MANUSCRIPT TITLE

Impacts of sequence and structure on pyrrolocytosine fluorescence in RNA

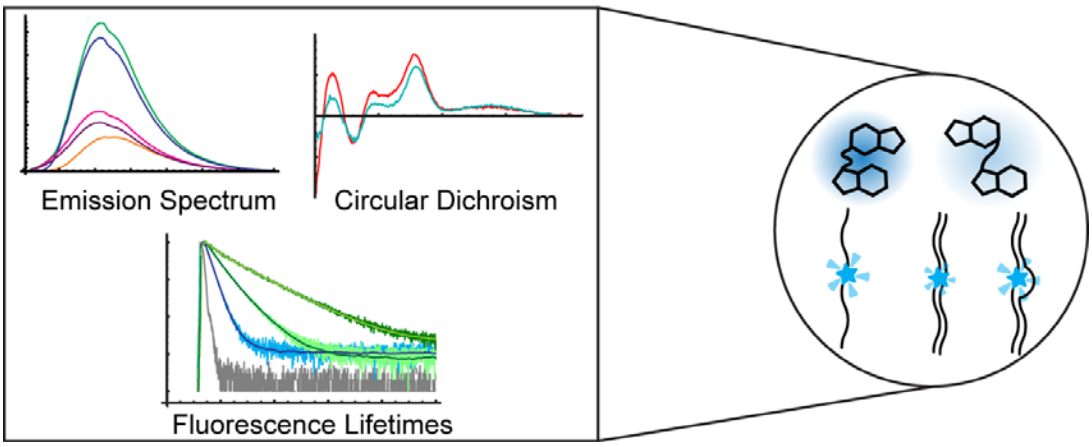
AUTHORS

Taylor L. Coulson[1], Julia R. Widom[1,*]

[1] Department of Chemistry and Biochemistry, University of Oregon, Eugene, Oregon, 97403 USA

* To whom correspondence should be addressed. Tel.: (541) 346-4670; Email: jwidom@uoregon.edu

GRAPHICAL ABSTRACT



ABSTRACT

Fluorescence spectroscopy encompasses many useful methods for studying the structures and dynamics of biopolymers. Applications to nucleic acids require the use of extrinsic fluorophores such as fluorescent base analogues (FBAs), which mimic the native bases but have enhanced fluorescence quantum yields. In this work, we use multiple complementary methods to systematically investigate the sequence- and structure-dependence of the fluorescence of the FBA pyrrolocytosine (pC) within RNA. We demonstrate that pC is typically brightest in conformations in which it is base-stacked but not base-paired, properties that distinguish it from more widely

used FBAs. This effect is strongly sequence-dependent, with adjacent adenosine and cytidine residues conferring the greatest contrast between stacked and unstacked structures. Structural heterogeneity was resolved in single-stranded RNA and fully complementary and mismatched double-stranded RNA using time-resolved fluorescence measurements and fluorescence-detected circular dichroism spectroscopy. Double-stranded contexts are distinguished from single-stranded contexts by the presence of inter-strand energy transfer from opposing bases, while base-paired pC is distinguished by its short excited state lifetime. This work will enhance the value of pC as a structural probe for biologically and medicinally significant RNAs by guiding the selection of labeling sites and interpretation of the resulting data.

INTRODUCTION

Fluorescence detection methods are valuable tools for elucidating the structures and dynamics of biopolymers both *in vitro* and *in vivo* (1, 2). For example, fluorescence-based experiments on nucleic acids can be designed to reveal changes in structural features such as base-stacking and hydrogen bonding in response to environmental stimuli (3, 4). Fluctuations in local base-stacking and base-pairing interactions typically occur on a faster timescale than global changes in secondary or tertiary structure, but these local structural reorganizations can be critical for binding interactions and large-scale structural shifts.

The native nucleic acid bases are practically non-fluorescent, with quantum yields on the order of 10^{-4} and excited state lifetimes of less than 1 ps due to rapid, non-radiative de-excitation (5). As a result, external probes must be introduced in order to exploit the structural sensitivity and specificity of fluorescence detection. Fluorescent base analogues (FBAs), such as the widely utilized 2-aminopurine (2-AP), are structural mimics of the native bases with increased fluorescence quantum yields that are designed to maintain the base-pairing properties of their cognate native bases. Certain FBAs, including 2-AP, exhibit environment-sensitive fluorescence

properties, which enables their use as probes of local nucleic acid structure (6–9). 2-AP has been extensively characterized and utilized in both DNA and RNA, but it is a structural mimic only of adenine and guanine and experiences a high degree of quenching upon base stacking (6, 10), in some cases rendering highly stacked conformations nearly undetectable (11). To expand the toolbox of probes available to study the ever-expanding list of biologically and medicinally relevant RNA structures, alternative FBAs must be characterized just as thoroughly. Systematic studies of the structure- and sequence-dependence of the fluorescence of any FBA are critical for optimal experimental design and data interpretation.

Pyrrolo-cytosine (pC; Fig. 1A), an analogue of cytosine first synthesized in 2004 (12), has been used sporadically in both DNA (13–15) and RNA (4, 16) and shows promise as a probe of structures and sequences where 2-AP is not optimal. Previous measurements in DNA have shown that pC experiences enhanced fluorescence when flanked by native bases, a property that was attributed to increased solvent shielding in base-stacked conformations (17). It was also observed that base pairing led to fluorescence quenching in pC-labeled DNA oligonucleotides (17), and that base-stacking with guanine opened new nonradiative de-excitation pathways (10, 17, 18). pC therefore shows promise as an alternative FBA that can probe interactions that 2-AP is less sensitive to. However, the mechanisms behind pC's fluorescence changes upon incorporation into larger nucleic acid structures are poorly understood and have been primarily studied in DNA.

In this work, we utilized a variety of spectroscopic methods to probe the impacts of local structure and sequence on pC's fluorescence, generating a map of its photophysical landscape within RNA that will enable the informed design of probe sequences and deeper interpretation of the data obtained on them. We show that its emissive properties are more complex in RNA than previously observed in DNA and that base-stacked conformations of pC are generally brighter than unstacked conformations, the opposite behavior to 2-AP. This work demonstrates

the rich structural information that will be obtainable using pC in biologically relevant RNA contexts, such as ligand binding pockets or regions of fluctuating hybridization status.

MATERIAL AND METHODS

Reagents and General Methods

All oligonucleotides were obtained from Dharmacon (Horizon Discovery) and deprotected and desalted by the manufacturer. 21-mers (Table 1) were additionally HPLC-purified by the manufacturer. Stock solutions were prepared in DEPC-treated water and their concentrations were determined by A260 using extinction coefficients provided by the manufacturer. pC riboside (pCr) was obtained from Biosearch Technologies (PYA11092-B010). In most cases, spectroscopic studies were performed with oligonucleotides at a concentration of 5 μM in a buffer consisting of 20 mM sodium phosphate (pH 7.4) and 100 mM NaCl. Where indicated, the buffer additionally included 30% v/v ethanol. All data on pC dinucleotide (pCpC) were recorded at a concentration of 1 μM. Spectra requiring deep UV excitation of 21-mers (absorbance, CD and fluorescence excitation) were recorded at 1 μM to reduce the inner filter effect. Double-stranded complexes were prepared by combining pC-labeled and unlabeled strands in a 1:1.1 ratio (slight excess of unlabeled strand) and annealing them by heating to 90°C for 2 minutes and allowing to cool in air for 10 minutes. Cuvettes were soaked in a solution of 1% Hellmanex at 35°C for 30 minutes and dried using N_2 before use. Unless otherwise noted, all spectroscopic measurements were done at 20°C.

Oligonucleotide	Sequence
Strand 1	5′ AUA AUU ACC (pC)CU CAU AUA UUA
Strand 2 GGG	5′ UAA UAU AUG AGG GGU AAU UAU
Strand 2 GUG	5′ UAA UAU AUG AGU GGU AAU UAU

Table 1. Sequences of the 21-mers used in this study

Purification of pCpC

pCpC (Fig. S1A) was obtained as a crude synthesis from Dharmacon and was purified via preparative thin-layer chromatography (TLC). A 20x20 cm 250-micron silica gel TLC plate (Millipore Sigma 1.05715.0001) was first run using pure methanol followed by drying using a gentle flow of N_2 overnight. Crude pCpC was then deposited in a line along the origin using a gel-loading pipette tip at a density of approximately 1.25 nmol/mm, leaving at least 3 cm empty on each edge (19). Based on established protocols for separating purine derivatives (20), a solvent system composed of *n*-butanol, acetone, 33% ammonia and water (50:40:3:15) was used for separation. The plate was run to a height of 1 cm below the top, dried under gentle N_2 flow on a hot plate set to 32°C for 30 minutes, and then run a second time to improve separation (19). Bands visible under UV illumination at 254 nm (Fig. S1B, Table S1) were cut out using a razor blade (~0.25 mL of silica gel) and transferred to 50 mL conical tubes. To extract each band, 1:1:1 water-methanol-acetonitrile was added in a volume based on Eq. 1 (19) and vortexed for 5 minutes before centrifuging at 1500 x *g* for 15 minutes.

$$\text{Eluent Volume} = (1.0 - \text{Retention Factor})(10)(\text{volume of silica}) \quad (1)$$

The supernatant was decanted and the extraction was repeated a second time to improve yield. Each extraction's supernatant was centrifuged three times at 1500 x *g* for 15 minutes and decanted between cycles to remove debris. The supernatant was distributed into 5 mL Eppendorf tubes and dried under vacuum at 45°C. The dry samples were redissolved in DEPC-treated water, filtered using an Ultrafree-MC Centrifugal Filter (0.1 µm pore size), and analyzed via mass spectrometry to identify bands and assess purity (Fig. S1C). Mass spectra were acquired on a Synapt G2-si mass spectrometer (Waters Corp.) in negative ion mode. The samples were sprayed from 3-5 µL of solution at a concentration of 25 µM using nano-Electrospray Ionization (nESI) from a borosilicate capillary with an inner diameter of 0.78 mm pulled to an inner diameter below 1 µm (~400 nm) and a capillary voltage of 0.3 kV.

A mass calibration was performed using Cesium Iodide with a resulting mass accuracy of ~7 ppm.

Spectroscopy and Statistical Analyses

Steady-State Absorbance and Fluorescence Spectroscopy. Steady-state spectra were collected on an Edinburgh FS5 spectrofluorometer with 1 nm steps and 1 s dwell time. Fluorescence excitation spectra were collected by scanning the excitation monochromator from 220 nm to 420 nm with a bandwidth of 2 nm while monitoring emission at 460 nm with a bandwidth of 3-5 nm, selected to avoid detector saturation. Absorbance spectra were collected in the same manner using the transmission detector on the FS5, with three sequential scans being collected and averaged together. Fluorescence emission spectra were collected by scanning the emission monochromator from 360 nm to 700 nm with a bandwidth of 2 nm while exciting at 340 nm with a bandwidth of 3-5 nm. Fluorescence excitation spectra of 21-mers were corrected for the inner filter effect using the corresponding absorbance spectra and then smoothed with a moving average over four data points. Tables 2 and 3 and Fig. 1C contain relative fluorescence quantum yields (FQYs) of pC and molar absorptivity at 347 nm (ε_{347}). Using pCr, which has an FQY of 0.023±0.002 (17), as a reference, the FQY of each oligonucleotide was calculated according to Eq. 2.

$$\Phi_S = \Phi_R \frac{I_S}{I_R} \frac{A_R}{A_S} \frac{\eta_S^2}{\eta_R^2} \qquad (2)$$

Φ is the fluorescence quantum yield while S and R represent sample and reference, respectively. I is the integrated fluorescence intensity, A is the absorbance at the excitation wavelength, and η is the refractive index of the solvent, which in this work was identical for the samples and reference. ε_{347} was calculated by dividing the absorbance at 347 nm by the concentration, which was determined *in situ* via A260 for single-stranded oligonucleotides. dsRNA samples were assumed to be 5 μM based on the measured concentrations of the ssRNA stocks they were prepared from. The mean and standard deviation of measurements on 3 separately prepared samples are reported.

