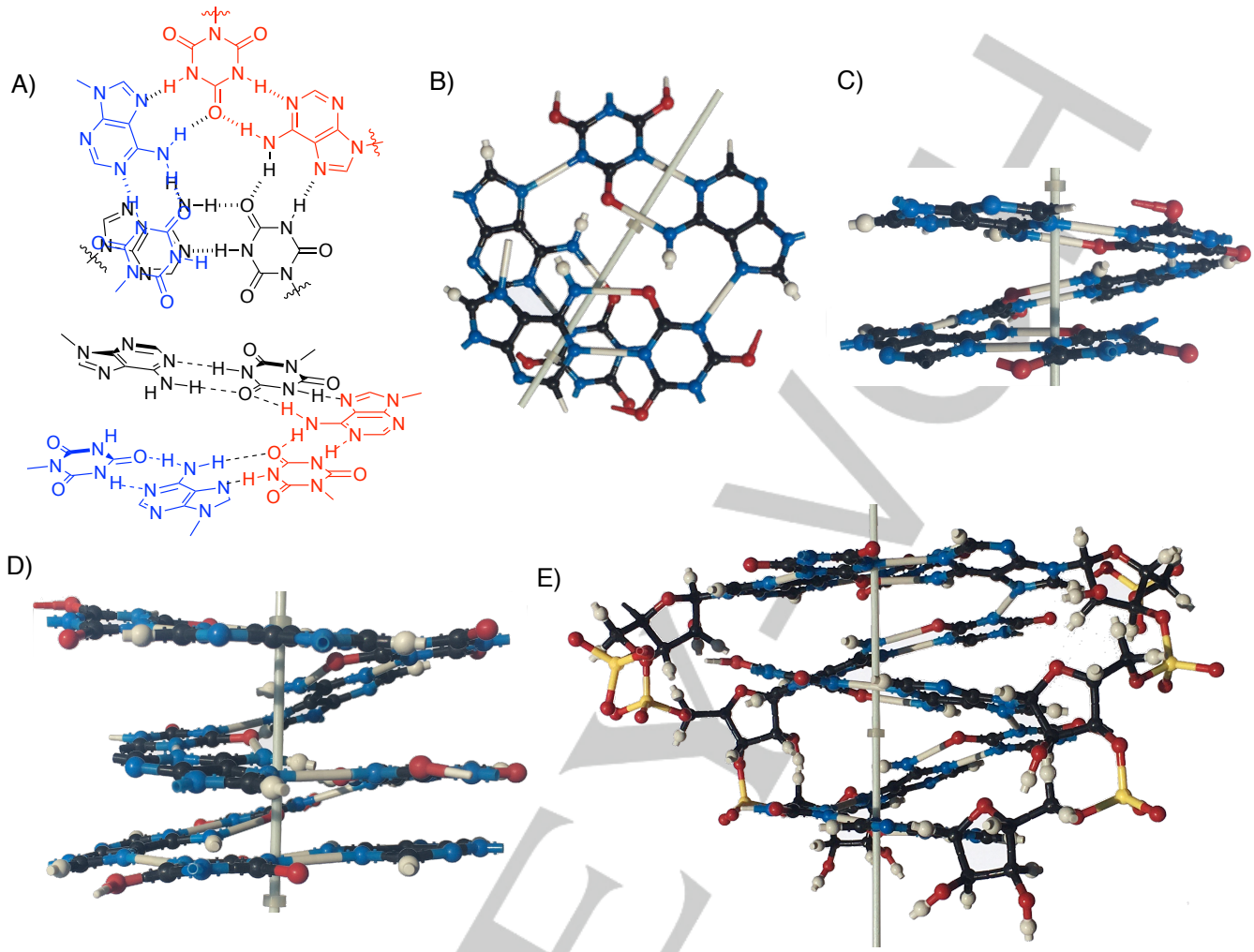


Figure 5. Developing the helicene-dyad model based on ChemDraw depiction and hand-held models. (A) The ChemDraw representation and (B and C) the hand-held molecular models showing the natural appearance of a (D) helicene type structure with three hydrogen-bonded adenine-CA units. (E) Side view of hand-held molecular model of an idealized CA-poly(rA) helicene-dyad structure; note the parallel orientation of the RNA strands as dictated by the anti-orientation of the nucleobases.

