

# SCIENTIFIC REPORTS

OPEN

## Fluorescent RNA cytosine analogue – an internal probe for detailed structure and dynamics investigations

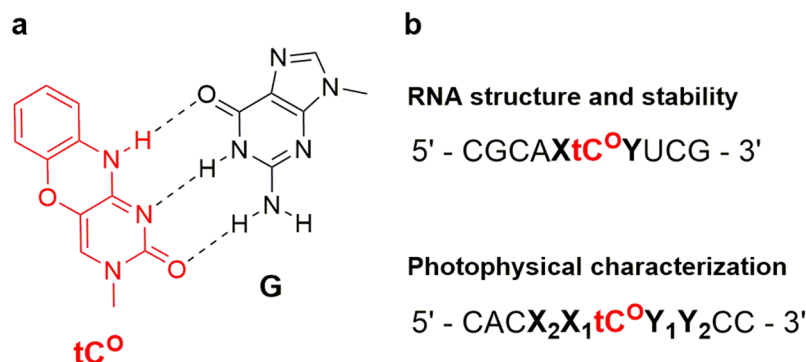
Anders Foller Füchtbauer<sup>1</sup>, Søren Preus<sup>2</sup>, Karl Börjesson<sup>3</sup>, Scott A. McPhee<sup>4</sup>, David M. J. Lilley<sup>4</sup> & L. Marcus Wilhelmsson<sup>1</sup>

The bright fluorescent cytosine analogue  $tC^O$  stands out among fluorescent bases due to its virtually unquenched fluorescence emission in duplex DNA. However, like most reported base analogues, it has not been thoroughly characterized in RNA. We here report on the first synthesis and RNA-incorporation of  $tC^O$ , and characterize its base-mimicking and fluorescence properties in RNA. As in DNA, we find a high quantum yield inside RNA duplexes ( $\langle \Phi_F \rangle = 0.22$ ) that is virtually unaffected by the neighbouring bases ( $\Phi_F = 0.20\text{--}0.25$ ), resulting in an average brightness of  $1900\text{ M}^{-1}\text{ cm}^{-1}$ . The average fluorescence lifetime in RNA duplexes is 4.3 ns and generally two lifetimes are required to fit the exponential decays. Fluorescence properties in ssRNA are defined by a small increase in average quantum yield ( $\langle \Phi_F \rangle = 0.24$ ) compared to dsRNA, with a broader distribution ( $\Phi_F = 0.17\text{--}0.34$ ) and slightly shorter average lifetimes. Using circular dichroism, we find that the  $tC^O$ -modified RNA duplexes form regular A-form helices and in UV-melting experiments the stability of the duplexes is only slightly higher than that of the corresponding natural RNA ( $\langle \Delta T_m \rangle = +2.3^\circ\text{C}$ ). These properties make  $tC^O$  a highly interesting fluorescent RNA base analogue for detailed FRET-based structural measurements, as a bright internal label in microscopy, and for fluorescence anisotropy measurements of RNA dynamics.

In recent years, we have come to understand that the roles of RNA in the cell are many and varied, and go well beyond those of the central dogma to include biocatalysis, transcription and genetic regulation. The discovery two decades ago that short RNA sequences can up- or downregulate gene expression through the RNA interference (RNAi)<sup>1</sup> pathway sparked high hopes for gene silencing with antisense oligonucleotides (ASOs)<sup>2,3</sup>, but while RNAi reagents such as siRNA have provided a wealth of information on gene function<sup>4</sup>, their utility in antisense therapy have so far been limited by stability and delivery challenges<sup>2–4</sup>.

The broad range of conformation and functions observed for RNA reflects the importance of the secondary and tertiary structure as well as dynamics, *i.e.* supertertiary structure<sup>5,6</sup>, for their function, and underline the need for tools that allow a better understanding of these parameters<sup>7,8</sup>. Traditionally, high-resolution structural insight into nucleic acids has been achieved using nuclear magnetic resonance (NMR) spectroscopy<sup>9</sup> or X-ray crystallography<sup>10</sup>, often complemented by lower-resolution techniques such as Förster resonance energy transfer (FRET)<sup>11,12</sup>. Single-molecule FRET opens up new possibilities when it comes to observing biomolecular structure and dynamics in live cells<sup>13,14</sup>. For in-cell single-molecule FRET measurements, external fluorophores such as Cy-dyes, Alexa-dyes and Atto-dyes have been employed because of their high brightness and photostability<sup>14–17</sup>. However, these probes are less useful for dynamics (anisotropy) and detailed, smaller structural features. If an internal, non-perturbing fluorescent probe is employed, real-time information on the structure and intrinsic dynamics of nucleic acids may also be obtained<sup>12,18</sup>, which could pave the way towards a deeper understanding of important cellular processes such as the genome editing CRISPR systems<sup>19–23</sup>. It may also support the

<sup>1</sup>Chemistry and Chemical Engineering/Chemistry and Biochemistry, Chalmers University of Technology, Gothenburg, SE-41296, Sweden. <sup>2</sup>Department of Chemistry, University of Copenhagen, Copenhagen, DK-2100, Denmark. <sup>3</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, Gothenburg, SE-41296, Sweden. <sup>4</sup>Cancer Research UK Nucleic Acid Structure Research Group, MSI/WTB Complex, The University of Dundee, Dow Street, Dundee, DD1 5EH, UK. Correspondence and requests for materials should be addressed to L.M.W. (email: [marcus.wilhelmsson@chalmers.se](mailto:marcus.wilhelmsson@chalmers.se))



**Figure 1.** (a) The structure of tC<sup>O</sup> (red) base-paired with guanine (G). (b) RNA sequences investigated in this study. Sites of variation in the 5'- and 3'-direction of tC<sup>O</sup> are denoted with X and Y, respectively. XY and X<sub>2</sub>X<sub>1</sub>Y<sub>1</sub>Y<sub>2</sub> are given as sequence names in Tables 1 and 2, respectively, and the full sequences can be found in Supplementary Table S1.