Time-Correlated Single Photon Counting (TCSPC). TCSPC measurements were performed on an Edinburgh FS5 spectrofluorometer equipped with an EPLED-320 (pulse width of 925.3 ps, wavelength of 325.5 nm, and FWHM bandwidth of 12.5 nm) and a high-speed PMT. Decays were collected to a peak count of 10,000 over a 100 ns window divided into 2048 channels with an LED pulse rate of 10 MHz. Reverse mode, in which the start and stop triggers of the counting mechanism are reversed, was utilized to decrease collection time. Emission was monitored at 460 nm with a bandwidth of 29.9 nm. Instrument response functions (IRFs) were recorded using buffer by tuning the emission monochromator to the LED wavelength. A typical IRF full-width at half maximum (FWHM) was ~1300 ps. In ideal conditions, time components up to 1/10 this length can be resolved with the time width of each channel (here, 50 ps) setting a lower limit. Data were analyzed in two ways, detailed below: using exponential reconvolution fits in Edinburgh's instrumental software, Fluoracle, and using lifetime distribution fits in open-source Matlab software (21) developed for use with other FBAs (22). For each sample, three decays were collected sequentially, and each was fitted individually. The mean and standard deviation of the resulting values across the three fits is reported.

Exponential Reconvolution Fit Analysis. This approach identifies an exponential function with one or more components that describes the decay of the molecule (Eq. 3), where $\boldsymbol{\tau}_i$ represents the time constant of component i and w_i represents its population weight.

$$\text{Fit} = \sum_{i=1}^{n} w_i \cdot e^{\frac{-t}{\tau_i}} \tag{3}$$

This fitting procedure iterates over time components of an initially proposed fit until it returns an equation that, when combined with the measured IRF, returns the measured decay. The average observed lifetime $\langle \boldsymbol{\tau} \rangle_{\text{obs}}$ is calculated as a weighted average of decay components (Eq. 4).

$$\langle \tau \rangle_{obs} = \sum_{i=1}^{n} w_i \cdot \tau_i \tag{4}$$

Lifetime Distribution Fit Analysis. Lifetime distribution analysis uses a similar approach to exponential reconvolution fit analysis where fitting parameters are iterated until

the fit returned, when combined with the IRF, matches the measured decay. Here, we use a previously published model in which the fitting equation allows both exponential decay components and components described by gamma probability distributions, where α_i and $\boldsymbol{\kappa}_i$ are the shape and scale parameters of the gamma distribution, respectively (21, 22; Eq. 5).

$$\text{Fit} = \sum_{i=1}^{n} \frac{w_i}{(1+\kappa_i t)^{\alpha_i+1}} + \sum_{i=n+1}^{n+m} w_i e^{\frac{-t}{\tau_1}} \qquad (5)$$

If a decay component is best described by a gamma distribution, the mean of the distribution (Eq. 6) is taken as its time constant.

$$\langle \tau \rangle_i = \frac{1}{\alpha_i \kappa_i} \qquad (6)$$

$\langle \boldsymbol{\tau} \rangle_{obs}$ is calculated as above using $\langle \boldsymbol{\tau} \rangle_i$ in place of $\boldsymbol{\tau}_i$ for any gamma components.

Selection of Fluorescence Decay Models. The quality of a fit is determined by two parameters: 1) residuals and 2) reduced χ^2. Residuals with a magnitude less than 5 and a random distribution combined with a reduced χ^2 of 1-1.3 are indicative of an acceptable fit (21). Mono-exponential fits were initially tested for all decays and additional exponential or gamma components were tested until the fit was acceptable. In cases where the decay is equally well fit by more than one model, the fit with the fewest number of parameters is selected.

Circular Dichroism and Fluorescence-Detected Circular Dichroism. CD and FDCD spectra were recorded as previously described for 2-AP (11) with minor modifications detailed below. Analysis was performed in Mathematica (Wolfram Alpha). Spectra were recorded on a Jasco J-1500 CD spectrometer equipped with a FDCD-551 attachment. A long-pass colored glass filter with a cut-on wavelength of 420 nm was placed in front of the detector. The measurement channels "FDCD" and "DC" were acquired with the gain on the detector selected such that the maximum DC signal was approximately 1 V. Scans were performed in continuous scanning mode from 200 to 410 nm with a 4 nm excitation bandwidth, a 4 s integration time, and a scan speed of 20 nm/minute. Due to its low fluorescence intensity, pCpC was collected in step-scan mode from 200 to 410 nm with a 4 nm excitation bandwidth, a 4 s

integration time, and a data pitch of 1 nm. Multiple scans were acquired, ranging from 5 for ApCA to 20 for pCpC, and averaged together to obtain adequate signal-to-noise ratio (SNR). CD spectra were collected using the same scan parameters as FDCD, though fewer cycles were typically needed to achieve adequate SNR, and were background-corrected using buffer scans. Sodium fluorescein was used as an achiral reference fluorophore to background-correct all FDCD spectra (23). For each measurement of a pC-containing sample, scans were also performed on buffer, water and an aqueous fluorescein solution at a concentration yielding a maximum DC signal of approximately 1 V at a similar detector gain setting to that used for the pC-containing sample (the required fluorescein concentrations ranged from 0.1-3 nM). To correct for stray or scattered excitation light, the FDCD and DC channels of the buffer scan were subtracted from the corresponding channels of the pC scan, and the FDCD and DC channels of the water scan were subtracted from the corresponding channels of the fluorescein scan. The quantity θ_F = FDCD/DC was then computed for both pC and fluorescein. θ_F for fluorescein was smoothed using a Wiener filter and then subtracted from θ_F for the pC-containing sample.

Established theory (24) was used to convert the background-corrected θ_F into more readily interpretable quantities. One such quantity is the dissymmetry factor $g_F = \Delta\varepsilon_F/\varepsilon_F$, where ε_F is the extinction coefficient of the fluorescent species in the sample and $\Delta\varepsilon_F$ is its difference in extinction coefficient under excitation with left- and right-handed circularly polarized light (ε_L-ε_R). g_F was calculated using Eq. 7 with θ_F and the CD (in millidegrees) and absorbance (A, in absorbance units) spectra of the sample as inputs:

$$g_f = \frac{\Delta\epsilon_F}{\epsilon_F} = \left(\frac{\theta_F}{14323} + \frac{CD}{32980*A} + \frac{\ln 10*CD*10^{-A}}{32980\left(1-10^{-A}\right)}\right) \quad (7)$$

The analogous quantity from standard CD spectra was calculated as g = CD/(32980*A).

CD spectra are most commonly displayed as $\Delta\varepsilon$, and obtaining an analogous quantity from FDCD, $\Delta\varepsilon_F$, requires an ε_F spectrum to multiply through Eq. 7. When multiple species and energy transfer (ET) are present, ε_F must be replaced with a population-, quantum yield- and ET efficiency-weighted sum over all species (24):

$$\epsilon_F \rightarrow \sum_i \phi_i \left(c_i \epsilon_i + \sum_j a_{ij} \epsilon_j c_j \right) \tag{8}$$

where i counts over all fluorescent species, j counts over all the species that can transfer energy to fluorescent species i, Φ_i is the fluorescence quantum yield of species i, a_{ji} is the efficiency of ET from species j to species i, ε_i and ε_j are the extinction coefficients of species i and ET donor j, respectively, and c_i and c_j are their concentrations. As previously noted (23), this quantity is reasonably approximated by the fluorescence excitation spectrum. To obtain fluorescence excitation spectra across the entire wavelength range probed by FDCD (200-420 nm), the DC signal at each wavelength was divided by the lamp intensity (measured with a power meter, Thorlabs S120VC) and the absorbance spectrum used to correct the DC spectrum for the inner filter effect. Fluorescence excitation spectra computed in this manner agreed well with spectra recorded on the FS5 over its accessible wavelength range (>230 nm). The excitation spectrum of each sample was converted to units of L/mol*cm by scaling it uniformly such that ε_{347} had the value listed in Table 2. g_F was multiplied by this spectrum in order to obtain $\Delta\varepsilon_F$.

Melting Curves

Melting curves were recorded on a Jasco J-1500 CD spectrometer in a low headspace, 1 cm pathlength cuvette (Starna 26.160/LHS). The samples were heated at 1°C/min from 4°C to 80°C and absorbance was monitored at 260 nm. T_M values were extracted by fitting the melting curves in Mathematica, as previously described (3).

Computational Methods.

Based on the work of Nguyen et al. (18), the ground state molecular geometries of G, C, and pCr nucleobases were optimized at the MP2/6-31G(d,p) level. These optimized monomers were combined into the trimers GpCG and CpCC in an A-form helix, the geometry of which was obtained from the literature (10). The vertical excitation energies for the first ten excited states were calculated using the configuration interaction singles with perturbative second order corrections (CIS(D)) at the aug-cc-pVDZ (25) level of theory, selected based on computational recommendations for pC-containing systems from Nguyen et al (18). Following excited state calculations, the natural transition orbitals (NTOs) (26) were calculated for each excited state at the same level of theory. All calculations were carried out in Gaussian09 (27).

RESULTS

We first investigated the impact of base-stacking and sequence context on the fluorescence of pC by comparing the steady-state and time-resolved emission properties of di- and trinucleotides containing pC adjacent to each native RNA base. Each dinucleotide studied consisted of pC to the 5′ side of a native base, and each trinucleotide consisted of pC flanked by two of the same native base. The fluorescence emission spectra of all dinucleotides exhibit a blue-shift in peak wavelength along with a change in lineshape relative to pCr (Fig. 1B). Minimal further changes in the spectrum are observed in the corresponding trinucleotides (Fig. S2). This hypsochromic shift in emission wavelength and suppression of a long-wavelength (red) vibronic shoulder is consistent across all sequence contexts, indicating that local changes in the solvation environment impact pC's emission spectrum regardless of the identity of the neighboring native base. When adjacent to one A or C residue, the raw fluorescence intensity (Fig. S3) and FQY (Fig. 1C; Table 2) of pC approximately double relative to pCr and increase further in the corresponding trinucleotides. In contrast, the presence of one adjacent G or U residue leaves pC's

FQY nearly unchanged, while it is enhanced by about 50% in the corresponding trinucleotides.

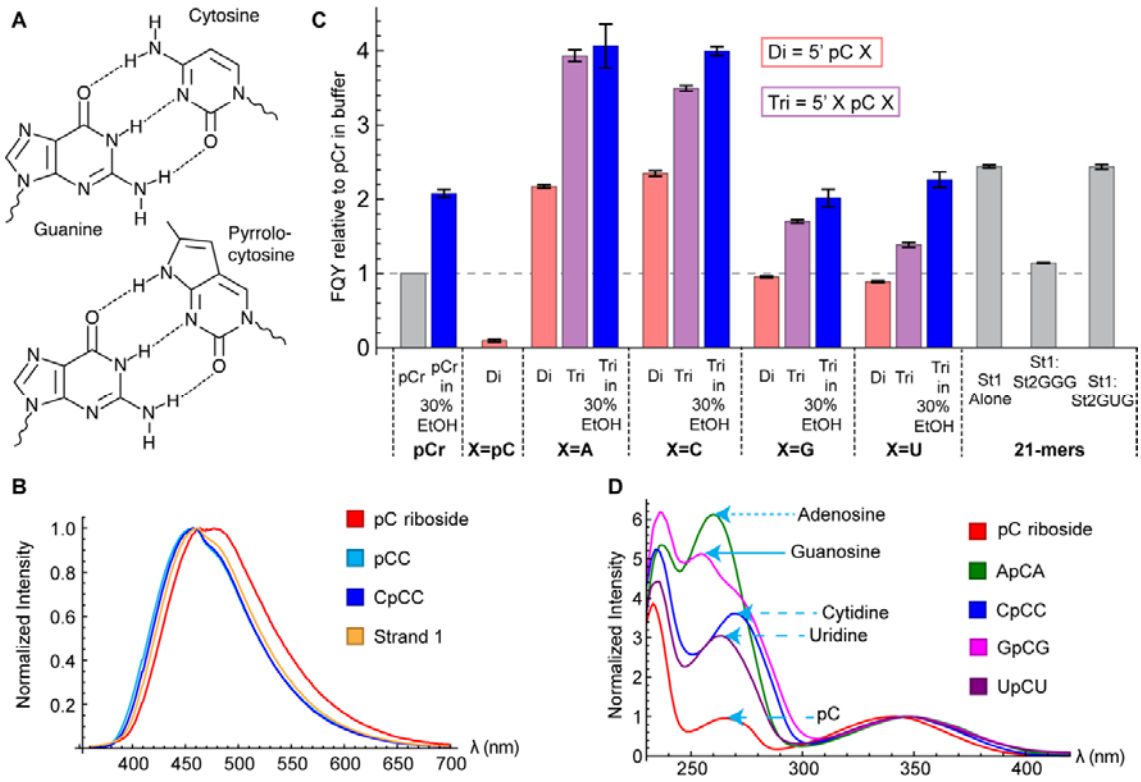


Figure 1. (A) Structures of cytosine (top) and pC (bottom) forming Watson-Crick base pairs with guanine. **(B)** Normalized fluorescence emission spectra of pCr (red), pCC (light blue), CpCC (dark blue), and Strand 1 (orange). **(C)** FQY of all oligonucleotides, normalized to value of 1 for pCr in buffer. Error bars show the standard deviation across three separately prepared samples. "Di" indicates a dinucleotide with the base indicated by "X" to the 3′ side of pC. "Tri" indicates a trinucleotide with the base indicated by "X" to the 5′ and 3′ sides of pC. Samples prepared in 30% ethanol are indicated by blue. **(D)** Normalized fluorescence excitation spectra of pCr (red), ApCA (green), CpCC (blue), GpCG (magenta), and UpCU (purple). Approximate peak absorbance wavelengths of each native base (260

