



Cite this: *Phys. Chem. Chem. Phys.*,  
2018, 20, 3543

# The structural diversity of C-rich DNA aggregates: unusual self-assembly of beetle-like nanostructures†

A. D. Protopopova,  ‡§ V. B. Tsvetkov,  ‡ A. M. Varizhuk, N. A. Barinov,  
V. V. Podgorsky, D. V. Klinov\* and G. E. Pozmogova  \*

We studied the ability of oligonucleotides  $C_nT_{25}$  ( $n = 2, 5, 7, 9, 12, 25$ ) to form an intermolecular i-motif using circular dichroism, ultra-violet spectroscopy, nuclear magnetic resonance, high-resolution atomic force microscopy, high-performance liquid chromatography, and molecular dynamics simulations. The arrangement of single-stranded oligonucleotides in multimer i-motifs was very unusual: C-tracts of different oligonucleotides followed each other consecutively in order to fold into a closed intermolecular i-motif core with minimal loops (one cytidine in a loop spanning over a minor groove, three cytidines in a loop over a major groove); intact T-tracts protruded from predefined loci allowing visualization of beetle-like nanostructures by atomic force microscopy. The same structures were formed from analogous biotinylated oligonucleotides demonstrating one of the potential applications of such structures as carriers of multiple functional groups. Our findings open up possibilities for the rational design of pH-sensitive DNA aggregates and evaluation of the efficiency of their assembly.

Received 8th August 2017,  
Accepted 19th November 2017

DOI: 10.1039/c7cp05380k

rsc.li/pccp

## Introduction

In recent years, non-canonical DNA secondary structures have attracted much attention. DNA molecules were found to fold into a considerably diverse range of structures *via* formation of non-canonical base pairs,<sup>1,2</sup> tetrads,<sup>3–5</sup> pentads, *etc.*<sup>6–9</sup> Experiments *in vitro* have suggested that such structures can regulate important biological processes including replication and transcription<sup>10–12</sup> and that they may be responsible for a variety of genetically inherited diseases.<sup>13–15</sup> Non-canonical sites are now reckoned to be promising drug-targets and nanoconstruction units. Therefore studies of folding mechanisms of unusual DNA structures formed by various natural and model sequences are important for their systematization and analysis.

Cytidine-rich oligonucleotides can fold into one of the most exotic non-canonical DNA secondary structures called i-motif.<sup>16</sup> It is a four-stranded structure made of two fully intercalated parallel-stranded duplexes, which are in antiparallel orientation. The duplexes are formed with cytidine—protonated cytidine base pairs. The i-motifs are predominantly formed at acidic pH, and usually fail at pH 7–8. They have characteristic spectra of circular dichroism<sup>17,18</sup> (CD, a positive band at 285 nm, a smaller negative

band around 260 nm, and a crossover at approximately 270 nm) and nuclear magnetic resonance<sup>19–21</sup> (NMR, imino protons between 14.5 and 16.5 ppm).

The i-motif core has a longitudinal twist and two distinct grooves: a minor (narrow) groove and a major (wide) groove. It is generally accepted that the i-motif can be formed with one, two, three or four DNA strands;<sup>20,21</sup> accordingly three, two, one or no loops are formed. The loops have a significant influence on the i-motif stability: in general, the longer the loops, the less stable the i-motif.<sup>22,23</sup> Structures with short loops are very stable at pH 5–6. This assures the use of i-motifs in DNA-origami nanodevices, especially when the three-dimensional structure or the pH responsiveness is needed.<sup>24–30</sup> The applications include hydrogels, sensors, molecular motors driven by pH changes, switchable hydrophilic/hydrophobic surfaces, and nano-containers for controlled release (see a recent review<sup>31</sup>).

A number of articles are dedicated to the atomic force microscopy (AFM) investigation of DNA nanostructures with embedded i-motifs.<sup>26,27,32</sup> In particular, one of the AFM studies has shown that long C-strands spontaneously form unusual compact nanospheres under mild acidic conditions<sup>32</sup> in contrast to long G-strands forming G-wires.<sup>33</sup> However, to date no systematical study on the structures of individual i-motifs has been performed using AFM.

To partially fill this gap, we synthesized a series of single-stranded oligonucleotides  $C_nT_{25}$  ( $n = 2, 5, 7, 9, 12, 25$ ) and studied their ability to form an intermolecular i-motif. We used CD and UV spectroscopy, NMR spectroscopy, high-resolution AFM, high-performance liquid chromatography (HPLC), and

Biophysics Department, Federal Research and Clinical Center of Physical-Chemical Medicine, Moscow, 119435, Russia. E-mail: klinov.dmitry@mail.ru, pozmge@gmail.com; Fax: +7 (499) 246-4409; Tel: +7 (499) 246-4843

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c7cp05380k

‡ The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint first authors.

§ Currently at University of Pennsylvania, School of Medicine, Philadelphia, PA, USA.



molecular dynamics (MD) simulations to clarify the mechanism of intermolecular complex folding under mildly acidic conditions. All model oligonucleotides carried an intact 3'-end fragment ( $T_{25}$ ), which was needed to allow distinguishing different i-motif complexes on AFM images by the number of DNA strands sticking out from the i-motif core.

We report a new principle of intermolecular i-motif folding that is not only interesting from the fundamental point of view, but also could be used to construct multi-arm junctions and/or functional blocks in DNA origami.<sup>34</sup> We describe a successful generation of five-, six-, and seven-arm junctions made of  $C_{12}T_{25}$  and  $C_{25}T_{25}$  oligonucleotides with or without biotinylation of their 3'-ends.

## Experimental

### Oligonucleotide synthesis, purification and MS analysis

Oligonucleotides were synthesized on a Biosset ASM-800 DNA synthesizer following standard phosphoramidite protocols using standard reagents (Glen Research, USA). 3'-BiotinTEG CPG (Glen Research, USA) was used to synthesize  $C_nT_{25}$ -biotin ( $n = 12, 25$ ). All results obtained with biotinylated oligonucleotides were indistinguishable from their non-biotinylated counterparts. Oligonucleotide purification was performed by preparative scale reverse-phase HPLC on a 250 mm  $\times$  4.0 mm Hypersil C18 column with detection at 260 nm and a 5–17.5% gradient of  $CH_3CN$  in 0.1 M ammonium acetate buffer at 55 °C. The dimethoxytrityl protection group was removed *via* treatment with 80% acetic acid (20 min), and detritylated oligonucleotides were further purified in a 5–17.5% gradient of  $CH_3CN$  in 0.1 M ammonium acetate buffer. The final purity of all oligonucleotides was determined to be 95% by HPLC.

### UV absorption and CD spectroscopy

Oligonucleotides were dissolved in 10 mM Tris-HCl buffer (pH 5.5 or 8.5) or 10 mM NaOAc buffer (pH 5.2) containing 10 mM KCl. Their concentrations were calculated from the absorbance measured at 90 °C and the extinction coefficients, which were approximated using a nearest-neighbour model. The samples were denatured at 95 °C for 5 min and snap cooled on ice (rapid annealing) or cooled to 5 °C at a rate of  $\sim 0.5$  °C  $min^{-1}$  (slow annealing) prior to measurements. UV absorption spectra, CD spectra, and CD melting curves were recorded on a Chirascan spectrophotometer (Applied Photophysics, UK) equipped with a thermostated cuvette holder. Cuvettes with 1 cm and 0.05 cm optical path lengths were used for the samples with 1–4  $\mu M$  and 2–20  $\mu M$  oligonucleotide concentrations, respectively. The ellipticity was registered every 1 °C across the 5–90/95 °C temperature range (heating/cooling rate 0.5 °C  $min^{-1}$ ).

### NMR spectroscopy

Oligonucleotides were diluted to 50  $\mu M$  final concentration with 10 mM NaOAc buffer (pH 5.2) containing 10 mM KCl. The samples were slowly annealed; 50  $\mu l$  of  $D_2O$  was added to 600  $\mu l$  aliquot of each sample for lock signal stabilization.