Sequence <sup>a</sup>	Sequence name	$T_m$ tC <sup>O</sup> [°C] <sup>b</sup>	$T_m^{UM}$ [°C] <sup>b</sup>	$\Delta T_m$ [°C]
5'-CGCAUtC <sup>O</sup> AUCG-3'	UA	56	55	+1
5'-CGCAUtC <sup>O</sup> UUCG-3'	UU	57	56	+1
5'-CGCACtC <sup>O</sup> UUCG-3'	CU	68	63	+5
5'-CGCAGtC <sup>O</sup> GUCG-3'	GG	69	67	+2

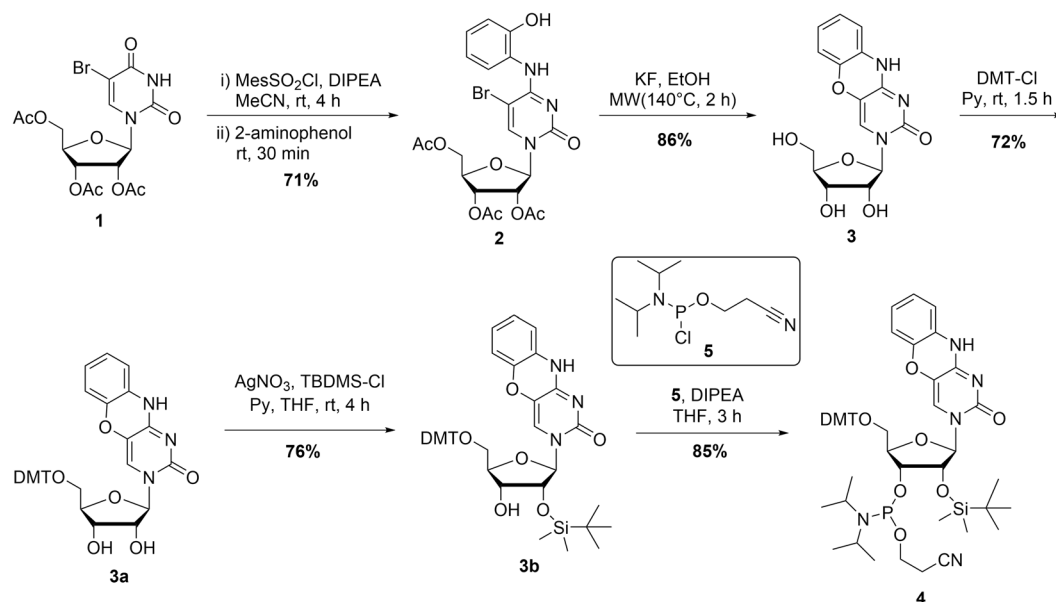
**Table 1.** Melting temperatures of tC<sup>O</sup>-containing RNA duplexes ( $T_m$ tC<sup>O</sup>), the corresponding unmodified duplexes ( $T_m^{UM}$ ) and the difference in melting temperature as a result of changing one C for tC<sup>O</sup> ( $\Delta T_m$ ). <sup>a</sup>In unmodified sequences, tC<sup>O</sup> is replaced by cytosine. <sup>b</sup>Measurements were performed in PBS buffer (100 mM Na<sup>+</sup>, pH 7.5).  $T_m$  values were determined as the average of the temperature values at the maximum of the first derivative and at half maximum of the melting curves.

Sequence name <sup>a</sup>	$\lambda_{Abs,max}$ [nm] <sup>b</sup>	$\lambda_{Em,max}$ [nm] <sup>b</sup>	$\Phi_F^{b,c}$	$\tau_1(\alpha_1)$ [ns] <sup>b,d</sup>	$\tau_2(\alpha_2)$ [ns] <sup>b,d</sup>	$\langle\tau\rangle$ [ns] <sup>e</sup>	$k_r$ [10 <sup>7</sup> s <sup>-1</sup> ] <sup>f</sup>	$k_{nr}$ [10 <sup>8</sup> s <sup>-1</sup> ] <sup>g</sup>
tC <sup>O</sup> monomer	359	457	0.30	3.4 (1.00)	—	3.4	8.8	2.1
GG	367	455	0.31	5.7 (0.71)	2.4 (0.29)	4.7	6.6	1.5
UG	364	456	0.21	4.1 (0.63)	1.6 (0.37)	3.2	6.6	2.5
GU	364	458	0.19	3.9 (0.57)	1.4 (0.43)	2.8	6.8	2.9
AA	368	458	0.34	5.8 (0.89)	2.0 (0.11)	5.4	6.3	1.2
AU	363	457	0.22	4.3 (0.64)	1.4 (0.36)	3.3	6.7	2.4
CC	368	453	0.33	5.6 (0.80)	2.2 (0.20)	4.9	6.7	1.4
GUUU	364	457	0.17	3.9 (0.50)	1.3 (0.50)	2.6	6.5	3.2
GAUU	364	457	0.22	4.4 (0.62)	1.5 (0.38)	3.3	6.7	2.4
UUUG	362	458	0.21	4.1 (0.63)	1.4 (0.37)	3.1	6.8	2.5

**Table 2.** Optical properties of the tC<sup>O</sup> monomer and tC<sup>O</sup>-containing RNA single strands. <sup>a</sup>Sequences are named after the bases flanking tC<sup>O</sup> (see Fig. 1b). Full sequences can be found as Supplementary Table S1. <sup>b</sup>Measurements were performed in PBS buffer (100 mM Na<sup>+</sup>, pH 7.5). <sup>c</sup>Fluorescence quantum yields are measured relative to the quantum yield of the potassium salt of the tC<sup>O</sup>-monomer in water ( $\Phi_F = 0.30$ )<sup>31</sup>. <sup>d</sup>The amplitudes are indicated in parenthesis. <sup>e</sup>Mean fluorescence lifetime  $\langle\tau\rangle = \sum \alpha_i \tau_i / \sum \alpha_i$ . <sup>f</sup>Radiative rate constant,  $k_r = \Phi_F / \langle\tau\rangle$ . <sup>g</sup>Non-radiative rate constant,  $k_{nr} = k_r / \Phi_F - k_r$ .

development of novel oligonucleotide-based therapeutics, such as ASOs, through visualisation of their intercellular transport and subcellular localisation<sup>2</sup>.