nm, 271 nm, 254 nm and 262 nm for A, C, G, and U, respectively) and pC itself (265 nm) are indicated by the tips of the arrows.

RNA	FQY[1]	ε_{347} (M^{-1} cm^{-1})	X^2	$\langle\tau\rangle_{obs}$ (ns)	w_1	τ_1 (ns)	w_2	τ_2 (ns)	w_3	τ_3 (ns)
pCr	0.023 ±0.002	3700[2]	1.14	1.963 ±0.001	0.991 ±0.002	1.930 ±0.005	0.009 ±0.002	5.7 ±0.6	--	--
pCpC[3]	0.0022 ±0.0006	7400[2]	1.22	2.43 ±0.02	0.962±0.006	0.03±0.01	0.035 ±0.006	2.07 ±0.03	0.003 ±0.004	6.8 ±0.3
pCA	0.050 ±0.005	3920 ±10	1.14	4.231 ±0.009	0.75 ±0.02	3.51 ±0.05	0.25 ±0.02	6.5 ±0.1	--	--
ApCA	0.09 ±0.01	2560 ±50	1.10	8.50 ±0.03	0.379 ±0.004	3.27 ±0.07	0.621 ±0.004	11.70 ±0.02	--	--
pCC	0.054 ±0.006	3780 ±60	1.07	4.52 ±0.02	0.701 ±0.005	3.24 ±0.03	0.299 ±0.005	7.52 ±0.06	--	--
CpCC	0.080 ±0.008	3530 ±50	1.08	7.12 ±0.02	0.355 ±0.003	3.33 ±0.06	0.645 ±0.003	9.201 ±0.009	--	--
pCG	0.022 ±0.002	3370 ±40	1.12	2.20 ±0.02	0.864 ±0.007	1.80 ±0.03	0.136 ±0.007	4.78 ±0.08	--	--
GpCG	0.039 ±0.004	2440 ±90	1.14	4.02 ±0.02	0.590 ±0.004	2.22 ±0.03	0.410 ±0.004	6.63 ±0.04	--	--
pCU	0.020 ±0.002	3860 ±60	1.08	2.01 ±0.03	0.58 ±0.03	0.98 ±0.06	0.35 ±0.02	2.8 ±0.1	0.077±0.009	6.2±0.2
UpCU	0.031 ±0.003	2950 ±20	1.06	3.20 ±0.02	0.32 ±0.01	1.36 ±0.05	0.64 ±0.01	3.78 ±0.04	0.035±0.003	9.5±0.3
St1	0.056 ±0.005	3109 ±6	1.12	6.770 ±0.006	0.493 ±0.007	4.46 ±0.03	0.507 ±0.007	9.01 ±0.04	--	--
1:2 GGG	0.026 ±0.002	3420 ±70	1.18	3.225 ±0.005	0.8682 ±0.0003	2.074 ±0.005	0.1318 ±0.0003	10.81 ±0.03	--	--

1:2 GUG	0.056 ±0.006	3460 ±40	1.13	6.21 ±0.01	0.65 ±0.01	4.72 ±0.06	0.35 ±0.01	8.92 ±0.08	--	--

Table 2. Fluorescence quantum yields (FQY), extinction coefficients (ε_{347}) and multi-exponential TCSPC fitting results for all oligonucleotides at 20 °C. Uncertainties for FQY and ε_{347} are the standard deviation across measurements on three samples. Uncertainties for fitting results are the standard deviation across three successive data acquisitions.

[1] Relative FQY determined using pCr (ref. 17) as a reference.

[2] Based on ε_{347} for pC reported at https://www.glenresearch.com/reports/gr15-13

[3] Shortest component omitted in calculation of $\langle\tau\rangle_{obs}$

Absorbance and fluorescence excitation spectra of the di- and trinucleotides all show the expected ~345 nm feature resulting from direct excitation of pC, but that feature's peak wavelength is slightly red-shifted and its extinction coefficient is lower in the trinucleotides relative to pCr (Fig. S3; Table 2). In fluorescence excitation spectra, significant differences in relative intensity and lineshape are observed in the 250-300 nm region (Fig. 1D). This could be attributed to energy transfer from the native bases such that excitation of an adjacent native base results in pC fluorescence. Indeed, the peak wavelength of this feature closely matches the known absorption maximum of the native base present in each oligonucleotide (Fig. 1D) (28). When normalizing by the 345 nm peak to account for differences in overall emission intensity, the variation in the intensity of the excitation peak in the 260 nm region suggests a moderately sequence-dependent energy transfer efficiency. The intensity is highest in pCA and ApCA, which could be explained by the fact that A has the largest extinction coefficient (15,400 $M^{-1}cm^{-1}$) of all four nucleoside monophosphates (NMPs). However, the peak intensity near 260 nm is almost identical in pCC and pCG despite CMP having an extinction coefficient of ~9,000 $M^{-1}cm^{-1}$ compared to GMP's 13,700 $M^{-1}cm^{-1}$, and is substantially lower in pCU despite UMP having an extinction coefficient (10,000 $M^{-1}cm^{-1}$) close to that of CMP (28). This

suggests that energy transfer from G or U is less efficient than from A and C, and/or that the most highly fluorescent conformations of pCG or pCU have low energy transfer efficiencies. In trinucleotides (Fig. 1D), the intensity of the ~260 nm feature in the excitation spectrum more closely tracks with extinction coefficient (A>G>>C), with the exception that it is again lower in UpCU than would be predicted by the extinction coefficient of UMP.

To probe the distribution of structures present in solution, we determined the excited state lifetimes of the oligonucleotides under study using TCSPC (Fig. 2). The resulting fluorescence decays were fitted with exponential or distribution models containing different numbers of decay components (see methods for details). Typically, each decay component is interpreted as a distinct conformational state or a distribution of similar or rapidly interconverting states, with its weight reflecting the prevalence of that state (21, 22). All pC-containing oligonucleotides have longer lifetimes than pCr and their decays exhibit evidence of conformational heterogeneity (Fig. 2). Specifically, the decays of all di- and trinucleotides require at least two components to be adequately fit (χ^2=1.06-1.14), which likely originate from a combination of base-stacked and unstacked states (Table 2). Unlike oligonucleotides containing A, C, and G, the decays of pCU and UpCU are significantly better fit with either a tri-exponential ("3E") model (χ^2=1.08 for pCU; 1.06 for UpCU) or a model containing a fast exponential component and a slower gamma distribution component ("1E+1G", χ^2=1.08 for pCU; 1.09 for UpCU) than with the simple bi-exponential ("2E") model that is adequate for all the others (χ^2=1.25 for pCU; 1.22 for UpCU). More detail about comparison of fitting models is given in Supplementary Note 1.

The shortest decay component is dominant in all dinucleotides, while the population distribution shifts away from the fastest decay component in trinucleotides (Table 2). In addition, the lifetime associated with the faster component is nearly unchanged between A-, C- and G-containing dinucleotides and their corresponding trinucleotides, while the slower component lengthens

significantly in the trinucleotides. We hypothesized that the longer decay component describes a base-stacked conformational state, as this state would be more protected from solvent in a trinucleotide and would be absent in pCr. 3E fits to the decays of pCU and UpCU show that the trinucleotide exhibits longer time constants in all decay components and a larger population associated with the intermediate τ~3 ns component, while the population of the longest decay component remains low (<10%) in both. Though pCU and UpCU sample a more complex distribution of conformational states, we hypothesized that, like the other di- and trinucleotides, the longer of their two heavily populated components describes a base-stacked structure.

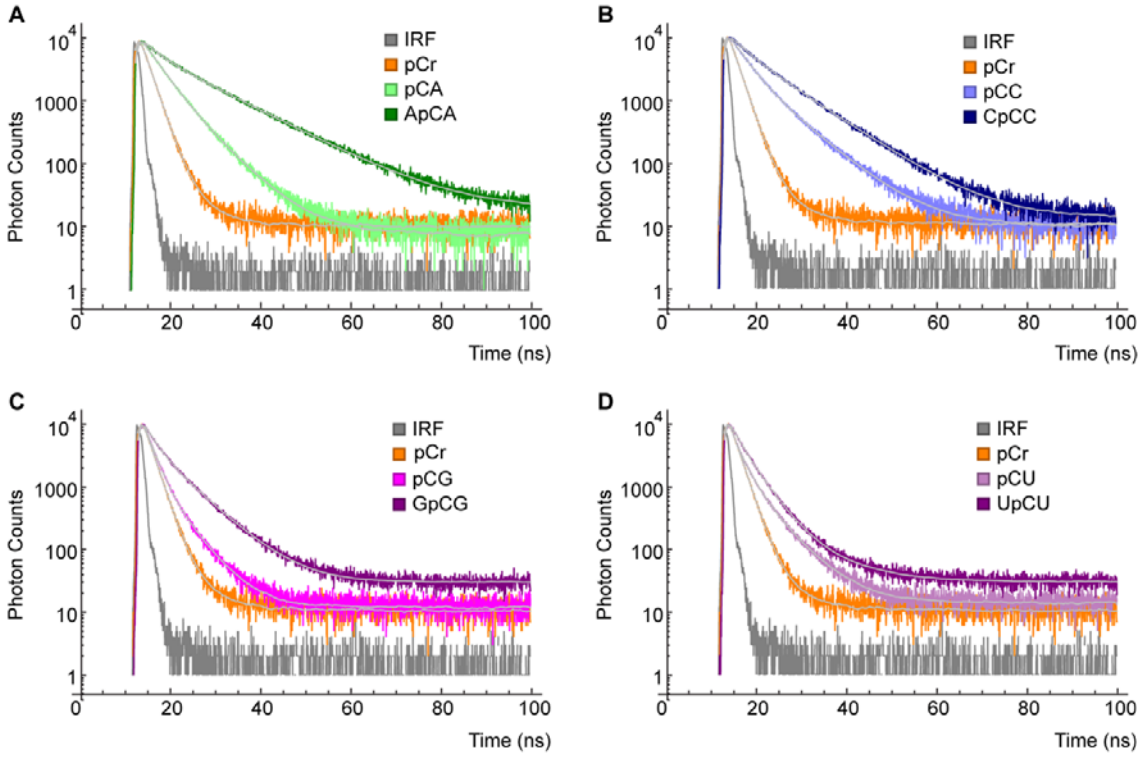


Figure 2. TCSPC measurements of pCr and di- and trinucleotides. **(A)** pCA and ApCA, **(B)** pCC and CpCC, **(C)** pCG and GpCG, **(D)** pCU and UpCU. Bi-exponential **(A, B, C)** or tri-exponential **(D)** fits are overlaid in light gray. A typical IRF is plotted in gray.