$^1H$  NMR common 1D spectra with WATERGATE water suppression (relaxation delay 1.2 s, 200  $\mu s$  delay for homospoil/gradient recovery and 20  $\mu s$  delay for binomial water suppression, 64k point digital resolution and 24 ppm sweep width in 128 transitions with 16 dummy scans) at 300 K were registered for each sample. The spectra were obtained on a Bruker Avance III 500 MHz NMR spectrometer (Bruker, USA) equipped with a triple-channel TCI Prodigy cryoprobe.

### Gel filtration chromatography

High-performance liquid chromatography (HPLC) was conducted on a Zorbax GF-250 column (Agilent Technologies, USA) (250  $\times$  4.6 mm) in the isocratic elution regime (10 °C; 0.3 ml  $min^{-1}$ ; injection volume 5  $\mu l$ ). Lyophilized oligonucleotides  $C_nT_{25}$  (100  $\mu M$ ) were diluted in a buffer containing 10 mM KCl, 25 mM Tris-HCl, 5% MeOH, pH 5.5 and slowly or rapidly annealed before the experiment as described above. A calibration curve (blue line in ESI,† Fig. S1a) was obtained using synthetic oligonucleotides of random sequence (15-mer, 50-mer, 64-mer), amplicons (180–210 bp), and plasmid DNA (2300–4800 bp).

### Atomic force microscopy

All experiments were performed on the surface of freshly cleaved highly oriented pyrolytic graphite (HOPG) rendered hydrophilic using a graphite modifier (GM)  $(CH_2)_n(NCH_2CO)_m-NH_2$  obtained from Nanotuning (Nanotuning, Russia). It was previously shown that the GM-HOPG surface promotes the adsorption of triple- and double-stranded DNA,<sup>35</sup> ssDNA,<sup>36</sup> and proteins<sup>37–39</sup> and helps in obtaining unique high-resolution images in an air environment. The surface was modified according to the standard protocol.<sup>35</sup> A drop of freshly prepared 0.01 mg  $ml^{-1}$  water solution of GM was placed on the surface of freshly cleaved HOPG for 5 to 7 minutes, and the excess liquid was then removed from the surface with a flow of nitrogen, making the GM-HOPG ready for application.

After annealing and right before the deposition on the surface, the samples were diluted with freshly prepared double-distilled water to a concentration of about 0.2  $\mu M$ . Then a drop of sample solution was applied on the GM-HOPG surface for 1–2 seconds and immediately removed from the surface with the flow of nitrogen. This sample preparation method eliminates salts and also minimizes both aggregation and denaturing artefacts arising during absorption and drying.

The AFM experiments were performed using a multimode AFM with a Ntegra Prima controller (NT-Mdt, Russia) in tapping mode with a typical scan rate of 1 Hz and typical free amplitude of several nm. All measurements were performed in air using super sharp silicon cantilevers<sup>40</sup> with a tip diameter of about 1 nm.

Standard algorithms of AFM image flattening were used (subtraction of surface and averaging by lines), no algorithms of resolution improvement were used, *i.e.*, raw AFM images are presented in the article. FemtoScan Online software (ATC, Russia; <http://www.femtoscanonline.com>) was used to filter, analyze, and present the AFM data.



## Molecular dynamics simulations

3D models of all i-motifs were based on a 1YBR PDB structure.<sup>41</sup> Creation of the starting models is shown schematically in Fig. S2 (ESI<sup>†</sup>). To create a monomeric structure (Fig. S2a, ESI<sup>†</sup>), eight adenines and five pairs of cytidines were removed from the 1YBR structure, three pairs of cytidines were used. Loops consisting of two or three cytidines were added across major grooves, a loop without additional bases was introduced over a minor groove, and a T<sub>5</sub> sequence was attached at the 3'-end of the oligonucleotide. The resulting structure was optimized using Sybyl-X software (Certara, USA) and Powell's method<sup>42</sup> with the following settings: parameters for intermolecular interactions and values of partial charges were taken from the Amber7ff99 force field; non-bonded cut-off distance was set to 8 Å; dielectric constant was equal to 4; number of iterations to 1000; simplex method was used for initial optimization; and we used a 0.05 kcal mol<sup>-1</sup> Å<sup>-1</sup> energy gradient convergence criterion. Note that the starting geometry of the outmost cytidine pairs was designed in such a way that hydrogen bonds between them were absent. This was done to test the stability of these pairs.

The following strategy was used to construct intermolecular i-motifs (Fig. S2b, ESI<sup>†</sup>). At the first stage, we created an i-motif core of a desired length using multiple copies of the cytidines from the 1YBR structure and taking into account the twist angle of the boundary cytidine pairs with respect to each other. One copy of the 1YBR structure was added at a time and attached covalently. Each time the core was optimized by the algorithm described above. If the number of cytidine pairs exceeded the required quantity, the extra pairs were deleted. At the second stage, loops made of one, two, or three cytidines were added across the grooves as needed for a particular model. At the third stage, T<sub>2</sub> or T<sub>5</sub> arms were added at the desired positions of the structure following the introduction of cuts in the core made to separate the neighbouring oligonucleotides from each other. The obtained structures were optimized one more time as described above.

MD simulations were carried out using Amber 14 software.<sup>43</sup> We used a TIP3P model of water molecules<sup>44</sup> and performed the simulations in a rectangular box with periodical boundary conditions. For neutralizing negative charge, K<sup>+</sup> ions were used. Energy of the system was minimized in three steps. First, coordinates of the solvent molecules were optimized by 500 steps of steepest descent followed by 500 steps of the conjugate gradient, while mobility of all solute atoms was restrained with a force constant of 500 kcal mol<sup>-1</sup> Å<sup>-2</sup>. The second step of energy minimization was realized without any restrictions using 1000 steps of steepest descent, and 1500 steps of the conjugate gradient. Finally, gradual heating to 300 K for 20 ps was performed. At this step, weak harmonic restraints were used with a force constant of 10 kcal mol<sup>-1</sup> Å<sup>-2</sup> for all atoms except solvent.

The SHAKE<sup>45</sup> algorithm was applied to constrain bonds to hydrogen atoms, which allowed the use of a 2 fs step in the simulation. Non-bonded van der Waals and electrostatic interactions were scaled with standard amber values (SCEE = 1.2, SCNB = 2.0). Parameters for interatomic energy calculations

were taken from the ff14SB force field.<sup>46</sup> The cutoff distance for non-bonded interactions was set to 10 Å, and the long-range electrostatics was calculated using a particle mesh Ewald method.<sup>47</sup> A production phase of the MD simulations was carried out using constant temperature ( $T = 300$  K) and pressure ( $p = 1$  atm) over 30 ns. We used a Langevin thermostat with a collision frequency of 1 ps<sup>-1</sup> to control the temperature.

Snapshot visualization and hydrogen bond analysis were performed using Visual Molecular Dynamics (VMD) software<sup>48</sup> with a donor-acceptor distance of 3.2 Å and an angle cutoff of 20 degrees. Analysis of the Coulomb interaction between the bases in cytidine pairs was performed using the cpptraj module of Amber with a dielectric constant equal to 2.<sup>49</sup>