The number of internal probes for nucleic acids, especially fluorescent base analogues, has increased considerably over the last decades<sup>24,25</sup>, and the search for new probes with improved photophysical properties is continuing. However, most of the probes reported to date are not synthesized for or characterized in RNA systems, and the vast majority of these probes are quenched considerably upon incorporation into nucleic acids, the effect being dependent on base sequence, position and whether the nucleic acid is single- or double-stranded<sup>26</sup>. This feature has been extremely useful in studies that probe *e.g.* RNA translation and catalysis<sup>27,28</sup> and DNA dynamics<sup>29</sup>. However, for structure and dynamics studies employing FRET or fluorescence anisotropy, or cellular tracking, a bright and stable fluorescent probe would be preferred.



**Figure 2.** Synthesis of the tC<sup>°</sup> ribonucleoside phosphoramidite monomer.

We have previously reported on two internal fluorescent probes based on a tricyclic cytosine scaffold, tC and tC<sup>°</sup><sup>30,31</sup>. These probes are unique in that they retain their high fluorescence quantum yield ( $\Phi_F \sim 0.20$ ) when incorporated into duplex DNA regardless of base sequence or position, positioning them among the brightest internal DNA probes reported to date<sup>31,32</sup>. We have shown that both probes are excellent cytosine analogues that are rigidly stacked within the duplex, where they have only limited effects on the native conformation of DNA<sup>30,31</sup>. Both compounds are good substrates for human DNA and RNA polymerases<sup>33,34</sup>, that have been used to investigate DNA-protein interactions<sup>35,36</sup> and to monitor the *i*-motif transition<sup>37,38</sup>. More recently, we developed the non-emissive FRET acceptor tC<sub>nitro</sub> that can be used with tC<sup>°</sup> to accurately distinguish distance- from orientation-changes through internucleobase FRET<sup>39,40</sup> and thereby quantitatively resolve solution structures<sup>41</sup>. Even small modifications of the DNA conformation may result in significant changes in the FRET efficiency of these probes, which have been used to gain insight into mammalian mitochondrial transcription<sup>42,43</sup>, the structure and dynamics of the DNA binding RAD4/XPC complex<sup>44</sup>, and the mechanism and kinetics of the B- to Z-DNA transition<sup>45</sup>.

Herein, we report the synthesis, RNA-incorporation and characterization of the tC<sup>°</sup> ribonucleoside, and characterize its base-mimicking and fluorescence properties as a probe for RNA systems. Two series of 10-mer RNA-sequences were studied (Fig. 1): First, the effects of tC<sup>°</sup>-incorporation into four RNA duplexes were investigated by circular dichroism and UV-melting experiments to ensure that tC<sup>°</sup> does not significantly perturb the RNA structure. A second series of nine decamer sequences were used to investigate the photophysical properties of tC<sup>°</sup> in single- and double-stranded RNA (ssRNA and dsRNA) in detail. Importantly, tC<sup>°</sup> has a high fluorescence quantum yield inside RNA duplexes that is virtually unaffected by the nature of the surrounding base pairs. These unique properties should make tC<sup>°</sup> a useful internal RNA probe for detailed FRET structure measurements, in fluorescence microscopy, and for probing the dynamics of subdomains of complex RNA 3D-structures with fluorescence anisotropy.

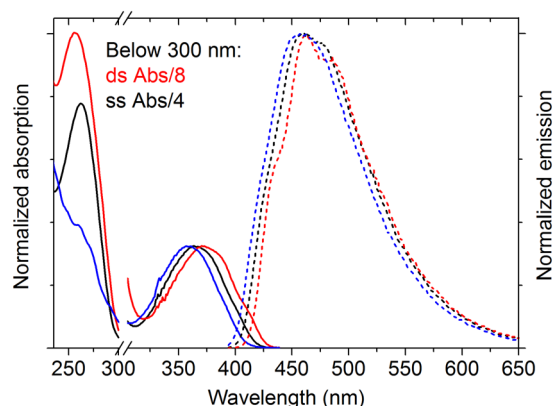
## Results and Discussion

### Synthesis of the tC<sup>°</sup> phosphoramidite monomer for incorporation into RNA oligonucleotides.

The tC<sup>°</sup>-ribonucleoside was synthesized for the first time in a two-step protocol, followed by three protection steps, to afford the tC<sup>°</sup> phosphoramidite monomer (4) (Fig. 2). This approach was influenced by the synthesis of the deoxy-tC<sup>°</sup> nucleoside by Lin and co-workers<sup>46</sup>. In short, activation of the O-4 of 2',3',5'-triacetyl-5-bromouridine (1) with 2-mesitylenesulfonyl chloride produced an active O-4-sulphonate ester, which was reacted with 2-aminophenol to produce 2. Deprotection and cyclization was achieved in one pot, using KF to facilitate the cyclization, to produce 3. Finally, the tC<sup>°</sup> ribonucleoside was protected with 5'-O-(4,4'-dimethoxytrityl) (DMT), followed by 2'-O-*tert*-butyldimethylsilyl (TBDMS) and a 3'-phosphoramidite for incorporation into RNA oligonucleotides using solid-phase synthesis (see Supplementary Table S1).

**Structure and stability of tC<sup>°</sup>-modified RNA oligonucleotides.** In a preliminary study, a small-scale batch of four modified 10-mer RNA sequences was synthesized, differing only in the bases neighbouring tC<sup>°</sup> (Fig. 1B and Table 1). To investigate the effect of tC<sup>°</sup>-incorporation on the structure and stability of RNA duplexes, circular dichroism (CD) as well as melting temperature was measured on both the modified RNA duplexes and the corresponding reference duplexes.

Circular dichroism spectra (see Supplementary Fig. S1) of all four tC<sup>°</sup>-modified sequences exhibit the general characteristics of A-form RNA, namely a positive band at 265 nm and a strong negative band at 210 nm, indicating



**Figure 3.** Isotropic absorption (solid) and emission (dashed) spectra of the  $tC^O$ -monomer (blue),  $tC^O$ -containing RNA single strand (AU, black) and  $tC^O$ -containing RNA duplex (AU, red). For better comparison of spectral envelopes, the spectra have been normalized at the maxima of the lowest energy absorption band (left) and of the emission (right), respectively.