CA-heterocycle that is known to form supramolecular assemblies with $r(A_{15})$[4]; (b) the free heterocycle-CA does base-pair with 40 μM $r(A_{16})$ at pH 6.0 (Figure 3C) and (c) the poly(rA) does not exhibit a homo-duplex behavior at pH 6.0 (Figures 4B and S72). Temperature-dependent UV spectra showed no indication of complex formation, as no shoulder was observed at 280-290 nm (Figure 4A). Thermal denaturation (UV-T_m) studies and temperature-dependent CD-spectroscopy likewise showed no indication of duplex or other complex formation (Figures 4B,C) – the mixtures displayed the same behavior of that of poly(rA_{16}) alone under identical conditions.

Since Sleiman and coworkers have shown that poly(dA) forms robust supramolecular assemblies with free CA, we also carried out studies with $d(A_{16})$ and measured the temperature-dependent UV- and CD-spectra at both pH 6.0 and 4.5 in the presence of 15mM CARC (Figure 4). Under both conditions the spectra were similar to that of $d(A_{16})$ alone, indicating that $d(A_{16})$ also does not interact with CARC. Thus, a simple ribose substitution on CA completely abolished the ability of CA to mediate supramolecular assemblies with poly(A) under comparable conditions. It is surprising, since both 2,4,6-triaminopyrimidine (TAP) and its ribose substituted derivative (TARC) are known to self-assemble with cyanuric acid.[2a, 2b] Efforts to investigate the base-pairing behavior of the corresponding CA-counterparts, TAP and TARC and their oligonucleotides (with purines and pyrimidines) are underway.

Finally, we increased the concentration of CARC (based on the suggestion that the minimum assembly concentration for CARC may be higher than for CA) and observed that at a concentration of 100 mM of CARC with 40 μM dA_{16}, a weak supramolecular assembly (UV-T_m ≈ 10° C) was observed as confirmed by UV- and CD-spectroscopy (Figures S84-S85). Increasing the concentration of CARC to 150 mM, increased the stability of the assembly slightly (CD-T_m ≈ 15° C, Figure S85) suggesting that this is indeed a supramolecular assembly mediated by CARC. The necessity for such high concentrations for CARC (≈ 7 fold greater than CA) highlights the detrimental effect of ribose substitution on CA.

Seeking to understand as to why such a simple ribose-substitution may have such a drastic effect on the cyanuric acid base-pairing behavior, we considered the pKa of the cyanuric acid heterocycle (≈ 6.9[4, 14]) and the pH of buffers (4.5-8) used in the study. It is known that a pKa-pH of less than 2 units weakens the base-pairing ability of a nucleobase in an oligonucleotide since the nucleobases become ionized by protonation or deprotonation,[15] and this may be the case of the cyanuric acid embedded in the RNA backbone.[12] However, at the monomer level, the pKa-pH difference does not explain satisfactorily the dichotomy between the free CA-heterocycle (pKa ≈ 6.9) and CARC (pKa ≈ 6.5) in their ability to hydrogen-bond and mediate supramolecular assembly. This leaves open the possibility that steric hindrance, caused by the substitution of ribose on the cyanuric acid, impedes the formation of supramolecular structures (Figure 1). In this context, it should be noted that *n*-alkanoic substituted cyanuric acids (Figure 1) were shown to form supramolecular assemblies with diaminopurine and adenine[5]. Furthermore, *N*-methyl substituted cyanuric acids have been shown to self-assemble with purine monomers resulting in tapes and sheets.[16] However, the ribose-substituent is much bulkier than these known examples and could be interfering with the stacking of the hexad (rosette)-structure.