To confirm the identities of these conformational states, we used differences in solvent polarity to alter the base-stacking equilibrium. Ethanol has previously been shown to enhance the raw fluorescence intensity of pCr (8) and shift the base-stacking equilibrium of di- and trinucleotides to favor the unstacked state (11, 29). We performed measurements on pCr and all four trinucleotides in the presence of 30% (v/v) EtOH, hypothesizing that ethanol would induce an increase in the population weights of decay components originating from unstacked conformations. In 30% EtOH, an increase in FQY is observed for pCr, CpCC, GpCG and UpCU (Table 3; Fig. 1C), and both decay time constants of all four trinucleotides lengthen (Table 3; Fig. 3; Fig. S4), attributed to the enhancement of fluorescence in a less polar environment (8). Notably, CpCC and ApCA show a significant increase in the population weight of the shorter decay component in 30% EtOH and a

RNA	**FQY**	**ε_{347} ($M^{-1}cm^{-1}$)**	**X^2**	**$\langle\tau\rangle_{obs}$ (ns)**	**w_1**	**τ_1 (ns)**	**w_2**	**τ_2 (ns)**
pCr	0.048 ±0.005	3990 ±70	1.12	3.511 ±0.004	1.00	3.511 ±0.004	--	--
ApCA	0.09 ±0.02	2670 ±80	1.07	8.298 ±0.008	0.494 ±0.004	4.36 ±0.03	0.506 ±0.004	12.15 ±0.04
CpCC	0.092 ±0.009	3740 ±20	1.14	7.63 ±0.02	0.57 ±0.02	4.9 ±0.1	0.43 ±0.02	11.2 ±0.2
GpCG	0.046 ±0.007	2248 ±5	1.08	4.452 ±0.002	0.614 ±0.006	2.68 ±0.02	0.386 ±0.006	7.28 ±0.04
UpCU	0.052 ±0.007	2420 ±90	1.14	4.06 ±0.02	0.68 ±0.01	3.00 ±0.06	0.32 ±0.01	6.28 ±0.08

Table 3. Relative FQY, ε_{347} and bi-exponential TCSPC fitting results for trinucleotides at 20°C in buffer containing 30% v/v ethanol. Uncertainties for FQY and ε_{347} are the standard deviation across measurements on three samples. Uncertainties for fitting results are the standard deviation across three successive data acquisitions.

compensatory decrease in the population weight of the longer one (Table 3). The same shift occurs to a lesser degree in GpCG. This shift in population weight supports the notion that the longer decay component originates from a base-stacked conformation and the shorter decay component, which is favored in less polar solvent, originates from an unstacked conformation. Unlike in buffer, the decay of UpCU in 30% EtOH is fit well (χ^2=1.14) with a 2E model that lacks the low population, slowly decaying component observed in buffer. EtOH impacts the two shorter components in the same manner as observed in the other trinucleotides, with their time constants lengthening and population weight shifting to the shortest decay component. Therefore, while adjacent uracil residues add complexity to the fluorescence decays of pC, the observed population shift suggests a similar structural interpretation.

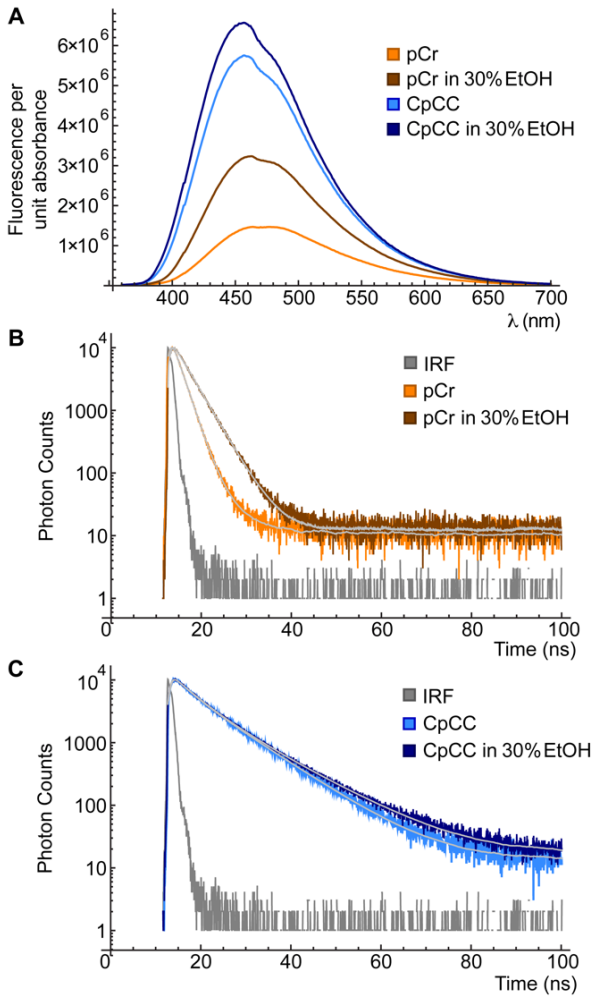


Figure 3. (A) Fluorescence per unit absorbance of pCr (orange) and CpCC (blue) in buffer (lighter shades) or buffer + 30% EtOH (darker shades). **(B)** TCSPC of pCr in buffer or buffer + 30% EtOH. **(C)** TCSPC of CpCC in buffer or buffer + 30% EtOH. Multi-exponential fits are overlaid in light gray (mono-exponential for pCr in 30% EtOH and bi-exponential for all others).

We further investigated the relationship between base-stacking, fluorescence quantum yield and energy transfer by comparing the oligonucleotides' circular dichroism (CD) and fluorescence-detected circular dichroism (FDCD) spectra (Fig. 4). FDCD spectra reveal the CD signals of fluorescent species in a sample, weighted by their populations and fluorescence quantum yields, by measuring the difference in emission intensity of a sample under excitation with left-handed and right-handed circularly polarized light (LCP and RCP, respectively) (23). A large FDCD signal indicates that the brightest species in a sample have large CD signals (typically associated with stacked structures), while a small FDCD signal indicates that the brightest species have small CD signals. Raw FDCD spectra were converted (details in methods) into the quantity $\Delta\varepsilon_F$, which is analogous to the quantity $\Delta\varepsilon$ commonly reported in standard CD (Fig. 4). $\Delta\varepsilon_F$ spectra will henceforth be referred to as "processed FDCD spectra". We also considered the dissymmetry factor $g_F=\Delta\varepsilon_F/\varepsilon_F$, analogous to $g=\Delta\varepsilon/\varepsilon$ obtained from standard CD, which can be computed directly from raw data (details in Supplementary Note 2 and Fig. S5) (24).

In the dinucleotide with the highest FQY, pCC, the CD and processed FDCD spectra agree roughly in lineshape and intensity, with the agreement being nearly quantitative in the long-wavelength band that originates from direct absorption by pC (Fig. 4). In CpCC, the FDCD and CD lineshapes are again very similar, but the longest-wavelength band is more intense in the FDCD spectrum. In both pCC and CpCC, the similarity between the FDCD and CD spectra below 300 nm is strong evidence of the energy transfer that was surmised from fluorescence excitation spectra, showing that differential absorption of LCP and RCP by adjacent C residues is efficiently transduced into differential emission by pC. Together, this indicates that the bright conformations of CpCC have large CD signals >300 nm and efficient energy transfer, both of which would be expected for stacked conformations. In contrast, the FDCD spectrum of pCU is far weaker than its CD spectrum across the

entire spectral range of 200-400 nm. Its FDCD and CD lineshapes are similar below 300 nm, indicating that the majority of this signal again originates from differential

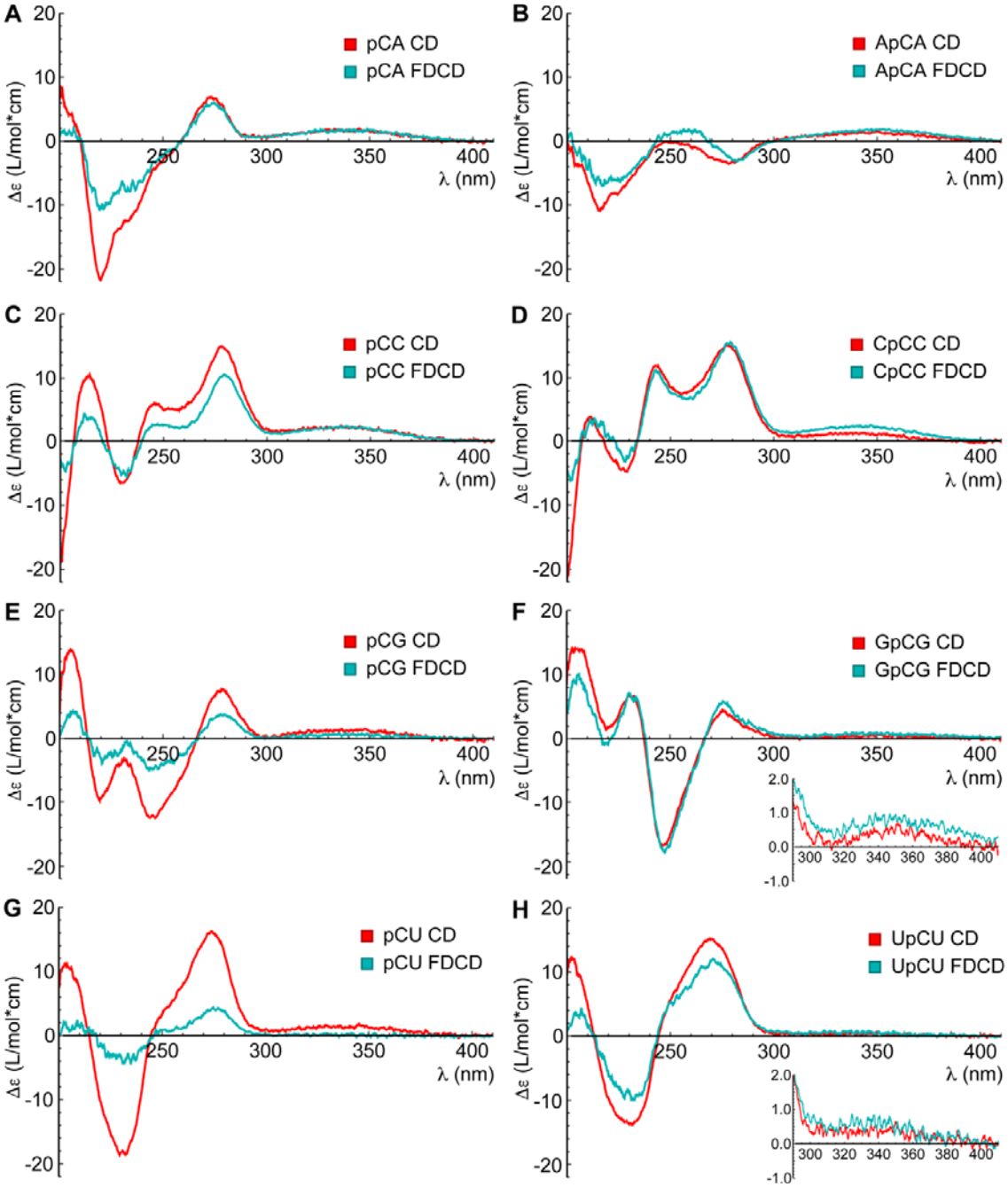


Figure 4. CD (red) and processed FDCD (cyan) spectra of di- and trinucleotides. **(A)** Spectra of pCA. **(B)** ApCA. **(C)** pCC. **(D)** CpCC. **(E)** pCG. **(F)** GpCG (inset: close-up of long-wavelength region). **(G)** pCU. **(H)** UpCU (inset: close-up of long-wavelength region).

absorption of LCP and RCP by U followed by energy transfer. The weak FDCD signal indicates that the bright conformations of pCU are those that have weak CD signals, and thus are likely unstacked, consistent with the fact that pCU is slightly quenched relative to pCr (Fig. 1C). In UpCU, the intensity of the FDCD spectrum approaches that of the CD spectrum, showing that stacked conformations are now bright, which is consistent with the increase in FQY of UpCU relative to pCr. Di- and trinucleotides containing A show very similar patterns of behavior to those containing C, and G exhibits similar patterns of behavior to U (Fig. 4).