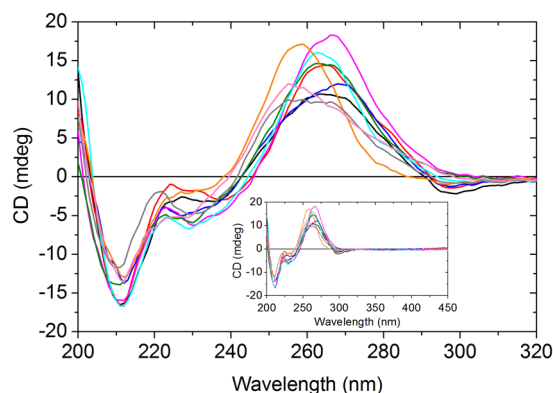
that when a cytosine is replaced with  $tC^O$  in a RNA duplex, the RNA adopts A-form geometry. In a similar manner to  $tC^O$  in DNA, no significant CD signal was found for the long wavelength absorption band of  $tC^O$ <sup>31</sup>. The thermal stability of the modified and the corresponding unmodified RNA duplexes is summarized in Table 1. On average, incorporation of  $tC^O$  increases the melting temperature by 2.3 °C. A similar increase in stability (2.7 °C) was observed for  $tC$ - and  $tC^O$ -incorporation into DNA, and presumably reflects that the increased base stacking interactions act to stabilize the duplex structure<sup>30,31</sup>. In the few reports on fluorescent base analogues in RNA, a slightly stabilizing effect was also reported for the emissive guanosine isomorph, <sup>th</sup>G<sup>47</sup>, whereas a negligible effect was reported for the uridine analogue  $U^{Dz}$ <sup>48</sup>, and a destabilizing effect was reported for benzo[b]thiophene-<sup>49</sup>, benzofuran-<sup>50</sup> and naphthalimide-conjugated<sup>51</sup> uridine analogues.

**Photophysical properties of  $tC^O$  in single-stranded RNA.** Having established that RNA duplexes retain their native A-helix conformation and are slightly stabilized after incorporating  $tC^O$ , we prepared nine new decamer sequences from a larger batch of  $tC^O$  ribonucleoside for investigation of the photophysical properties of  $tC^O$  in single- and double-stranded RNA (ssRNA and dsRNA). Three of the sequences (GUUU, GAUU and UUUU) were chosen to investigate possible effects from guanines positioned two bases away from  $tC^O$ , as a slight quenching effect of  $tC^O$  by guanines have been observed in DNA<sup>31</sup>.

Figure 3 shows representative absorption and emission spectra of  $tC^O$  in single- and double-stranded RNA together with the corresponding spectra of  $tC^O$  as a monomer. As is the case for DNA, the long-wavelength absorption maximum is slightly red-shifted upon incorporation in ssRNA (362–368 nm) compared to the  $tC^O$  monomer (359 nm)<sup>31</sup>. The emission spectrum of  $tC^O$  in ssRNA shows slightly more vibrational fine structure than that of the monomer, with an emission maximum (453–458 nm) very close to that of the monomer (457 nm). Similar spectral effects were observed for  $tC^O$  in DNA and is common for fluorescent base analogues incorporated into nucleic acids<sup>24,25</sup>.

Table 2 lists the fluorescence quantum yield,  $\Phi_F$ , mean fluorescence lifetimes,  $\langle\tau\rangle$ , and calculated radiative rate constants,  $k_r$ , and non-radiative rate constants,  $k_{nr}$  of  $tC^O$  in ssRNA. The fluorescence decays could all be fitted with two fluorescence lifetimes (see Supplementary Fig. S2). When comparing the mean fluorescence lifetimes of all sequences, the six sequences where  $tC^O$  is flanked by a uracil have a shorter lifetime (2.6–3.3 ns) than both the free monomer (3.4 ns) and the other three single-strands, GG, AA and CC (4.7–5.4 ns). These six sequences also have a reduced fluorescence quantum yield and a higher non-radiative rate constant ( $\Phi_F = 0.17$ – $0.22$ ,  $k_{nr} = 2.4$ – $3.2 \cdot 10^8 \text{ s}^{-1}$ ) compared to the free monomer ( $\Phi_F = 0.30$ ,  $k_{nr} = 2.1 \cdot 10^8 \text{ s}^{-1}$ ) as well as the other three sequences ( $\Phi_F = 0.31$ – $0.34$ ,  $k_{nr} = 1.2$ – $1.5 \cdot 10^8 \text{ s}^{-1}$ ). Interestingly, this quenching effect appears to be similar regardless of whether the uracil is positioned on the 3' side or the 5' side (*cf.* sequences UG and GU). The effect is not increased significantly in the two sequences where  $tC^O$  is positioned between two uracils (GUUU and UUUU), and the differences observed in quantum yield and mean fluorescence lifetime between them is too small for us to conclude whether the quenching is dependent on the relative orientation and interaction between  $tC^O$  and its flanking uracils. Comparing sequences AU with GAUU and GUUU with UUUU, we find that replacing a uracil with guanine two steps away from  $tC^O$  only has a small or negligible effect on the fluorescence quantum yield and average fluorescence lifetime. In ssDNA, sequences containing a guanine on the 5' side to  $tC^O$  were shown to have shorter lifetimes (2.8–3.0 ns) and quantum yield ( $\Phi_F = 0.14$ – $0.17$ ) compared to other sequences ( $\langle\tau\rangle = 4.4$ – $5.8$  ns and  $\Phi_F = 0.29$ – $0.41$ ), whereas sequences with guanine on the 3' side were among the brightest, suggesting that subtle structural differences such as relative orientation and stacking plays an important role<sup>31</sup>.

**Structural and photophysical properties of  $tC^O$  in double-stranded RNA.** Representative absorbance and emission spectra of  $tC^O$  in double-stranded RNA is included in Fig. 3. The long-wavelength absorption maximum is slightly red-shifted (368–373 nm) compared to that of single strands (362–368 nm), and there is a very weak vibrational fine structure on the long wavelength side of the absorption spectra. Although the overall



**Figure 4.** Circular dichroism spectra of nine RNA duplexes containing  $tC^O$ . Duplexes are denoted by the bases neighbouring  $tC^O$  according to Fig. 1b, and consist of the modified strands GG (black), UG (red), GU (blue), AA (orange), AU (pink), CC (gray), GUUU (purple), GAUU (green) and UUUG (cyan) hybridized to the complementary natural RNA strand. In the insert, the wavelength range is extended to 450 nm.