With this in mind, as a first step, we drew the hexad structures in ChemDraw and made molecular models (from HGS Biochemistry RNA, DNA kits) to mimic the cyanuric acid-adenine hexad structure as drawn in the literature[4-5]. Much to our consternation we found that it was impossible to either draw or construct the cyanuric acid-adenine planar hexad-models while maintaining the rules of regular H-bonding patterns (Figures 5A and 5B). The natural constraints imposed by the hydrogen bonds in the hand-held models or ChemDraw did not allow for a strain-free planar-hexad structure in accommodating the third aden(os)ine-CA unit while maintaining the same Watson-Crick and Hoogsteen hydrogen-bond pattern between the first two and the third adenosine-CA or adenine-CA base-pairs – which was not the case for the CA-TAP (triazine-pyrimidine) model that was a perfect fit (Figure 1). For the purine-CA, every time, the third aden(os)ine-CA base-pair ended up either above or below the plane of the first two CA-adenosine base-pairs, leading to a helicene-type structure (Figure 5C). If we followed this hydrogen-bond mediated helicene-structure to its logical conclusion, then it led naturally to a supramolecular structure where the aden(os)ine-CA base pairs are propagated in a twisted-ribbon helicene structure (Figures 5C and 5D). And, as a consequence of this helicene structure, the RNA-backbones attached to the N9-position of adenine are positioned diagonally opposite to each other leading to a dyad of backbones (Figure 5E). That is, there are two RNA backbones as opposed to three RNA strands as implied in the hexad-rosette structure. Constructing such a poly(rA)-cyanuric acid model (Figure 5E) showed that this helicene-dyad model can be a viable alternative to the planar hexad-structure that has been proposed in previous works.[4-5] The resulting helicene model predicts that the two oligo-RNA strands are in the parallel orientation with an interesting property – while the sugar-phosphate backbones form a right-hand helix, the overall direction of the base-pairs within the helicene are left-handed (Figure 5E). The overall right-handedness resulting from the D-ribose sugar configuration naturally dictates the left-hand orientation of the nucleobases.

Many of the biophysical data presented in the previous work of Sleiman group[4, 17] seem to be consistent with the helicene-dyad model for the CA-mediated supramolecular assemblies of poly(A). The height of the fibers as seen by the AFM data[4] would also be consistent with a helicene model. As predicted by the helicene-dyad model, the system would have a parallel arrangement of backbones with a 1:1 CA:A ratio, which is what is observed.[4, 17] Furthermore, the robust elongation into micron-long polymers[4, 17] fits well with the helicene model which naturally results in overhangs (sticky ends) which would lead to (a) a staggered assembly[4] as opposed to a possibly blunt-end assembly from a hexad-rosette structure and (b) overlap between strands creating distribution of fiber lengths.[17] The opposing helicity of the backbone (right-hand) versus the helicene of the nucleobases (left-hand) may also explain the dramatic CD-characteristic of CA-poly(A) systems (strong negative bands at 212 and 252 nm with weak positive bands at 267 and 285 nm) as noted by Sleiman and coworkers.[4] Our helicene model suggests that a significant structural change has taken place and should not be interpreted as implying anything more (especially with the overall handedness of the system). Moreover, the identical CD of CA+poly(dA) and CA+poly(rA)[4] suggests that the DNA backbone in the CA-mediated supramolecular assembly has an overall backbone form that is closely similar to that of A-RNA.