## Results and discussion

### CD, absorption, and NMR spectroscopy

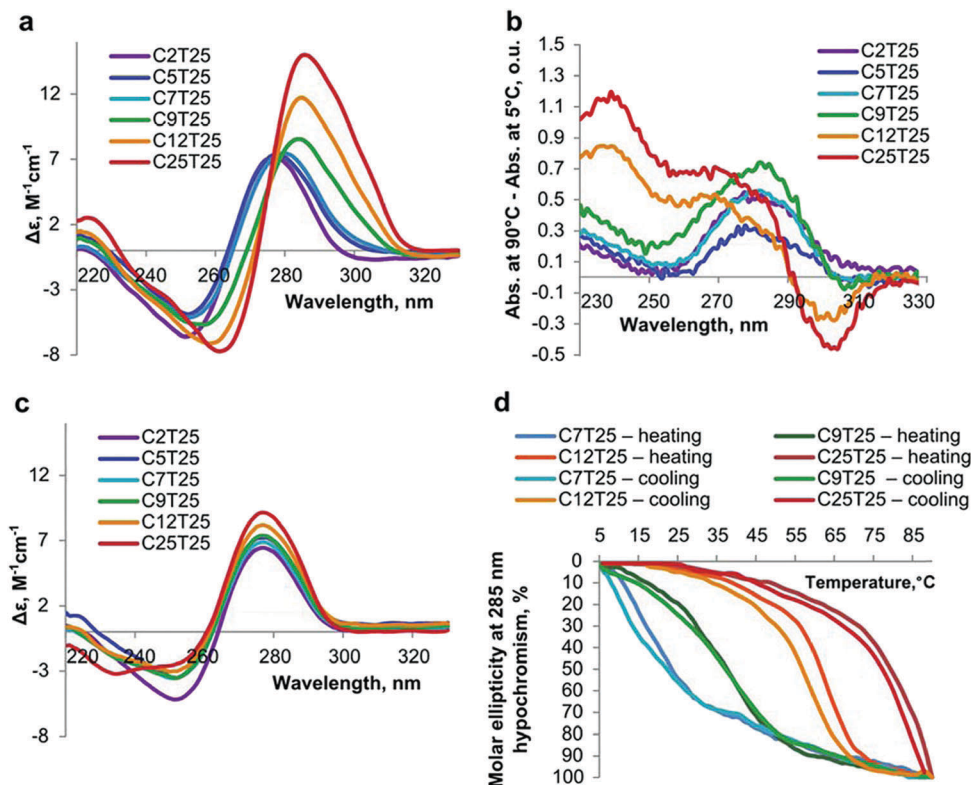
The CD and absorption spectra of slowly annealed C<sub>9</sub>T<sub>25</sub>-C<sub>25</sub>T<sub>25</sub> contained specific signatures presented by classical i-motif structures, *i.e.*, a strong positive band at 285 nm and a smaller negative band around 260 nm in the CD spectra (Fig. 1a) and a negative band at approximately 295–305 nm in the thermal difference spectra (TDS, Fig. 1b). These oligonucleotides also demonstrated a typical i-motif behaviour in response to pH (Fig. 1c) and temperature alteration (Fig. 1d). The <sup>1</sup>H NMR spectra contained i-motif signatures in the characteristic imino proton region (14.5–16.5 ppm, Fig. 2).

The CD spectrum of C<sub>7</sub>T<sub>25</sub> (Fig. 1a) had some features that might be attributed to a thermodynamically rather unstable i-motif. The positive maximum was slightly shifted from 275 nm towards 285 nm, and a hypsochromic shift of the positive CD band at pH 8.5 indicated denaturation of the structure (ESI<sup>†</sup>, Fig. S3a). In addition, minor signs of imino proton signals were present in the <sup>1</sup>H NMR spectrum of C<sub>7</sub>T<sub>25</sub> (Fig. 2); however, according to NMR data, the yield of the i-motif at room temperature was small.

According to both CD and NMR spectroscopy data, C<sub>2</sub>T<sub>25</sub> and C<sub>5</sub>T<sub>25</sub> did not form i-motifs at 20 μM or 50 μM concentration. The respective CD melting curves bore no sigma-curve resemblance, and hyper/hypochromism at 270–290 nm was linear under heating and cooling in the range 5–90 °C. The <sup>1</sup>H NMR spectra contained no signals in the imino proton region (14.5–16.5 ppm, Fig. 2).

The effects of concentration on the secondary structures of C<sub>2</sub>T<sub>25</sub>-C<sub>25</sub>T<sub>25</sub> at different pH were analyzed and summarized in ESI<sup>†</sup>, Fig. S3. Characteristic pH-induced shifts of C<sub>7</sub>T<sub>25</sub>-C<sub>25</sub>T<sub>25</sub> CD maxima at 5 °C were observed throughout the 1–50 μM concentration range (Fig. S3a, ESI<sup>†</sup>). Melting and annealing curves of C<sub>9</sub>T<sub>25</sub>-C<sub>25</sub>T<sub>25</sub> demonstrated significant concentration dependency with minor to zero hysteresis at 50 μM and substantial hysteresis at 1–4 μM (Fig. S3b, ESI<sup>†</sup>). This argued strongly for the presence of intermolecular structures. All melting/annealing curves (Fig. 1d and Fig. S3b, ESI<sup>†</sup>) had one apparent inflection, suggesting domination of a particular i-motif type in each case. However, overlapping sigmoids with





**Fig. 1** CD and UV spectra of slowly annealed  $C_nT_{25}$  oligonucleotides. (a) CD spectra obtained at 5 °C (pH 5.5). Ellipticity is given per mole of nucleotide. (b) TDS spectra (pH 5.5). (c) CD spectra obtained at 5 °C (pH 8.5). Ellipticity is given per mole of nucleotide. (d) Thermal denaturation/renaturation profiles (pH 5.5). All spectra were obtained from 20  $\mu\text{M}$  oligonucleotide solutions annealed in 10 mM Tris-HCl buffer (pH 5.5 or 8.5) supplemented with 10 mM KCl.

relatively close transition points cannot be excluded. For instance, the curves of  $C_{25}T_{25}$  could be fitted with both single and double sigmoid functions. Thus, other methods were required to clarify whether individual structures or equilibrium mixtures were present in solution and to determine the stoichiometry of i-motif formation.

According to CD and TDS spectroscopy, the annealing rate had no significant effect on the secondary structures of  $C_2T_{25}$ – $C_{25}T_{25}$  at pH 5.5. The CD spectra obtained after rapid annealing were similar to the CD spectra obtained after slow annealing and are shown in Fig. S4 (ESI†). The TDS spectra obtained after rapid annealing are not shown, but they were identical to those obtained from the slowly annealed samples.

Altogether, the spectral data confirmed that the i-motif structures were formed in  $C_7T_{25}$ – $C_{25}T_{25}$  (although the  $C_7T_{25}$  i-motif was unstable), and that  $C_9T_{25}$ – $C_{25}T_{25}$  were prone to intermolecular folding.

### Atomic force microscopy

Representative images of the intra- and intermolecular structures formed from  $C_nT_{25}$  ( $n = 2, 3, 5, 7, 9, 12, 25$ ) oligonucleotides at pH 5.5 are shown in Fig. 3. We identified assembled structures by their characteristic morphology: an elevated globular or cylindrical core with one or several thin extensions ( $T_{25}$  arms) sticking out from it. A number of extensions corresponded to the degree of multimerization. The average height of the i-motif core was about 1.5–2 nm; a detailed analysis of i-motif heights

is presented in the ESI,† Fig. S5. The AFM study confirmed that these nanostructures disassembled under pH 8.5 (ESI,† Fig. S6).

$C_2T_{25}$  and  $C_3T_{25}$  did not form any intermolecular i-motif structure, only extended single strands and coiled globules of DNA with an average height of  $0.5 \pm 0.1$  nm ( $N = 1666$ ) were observed on these samples (Fig. 3a).  $C_5T_{25}$  oligonucleotides were also predominantly single-stranded under microscopy conditions. Nevertheless, we did find some tetramers (M-4, a few of them are marked with arrows in Fig. 3b), but the number of such structures was vanishingly small.

On the AFM images of  $C_7T_{25}$  and  $C_9T_{25}$ , we observed two major types of structures: tetramers M-4 and dimers M-2 (Fig. 3c and d). The percentage of these structures depended on the annealing conditions and was difficult to determine using AFM. However, we noticed that M-4 dominated in both samples at a concentration of 100  $\mu\text{M}$ , while the fractions of M-4 and M-2 were roughly equal at a concentration of 35  $\mu\text{M}$ . In addition to the major structures, M-3 trimers and unfolded ssDNAs were found on the sample surfaces at both concentrations.