Numerous studies utilizing 2-AP have placed probes at two adjacent sites (30–35) such that a base-stacked structure forms a classic "H-type" exciton-coupled homodimer in which the lowest-energy excited state is dark (30, 36). We investigated whether a pC homodimer would exhibit similar properties, given the competing factor that base-stacking in some cases enhances the fluorescence of pC. We found that the dinucleotide pCpC was quenched approximately 10-fold compared to pCr, greatly exceeding any pC monomer-containing oligonucleotide (Table 2). As seen in 2-AP dinucleotide (11, 29), pCpC exhibits a strong sigmoidal feature in its CD spectrum at short wavelength and an FDCD spectrum that is far weaker (Fig. S6). This suggests that the base-stacked conformations that give rise to large CD signals are strongly quenched, and thus contribute minimally to the FDCD signal. Furthermore, in contrast to all pC monomer-containing oligonucleotides, fitting the fluorescence decays of pCpC required a time component too short to be accurately resolved with our benchtop TCSPC instrument (Table 2). Together, these results suggest that pC's distinctive emissive properties may be best exploited using single labels, as the nuanced dependence of its fluorescence on local structure is obscured by the strong self-quenching that occurs in a homodimer.

To further investigate the mechanism of pC's observed sequence-dependent fluorescence enhancement, we calculated excitation energies, oscillator strengths and natural transition orbitals (NTOs) at the CIS(D)/aug-cc-pVDZ level of theory for

CpCC and GpCG (representing strongly and weakly fluorescence-enhancing native bases, respectively) and their constituent nucleobases (Table 4; Fig. S7). Each trinucleotide was modeled by arranging geometry-optimized monomers in an A-form helical geometry with coordinates obtained from the literature (10). The lowest-energy transition in both trimers is a $\pi \rightarrow \pi^*$ transition localized on the pC residue (Fig. S7), with a decreased oscillator strength as compared to the equivalent transition in pC base (34% lower for CpCC and 39% lower for GpCG; Table 4). The $S_0 \rightarrow S_1$ excitation energy decreases relative to pC base in both trimers (by ~0.1 eV in CpCC vs. ~0.05 eV in GpCG). These results are directionally consistent with the experimentally observed reductions in extinction coefficient (Table 2) and redshifts of the longest-wavelength absorbance band (Fig. S3) in the trinucleotides relative to pCr.

Slight differences in behavior due adjacent native base identity are also observed in higher-energy excited states. The S_4 and S_5 states of CpCC show oscillator strengths close to that of S_1 while both being approximately 1 eV greater in energy. The S_4 state shows full charge transfer from the pC base to the 3′ C, while the S_5 state shows localized excitation of the 5′ C (Fig. S7). Likewise, the S_8 and S_9 states of GpCG lie approximately 1 eV higher in energy than S_1 and exhibit high oscillator strengths and excitations localized on the 3′ and 5′ native base, respectively. Unlike the S_4 state of CpCC, the S_8 state of GpCG lacks charge-transfer character. However, its S_{10} state is only slightly higher in energy (~0.01 eV), has an oscillator strength ~1/3 that of S_1 state, and does show charge transfer from pC to the 3′ G. Excitation into these native-base-localized states followed by energy transfer to pC is likely responsible for the ~260 nm features observed in fluorescence excitation and FDCD spectra. With large energy gaps between the bright S_1 state and the other excited states, GpCG and CpCC lack pathways to quenching such as the low-lying charge-transfer and dark states that have been observed in computational studies of other FBAs such as 2-AP (37) and thienoguanosine (thG) (38).

Model	Excited State	Excitation Energy (eV)	Wavelength (nm)	Oscillator Strength	Character of Transition
pC base	S_1	3.77	329	0.2555	π → π*
CpCC	S_1	3.66	338	0.1689	π → π* Localized on pC
	S_2	4.21	294	0.0009	Localized on pC
	S_3	4.53	273	0.0079	Localized on pC
	S_4	4.59	269	0.0954	CT pC → 3′ C
	S_5	4.66	266	0.1480	Localized on 5′ C
	S_6	4.67	265	0.0013	CT pC → 3′ C
	S_7	4.85	255	0.0151	CT pC → 3′ C
	S_8	5.10	243	0.0046	Localized on pC
	S_9	5.17	240	0.0147	Localized on 5′ C
	S_{10}	5.27	235	0.0029	Delocalized
GpCG	S_1	3.70	335	0.1572	π → π* Localized on pC
	S_2	4.32	287	0.0062	Localized on pC
	S_3	4.72	263	0.0049	Localized on pC
	S_4	4.77	260	0.0159	Localized on 3′ G
	S_5	4.79	259	0.0014	Localized on 5′ G
	S_6	4.83	257	0.0106	Localized on 5′ G
	S_7	4.90	253	0.0059	CT pC → 5′ G
	S_8	4.97	249	0.0886	Localized on 3′ G
	S_9	5.02	247	0.1477	Localized on 5′ G
	S_{10}	5.04	246	0.0574	CT pC → 3′ G

Table 4. Summary of excited state calculations for CpCC and GpCG with the S_1 state of pC base for comparison. Charge transfer states are indicated as "CT".

To investigate pC's potential to report on aspects of RNA structure other than base-stacking, we compared a single-stranded (ss) 21-nucleotide RNA with a central CpCC sequence to a fully complementary double-stranded (ds) RNA containing a G opposite pC and a duplex containing a U opposite pC (mismatch; mm) (Table 1). Coupling steady-state and time-resolved spectroscopic methods again proved to be useful as they are sensitive to different aspects of pC's local environment. As previously observed in DNA and RNA (4, 17), pC's fluorescence is quenched significantly when opposite a complementary G (Fig. 1C; Table 2). This is accompanied by the appearance of a dominant (>80% population weight) ~2 ns component in its fluorescence decay (Table 2). In contrast, mm dsRNA recovers nearly the same fluorescence intensity as ssRNA with comparable decay time constants and weights when analyzed with a bi-exponential model (χ^2=1.12 for ss; 1.13 for mm). Interestingly, the decay of mm dsRNA is also well described (χ^2= 1.11) by a single broad distribution of conformations and associated lifetimes (Table 5). This "1G" model fits the decay of ssRNA only moderately well (χ^2= 1.23) and fits the decay of fully complementary dsRNA very poorly (χ^2>2), indicating that mm dsRNA can access a more continuous range of conformational states. While base-paired structures are strongly distinguished from unpaired (ss or mm) structures by their fluorescence intensity and decays, we found that single-stranded environments are clearly distinguished from ds or mm by their excitation and emission spectra (Fig. 5). Specifically, fully complementary and mismatched duplexes both show evidence of inter-strand energy transfer in their fluorescence excitation spectra through a blueshift of the ~260 nm feature toward the absorbance maximum of G. Their emission spectra are blueshifted relative to ssRNA and have a less prominent red vibronic shoulder (Fig. 5).

Unsurprisingly, the CD spectra of fully complementary and mismatched dsRNA are very similar, while the CD spectrum of ssRNA is weaker and has a different lineshape (Fig. 6). The processed FDCD spectra are far weaker than the CD spectra in

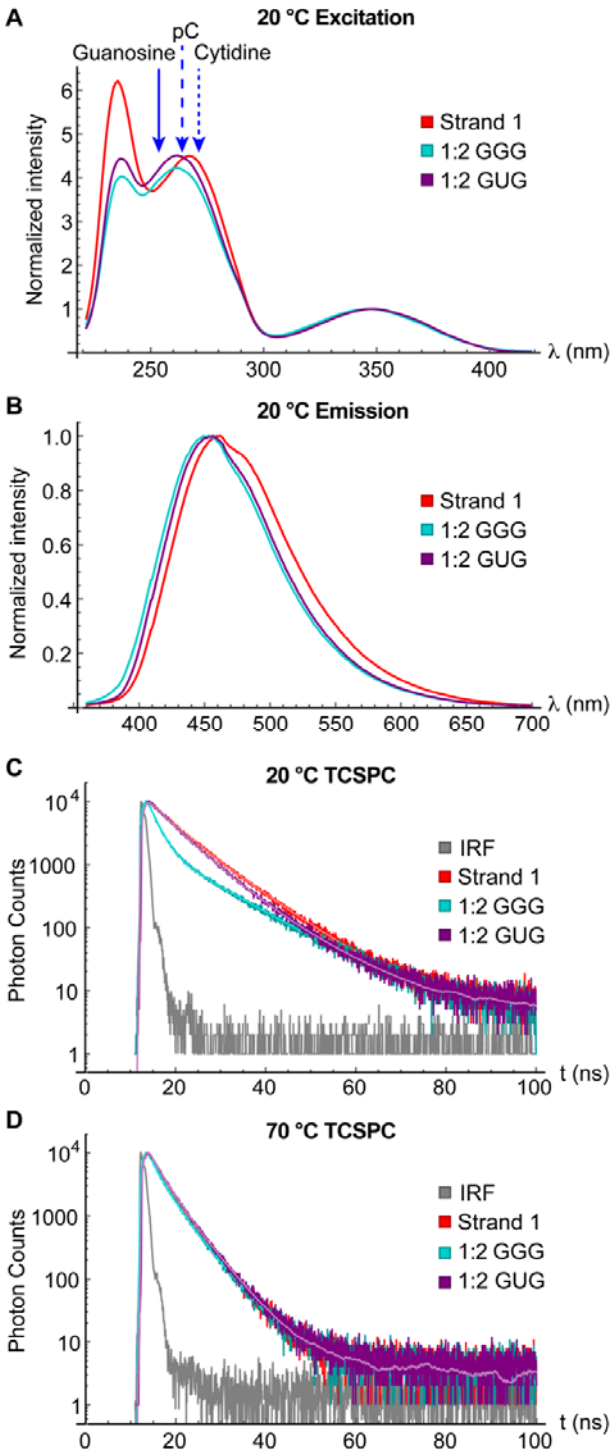


Figure 5. Steady-state fluorescence spectra and TCSPC of single-stranded ("Strand 1"; red), double-stranded ("1:2 GGG"; cyan) and mismatched ("1:2 GUG"; purple) 21-mers. **(A)** Normalized fluorescence excitation spectra of ss, ds and mm RNA at 20 °C. Emission was detected at 460 nm. The approximate peak absorption wavelengths of guanosine (solid blue arrow), cytidine (dotted arrow) and a secondary absorption peak of pC (dashed arrow) are indicated. **(B)** Normalized fluorescence emission spectra of ss, ds and mm RNA at 20 °C under 340 nm excitation. **(C)** IRF (gray) and fluorescence decays of ss, ds and mm RNA at 20 °C. Fits (bi-exponential for Strand 1 and 1:2 GGG; 1 gamma for 1:2 GUG) are overlaid in lighter color. **(D)** IRF (gray) and fluorescence decays of ss, ds and mm RNA at 70 °C. 1 exponential +1 gamma fits are overlaid in lighter color.

all cases, which is unsurprising given that they arise only from excitation of pC and bases that transfer energy to it. A positive CD signal is observed at long wavelength in ds and mm RNA, as seen in all di- and trinucleotides (Fig. 4). Notably, the FDCD signal of fully complementary dsRNA is negative at long wavelength, while the long-wavelength features of ss and mm RNAs are positive. This agrees with the observation via TCSPC that pC experiences similar local environments in ss and mm 21-mers, sampling states with lifetimes of approximately 4.5 and 9 ns in both structural contexts when analyzed with a bi-exponential model (Table 2). This stands in contrast to the distinctive (τ=2.1 ns) emissive state of pC in dsRNA. The FDCD signal in the region around 260 nm primarily reflects the optical activity of the native bases that transfer energy to pC (as seen in di- and trinucleotides), and these are observed to adopt a similar geometry in both fully complementary and mismatched duplexes (Fig. 6B).