Is a hydrogen-bond mediated/propagated helicene structure possible? A survey of the literature of helicenes reveals that they are entirely composed of covalent structures comprised of polyaromatic compounds which can be left- (M) or right- (P) handed.[18] While a small molecule mediated hydrogen bond mediated helicene structure has not been reported, there are studies[19] along with a theoretical work[20] that suggests that hydrogen-bonded helicenes should be feasible. When CARC is substituted for the cyanuric acid in the hand-held helicene-dyad model, it shows that the ribose moiety is close to the RNA/DNA phosphodiester backbone attached to the adenosine, suggesting a steric hindrance. Thus, a steric clash may account for the inability of CARC to mediate robust supramolecular self-assemblies with poly(A). That a substitution on CA does have an impact on its ability to mediate self-assembly with purines is also in line with the observations reported by Li et al who showed decrease in the stabilities of the assemblies of adenine with butanoic substituted CA.[5] Moreover, Ganesh and coworkers have shown the N-9 ethyl substituted adenine self-assembles with N-methyl substituted CA and forms planar sheets as confirmed by X-ray studies.[16] Thus, it is clear that the substitution on the CA can have a remarkable effect on the outcome of the nature of the supramolecular assembly mediated by CA with a bulky substitution being detrimental. Moreover, if the proposed hydrogen-bonded helicene structure is indeed possible and confirmed, then it would represent a different structural paradigm, also for the reprogramming of nucleic acid structure induced by small molecules with implications for nucleic acid based nanostructures and origami.[21]

In order to understand the details of the differences between the hexad-rosette model and the helicene model, we initiated collaborative *in silico* molecular dynamics simulations led by the Sherrill group. The results of this investigation are described and discussed in detail in an accompanying paper.[7] Briefly, the simulations show that the triazine-purine hexad-rosette structure is unstable and readily, and naturally, tends toward the helicene model, where the base-pairs are out of plane as predicted by the

ChemDraw figures and molecular models (Figure 5). The simulations also show that the helicene-dyad model can accommodate the RNA (or DNA) strands with minimal deviation from the canonical A-form conformation, with the helicene-helicene model similarly adopting the A-form conformation but with a greater deviation.[7] Substituting CARC instead of CA was confirmed to destabilize the assembly due to steric hindrance between CARC and the RNA backbone.[7]

The observation that cyanuric acid is able to mediate supramolecular assemblies as opposed to CARC or cyanuric acid oligoribonucleotide shows that a heterocycle may not function exactly the same way when it is incorporated into a backbone. And this reveals a limitation of extrapolating the base-pairing behavior all the way from a monomer to the supramolecular assembly. The fact that cyanuric acid RNA-oligomer is compromised in base-pairing, in the conventional oligomeric pre-RNA to oligomeric RNA scenario[22], would be considered problematic since the cyanuric acid RNA-oligomer would be hampered in function and, importantly, of base-pair mediated information transfer to its successor. It is well known that the nature of the backbone can affect the base-pairing properties of the heterocycles.[23] The fact that a single cyanuric acid substitution in a PNA backbone forms a stable triplex with DNA[10] and seems not to suffer the same fate in a RNA backbone suggests that cyanuric acid may still be able to function in selective places in alternative backbones. However, the idea that pre-RNA to RNA transition has to take place while maintaining the same pairing paradigms (e.g. hexad vs. dyad), or only between informational oligomers[24] (resembling biological processes between different genetic systems) may be too restrictive from the prevailing prebiotic perspective.[25] The results from this investigation not only suggest that the conventional paradigm[22, 26] of pre-RNA oligomer giving rise to RNA needs to be rethought and expanded in possibilities, while also offering a simpler alternative paradigm. The fact that cyanuric acid by itself is able to mediate the self-assemblies of adenine based systems[4-5, 16] suggests another possible role of such alternative heterocycles (at the monomer level) in a pre-RNA world, in addition to potentially serving as the pre-RNA bases[3] (at the oligomer level). In this context, it is pertinent to note that a "paradox of base-pairing" argument against RNA being first has been proposed, which essentially states that since the RNA nucleobases/nucleos(t)ides are unable to base-pair as monomers, it is hard to envision how their base-pairing polymers would have been formed through the self-assemblies of such non-associating monomeric units.[27] With cyanuric acid (CA), CARC and CA-oligonucleotides, there is the reversal of this paradox! While CA can base-pair as a monomer, it loses its base-pairing capacity at the oligomeric stage in the RNA backbone. However, this result should not be construed to suggest that CA should be excluded as a plausible pre-RNA base, since it could still function in a different polymeric backbone (e.g. PNA[28]) or via different paradigms such as hexad-based[3] informational systems. Nevertheless, putting these two contradictions together, a solution to both of these paradoxes comes out naturally – one that is congruent with ideas suggested before.[27, 29] Cyanuric acid is functional at the monomer level, base-paring with adenine and adenosine nucleotide. While at the oligomer level the base-paring property of adenine/adenosine nucleotide naturally emerges and takes over. Thus, the paradigm of the role of "pre-RNA" in a prebiotic scenario need not be exactly the same as how information transfer takes place between oligomers but can be simpler and straightforward at the monomeric level itself. This realization allows for a shift in the paradigm for a "pre-RNA world" scenario suggesting a wider role for small molecules[3-4, 30] to operate as "pre-RNA" candidates – as enablers of assembly of RNA/DNA nucleos(t)ide units that are unable to assemble by themselves. Thus, there is no need to confine oneself to the widely accepted model of a fully formed pre-RNA oligomer to an RNA/DNA oligomer transition[22, 24b, 31]. Rather, it could be expanded to include other possibilities – one that is as simple as small molecules ("midwives"[27]) mediating supramolecular assemblies that may be able to directly give rise to RNA/DNA – thus bypassing the requirement of an informational and functional pre-RNA oligomer that gives rise to RNA/DNA.