Further lengthening of a C-tract led to an increase in the polymorphism of the structures in  $C_{12}T_{25}$  and  $C_{25}T_{25}$  and to the formation of unusual multimers: M-5 pentamers, M-6 hexamers, and M-7 heptamers (Fig. 3e and f). A gallery of frequently observed structures is shown in Fig. 4. Additionally, one example of a longer M-9 associate was found in  $C_{25}T_{25}$  (ESI,† Fig. S7). Although we could not resolve the fine structural details of the i-motifs using AFM, the spatial location of  $T_{25}$  arms was





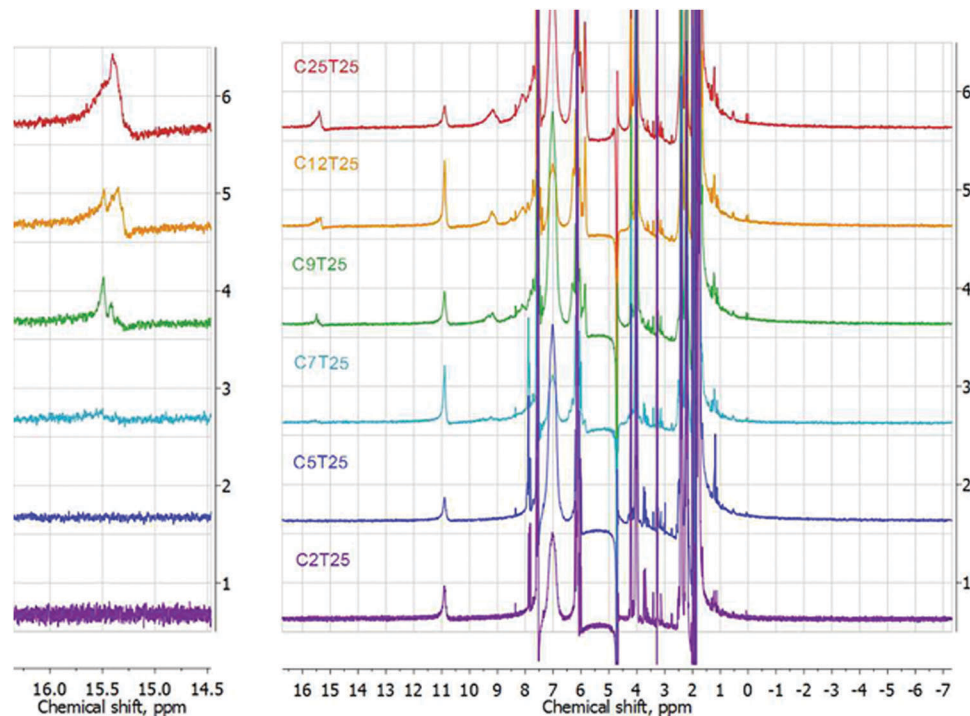


Fig. 2  $^1\text{H}$  NMR spectra of  $C_n\text{T}_{25}$  oligonucleotides at pH 5.2. Signals in the characteristic imino proton region (14.5–16.5 ppm) are magnified in the left panel. All spectra were obtained at 300 K in 10 mM NaOAc (pH 5.2) and 10 mM KCl.

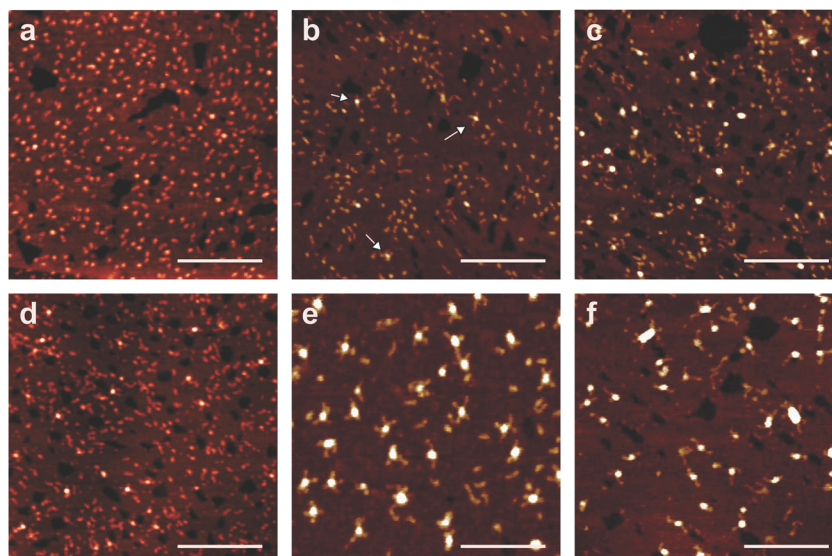


Fig. 3 Representative AFM images of i-motif based nanostructures formed by  $C_n\text{T}_{25}$  oligonucleotides at pH 5.5. (a)  $C_2\text{T}_{25}$ ; (b)  $C_5\text{T}_{25}$ ; (c)  $C_7\text{T}_{25}$ ; (d)  $C_9\text{T}_{25}$ ; (e)  $C_{12}\text{T}_{25}$ ; (f)  $C_{25}\text{T}_{25}$ . The length of the scale bars is 100 nm.

characteristic for each type of multimer (see generalized coloured folding schemes in Fig. 4). The arrangement of the arms in M-2 and M-4 was consistent with a generally accepted view on the i-motif formation.<sup>31</sup> Namely, M-2 consisted of two hairpins; M-4 consisted of four parallel strands (Fig. 5).

Interestingly, it was previously shown that guanidine-rich oligonucleotides  $A_{15}G_{15}$  self-assemble into indefinitely lengthening frayed wires<sup>30</sup> composed of a G-quadruplex core with

single-stranded adenosine tracts emanating from the core. Importantly, our intermolecular i-motifs do not grow into wires but form closed nanostructures with a fixed arrangement of strands.

We supposed that the C tracts of several molecules dock head to tail to form M-3, M-5, M-6, and M-7, while the  $T_{25}$  arms extend from the middle of the i-motif core (Fig. 5). This way a number of cytidines that are not included into the i-motif core



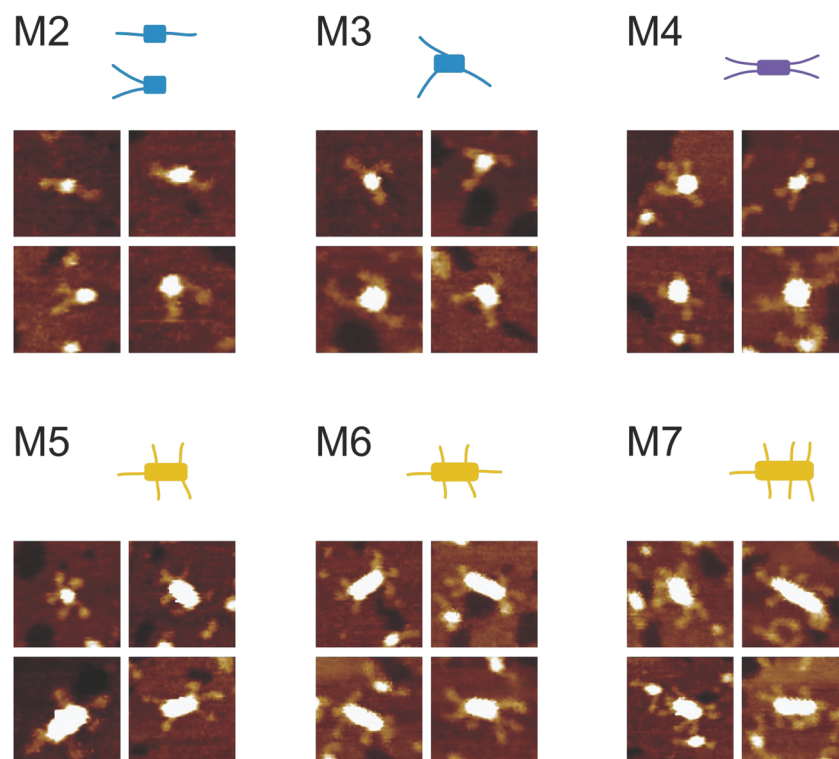


Fig. 4 Generalized structures of frequently observed multimer i-motifs and corresponding AFM images. Colours correspond to the colour scheme used for the discrimination of multimer structures on the other images of the article.