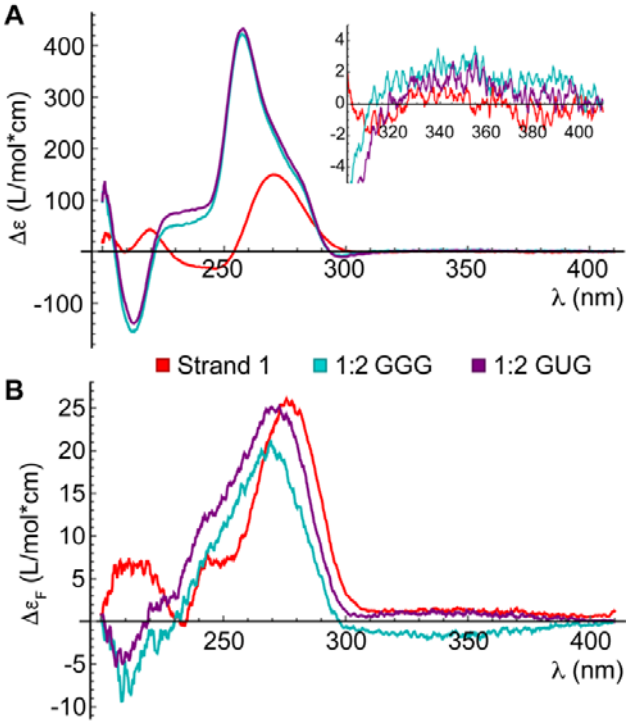


Fig 6. (A) CD spectra of ss (red), ds (cyan), and mm (purple) 21-mers. The inset shows a close-up of the 300-410 nm range. **(B)** Processed FDCD spectra of ss (red), ds (cyan), and mm (purple) 21-mers.

We next investigated ss, ds and mm 21-mers at elevated temperature. In all three cases, increasing the temperature leads to a decrease in the overall emission

intensity (Fig. S8) and a shortening of the time constant of each decay component (Table 5, Fig. S9). ssRNA shows a population shift from the stacked (longer lifetime) to the unstacked conformation as the temperature is raised. Fully complementary dsRNA shows a population shift from its unique ~2 ns (presumably base-paired)

RNA	**°C**	χ^2	$\langle\tau\rangle_{obs}$ **(ns)**	**Exponential Components**				**Gamma Components**	
				w_1	τ_1 **(ns)**	w_2	τ_2 **(ns)**	w_3	$\langle\tau\rangle_3$ **(ns)**
St1	20	1.12	6.770 ±0.006	0.493 ±0.007	4.46 ±0.03	0.507 ±0.007	9.01 ±0.04	--	--
	30	1.15	6.10 ±0.04	0.442 ±0.007	3.68 ±0.08	0.558 ±0.007	8.02 ±0.06	--	--
	40	1.14	5.38 ±0.01	0.478 ±0.004	3.27 ±0.04	0.522 ±0.004	7.31 ±0.02	--	--
	50	1.16	4.612 ±0.009	0.39 ±0.01	6.32 ±0.04	--	--	0.61 ±0.01	3.54 ±0.03
	60	1.16	3.90 ±0.02	0.35 ±0.02	5.5 ±0.1	--	--	0.65 ±0.02	3.06 ±0.05
	70	1.17	3.35 ±0.03	0.27 ±0.03	4.9 ±0.3	--	--	0.73 ±0.03	2.78 ±0.04
1:2 GGG	20	1.18	3.225 ±0.005	0.8682 ±0.0003	2.074 ±0.005	0.1318 ±0.0003	10.81 ±0.03	--	--
	30	1.15	2.668 ±0.003	0.838 ±0.001	1.629 ±0.003	--	--	0.162 ±0.001	8.04 ±0.05
	40	1.15	2.204 ±0.009	0.834 ±0.004	1.36 ±0.02	--	--	0.166 ±0.004	6.45 ±0.09
	50	1.18	1.744	0.77	1.09	--	--	0.23	4.0

			±0.004	±0.01	±0.01			±0.01	±0.1
	60	1.18	1.54 ±0.05	0.674 ±0.007	0.83 ±0.08	--	--	0.326 ±0.007	3.01 ±0.03
	70	1.14	2.92 ±0.03	0.29 ±0.03	4.6 ±0.1	--	--	0.71 ±0.03	2.22 ±0.08
1:2 GUG	20[1]	1.13	6.21 ±0.01	--	--	--	--	1.00	6.234 ±0.004
	30	1.22	5.41 ±0.01	--	--	--	--	1.00	5.41 ±0.01
	40	1.18	4.693 ±0.009	--	--	--	--	1.00	4.693 ±0.009
	50	1.22	4.10 ±0.01	--	--	--	--	1.00	4.10 ±0.01
	60	1.09	3.827 ±0.09	0.24 ±0.02	5.6 ±0.1	--	--	0.76 ±0.02	3.27 ±0.04
	70	1.16	3.37 ±0.01	0.218 ±0.008	5.0 ±0.1	--	--	0.782 ±0.008	2.93 ±0.02

Table 5. TCSPC fitting results for ss, ds, and mm 21-mers. The optimal fitting model for each circumstance (bi-exponential, 1 exponential+1 gamma, or 1 gamma) is indicated by which columns are occupied. Uncertainties are the standard deviation across three successive data acquisitions.

[1]Bi-exponential fit provided in Table 2.

conformation to a conformation with a lifetime similar to the longer one observed in ssRNA and mm dsRNA. Both duplexes exhibit two transitions in UV melting curves (Fig. S10). A transition at 47 °C for fully complementary dsRNA and 45 °C for mm dsRNA likely corresponds to melting of the AU-rich tails, while a second transition at

65 °C for fully complementary dsRNA and 55 °C for mm dsRNA results in full strand dissociation. Just above its higher T_M, both duplexes attain nearly identical fluorescence spectra (Fig. S8) and decay properties (Table 5; Figs. 5 and S9) to ssRNA. At all temperatures at which mm dsRNA is intact, its decays are best fit with a single distribution of states (discussed further in Supplementary Note 1). From 30 to 60 °C, the slower decay component of fully complementary dsRNA is best described with a distribution rather than an exponential component, suggesting a disordered ensemble of non-base-paired conformations. The faster decay component of ssRNA is best described with a distribution from 50 to 70 °C (Table 5, Table S3). The fluorescence excitation and emission spectra of mm and fully complementary dsRNA gradually approach those of ssRNA as the temperature is increased (Fig. S8). Specifically, the signature of energy transfer from G disappears from the excitation spectrum and the emission spectrum exhibits a redshift and increase in the red vibronic shoulder characteristic of increased solvent exposure.

DISCUSSION

In this study, we performed a detailed analysis of the photophysical behavior of pC in RNA using multiple complementary fluorescence-detection methods. Steady-state fluorescence spectroscopy showed that the fluorescence properties of pC in RNA are altered relative to that of monomeric pCr in a sequence-dependent manner. In dinucleotides, the FQY of pC is enhanced relative to pCr when adjacent to an A or C residue and is nearly unchanged when adjacent to a G or U residue. In trinucleotides, the FQY of pC is further enhanced when flanked by A or C residues and enhanced to a much lesser degree when flanked by G or U. The emission spectra of oligonucleotides of all sequences show a consistent hypsochromic shift in emission wavelength and suppression of a long-wavelength vibronic shoulder relative to pCr. The sequence-independent nature of these alterations to the emission spectrum, as well as their appearance when pCr is measured in 30% EtOH, suggest that they predominantly reflect the polarity of the local environment

experienced by pC. Indeed, the extent of the changes in both λ_{max} and lineshape in fully aqueous solvent is correlated to the degree of protection from the solvent, where the spectra of dinucleotides deviate the least from pCr and the spectrum of fully complementary dsRNA deviates the most (dsRNA ≈ mismatched duplex >> ssRNA ≈ CpCC ≈ pCC >> pCr). In contrast, 2-AP experiences minimal changes in lineshape or λ_{max} across the analogous series of 2-APr, (2-AP)C and C(2-AP)C (Fig. S11) (11). Observation of inter-strand energy transfer in fluorescence excitation spectra of pC further differentiates double-stranded from single-stranded environments. Facile measurement of steady-state fluorescence spectra can therefore give a general identification of the local environment of pC in RNA of unknown structure.

While comparison to 2-AP is illuminating, other FBAs exhibit properties more similar to pC such as moderate variation in FQY and environment-sensitive emission spectra. A standout among "isomorphic" FBAs (those that at least roughly retain the size and shape of their cognate native bases) is thG (39). It has been widely used as a probe of conformational changes (40, 41), including those induced by protein binding (42–45) and catalysis (46), and as a scaffold for the development of more specialized FBA-based probes (47, 48). The emission of thG is sensitive to both base stacking, which induces a moderate reduction in FQY from 0.46 for the free nucleoside to ~0.1 when flanked by Gs or Ts, and base pairing (39, 49).This ~5-fold difference in FQY is similar to the contrast we observe between the brightest and darkest pC-labeled oligonucleotides, though the maximum FQY we observe for pC (0.09) is substantially lower.

The most widely used and studied fluorescent cytosine analogues include an expansive family of tricyclic probes such as tC (50), tC^0 (51), DEA-tC (52), and ABN (53). tC and tC^0 are noteworthy for having nearly identical FQY (~0.2) as free nucleosides and in single-stranded and double-stranded nucleic acids (50, 51), while DEA-tC has the unusual property of being brightest in double-stranded nucleic acids (52, 54). The minimally fluorescent tC_{nitro} (55), has been used as a FRET acceptor with

tC^0 as the donor (56), and the exceptionally bright ABN is readily detectable at the single molecule level (53). pC exists in a space complementary to these advanced FBAs, with its small size and environment-sensitive fluorescence justifying its selection when a general probe of nucleic acid structure and/or protein interactions is desired.

Time-resolved fluorescence measurements demonstrated that pC typically samples two distinct environments with different emissive properties in short RNA oligonucleotides. This stands in contrast to previous measurements in which the decays of the DNA dinucleotides deoxyribo-pCT (dpCT), dpCC, dpCA, dpCG and trinucleotide dTpCT were all adequately fit with mono-exponential models, and only dGpCG required a bi-exponential model (17). The sequence dependence of both intensity and lifetime observed in DNA was far more mild than we observe in RNA (17). The previously studied DNA oligonucleotides may interconvert between stacked and unstacked conformations at a rate on the order of or faster than the relaxation of pC, which would prevent their resolution into distinct decay components. Both computational (57) and experimental (58) results have suggested that this factor impacts the fluorescence decays of 2-AP in short DNA oligonucleotides. It is also possible that, in contrast to our observations in RNA, pC has similar lifetimes in the stacked and unstacked conformations of certain DNA dinucleotides. Indeed, it was proposed based on molecular dynamics simulations that dpCT exists on average in a stacked conformation, suggesting that a T residue stacked on only one side of pC is insufficient to alter its photophysics (17). Our observation that the decays of pCU and UpCU are best fit with a tri-exponential or distribution model indicates that they lack a single, well-defined base-stacked structure. This may reflect the fact that base-stacked structures of U-containing oligonucleotides are generally less stable than those of other native bases (59–61).

Through comparison of the fluorescence decays of different RNA species, comparison of fully aqueous solvent to 30% ethanol, and comparison of CD and FDCD spectra, we provide strong evidence that the brighter, more slowly decaying

species occupies a base-stacked structure. FDCD previously revealed the exact opposite for 2-AP, with the bright conformations being so drastically unstacked that their CD spectra resemble free 2-AP riboside (11). The bright conformations that FDCD reports on additionally exhibit, with the exception of pCU, efficient energy transfer from native bases to pC. We note that a previous publication reported attempting FDCD measurements on pC-labeled DNA (in a TpCA sequence context) in the 300-420 nm range, but no signal was observed with the instrumentation used in that study even at higher concentrations than those used here (62).