Conclusion

In summary, we have shown that a ribose-substitution on cyanuric acid (CARC), unexpectedly, interferes with the ability of cyanuric acid (CA) to mediate supramolecular self-assemblies with purines. This impairment extends to the cyanuric acid RNA oligoribonucleotides as well. These observations led us to propose (a) an alternative helicene-dyad model that accounts for the difference in the behavior between CA and CARC and (b) that the usually considered model of informational transfer between a fully assembled pre-RNA oligomer to RNA oligomer can be expanded to include a simpler model wherein small molecules themselves can play a role in the self-assembly and emergence of oligomers of extant nucleic acids. Additionally, the hydrogen bond mediated helicene could have implications for small molecule mediated structural control and reorganization of nucleic acids and similar polymers.[32]

Acknowledgements

This work was jointly supported by NSF and the NASA Astrobiology Program under the NSF Center for Chemical Evolution, CHE-1504217. We thank Prof. Gary Schuster (GIT) for helpful discussions and Prof. Luke Leman (TSRI) for feedback on the manuscript. We are grateful to the anonymous reviewer for suggesting the increased CARC concentration experiments.

Keywords: Cyanuric acid • DNA • Hydrogen-bonded Helicene • RNA • RNA World

[1] A. E. Engelhart, N. V. Hud, in Cold Spring Harbor Perspectives in Biology, Vol. 2, **2010**.

[2] a) B. J. Cafferty, I. Gállego, M. C. Chen, K. I. Farley, R. Eritja, N. V. Hud, *J. Am. Chem. Soc.* **2013**, *135*, 2447-2450; b) M. C. Chen, B. J. Cafferty, I. Mamajanov, I. Gállego, J. Khanam, R. Krishnamurthy, N. V. Hud, *J. Am. Chem. Soc.* **2013**, *136*, 5640-5646; c) B. J. Cafferty, D. M. Fialho, J. Khanam, R. Krishnamurthy, N. V. Hud, *Nat. Commun.* **2016**, 7:11328; d) G. Whitesides, J. Mathias, C. Seto, *Science*, **1991**, *254*, 1312-1319.

[3] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Che. Biol.* **2013**, *20*, 466-474.