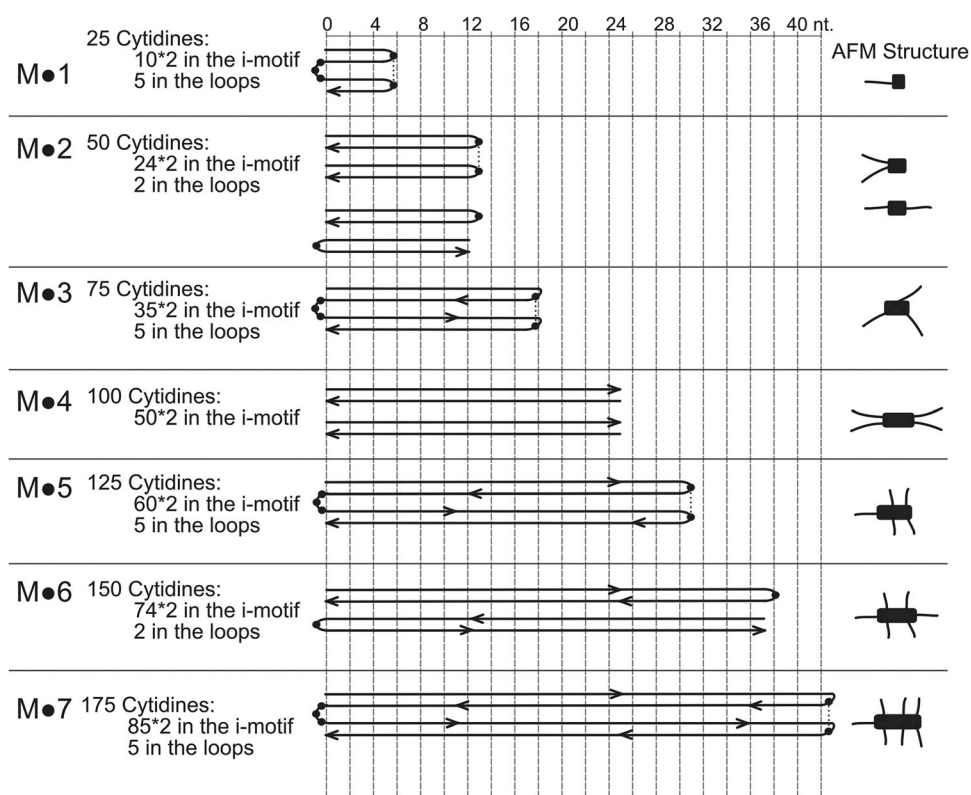


Fig. 5 Intermolecular i-motifs folding schemes for  $C_{25}T_{25}$  and their predicted AFM structures. Cytidines included in the loops are shown with filled circles. The  $T_{25}$  3'-ends of oligonucleotides are shown with arrows. Potential hydrogen bonds between the oligonucleotides in the minor loops are shown with black dotted lines.



are minimized. To prove this theoretical folding scheme, clarify the structure of the loops, and study stability and an overall structure of the intermolecular i-motif core, we performed MD simulations as described below.

### MD simulations of i-motifs

In an intermolecular i-motif, one or two loops span over a minor groove of the core, and two or one, respectively, cross a major groove; the loops over the major groove should be longer. First, we determined a number of bases that can form a loop over a major groove. We built two monomers with three pairs of cytidines in the i-motif core and two or three cytidines in the major loops (Fig. S2a and Fig. 6a–f). The structure with three cytidines was stable over the whole 30 ns simulation as can be seen from the graphs in Fig. 6e, f and Fig. S8a, b (ESI†). The structure with two cytidines in the loops was unstable: hydrogen bonds in the outermost cytidine pair broke in the very beginning of the simulation and were never restored (Fig. 6a–c and Fig. S8c, d). Thus, we conclude that loops spanning the major groove of the i-motif core likely contain three cytidines.

Second, we determined a number of bases that can form a loop over a minor groove. We built a model of a dimer with eight pairs of cytidines in the i-motif core and only one cytidine forming each of the loops (Fig. 6g–i). This structure was not only stable over the whole 30 ns simulation, but cytidines from the loops quickly formed a ninth pair of the core (Fig. 6i, green line). However, judging by the number of hydrogen bonds, the energy of Coulomb interaction, angle and distance between the bases this pair was less stable than the others (Fig. 6h, i and Fig. S8e, f, ESI†).

To verify the prediction of the model about the formation of one-letter loops, we synthesized a series of oligonucleotides in which 3' and 5' terminal parts were complementary and formed a hairpin, while the central part could participate in the formation of an i-motif dimer with two loops over the minor grooves. The sequence CTGACATCTG-TC<sub>13</sub>T-CAGATGTCAG mimicked the oligonucleotides from C<sub>n</sub>T<sub>25</sub> series; in CTGACATCTG-TC<sub>6</sub>TC<sub>6</sub>T-CAGATGTCAG the central cytidine, potentially forming a one-letter loop, was replaced with thymidine; in CTGACATCTG-TC<sub>5</sub>T<sub>3</sub>C<sub>5</sub>T-CAGATGTCAG three cytidines were replaced with