Sequence <sup>a</sup>	$\lambda_{Abs,max}$ [nm] <sup>b</sup>	$\lambda_{Em,max}$ [nm] <sup>b</sup>	$\Phi_F^{b,c}$	$\tau_1(\alpha_1)$ [ns] <sup>b,d</sup>	$\tau_2(\alpha_2)$ [ns] <sup>b,d</sup>	$\langle \tau \rangle$ [ns] <sup>e</sup>	$k_r$ [10 <sup>7</sup> s <sup>-1</sup> ] <sup>f</sup>	$k_{nr}$ [10 <sup>8</sup> s <sup>-1</sup> ] <sup>g</sup>
GG	373	453	0.22	4.9 (0.87)	1.9 (0.13)	4.5	4.9	1.7
UG	368	452	0.20	4.1 (0.88)	2.1 (0.12)	3.8	5.3	2.1
GU	373	460	0.22	4.8 (0.88)	2.2 (0.12)	4.5	4.9	1.7
AA	371	457	0.23	4.6 (0.89)	2.1 (0.11)	4.3	5.3	1.8
AU	370	459	0.23	4.4 (0.90)	2.0 (0.10)	4.2	5.5	1.8
CC	371	452	0.25	4.9 (0.91)	2.1 (0.09)	4.7	5.3	1.6
GUUU	370	456	0.22	4.5 (0.87)	2.2 (0.13)	4.2	5.2	1.9
GAUU	373	459	0.22	4.6 (0.88)	2.2 (0.12)	4.3	5.1	1.8
UUUG	371	456	0.23	4.3 (0.89)	1.9 (0.11)	4.1	5.6	1.9

**Table 3.** Optical properties of  $tC^O$ -containing RNA double strands. <sup>a</sup>Sequences are named after the bases flanking  $tC^O$  (see Fig. 1b). Full sequences can be found in Supplementary Table S1. <sup>b</sup>Measurements were performed in PBS buffer (100 mM Na<sup>+</sup>, pH 7.5). <sup>c</sup>Fluorescence quantum yields are measured relative to the quantum yield of the potassium salt of the  $tC^O$ -monomer in water ( $\Phi_F = 0.30$ )<sup>31</sup>. <sup>d</sup>The amplitudes are indicated in parenthesis. <sup>e</sup>Mean fluorescence lifetimes  $\langle \tau \rangle = \sum \alpha_i \tau_i / \sum \alpha_i$ . <sup>f</sup>Radiative rate constant,  $k_r = \Phi_F / \langle \tau \rangle$ . <sup>g</sup>Non-radiative rate constant,  $k_{nr} = k_r / \Phi_F - k_r$ .

emission peak is slightly red-shifted, the vibrational fine structure observed in the emission spectra results in an emission maximum (452–460 nm) that is not shifted significantly compared to the monomer (457 nm). The vibrational fine structure in both the absorbance and emission spectra of  $tC^O$  in dsRNA was also observed for  $tC^O$  in dsDNA, indicating that  $tC^O$  is firmly stacked inside the RNA helix as has previously been found in DNA<sup>31</sup>.

CD spectra (Fig. 4) on all nine  $tC^O$ -modified RNA duplexes show the general characteristics of A-form RNA, namely a positive band at 265 nm and a strong negative band at 210 nm. The long-wavelength band of  $tC^O$  is not observed in any of the CD spectra (see insert Fig. 4). For other base analogues such as the parent compound tC and 2-AP, the long wavelength absorption band can indeed be observed in CD. So far no satisfactory explanation for this has been found, but the same behaviour was observed in the CD spectra of  $tC^O$  in dsDNA<sup>31</sup>.

Table 3 lists the photophysical properties of  $tC^O$  when incorporated in double-stranded RNA. Overall, there is less variation between the fluorescence quantum yields and the mean fluorescence lifetimes of the double-stranded sequences compared to the single-stranded ones. The fluorescence quantum yield in dsRNA is slightly decreased compared to the free monomer and is relatively insensitive to the neighbouring bases ( $\Phi_F = 0.20$ – $0.25$ ). A similar trend has been observed for  $tC^O$  in dsDNA ( $\Phi_F = 0.22 \pm 0.05$ )<sup>31</sup>. Inside dsDNA, tC and  $tC^O$  are one of the few internal base analogues that have a single lifetime ( $\langle \tau \rangle = 3.4$ – $4.8$  ns for  $tC^O$ ), which is a great advantage in techniques such as fluorescence anisotropy and FRET. However, while the fluorescence decays of  $tC^O$  in dsRNA can be fitted using a single lifetime, slightly better fits are obtained when using two lifetimes: a large component of 4.1–4.9 ns and a smaller component of  $\sim 2$  ns, resulting in mean fluorescence lifetimes ranging between 3.8 and 4.7 ns (see Supplementary Fig. S3). The calculated radiative and non-radiative rate constants for the  $tC^O$ -containing double strands range between  $4.9 \cdot 10^7$  and  $5.6 \cdot 10^7$  s<sup>-1</sup>, and between  $1.7 \cdot 10^8$  and  $2.1 \cdot 10^8$  s<sup>-1</sup>, respectively. These values are comparable with the values for  $tC^O$  in dsDNA ( $k_r = 4.6$ – $5.6 \cdot 10^7$  s<sup>-1</sup>;  $k_{nr} = 1.5$ – $2.3 \cdot 10^8$  s<sup>-1</sup>). Much like in dsDNA, the radiative rate constant is reduced by  $\sim 40\%$  compared to the  $tC^O$  monomer ( $k_r = 8.8 \cdot 10^7$  s<sup>-1</sup>), most likely a consequence of the electronic interaction between  $tC^O$  and other bases in the DNA, leading to a hypochromic effect<sup>31</sup>. This would result in a reduction of the radiative rate constant,



since it is proportional to the extinction coefficient according to the Strickler-Berg relation<sup>52</sup>. The observed change in the radiative rate constant is similar to the hypochromicity of dsDNA as compared to a free nucleoside, which is typically <40%<sup>53</sup>. Altogether, our investigations of the structural and photophysical properties in double-stranded RNA suggest that  $tC^O$  is a very promising non-perturbing internal RNA fluorophore.