[4] N. Avakyan, A. A. Greschner, F. Aldaye, C. J. Serpell, V. Toader, A. Petitjean, H. F. Sleiman, *Nat. Chem.* **2016**, *8*, 368-376.

[5] C. Li, B. J. Cafferty, S. C. Karunakaran, G. B. Schuster, N. V. Hud, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20091-20096.

[6] H. Kashida, Y. Hattori, K. Tazoe, T. Inoue, K. Nishikawa, K. Ishii, S. Uchiyama, H. Yamashita, M. Abe, Y. Kamiya, H. Asanuma, *J. Am. Chem. Soc.* **2018**, *140*, 8456-8462.

[7] A. Alenaizan, K. Fauché, R. Krishnamurthy, C. D. Sherrill, (submitted) **2020**.

[8] a) A. R. Cimpoia, P. J. Hunter, C. A. Evans, H. Jin, T. Breining, T. S. Mansour, *J. Carbohyd. Chem.* **1994**, *13*, 1115-1119; b) A. Khaled, T. Ivannikova, C. Augé, *Carbohyd. Res.* **2004**, 339, 2641-2649.

[9] a) P. Kumar, N. N. Ghosh, K. L. Sadana, B. S. Garg, K. C. Gupta, *Nucleosides and Nucleotides* **1993**, *12*, 565-584; b) R. T. Pon, S. Yu, *Nucl. Acids Res.* **1997**, *25*, 3629-3635; c) R. T. Pon, N. Usman, K. K. Ogilvie, *Biotechniques* **1988**, *6*, 768-775.

[10] R. Vysabhattar, K. N. Ganesh, *Tet. Lett.* **2008**, *49*, 1314-1318.

[11] a) M. I. Zarudnaya, I. M. Kolomiets, A. L. Potyahaylo, D. M. Hovorun, *J. Biomol.Struct. and Dyn.* **2019**, *37*, 2837-2866; b) N. Safaee, A. M. Noronha, D. Rodionov, G. Kozlov, C. J. Wilds, G. M. Sheldrick, K. Gehring, *Angew. Chem. Int. Ed.* **2013**, *52*, 10370-10373; c) M. L. Gleghorn, J. Zhao, D. H. Turner, L. E. Maquat, *Nucl. Acids Res.* **2016**, *44*, 8417-8424.

[12] D. Gasparutto, S. Da Cruz, A.-G. l. Bourdat, M. Jaquinod, J. Cadet, *Chem. Res. Toxicol.* **1999**, *12*, 630-638.

[13] M. Ma, D. Bong, *Langmuir* **2011**, *27*, 8841-8853.

[14] W. M. Haynes, CRC Handbook of Chemistry and Physics, CRC Press, **2016**.

[15] R. Krishnamurthy, *Acc. Chem. Res.* **2012**, *45*, 2035-2044.

[16] V. R. Pedireddi, A. Ranganathan, K. N. Ganesh, *Org. Lett.* **2001**, *3*, 99-102.

[17] R. W. Harkness V, N. Avakyan, H. F. Sleiman, A. K. Mittermaier, *Nat. Commun.* **2018**, *9*, 3152.

[18] a) Y. Shen, C.-F. Chen, *Chem. Rev.* **2012**, *112*, 1463-1535; b) M. Gingras, *Chem. Soc. Rev.* **2013**, *42*, 968-1006; c) M. Gingras, G. Félix, R. Peresutti, *Chem. Soc. Rev.* **2013**, *42*, 1007-1050; d) M. Gingras, *Chem. Soc. Rev.* **2013**, *42*, 1051-1095; e) S. K. Pedersen, K. Eriksen, M. Pittelkow, *Angew. Chem. Int. Ed.* **2019**, 58, 18419-18423.