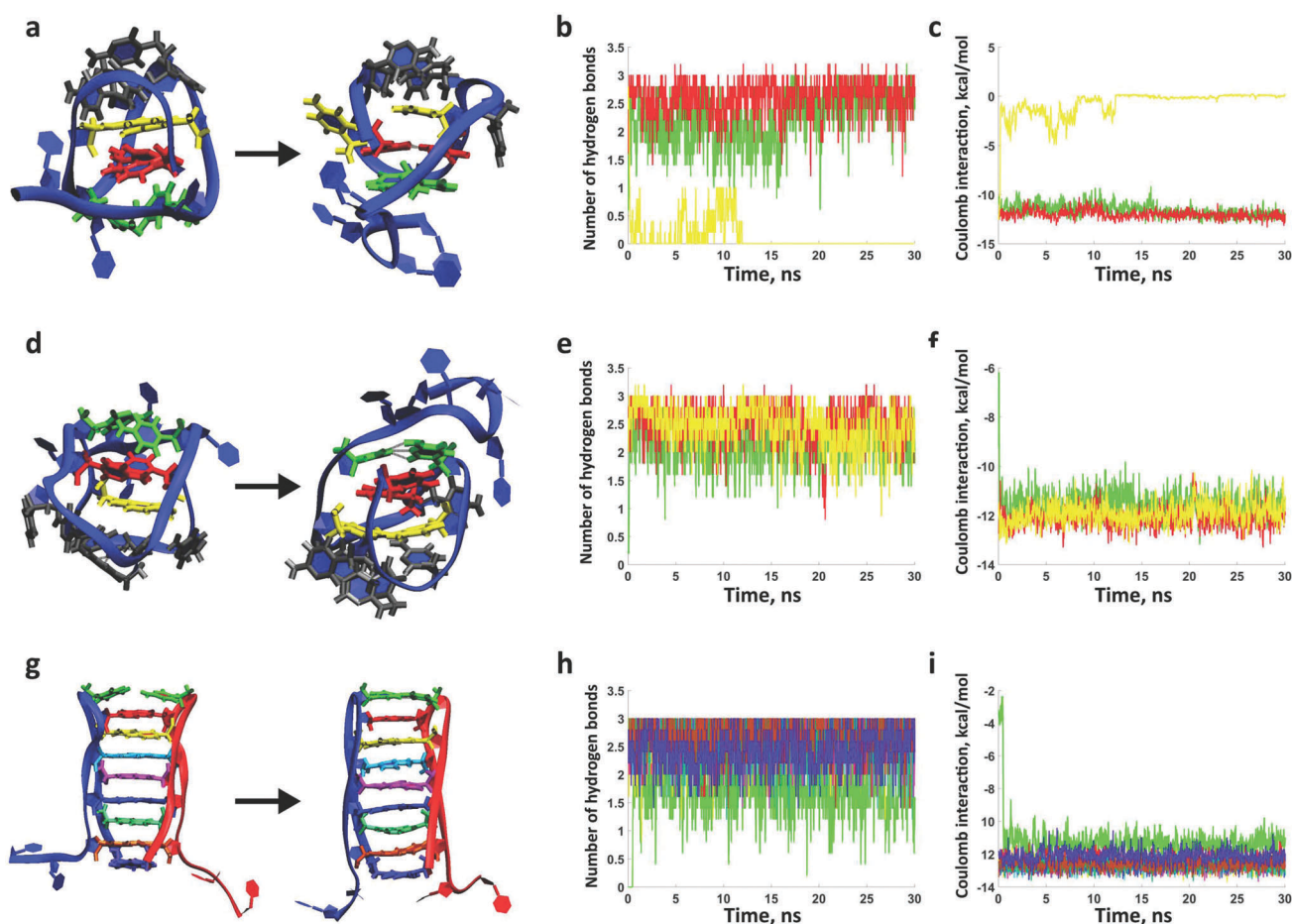


Fig. 6 MD simulations of intermolecular i-motifs clarifying the structure of the loops: (a–c) C<sub>10</sub>T<sub>5</sub> monomer; (d–f) C<sub>12</sub>T<sub>5</sub> monomer; (g–i) C<sub>9</sub>T<sub>2</sub> dimer. All simulations were performed for 30 ns and are presented with start and end conformations, dynamics of a number of hydrogen bonds between the bases forming each of the cytidine pairs, and dynamics of the energy of electrostatic interaction between the bases in the pairs. In the schemes, cytidines that do not form hydrogen bonds are shown in black, cytidines connected with hydrogen bonds are colored in pairs, and colours of the pairs correspond to the graphs of parameters.





thymidines, potentially disturbing two cytidine pairs in the core; and the CTGACATCTG-T<sub>15</sub>-CAGATGTCAG sequence was used as control. The CD spectra of the first two oligonucleotides were almost identical containing specific signatures presented by classical i-motif structures (Fig. S9a, ESI†). The oligonucleotide with three thymidines in the central region also had a characteristic i-motif CD spectrum, but with a smaller amplitude (Fig. S9a, ESI†) indicating that the i-motif core of this oligonucleotide was shorter. All three oligonucleotides demonstrated a combined melting of a hairpin and an i-motif in response to temperature alteration (Fig. S9b and c, ESI†). These results confirm the formation of the loop from one oligonucleotide, and agree with a recent study of stability of C<sub>n</sub> (n = 15, 19, 23, 27) molecules.<sup>51</sup>

Finally, we built models of a trimer (Fig. S2b, ESI†) and a tetramer based on the folding scheme suggested by the AFM results and the length of the loops suggested by the MD simulations. We confirmed the stability of the model structures with MD. The cores of both structures were stable over the 30 ns simulations (Fig. 7). In particular, the base pairs situated at the junctions between the strands were stable even if the junction was at the last pair of the core (Fig. 7b). Hydrogen bonds between cytidines in the loops spanning minor grooves were fluctuating over time and with the increase in the length of T-tracts became less stable (ESI,† Fig. S10). Currently, we cannot prove that weak hydrogen bonds between the minor loops are indeed formed in C<sub>n</sub>T<sub>25</sub> oligonucleotide series.

### Putative folding schemes of intermolecular i-motifs made of C<sub>n</sub>T<sub>25</sub> oligonucleotides

Based on the AFM data and the results of the MD simulations we built simplified principal schemes of the prevalent i-motif structures formed from C<sub>5</sub>T<sub>25</sub>–C<sub>25</sub>T<sub>25</sub> oligonucleotides (Fig. 5 and ESI,† Fig. S11–S14). First, we calculated a total number of cytidines in the structure (for example, for a M-7 multimer made of C<sub>25</sub>T<sub>25</sub> this is 7 × 25 = 175 nucleotides). Second, we identified a number of cytidines required to form the loops (5 in the case of the M-7). The remaining cytidines (175 – 5 = 170) constitute a core of 85 × 2 nucleotides. Using the described principle, we built schemes of all possible multimers (Fig. 5 and ESI,† Fig. S11–S14).

From the AFM experiments, C<sub>2</sub>T<sub>25</sub> and C<sub>3</sub>T<sub>25</sub> did not fold into stable M-4; therefore we supposed that in our system two and three tetrads (up to six cytidine pairs) were insufficient to stabilize the i-motif. Hence, C<sub>12</sub>T<sub>25</sub>, C<sub>9</sub>T<sub>25</sub>, and C<sub>7</sub>T<sub>25</sub> should not

fold into stable M-1, and C<sub>5</sub>T<sub>25</sub> should not fold into M-1, M-2, and M-3. On the other hand, formation of long loops from short oligonucleotides is problematic; therefore C<sub>5</sub>T<sub>25</sub> should not form M-5.

### HPLC analysis of i-motif structures

The proposed folding schemes were verified by gel filtration HPLC analysis. Gel chromatography allowed us to characterize quantitatively the impact of C-block length on multimer fractions. C<sub>2</sub>T<sub>25</sub> and C<sub>3</sub>T<sub>25</sub> were eluted from the column as monomers; AFM, CD, and NMR data suggest that these monomers were unfolded. C<sub>5</sub>T<sub>25</sub> was the shortest oligonucleotide that demonstrated multimer formation by HPLC. As expected, it did not fold into M-2, M-3, and M-5, but was able to form M-4 (Fig. 8a). This was in agreement with the folding schemes (Fig. S14, ESI†), but in a seeming contradiction with CD and NMR data. We attribute that the lack of correspondence to the difference in experimental conditions: we performed HPLC in the presence of MeOH to avoid nonspecific sorption (all attempts to select optimal conditions for separations without MeOH were unsuccessful). A number of co-solvents, including methanol, have been reported previously to induce noncanonical nucleic acid folding<sup>52</sup> – presumably, due to the crowding effect.<sup>53,54</sup>

In C<sub>7</sub>T<sub>25</sub>–C<sub>12</sub>T<sub>25</sub> the multimer fraction gradually increased in proportion to the thermodynamic profitability of the intermolecular i-motif formation. This was accompanied by a decrease in the

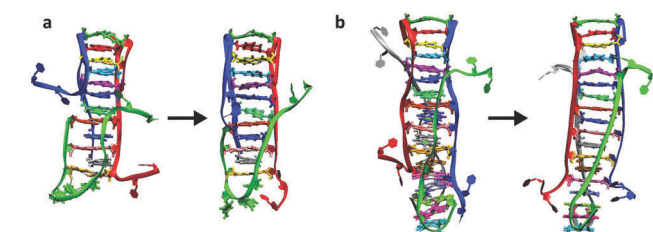


Fig. 7 MD simulations of intermolecular i-motifs: start and end conformations of (a) C<sub>9</sub>T<sub>2</sub> trimer; (b) C<sub>9</sub>T<sub>2</sub> tetramer. Simulation time 30 ns.