The enhancement of pC fluorescence in trinucleotides of all sequence contexts compared to the corresponding dinucleotides is accompanied by both a shift toward the stacked conformation and a slowing of this conformation's decay rate, suggesting that the stacked conformation is both more prevalent and more emissive in trinucleotides. Previous computational studies have found that the $S_0 \rightarrow S_1$ transition of pC exhibits reduced oscillator strength when it is base-stacked with G on one or both sides in the geometry of an A- or B-form helix, reducing the predicted rate of radiative relaxation (10, 18, 63). Our computational studies replicate this finding and show that it holds to a slightly lesser degree in CpCC, which is consistent with the experimentally observed decrease in the molar absorptivity of pC in trinucleotides relative to pCr (Table 2). This suggests that the enhancement of pC fluorescence observed in both DNA (17) and RNA trinucleotides relative to a pC monomer originates from a decrease in the overall rate of nonradiative decay. The rigidity and solvent protection afforded to pC in a trinucleotide as compared to pCr, combined with the conspicuous absence of low-lying charge transfer or dark states, could enable the non-radiative rate to decrease in trinucleotides to a greater degree than the radiative rate, giving rise to the observed enhancement of fluorescence.

It is worth noting that among pC monomer-containing oligonucleotides, sub-nanosecond components were found only for the 3E and 1E+1G fits to the decays of pCU and UpCU in buffer, respectively (Tables 2 and S2). This suggests that our benchtop TCSPC instrumentation is sufficient to resolve all decay components of the

pC-containing oligonucleotides investigated. If the decays of all conformations are indeed being successfully resolved, the intensity ratio between pCr and an oligonucleotide should be roughly equal to the ratio of their $\langle\tau\rangle_{obs}$ values (64). In Table 6, these ratios are compared for two pC- and 2-AP-labeled oligonucleotides of analogous sequence, measured under identical experimental conditions. It is clearly evident that the observed changes in lifetime are adequate to explain the changes in intensity of these pC-containing oligonucleotides, while the same is not true of 2-AP (11). This is consistent with the repeated observation of decay components in the 10s of picoseconds or shorter for 2-AP-containing oligonucleotides in ultrafast measurements (65, 66). Thus, our results suggest that pC can be more thoroughly characterized with readily accessible instrumentation than 2-AP, a benefit that was also noted when oligonucleotides containing thG were compared to equivalent sequences containing 2-AP (67).

	X=2-AP[1]		**X=pC[2]**	
Sequence	**FQY/FQY$_{riboside}$**	**$\langle\tau\rangle_{obs}/\langle\tau\rangle_{obs,riboside}$**	**FQY/FQY$_{riboside}$**	**$\langle\tau\rangle_{obs}/\langle\tau\rangle_{obs,riboside}$**
XC	0.25	0.55	2.35	2.25
CXC	0.10	0.44	3.49	3.56

Table 6. FQY and observed average lifetime ratios (oligonucleotide : riboside) for 2-AP and pC.

[1]Recalculated from data in ref. 11

[2]This work

pC proved to be a sensitive probe of base-pairing, as previously observed in both systematic investigation of model DNA oligonucleotides (17) and benchmarking using variants of biologically relevant RNA (4). We found that a fully complementary duplex exhibited significantly reduced emission intensity relative to ssRNA and

mismatched dsRNA, showing that base-pairing facilitates nonradiative relaxation of pC. Tautomerization in the excited state via double proton transfer between pC and G has been identified computationally as a plausible quenching mechanism (68). We further observed a dominant ~2 ns decay component unique to fully complementary dsRNA, which is comparable to the reported average lifetimes of pC in a variety of dsRNA fragments designed to mimic segments of a T-box riboswitch-tRNA complex (4). We observed that dsRNA and mismatched duplex have very similar emission spectra that indicate that pC occupies a similarly nonpolar environment in both cases, while ssRNA and mismatched duplex have similar emission intensities and decays (Fig. 5). The quenching of pC in fully complementary dsRNA therefore seems to be a direct consequence of base-pairing rather than the protected and base-stacked structural context of dsRNA.

In this work, we take a multi-pronged approach to empirically characterize the structure-sequence-photophysics relationships of pC in RNA. Further experimental and computational investigation is required to identify the precise photophysical mechanisms underlying the results reported here. Previous theoretical studies of base-stacked complexes have identified mechanisms by which pC could be quenched by G, such as electron transfer from pC (17, 18) and reductions in the oscillator strength of the fluorescence transition (10, 63), but do not explain the overall enhancement of pC fluorescence observed in GpCG. While our computational results also do not reveal an obvious source of fluorescence enhancement, even in CpCC, they do show that certain quenching mechanisms that impact other FBAs are absent in pC. In future work, we plan to dissect the relaxation pathways that occur prior to emission by utilizing ultrafast fluorescence upconversion spectroscopy to measure time-resolved emission spectra at a sub-picosecond timescale.

The wide-ranging empirical characterization reported here will facilitate informed design of biologically relevant pC-containing RNA sequences and nuanced interpretation of the ensuing findings. We demonstrate that steady-state emission spectra reflect the polarity of the local environment of pC, while the energy transfer

observed in excitation spectra distinguishes single- from double-stranded environments. TCSPC resolves heterogeneity in base stacking and hydrogen bonding and FDCD confirms that base-stacked structures are generally brighter and exhibit efficient energy transfer. The findings reported here will be applied to further develop pC as a reporter of RNA conformational changes in response to environmental stimuli.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

AUTHOR CONTRIBUTIONS

T.L.C.: Conceptualization, Investigation, Visualization, Formal Analysis, Writing - original draft. J.R.W.: Conceptualization, Investigation, Visualization, Supervision, Funding Acquisition, Writing - review and editing.

ACKNOWLEDGEMENTS

The authors acknowledge Dr. Jim Prell of the University of Oregon and lab member Austin Green for contributing mass spectrometry instrumentation and expertise.

FUNDING

This work was supported by the National Science Foundation [2338251 to J.R.W.].

CONFLICT OF INTEREST

None declared

REFERENCES

1. Cantor,C.R. (1971) Fluorescence Studies of Biopolymer Structure. *Trans. N. Y. Acad. Sci.*, **33**, 576–585.

2. Dirks,R.W. and Tanke,H.J. (2006) Advances in Fluorescent Tracking of Nucleic Acids in Living Cell. *BioTechniques*, **40**, 489–495.

3. Hoeher,J.E., Sande,N.E. and Widom,J.R. (2024) Probing and perturbing riboswitch folding using a fluorescent base analogue. *Photochem. Photobiol.*, **100**, 419–433.

4. Zhang,J. and Ferré-D'Amaré,A.R. (2014) Direct Evaluation of tRNA Aminoacylation Status by the T-Box Riboswitch Using tRNA-mRNA Stacking and Steric Readout. *Mol. Cell*, **55**, 148–155.

5. Daniels,M. and Hauswirth,W. (1971) Fluorescence of the Purine and Pyrimidine Bases of the Nucleic Acids in Neutral Aqueous Solution at 300 ° K. *Science*, **171**, 675–677.

6. Jean,J.M. and Hall,K.B. (2001) 2-Aminopurine fluorescence quenching and lifetimes: Role of base stacking. *Proc. Natl. Acad. Sci.*, **98**, 37–41.

7. Hall,K.B. (2009) 2-Aminopurine as a Probe of RNA Conformational Transitions. In *Methods in Enzymology*. Elsevier, Vol. 469, pp. 269–285.

8. Tinsley,R.A. and Walter,N.G. (2006) Pyrrolo-C as a fluorescent probe for monitoring RNA secondary structure formation. *RNA*, **12**, 522–529.

9. Bharill,S., Sarkar,P., Ballin,J.D., Gryczynski,I., Wilson,G.M. and Gryczynski,Z. (2008) Fluorescence Intensity Decays of 2-Aminopurine solutions: Lifetime Distribution Approach. *Anal. Biochem.*, **377**, 141–149.

10. Hardman,S.J.O. and Thompson,K.C. (2006) Influence of Base Stacking and Hydrogen Bonding on the Fluorescence of 2-Aminopurine and Pyrrolocytosine in Nucleic Acids. *Biochemistry*, **45**, 9145–9155.

11. Widom,J.R. and Hoeher,J.E. (2022) Base-Stacking Heterogeneity in RNA Resolved by Fluorescence-Detected Circular Dichroism Spectroscopy. *J. Phys. Chem. Lett.*, **13**, 8010–8018.

12. Berry,D.A., Jung,K.-Y., Wise,D.S., Sercel,A.D., Pearson,W.H., Mackie,H., Randolph,J.B. and Somers,R.L. (2004) Pyrrolo-dC and pyrrolo-C: fluorescent analogs of cytidine and 2′-deoxycytidine for the study of oligonucleotides. *Tetrahedron Lett.*, **45**, 2457–2461.

13. Johnson,N.P., Baase,W.A. and von Hippel,P.H. (2005) Investigating local conformations of double-stranded DNA by low-energy circular dichroism of pyrrolo-cytosine. *Proc. Natl. Acad. Sci. U. S. A.*, **102**, 7169–7173.

14. Dash,C., Rausch,J.W. and Le Grice,S.F.J. (2004) Using pyrrolo-deoxycytosine to probe RNA/DNA hybrids containing the human immunodeficiency virus type-1 3′ polypurine tract. *Nucleic Acids Res.*, **32**, 1539–1547.

15. Zhang,X. and Wadkins,R.M. (2009) DNA Hairpins Containing the Cytidine Analog Pyrrolo-dC: Structural, Thermodynamic, and Spectroscopic Studies. *Biophys. J.*, **96**, 1884–1891.

16. Buskiewicz,I.A. and Burke,J.M. (2012) Folding of the hammerhead ribozyme: Pyrrolo-cytosine fluorescence separates core folding from global folding and reveals a pH-dependent conformational change. *RNA*, **18**, 434–448.

17. Hardman,S.J.O., Botchway,S.W. and Thompson,K.C. (2008) Evidence for a Nonbase Stacking Effect for the Environment-sensitive Fluorescent Base Pyrrolocytosine—Comparison with 2-Aminopurine. *Photochem. Photobiol.*, **84**, 1473–1479.

18. Nguyen,Q.L., Spata,V.A. and Matsika,S. (2016) Photophysical properties of pyrrolocytosine, a cytosine fluorescent base analogue. *Phys. Chem. Chem. Phys.*, **18**, 20189–20198.

19. Fried,B. and Sherma,J. (1994) Thin-layer chromatography: techniques and applications 3rd ed., rev.expanded. M. Dekker, New York.

20. Kolassa,N., Roos,H. and Pfleger,K. (1972) A separation of purine derivatives by thin-layer chromatography on silica gel plates suitable for metabolic studies. *J. Chromatogr. A*, **66**, 175–177.

21. Smith,D.A., McKenzie,G., Jones,A.C. and Smith,T.A. (2017) Analysis of time-correlated single photon counting data: a comparative evaluation of deterministic and probabilistic approaches. *Methods Appl. Fluoresc.*, **5**, 042001.

22. Fogarty,A.C., Jones,A.C. and Camp,P.J. (2011) Extraction of lifetime distributions from fluorescence decays with application to DNA-base analogues. *Phys. Chem. Chem. Phys.*, **13**, 3819–3830.

23. Turner,D.H., Tinoco,I.Jr. and Maestre,M. (1974) Fluorescence detected circular dichroism. *J. Am. Chem. Soc.*, **96**, 4340–4342.

24. Tinoco,I. and Turner,D.H. (1976) Fluorescence detected circular dichroism. Theory. *J. Am. Chem. Soc.*, **98**, 6453–6456.

25. Dunning,T.H.,Jr. (1989) Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.*, **90**, 1007–1023.

26. Martin,R.L. (2003) Natural transition orbitals. *J. Chem. Phys.*, **118**, 4775–4777.

27. Frisch,M.J., Trucks,G.W., Cheeseman,J.R., Scalmani,G., Caricato,M., Hratchian,H.P., Li,X., Barone,V., Bloino,J., Zheng,G., *et al.* Gaussian 09.