[19] a) E. Murguly, R. McDonald, N. R. Branda, Org. Lett. 2000, 2, 3169-3172; b) T. Giorgi, S. Lena, P. Mariani, M. A. Cremonini, S. Masiero, S. Pieraccini, J. P. Rabe, P. Samorì, G. P. Spada, G. Gottarelli, *J. Am. Chem. Soc.* **2003**, *125*, 14741-14749; c) M. Gellert, M. N. Lipsett, D. R. Davies, *Proc. Natl. Acad. Sci.* **1962**, *48*, 2013-2018; d) V. Sasisekharan, S. Zimmerman, D. R. Davies, *J. Mol. Biol.* **1975**, *92*, 171-179.

[20] H. S. Choi, K. S. Kim, *J. Phys. Chem. B* **2000**, *104*, 11006-11009.

[21] N. C. Seeman, H. F. Sleiman, *Nat. Rev. Mat.* **2017**, *3*, 17068.

[22] G. F. Joyce, L. E. Orgel, Cold Spring Harbor Monograph Series **2006**, *43*, 23-56.

[23] R. Krishnamurthy, *Synlett* **2014**, *25*, 1511-1517.

[24] a) L. E. Orgel, *Crit. Rev. Biochem. Molecul. Biol.* **2004**, *39*, 99-123; b) B. Alberts, A. Johnson, J. Lewis, Martin Raff, K. Roberts, P. Walter, Molecular Biology of the Cell, 4th ed., Garland Science, New York, **2002**.

[25] R. Krishnamurthy, *Israel J. Chem.* **2015**, *55*, 837-850.

[26] L. E. Orgel, NATO ASI Series, Series A: Life Sciences **1989**, *169*, 215-224.

[27] N. V. Hud, F. A. L. Anet, *J. Theor. Biol.* **2000**, *205*, 543-562.

[28] K. N. Ganesh, A. Gourishankar, R. Vysabhattar, P. Bokil, Nucleic Acids Symposium Series **2007**, *51*, 17-18.

[29] N. V. Hud, S. S. Jain, X. Li, D. G. Lynn, *Chem.* & *Biodiver.* **2007**, *4*, 768-783.

[30] a) B. J. Cafferty, C. Musetti, K. Kim, E. D. Horowitz, R. Krishnamurthy, N. V. Hud, *Chem. Commun.* **2016**, *52*, 5436-5439; b) B. J. Cafferty, N. V. Hud, *Israel J. Chem.* **2015**, *55*, 891-905.

[31] P. E. Nielsen, *Chem.* & *Biodiver.* **2007**, *4*, 1996-2002.

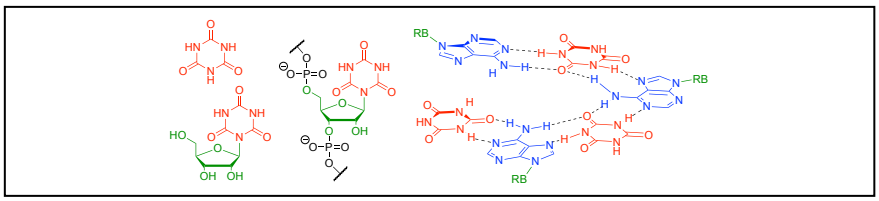
[32] a) E. Yashima, N. Ousaka, D. Taura, K. Shimomura, T. Ikai, K. Maeda, *Chem. Rev.* **2016**, *116*, 13752-13990; b) T. Aizawa, K. Aratsu, S. Datta, T. Mashimo, T. Seki, T. Kajitani, F. Silly, S. Yagai, *Chem. Commun.* **2020**, *56*, 4280-4283.

10.1002/chem.202004397

WILEY-VCH

RESEARCH ARTICLE

Entry for the Table of Contents



Not a Hexad but a Helicene. An unprecedented hydrogen-bonded helicene model is proposed to explain the surprising loss of base-pairing properties of the cyanuric acid heterocycle when tagged with ribonucleoside and oligonucleotide. The results have wide ranging implications for creating novel nucleic acid structures, synthesis of long helicenes and the role of small molecules as pre-RNA.