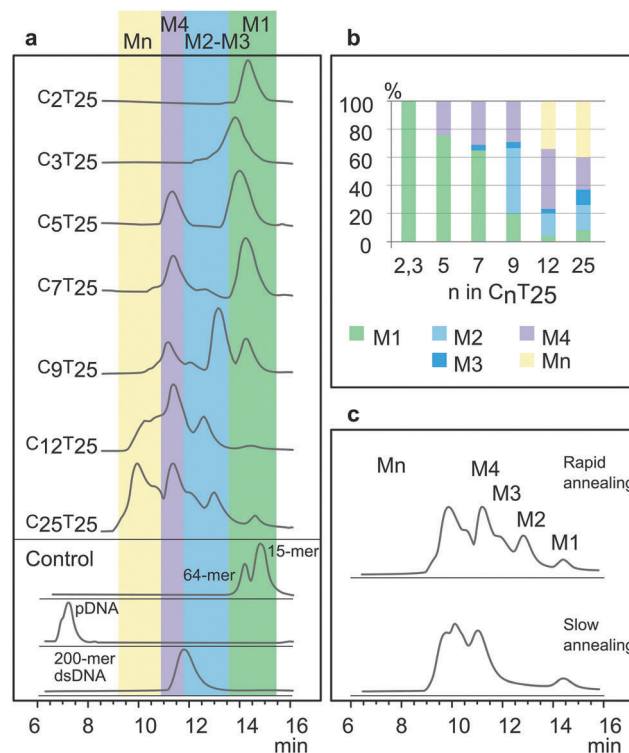


Fig. 8 HPLC Analysis of the C<sub>n</sub>T<sub>25</sub> multimers. (a) HPLC chromatograms obtained from rapidly annealed samples of C<sub>n</sub>T<sub>25</sub> oligonucleotides and control oligonucleotides. (b) Percentage of oligonucleotides involved in the formation of monomeric and multimer structures formed after rapid annealing of C<sub>n</sub>T<sub>25</sub>. (c) An influence of annealing rate on the multimer composition of the solution of C<sub>25</sub>T<sub>25</sub> oligonucleotides.





monomer fraction: in  $C_{12}T_{25}$  monomers comprised only 3% of the total oligonucleotide content (Fig. 8a and b). However, the HPLC chromatogram of  $C_{25}T_{25}$  showed a rescued M-1 peak, which is probably dictated by the accumulation of assembled i-motif monomers (Fig. 8a and b).

A noticeable relation to the annealing rate was observed in the HPLC data for  $C_{12}T_{25}$  and  $C_{25}T_{25}$  only. Slow annealing leads to the redistribution of M-2 and M-3 structures to M-4 and multimers (Fig. 8c).

## Conclusions

In summary, we performed a systematic study of i-motifs formed from  $C_nT_{25}$  oligonucleotides ( $n = 2, 5, 7, 9, 12, 25$ ). We used a modified graphite substrate and super-sharp cantilevers to improve AFM image resolution. We also used an original approach to study the i-motif structure by microscopy: the intact T-tracts allowed us to visualize beetle-like nanostructures with distinct number and unique distribution of  $T_{25}$  arms around the i-motif core. We developed an original technique for a step-by-step assembly of a custom i-motif structure for MD simulation and analysis. The results of AFM, molecular modelling, HPLC, optical experiments, and NMR showed the influence of the C-track length on the structures and relative contributions of various associates.  $C_nT_{25}$  oligomers formed mainly dimers at  $n = 9$ ; tetramers at  $n = 5-7$ , and higher-order multimers at  $n \geq 12$ . Based on the AFM imaging and the results of molecular modelling, we suggested a fundamental principle of intermolecular i-motif folding. We supposed that long C-tracts of different oligonucleotides followed each other consecutively in order to fold into the closed intermolecular i-motif core with minimal loops. MD simulations have shown that such structures are stable and that loops spanning over minor grooves contain one cytidine, and loops over major grooves contain three cytidines. Similar structures were formed from biotinylated oligonucleotides  $C_nT_{25}$ -biotin ( $n = 12, 25$ ) demonstrating one of the potential applications of such structures as carriers of multiple functional groups.

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

This work was supported by the Russian Science Foundation [14-25-00013 to G. E. P.].

## Notes and references

- H. S. Alvey, F. L. Gottardo, E. N. Nikolova and H. M. Al-Hashimi, Widespread transient Hoogsteen base pairs in canonical duplex DNA with variable energetics, *Nat. Commun.*, 2014, **5**, 4786.
- A. K. Shchlyolkina, O. F. Borisova, M. A. Livshits, G. E. Pozmogova, B. K. Chernov, R. Klement and T. M. Jovin, Parallel-stranded DNA with mixed AT/GC composition: role of trans G-C base pairs in sequence dependent helical stability, *Biochemistry*, 2000, **39**, 10034-10044.
- B. Zamiri, M. Mirceta, K. Bomsztyk, R. B. Macgregor and C. E. Pearson, Quadruplex formation by both G-rich and C-rich DNA strands of the C9orf72 (GGGGCC)<sub>8</sub>-(GGCCCC)<sub>8</sub> repeat: effect of CpG methylation, *Nucleic Acids Res.*, 2015, **43**, 10055-10064.
- J. Viladoms, N. Escaja, M. Frieden, I. Gómez-Pinto, E. Pedroso and C. González, Self-association of short DNA loops through minor groove C:G:G:C tetrads, *Nucleic Acids Res.*, 2009, **37**, 3264-3275.
- N. Escaja, J. Viladoms, M. Garavís, A. Villasante, E. Pedroso and C. González, A minimal i-motif stabilized by minor groove G:T:G:T tetrads, *Nucleic Acids Res.*, 2012, **40**, 11737-11747.
- A. Kettani, A. Gorin, A. Majumdar, T. Hermann, E. Skripkin, H. Zhao, R. Jones and D. J. Patel, A dimeric DNA interface stabilized by stacked A(GGGG)A hexads and coordinated monovalent cations, *J. Mol. Biol.*, 2000, **297**, 627-644.
- A. Matsugami, K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi and M. Katahira, An intramolecular quadruplex of (GGA)<sub>4</sub> triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction, *J. Mol. Biol.*, 2001, **313**, 255-269.
- A. T. Phan, V. Kuryavii, J.-B. Ma, A. Faure, M.-L. Andréola and D. J. Patel, An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 634-639.
- A. Varizhuk, D. Ischenko, V. Tsvetkov, R. Novikov, N. Kulemin, D. Kaluzhny, M. Vlasenok, V. Naumov, I. Smirnov and G. Pozmogova, The expanding repertoire of G4 DNA structures, *Biochimie*, 2017, **135**, 54-62.
- N. Saini, Y. Zhang, K. Usdin and K. S. Lobachev, When secondary comes first-The importance of non-canonical DNA structures, *Biochimie*, 2013, **95**, 117-123.
- H. J. Kang, T. V. T. Le, K. Kim, J. Hur, K. K. Kim and H. J. Park, Novel interaction of the Z-DNA binding domain of human ADAR1 with the oncogenic c-Myc promoter G-quadruplex, *J. Mol. Biol.*, 2014, **426**, 2594-2604.
- Y. Du and X. Zhou, Targeting non-B-form DNA in living cells, *Chem. Rec.*, 2013, **13**, 371-384.
- A. R. Haeusler, C. J. Donnelly, G. Periz, E. A. J. Simko, P. G. Shaw, M.-S. Kim, N. J. Maragakis, J. C. Troncoso, A. Pandey, R. Sattler, J. D. Rothstein and J. Wang, C9orf72 nucleotide repeat structures initiate molecular cascades of disease, *Nature*, 2014, **507**, 195-200.
- Y. Wu and R. M. Brosh, G-Quadruplex nucleic acids and human disease, *FEBS J.*, 2010, **277**, 3470-3488.
- R. Simone, P. Fratta, S. Neidle, G. N. Parkinson and A. M. Isaacs, G-Quadruplexes: emerging roles in neurodegenerative diseases and the non-coding transcriptome, *FEBS Lett.*, 2015, **589**, 1653-1668.
- K. Gehring, J.-L. Leroy and M. Gueron, A tetrameric DNA structure with protonated cytosine-cytosine base pairs, *Nature*, 1993, **363**, 561-565.
- J. M. Dettler, R. Buscaglia, J. Cui, D. Cashman, M. Blynn and E. a. Lewis, Biophysical characterization of an ensemble of