28. Cavaluzzi,M.J. and Borer,P.N. (2004) Revised UV extinction coefficients for nucleoside-5′-monophosphates and unpaired DNA and RNA. *Nucleic Acids Res.*, **32**, e13.

29. Johnson,N.P., Baase,W.A. and von Hippel,P.H. (2004) Low-energy circular dichroism of 2-aminopurine dinucleotide as a probe of local conformation of DNA and RNA. *Proc. Natl. Acad. Sci.*, **101**, 3426–3431.

30. Widom,J.R., Johnson,N.P., von Hippel,P.H. and Marcus,A.H. (2013) Solution conformation of 2-aminopurine (2-AP) dinucleotide determined by ultraviolet 2D fluorescence spectroscopy (UV-2D FS). *New J. Phys.*, **15**.

31. Jose,D., Weitzel,S.E. and von Hippel,P.H. (2012) Breathing fluctuations in position-specific DNA base pairs are involved in regulating helicase movement into the replication fork. *Proc. Natl. Acad. Sci.*, **109**, 14428–14433.

32. Datta,K., Johnson,N.P. and von Hippel,P.H. (2010) DNA conformational changes at the primer-template junction regulate the fidelity of replication by DNA polymerase. *Proc. Natl. Acad. Sci.*, **107**, 17980–17985.

33. Datta,K. and von Hippel,P. (2008) Direct Spectroscopic Study of Reconstituted Transcription Complexes Reveals That Intrinsic Termination Is Driven Primarily by Thermodynamic Destabilization of the Nucleic Acid Framework. *J. Biol. Chem.*, **283**, 3537–3549.

34. Datta,K., Johnson,N.P. and von Hippel,P.H. (2006) Mapping the Conformation of the Nucleic Acid Framework of the T7 RNA Polymerase Elongation Complex in Solution Using Low-energy CD and Fluorescence Spectroscopy. *J. Mol. Biol.*, **360**, 800–813.

35. Jose,D., Datta,K., Johnson,N.P. and von Hippel,P.H. (2009) Spectroscopic studies of position-specific DNA "breathing" fluctuations at replication forks and primer-template junctions. *Proc. Natl. Acad. Sci. USA*, **106**, 4231–4236.

36. Kasha,M., Rawls,H.R. and El-Bayoumi,M.A. (1965) The exciton model in molecular spectroscopy. *Pure Appl. Chem.*, **11**, 371–392.

37. Jean,J.M. and Hall,K.B. (2002) 2-Aminopurine Electronic Structure and Fluorescence Properties in DNA. *Biochemistry*, **41**, 13152–13161.

38. Kuchlyan,J., Martinez-Fernandez,L., Mori,M., Gavvala,K., Ciaco,S., Boudier,C., Richert,L., Didier,P., Tor,Y., Improta,R., *et al.* (2020) What Makes Thienoguanosine an Outstanding Fluorescent DNA Probe? *J. Am. Chem. Soc.*, **142**, 16999–17014.

39. Shin,D., Sinkeldam,R.W. and Tor,Y. (2011) Emissive RNA Alphabet. *J. Am. Chem. Soc.*, **133**, 14912–14915.

40. Park,S., Otomo,H., Zheng,L. and Sugiyama,H. (2014) Highly emissive deoxyguanosine analogue capable of direct visualization of B–Z transition. *Chem. Commun.*, **50**, 1573–1575.

41. Sholokh,M., Sharma,R., Grytsyk,N., Zaghzi,L., Postupalenko,V.Y., Dziuba,D., Barthes,N.P.F., Michel,B.Y., Boudier,C., Zaporozhets,O.A., *et al.* (2018) Environmentally Sensitive Fluorescent Nucleoside Analogues for Surveying Dynamic Interconversions of Nucleic Acid Structures. *Chem. – Eur. J.*, **24**, 13850–13861.

42. Grytsyk,N., Richert,L., Didier,P., Dziuba,D., Ciaco,S., Mazzoleni,V., Lequeu,T., Mori,M., Tor,Y., Martinez-Fernandez,L., *et al.* (2022) Thienoguanosine, a unique non-perturbing reporter for investigating rotational dynamics of DNA duplexes and their complexes with proteins. *Int. J. Biol. Macromol.*, **213**, 210–225.

43. Bhujbalrao,R., Gavvala,K., Singh,R.K., Singh,J., Boudier,C., Chakrabarti,S., Patwari,G.N., Mély,Y. and Anand,R. (2022) Identification of allosteric hotspots regulating the ribosomal RNA binding by antibiotic resistance-conferring Erm methyltransferases. *J. Biol. Chem.*, **298**, 102208.

44. Zaayter,L., Mori,M., Ahmad,T., Ashraf,W., Boudier,C., Kilin,V., Gavvala,K., Richert,L., Eiler,S., Ruff,M., *et al.* (2019) A Molecular Tool Targeting the Base-Flipping Activity of Human UHRF1. *Chem. – Eur. J.*, **25**, 13363–13375.

45. Kilin,V., Gavvala,K., Barthes,N.P.F., Michel,B.Y., Shin,D., Boudier,C., Mauffret,O., Yashchuk,V., Mousli,M., Ruff,M., *et al.* (2017) Dynamics of Methylated Cytosine Flipping by UHRF1. *J. Am. Chem. Soc.*, **139**, 2520–2528.

46. Cong,D., Li,Y., Ludford III,P.T. and Tor,Y. (2022) Isomorphic Fluorescent Nucleosides Facilitate Real-Time Monitoring of RNA Depurination by Ribosome Inactivating Proteins. *Chem. – Eur. J.*, **28**, e202200994.

47. Steinbuch,K.B., Cong,D., Rodriguez,A.J. and Tor,Y. (2024) Emissive Guanosine Analog Applicable for Real-Time Live Cell Imaging. *ACS Chem. Biol.*, **19**, 1836–1841.

48. Fillion,A., Pinto,J.F. and Granzhan,A. (2022) Harnessing an emissive guanine surrogate to design small-molecule fluorescent chemosensors of O6-methylguanine-DNA-methyltransferase (MGMT). *Org. Biomol. Chem.*, **20**, 1888–1892.

49. Sholokh,M., Sharma,R., Shin,D., Das,R., Zaporozhets,O.A., Tor,Y. and Mély,Y. (2015) Conquering 2-Aminopurine's Deficiencies: Highly Emissive Isomorphic Guanosine Surrogate Faithfully Monitors Guanosine Conformation and Dynamics in DNA. *J. Am. Chem. Soc.*, **137**, 3185–3188.

50. Wilhelmsson,L.M., Holmén,A., Lincoln,P., Nielsen,P.E. and Nordén,B. (2001) A Highly Fluorescent DNA Base Analogue that Forms Watson−Crick Base Pairs with Guanine. *J. Am. Chem. Soc.*, **123**, 2434–2435.

51. Sandin,P., Börjesson,K., Li,H., Mårtensson,J., Brown,T., Wilhelmsson,L.M. and Albinsson,B. (2008) Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res.*, **36**, 157–167.

52. Burns,D.D., Teppang,K.L., Lee,R.W., Lokensgard,M.E. and Purse,B.W. (2017) Fluorescence Turn-On Sensing of DNA Duplex Formation by a Tricyclic Cytidine Analogue. *J. Am. Chem. Soc.*, **139**, 1372–1375.

53. Samaan,G.N., Wyllie,M.K., Cizmic,J.M., Needham,L.-M., Nobis,D., Ngo,K., Andersen,S., Magennis,S.W., Lee,S.F. and Purse,B.W. (2021) Single-molecule fluorescence detection of a tricyclic nucleoside analogue. *Chem. Sci.*, **12**, 2623–2628.

54. Turner,M.B., Cizmic,J.M., Rosansky,D.B., Ceja,J., Patterson,M., Kilcoyne,S., Thurber,K., Kim,G., Dwyer,T.J. and Purse,B.W. (2023) Sequence-Specific Fluorescence Turn-On Sensing of RNA by DNA Probes Incorporating the Tricyclic Cytidine Analogue DEAtC. *Bioconjug. Chem.*, **34**, 1061–1071.

55. Preus,S., Börjesson,K., Kilså,K., Albinsson,B. and Wilhelmsson,L.M. (2010) Characterization of Nucleobase Analogue FRET Acceptor tCnitro. *J. Phys. Chem. B*, **114**, 1050–1056.

56. Börjesson,K., Preus,S., El-Sagheer,A.H., Brown,T., Albinsson,B. and Wilhelmsson,L.M. (2009) Nucleic Acid Base Analog FRET-Pair Facilitating Detailed Structural Measurements in Nucleic Acid Containing Systems. *J. Am. Chem. Soc.*, **131**, 4288–4293.

57. Remington,J.M., McCullagh,M. and Kohler,B. (2019) Molecular Dynamics Simulations of 2-Aminopurine-Labeled Dinucleoside Monophosphates Reveal Multiscale Stacking Kinetics. *J. Phys. Chem. B*, **123**, 2291–2304.

58. Jean,J.M. and Hall,K.B. (2004) Stacking–Unstacking Dynamics of Oligodeoxynucleotide Trimers. *Biochemistry*, **43**, 10277–10284.

59. Lee,C.-H., Ezra,F.S., Kondo,N.S., Sarma,R.H. and Danyluk,S.S. (1976) Conformational properties of dinucleoside monophosphates in solution: dipurines and dipyrimidines. *Biochemistry*, **15**, 3627–3639.

60. Ezra,F.S., Lee,C.-H., Kondo,N.S., Danyluk,S.S. and Sarma,R.H. (1977) Conformational properties of purine-pyrimidine and pyrimidine-purine dinucleoside monophosphates. *Biochemistry*, **16**, 1977–1987.

61. Frechet,D., Ehrlich,R., Remy,P. and Gabarro-Arpa,J. (1979) Thermal perturbation differential spectra of ribonucleic acids. U. Nearest neighbour interactions. *Nucleic Acids Res.*, **7**, 1981–2001.

62. Bhavsar,Y.P., Reilly,S.M. and Wadkins,R.M. (2011) Evaluation of Fluorescent Analogs of Deoxycytidine for Monitoring DNA Transitions from Duplex to Functional Structures. *J. Nucleic Acids*, **2011**, 986820.

63. Thompson,K.C. and Miyake,N. (2005) Properties of a New Fluorescent Cytosine Analogue, Pyrrolocytosine. *J. Phys. Chem. B*, **109**, 6012–6019.

64. Greiner,V.J., Kovalenko,L., Humbert,N., Richert,L., Birck,C., Ruff,M., Zaporozhets,O.A., Dhe-Paganon,S., Bronner,C. and Mély,Y. (2015) Site-Selective Monitoring of the Interaction of the SRA Domain of UHRF1 with Target DNA Sequences Labeled with 2-Aminopurine. *Biochemistry*, **54**, 6012–6020.

65. Gelot,T., Tourón-Touceda,P., Crégut,O., Léonard,J. and Haacke,S. (2012) Ultrafast site-specific fluorescence quenching of 2-aminopurine studied by femtosecond down-conversion. *J. Phys. Chem. A*, **116**, 2819–2825.

66. Nguyen,H.-N., Zhao,L., Gray,C.W., Gray,D.M. and Xia,T. (2011) Ultrafast Fluorescence Decay Profiles Reveal Differential Unstacking of 2-Aminopurine from Neighboring Bases in Single-Stranded DNA-Binding Protein Subsites. *Biochemistry*, **50**, 8989–9001.

67. Sholokh,M., Sharma,R., Shin,D., Das,R., Zaporozhets,O.A., Tor,Y. and Mély,Y. (2015) Conquering 2-Aminopurine's Deficiencies: Highly Emissive Isomorphic Guanosine Surrogate Faithfully Monitors Guanosine Conformation and Dynamics in DNA. *J. Am. Chem. Soc.*, **137**, 3185–3188.

68. Bull,G.D. and Thompson,K.C. (2018) Proton Transfer and Tautomerism in 2-Aminopurine–Thymine and Pyrrolocytosine–Guanine Base Pairs. *Biochemistry*, **57**, 4547–4561.