## Conclusion

We have, for the first time, shown that the base analogue  $tC^O$  can be incorporated into RNA, where it preserves the A-form duplex and slightly stabilizes it. The  $tC^O$  ribonucleoside can be synthesized through a simple and straight-forward synthetic protocol that affords the phosphoramidite-protected  $tC^O$  ribonucleoside in high yield. As is the case for DNA,  $tC^O$  retains its high fluorescence quantum yield inside RNA duplexes, and is virtually unaffected by the nature of the surrounding base pairs. Because of these properties, the  $tC^O$ -ribonucleoside is potentially a valuable fluorescent RNA base analogue whose average brightness ( $\epsilon \cdot \Phi_F$ ) of  $1900 \text{ M}^{-1} \text{ cm}^{-1}$  in dsRNA is the greatest for any internal RNA analogue reported to date. As we have previously demonstrated for DNA-systems, we envisage its main advantages in RNA to be in detailed FRET-based structural measurements, where there is a current lack of internal FRET-probes, as a bright label in microscopy, and in fluorescence anisotropy probing the dynamics of subdomains of complex RNA 3D-structures.

## Methods

**Materials and instruments.** Commercially available reagents were used without further purification. Deoxygenation of reaction mixtures was achieved by bubbling argon through the solution for 30 min. Column chromatography were performed using silica gel (Matrex, LC 60 Å/35–70 µm).  $^1\text{H}$  (400 MHz) and  $^{13}\text{C}$  (100.6 MHz) NMR spectra were recorded at room temperature using a Jeol Eclipse 400 NMR spectrometer. All shifts are recorded in ppm relative to the deuterated solvent ( $\text{CDCl}_3$ ,  $\text{THF}-d_8$  or  $\text{DMSO}-d_6$ ). Positive FAB high resolution mass spectra were obtained on a JEOL SX102 mass spectrometer at Instrumentstationen, Lund University, Sweden. Samples were desorbed from a 3-NBA matrix using 6 kV xenon atoms. Synthesis and characterization data of the synthesized compounds can be found in the Supplementary Information online.

**Incorporation of  $tC^O$  in RNA-oligonucleotides and their purification.** Oligoribonucleotides were synthesized using UltraMILD ribonucleotide phosphoramidites (Link Technologies) with 2'-*O*-*tert*-butyldimethylsilyl (TBDMS) protection implemented on an Applied Biosystems 394 synthesizer<sup>1</sup>. Oligoribonucleotides were cleaved from the support and base deprotected in 25% ethanol/ammonia solution at 20 °C for 3 h, and evaporated to dryness. Removal of TBDMS protecting groups was achieved by redissolving oligoribonucleotides in 115 µL dimethyl sulfoxide to which was added 125 µL 1 M triethylamine trihydrofluoride (Sigma-Aldrich) and incubated at 65 °C for 2.5 h prior to butanol precipitation. All oligonucleotides were purified by gel electrophoresis in 20% polyacrylamide under denaturing conditions (7 M urea) in 90 mM Tris-borate (pH 8.3), 10 mM EDTA (TBE buffer). The full-length RNA product was visualized by brief ultraviolet shadowing. The band was excised and electroeluted using an Elutrap (Whatman) into 45 mM Tris-borate (pH 8.5), 5 mM EDTA buffer, 8 M ammonium chloride at 200 V. The RNA was precipitated with ethanol, washed with 70% ethanol, dried and resuspended in water. Oligoribonucleotides were subjected to further purification by reversed-phase HPLC (ACE C18-AR, Advanced Chromatography Technologies), using an acetonitrile gradient with an aqueous phase of 100 mM triethylammonium acetate (pH 7.0).

**Oligonucleotide preparation.** All samples used in this study, unless stated otherwise, were prepared in a sodium phosphate buffer, pH 7.5 with 100 mM added NaCl and 1 mM EDTA. All samples were mixed and handled in sterile, RNase-free environments. The oligonucleotide concentration was determined by measuring the absorption at 260 nm. The molar absorptivities of the unmodified oligonucleotide single strands at 260 nm were calculated in IDT's online oligonucleotide analyzer<sup>54</sup>. Molar absorptivities of modified strands were calculated in the same way, with the modified base replaced by cytosine, and correcting for the molar absorptivity difference between  $tC^O$  ( $\epsilon_{tC^O} = 11000 \text{ M}^{-1} \text{ cm}^{-1}$ ) and cytosine ( $\epsilon_C = 7400 \text{ M}^{-1} \text{ cm}^{-1}$ ) at 260 nm. The total extinction coefficient of each sequence can be found in Supplementary Table S1. Double-stranded oligonucleotides were formed by mixing equimolar amounts of complementary single strands in phosphate buffer at room temperature. In fluorescence measurements, an excess of 20% of the non-fluorescent complementary strand was used to ensure that no single-stranded modified sequences were present after hybridization. Hybridization was performed by heating the samples to 85 °C followed by cooling to 5 °C at a rate of 1 °C/min.

**UV-detected thermal melting.** Melting curves were recorded on a Cary 4000 spectrophotometer (Varian Technologies) with a programmable multi-cell temperature block, by heating the samples to 85 °C at 0.5 °C/min, followed by cooling to 5 °C at a rate of 0.5 °C/min. The temperature was kept at 85 °C for 5 min between heating and cooling. For unmodified duplexes a temperature range of 15 °C to 92 °C was used. The absorption at 260 nm was recorded every 0.5 °C. Melting temperatures presented in this article are averages of the temperature values at the maximum of the first derivative and at half maximum of the melting curves, and were measured at least twice.

**Circular dichroism.** Circular dichroism (CD) spectra were recorded on a Chirascan CD spectrometer (Applied Photophysics) at 25 °C. Spectra of solutions containing 7 mM RNA duplexes, prepared as described above, were recorded between 200 and 450 nm at a scan rate of 1 nm/s. All spectra were corrected for background contributions, and smoothed (3-point adjacent-averaging).