- intramolecular i-motifs formed by the human c-MYC NHE III P1 promoter mutant sequence, *Biophys. J.*, 2010, **99**, 561–567.
- 18 J. L. Mergny, J. Li, L. Lacroix, S. Amrane and J. B. Chaires, Thermal difference spectra: a specific signature for nucleic acid structures, *Nucleic Acids Res.*, 2005, **33**, e138.
- 19 X. Han, J. L. Leroy and M. Guéron, An intramolecular i-motif: the solution structure and base-pair opening kinetics of d(5mCCT3CCT3ACCT3CC), *J. Mol. Biol.*, 1998, **278**, 949–965.
- 20 J. Dai, A. Ambrus, L. H. Hurley and D. Yang, A direct and nondestructive approach to determine the folding structure of the i-motif DNA secondary structure by NMR, *J. Am. Chem. Soc.*, 2009, **131**, 6102–6104.
- 21 A. Laisné, D. Pompon and J. L. Leroy, [C7GC4]<sub>4</sub> Association into supra molecular i-motif structures, *Nucleic Acids Res.*, 2010, **38**, 3817–3826.
- 22 S. M. Reilly, R. K. Morgan, T. A. Brooks and R. M. Wadkins, Effect of interior loop length on the thermal stability and pK<sub>a</sub> of i-motif DNA, *Biochemistry*, 2015, **54**, 1364–1370.
- 23 S. P. Gurung, C. Schwarz, J. P. Hall, C. J. Cardin and J. Brazier, The importance of loop length on the stability of i-motif structures, *Chem. Commun.*, 2015, **51**, 5630–5632.
- 24 Y. Yang, C. Zhou, T. Zhang, E. Cheng, Z. Yang and D. Liu, DNA pillars constructed from an i-motif stem and duplex branches, *Small*, 2012, **8**, 552–556.
- 25 H. Mei, S. Budow and F. Seela, Construction and assembly of chimeric DNA: oligonucleotide hybrid molecules composed of parallel or antiparallel duplexes and tetrameric i-motifs, *Biomacromolecules*, 2012, **13**, 4196–4204.
- 26 T. Li and M. Famulok, I-Motif-programmed functionalization of DNA nanocircles, *J. Am. Chem. Soc.*, 2013, **135**, 1593–1599.
- 27 K. Akinori, R. Watanabe, Y. Yamanaka, T. Tamaki, M. Kaino and Y. Ohya, Nanomechanical DNA origami pH sensors, *Sensors*, 2014, **14**, 19329–19335.
- 28 S. Son, J. Nam, J. Kim, S. Kim and W. J. Kim, I-Motif-driven Au nanomachines in programmed siRNA delivery for gene-silencing and photothermal ablation, *ACS Nano*, 2014, **8**, 5574–5584.
- 29 S. Modi, M. G. Swetha, D. Goswami, G. D. Gupta, S. Mayor and Y. Krishnan, A DNA nanomachine that maps spatial and temporal pH changes inside living cells, *Nat. Nanotechnol.*, 2009, **4**, 325–330.
- 30 S. Modi, C. Nizak, S. Surana, S. Halder and Y. Krishnan, Two DNA nanomachines map pH changes along intersecting endocytic pathways inside the same cell, *Nat. Nanotechnol.*, 2013, **8**, 459–467.
- 31 Y. Dong, Z. Yang and D. Liu, DNA nanotechnology based on i-motif structures, *Acc. Chem. Res.*, 2014, **47**, 1853–1860.
- 32 D. Zillich, K. Liu, L. Sagiv, D. Porath and A. Kotlyar, I-Motif nanospheres: unusual self-assembly of long cytosine strands, *Small*, 2011, **7**, 1029–1034.
- 33 A. B. Kotlyar, N. Borovok, T. Molotsky, H. Cohen, E. Shapir and D. Porath, Long, monomolecular guanine-based nanowires, *Adv. Mater.*, 2005, **17**, 1901–1905.
- 34 N. C. Seeman, Nucleic acid junctions and lattices, *J. Theor. Biol.*, 1982, **99**, 237–247.
- 35 D. Klinov, B. Dwir, E. Kapon, N. Borovok, T. Molotsky and A. Kotlyar, High-resolution atomic force microscopy of duplex and triplex DNA molecules, *Nanotechnology*, 2007, **18**, 225102.
- 36 J. Adamcik, D. V. Klinov, G. Witz, S. K. Sekatskii and G. Dietler, Observation of single-stranded DNA on mica and highly oriented pyrolytic graphite by atomic force microscopy, *FEBS Lett.*, 2006, **580**, 5671–5675.
- 37 A. D. Protopopova, N. A. Barinov, E. G. Zavyalova, A. M. Kopylov, V. I. Sergienko and D. V. Klinov, Visualization of fibrinogen  $\alpha$ C regions and their arrangement during fibrin network formation by high-resolution AFM, *J. Thromb. Haemostasis*, 2015, **13**, 570–579.
- 38 N. A. Barinov, V. V. Prokhorov, E. V. Dubrovin and D. V. Klinov, AFM visualization at a single-molecule level of denaturated states of proteins on graphite, *Colloids Surf., B*, 2016, **146**, 777–784.
- 39 A. D. Protopopova, R. I. Litvinov, D. K. Galanakis, C. Nagaswami, N. A. Barinov, A. R. Mukhitov, D. V. Klinov and J. W. Weisel, Morphometric characterization of fibrinogen's  $\alpha$ C regions and their role in fibrin self-assembly and molecular organization, *Nanoscale*, 2017, **9**, 13707–13716.
- 40 D. V. Klinov, I. V. Lagutina, V. V. Prokhorov, T. Neretina, P. P. Khil, Y. B. Lebedev, D. I. Cherny, V. V. Demin and E. D. Sverdlov, High resolution mapping DNAs by R-loop atomic force microscopy, *Nucleic Acids Res.*, 1998, **26**, 4603–4610.
- 41 N. Esmaili and J. L. Leroy, i-Motif solution structure and dynamics of the d(AACCCC) and d(CCCCAA) tetrahymena telomeric repeats, *Nucleic Acids Res.*, 2005, **33**, 213–224.
- 42 M. J. D. Powell, Restart procedures for the conjugate gradient method, *Math. Program.*, 1977, **12**, 241–254.
- 43 D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu and P. A. Kollman, *Amber 14*, University of California, San Francisco, 2014.
- 44 P. Mark and L. Nilsson, Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.
- 45 A. Pérez, I. Marchán, D. Svozil, J. Spöner, T. E. Cheatham, C. a. Loughton and M. Orozco, Refinement of the AMBER force field for nucleic acids: improving the description of alpha/gamma conformers, *Biophys. J.*, 2007, **92**, 3817–3829.
- 46 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, Comparison of multiple Amber force fields and development of improved protein backbone parameters, *Proteins*, 2006, **65**, 712–725.
- 47 T. Darden, D. York and L. Pedersen, Particle mesh Ewald: An  $N\log(N)$  method for Ewald sums in large systems, *J. Chem. Phys.*, 1993, **98**, 10089–10092.
- 48 W. Humphrey, A. Dalke and K. Schulten, VMD – Visual Molecular Dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.



- 49 J. Zavadlav, R. Podgornik and M. Praprotnik, Adaptive Resolution Simulation of a DNA Molecule in Salt Solution, *J. Chem. Theory Comput.*, 2015, **11**, 5035–5044.
- 50 M. A. Batalia, E. Protozanova, R. B. Macgregor and D. A. Erie, Self-assembly of frayed wires and frayed-wire networks: nanoconstruction with multistranded DNA, *Nano Lett.*, 2002, **2**, 269–274.
- 51 A. M. Fleming, Y. Ding, R. A. Rogers, J. Zhu, J. Zhu, A. D. Burton, C. B. Carlisle and C. J. Burrows,  $4n - 1$  is a 'sweet spot' in DNA i-motif folding of 2'-deoxycytidine homopolymers, *J. Am. Chem. Soc.*, 2017, **139**, 4682–4689.
- 52 R. Shiman and D. E. Draper, Stabilization of RNA tertiary structure by monovalent cations, *J. Mol. Biol.*, 2000, **302**, 79–91.
- 53 J. Cui, P. Waltman, V. H. Le and E. A. Lewis, The effect of molecular crowding on the stability of human c-MYC promoter sequence i-motif at neutral pH, *Molecules*, 2013, **18**, 12751–12767.
- 54 A. Rajendran, S. Nakano and N. Sugimoto, Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH, *Chem. Commun.*, 2010, **46**, 1299–1301.