**Steady-state fluorescence.** Quantum yields ( $\Phi_F$ ) of the different tC<sup>O</sup>-modified RNA oligonucleotides were determined relative to the quantum yield of the potassium salt of a tC<sup>O</sup>-monomer in H<sub>2</sub>O ( $\Phi_F = 0.30$ )<sup>31</sup>. Samples containing duplex RNA were prepared as described above. The single-stranded RNA oligonucleotides were set to have an absorption around 0.05 at the excitation wavelength. Spectra were recorded on a SPEX fluorolog 3 spectrofluorimeter (JY Horiba). The samples were excited at 358 nm, and emission spectra were recorded between 385 and 710 nm.

**Time-resolved fluorescence.** Fluorescence lifetimes were measured using time-correlated single-photon counting (TCSPC). The excitation pulse was generated by a PicoQuant pulsed (10 MHz) laser diode, emitting at 377 nm. The emission was monitored at 470 nm. Photons were collected by a microchannel-plate photomultiplier tube (MCP-PMT R3809U-50; Hamamatsu) and fed into a multichannel analyzer (Edinburgh Analytical Instruments) with 4096 channels. 10000 counts were recorded in the top channel. The intensity data were convoluted with the instrument response function and fitted to mono-, bi- or tri-exponential expressions using Fluofit Pro v.4 software (PicoQuant GmbH). The average lifetimes were amplitude-weighted:  $\langle \tau \rangle = \sum \alpha_i \tau_i / \sum \alpha_i$ , where  $\langle \tau \rangle$  is the average lifetime,  $\tau_i$  is the  $i^{\text{th}}$  lifetime and  $\alpha_i$  is the amplitude of the  $i^{\text{th}}$  lifetime.

## References

1. Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806–811 (1998).
2. Juliano, R. L. The delivery of therapeutic oligonucleotides. *Nucleic Acids Res.* **44**, 6518–6548 (2016).
3. Haussecker, D. & Kay, M. A. Drugging RNAi. *Science* **347**, 1069 (2015).
4. Wittrup, A. & Lieberman, J. Knocking down disease: a progress report on siRNA therapeutics. *Nat. Rev. Genet.* **16**, 543–552 (2015).
5. Tompa, P. On the supertertiary structure of proteins. *Nat. Chem. Biol.* **8**, 597–600 (2012).
6. Dethoff, E. A., Chugh, J., Mustoe, A. M. & Al-Hashimi, H. M. Functional complexity and regulation through RNA dynamics. *Nature* **482**, 322–330 (2012).
7. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).
8. Aitken, C. E., Petrov, A. & Puglisi, J. D. Single ribosome dynamics and the mechanism of translation. *Annu. Rev. Biophys.* **39**, 491–513 (2010).
9. Foster, M. P., McElroy, C. A. & Amero, C. D. Solution NMR of large molecules and assemblies. *Biochemistry* **46**, 331–340 (2007).
10. Holbrook, S. R. Structural principles from large RNAs. *Annu. Rev. Biophys.* **37**, 445–464 (2008).
11. Preus, S. & Wilhelmsson, L. M. Advances in quantitative FRET-based methods for studying nucleic acids. *ChemBioChem* **13**, 1990–2001 (2012).
12. Stephenson, J. D., Kenyon, J. C., Symmons, M. F. & Lever, A. M. L. Characterizing 3D RNA structure by single molecule FRET. *Methods* **103**, 57–67 (2016).
13. Sustarsic, M. & Kapanidis, A. N. Taking the ruler to the jungle: single-molecule FRET for understanding biomolecular structure and dynamics in live cells. *Curr. Opin. Struct. Biol.* **34**, 52–59 (2015).
14. König, I. *et al.* Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nat. Methods* **12**, 773–779 (2015).
15. Sakon, J. J. & Weninger, K. R. Detecting the conformation of individual proteins in live cells. *Nat. Methods* **7**, 203–205 (2010).
16. Fessl, T. *et al.* Towards characterization of DNA structure under physiological conditions *in vivo* at the single-molecule level using single-pair FRET. *Nucleic Acids Res.* **40**, e121 (2012).
17. Crawford, R. *et al.* Long-lived intracellular single-molecule fluorescence using electroporated molecules. *Biophys. J.* **105**, 2439–2450 (2013).
18. Sindbert, S. *et al.* Accurate distance determination of nucleic acids via Förster resonance energy transfer: Implications of dye linker length and rigidity. *J. Am. Chem. Soc.* **133**, 2463–2480 (2011).
19. Shalem, O., Sanjana, N. E. & Zhang, F. High-throughput functional genomics using CRISPR-Cas9. *Nat. Rev. Genet.* **16**, 299–311 (2015).
20. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816 (2012).
21. Li, X., Wu, R. & Ventura, A. The present and future of genome editing in cancer research. *Hum. Genet.* **135**, 1083–1092 (2016).
22. Stone, D., Niyonzima, N. & Jerome, K. R. Genome editing and the next generation of antiviral therapy. *Hum. Genet.* **135**, 1071–1082 (2016).
23. Haeussler, M. & Concordet, J.-P. Genome editing with CRISPR-Cas9: Can it get any better? *J. Genet. Genomics* **43**, 239–250 (2016).
24. Sinkeldam, R. W., Greco, N. J. & Tor, Y. Fluorescent analogs of biomolecular building blocks: Design, properties, and applications. *Chem. Rev.* **110**, 2579–2619 (2010).
25. Wilhelmsson, L. M. Fluorescent nucleic acid base analogues. *Q. Rev. Biophys.* **43**, 159–183 (2010).
26. Mata, G., Schmidt, O. P. & Luedtke, N. W. A fluorescent surrogate of thymidine in duplex DNA. *Chem. Commun.* **52**, 4718–4721 (2016).
27. McCoy, L. S., Shin, D. & Tor, Y. Isomorphic emissive GTP surrogate facilitates initiation and elongation of *in vitro* transcription reactions. *J. Am. Chem. Soc.* **136**, 15176–15184 (2014).
28. Liu, W., Shin, D., Tor, Y. & Cooperman, B. S. Monitoring translation with modified mRNAs strategically labeled with isomorphic fluorescent guanosine mimetics. *ACS Chem. Biol.* **8**, 2017–2023 (2013).
29. Sholokh, M. *et al.* 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 (2015).
30. Engman, K. C. *et al.* DNA adopts normal B-form upon incorporation of highly fluorescent DNA base analogue tC: NMR structure and UV-Vis spectroscopy characterization. *Nucleic Acids Res.* **32**, 5087–5095 (2004).
31. Sandin, P. *et al.* Characterization and use of an unprecedentedly bright and structurally non-perturbing fluorescent DNA base analogue. *Nucleic Acids Res.* **36**, 157–167 (2008).
32. Sandin, P. *et al.* Fluorescent properties of DNA base analogue tC upon incorporation into DNA – negligible influence of neighbouring bases on fluorescence quantum yield. *Nucleic Acids Res.* **33**, 5019–5025 (2005).
33. Stengel, G., Urban, M., Purse, B. W. & Kuchta, R. D. Incorporation of the fluorescent ribonucleotide analogue tCTP by T7 RNA polymerase. *Anal. Chem.* **82**, 1082–1089 (2010).
34. Stengel, G., Urban, M., Purse, B. W. & Kuchta, R. D. High density labeling of polymerase chain reaction products with the fluorescent base analogue tCo. *Anal. Chem.* **81**, 9079–9085 (2009).
35. Kuznetsov, N. A. *et al.* Conformational dynamics of DNA repair by *Escherichia coli* endonuclease III. *J. Biol. Chem.* **290**, 14338–14349 (2015).
36. Kuznetsova, A. A. *et al.* New environment-sensitive multichannel DNA fluorescent label for investigation of the protein-DNA interactions. *PLoS One* **9**, e100007 (2014).

37. Bielecka, P. & Juskowiak, B. Fluorescent sensor for pH monitoring based on an i-Motif-switching aptamer containing a tricyclic cytosine analogue (tC). *Molecules* **20**, 18511–18525 (2015).
38. Reilly, S. M. *et al.* Folding and Hydrodynamics of a DNA i-Motif from the c-MYC Promoter Determined by Fluorescent Cytidine Analogs. *Biophys. J.* **107**, 1703–1711 (2014).
39. Preus, S., Börjesson, K., Kilså, K., Albinsson, B. & Wilhelmsson, L. M. Characterization of nucleobase analogue FRET acceptor tC(nitro). *J. Phys. Chem. B* **114**, 1050–1056 (2010).
40. Börjesson, K. *et al.* Nucleic acid base analog FRET-pair facilitating detailed structural measurements in nucleic acid containing systems. *J. Am. Chem. Soc.* **131**, 4288–4293 (2009).
41. Preus, S., Kilså, K., Miannay, F. A., Albinsson, B. & Wilhelmsson, L. M. FRETmatrix: a general methodology for the simulation and analysis of FRET in nucleic acids. *Nucleic Acids Res.* **41**, e18 (2013).
42. Shi, Y. H. *et al.* Mammalian transcription factor A is a core component of the mitochondrial transcription machinery. *Proc. Natl. Acad. Sci. USA* **109**, 16510–16515 (2012).
43. Posse, V. *et al.* The amino terminal extension of mammalian mitochondrial RNA polymerase ensures promoter specific transcription initiation. *Nucleic Acids Res.* **42**, 3638–3647 (2014).
44. Velmurugu, Y., Chen, X., Slogoff Sevilla, P., Min, J.-H. & Ansari, A. Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Natl. Acad. Sci. USA* **113**, E2296–E2305 (2016).
45. Dumat, B., Larsen, A. F. & Wilhelmsson, L. M. Studying Z-DNA and B- to Z-DNA transitions using a cytosine analogue FRET-pair. *Nucleic Acids Res.* **44**, e101 (2016).
46. Lin, K.-Y., Jones, R. J. & Matteucci, M. Tricyclic 2'-deoxycytidine analogs: Syntheses and incorporation into oligodeoxynucleotides which have enhanced binding to complementary RNA. *J. Am. Chem. Soc.* **117**, 3873–3874 (1995).
47. Shin, D., Sinkeldam, R. W. & Tor, Y. Emissive RNA alphabet. *J. Am. Chem. Soc.* **133**, 14912–14915 (2011).
48. Smith, C. C., Hollenstein, M. & Leumann, C. J. The synthesis and application of a diazine-modified uridine analogue for investigating RNA-protein interactions. *RSC Adv.* **4**, 48228–48235 (2014).
49. Pawar, M. G. & Srivatsan, S. G. Synthesis, photophysical characterization, and enzymatic incorporation of a microenvironment-sensitive fluorescent uridine analog. *Org. Lett.* **13**, 1114–1117 (2011).
50. Tanpure, A. A. & Srivatsan, S. G. A microenvironment-sensitive fluorescent pyrimidine ribonucleoside analogue: Synthesis, enzymatic incorporation, and fluorescence detection of a DNA abasic site. *Chem. Eur. J.* **17**, 12820–12827 (2011).
51. Tanpure, A. A. & Srivatsan, S. G. Synthesis, Photophysical Properties and Incorporation of a Highly Emissive and Environment-Sensitive Uridine Analogue Based on the Lucifer Chromophore. *ChemBioChem* **15**, 1309–1316 (2014).
52. Strickler, S. J. & Berg, R. A. Relationship between absorption intensity and fluorescence lifetime of molecules. *J. Chem. Phys.* **37**, 814–822 (1962).
53. Bloomfield, V. A., Crothers, D. M. & Tinoco, I. *Nucleic acids: Structures, properties, and functions*. (University Science Books, 1999).
54. Integrated DNA Technologies, I. Oligo analyzer 3.1. <http://eu.idtdna.com/calc/analyzer> (accessed August 5th 2016).

## Acknowledgements

The authors thank the Swedish Research Council (VR) [2008–4990 to L.M.W.], Olle Engkvist Byggmästare Foundation (L.M.W.) and Cancer Research UK (D.M.J.L.) for financial support. Prof. Bo Albinsson is acknowledged for fruitful discussions and for giving S.P. and K.B. the opportunity to spend their time participating in this project.

## Author Contributions

K.B. and A.F.F. performed the chemical synthesis and characterisation. RNA incorporation and purification was performed by S.A.M. and D.M.J.L. The photophysical measurements were performed and partially analyzed by S.P., and mainly analyzed by A.F.F. L.M.W. supervised the project, analyzed data and contributed in the manuscript writing. A.F.F. was responsible for writing and structuring the manuscript, but all authors contributed to the discussion of the results and commented on and adjusted the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-02453-1

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017