Access to a stabilized i-motif DNA structure through four successive ligation reactions on a cyclopeptide scaffold†

Alexandre Devaux, Laureen Bonnat, Thomas LavergneID* and Eric DefrancqID*

†Electronic supplementary information (ESI) available: RP-HPLC profiles of products and ESI-MS and MALDI TOF-MS spectra of purified ODNs, CD spectra. See DOI: 10.1039/d0ob01311k

i-Motifs are largely underexplored tetraplex nucleic acid structures which have been suggested to perform essential biological functions and might constitute future therapeutic targets. i-Motifs generally require acidic conditions to fold in vitro, a particularity which significantly complicates the use of native i-motif forming sequences for interaction studies with potential ligands and biological components (e.g. proteins). In this context, we report herein on the assembly of a peptide–DNA conjugate capable of folding at room temperature into a stable i-motif structure at neutral pH. To achieve the controlled assembly of the i-motif forming conjugate, we developed a new synthetic pathway of four successive orthogonal ligation reactions between bifunctional C-rich DNA strands and a tetrafunctional cyclopeptide scaffold.

Introduction

Over half a century ago, the seminal contributions of Watson, Crick, Wilkins and Franklin established the double-helical structure of DNA in which two antiparallel strands are held together through canonical A/T and G/C base pairing. However, the past decades have brought accumulating evidence of the existence of a wide variety of alternative nucleic acid topologies including hairpin, triplex, cruciform and tetraplex structures. In particular, G- and C-rich sequences have been shown to form tetraplex architectures, respectively, G-quadruplex (G4) and i-motif structures.1,2 Unlike duplex DNA, G4 and i-motif formation does not involve canonical Watson–Crick base pairing: indeed a G4 DNA is formed from the stacking of Hoogsteen hydrogen bonded guanine tetrads whereas for i-motif formation two anti-parallel duplexes are held together by hemi-protonated C–C+ base pairs intercalated with each other (Fig. 1).

G4 DNA structures have been the subject of intensive investigations. Sequencing and bioinformatics analyses of the human genome indicate that it contains as many as 700 000 sequences having the potential to form stable G-quadruplex structures3 and a large number of studies have reported on their central biological functions within physiological and pathological processes.4 Although i-motif DNA structures have attracted less attention, several recent studies suggest that they may also perform essential biological functions. By using the Quadparser algorithm, more than 5000 putative i-motif forming sequences have been found in the human genome. Interestingly, more than 600 were found to be located in promoter sequences of genes such as c-myc, bcl-2 or c-kit.5 Recent studies have proposed that i-motifs are involved in the regulation of transcription6 and the integrity of telomeric DNA.7 Proteins such as hnRNP LL have been identified to interact with i-motif structures acting as an activating transcription factor.8 More recently, two independent studies have demonstrated the presence of i-motif structures in human cells through in cellulo NMR spectroscopy9 and immunofluorescence by using a specific antibody against the i-motif.10 In addition to the regulation of cellular processes in human cells, i-motif forming sequence has been identified in the long terminal repeat promoter of the HIV-1 proviral genome where it has been shown to modulate viral transcription.11

Encouraged by these inspiring reports, several groups have engaged in successful efforts to identify small molecule ligands that interact with C-rich sequences, stabilizing or destabilizing the corresponding i-motif structures.12 Some of these ligands have been shown to modulate gene regulation and expression and they now constitute promising probes and drug candidates targeting i-motif sequences within cancer-related genes or viral genomes.6

Beyond their biological occurrences, i-motifs have also attracted huge interest in the field of functional nano-
structures. *i*-Motif structures which can fold and unfold reversibly depending on their environments have been manipulated by chemists to create nanotools such as in cell pH sensors \(^{13}\) or pH responsive shape-memory hydrogels. \(^{14}\) Indeed, the formation and the stability of *i*-motif DNA structures are affected in vitro by a number of parameters such as the length and the nature of the sequence, the solute concentration and the pH. In particular, *i*-motifs are generally formed in vitro at acidic pH, a particularity that complicates the use of native *i*-motif sequences to study the interactions of potential ligands. As a consequence, numerous strategies have been explored to stabilize the *i*-motif structure and to promote its formation at physiological pH. Extending the number of C–C+ base pairs and adjusting the sequence of the loop region can help to stabilize the folded structure. \(^{15}\) For a given sequence of interest, modifications of the sugar, \(^{16}\) base \(^{17}\) and phosphate moieties \(^{18}\) can promote the formation of the *i*-motif. However, all of these modifications are prone to affect the recognition of the native structure of interest by putative targeting ligands.

In this context, the design of biomolecular conjugates capable of folding into well-defined and stable native *i*-motif structures would be of interest as they would find applications in the development of structure-specific *i*-motif binding ligands as well as in the identification of their structure-specific binding partners (e.g., proteins) via pull-down strategies.

In the past, our group, along with others, have proposed the use of peptide scaffolds to assemble stable mimics of tetraplex DNA structures (as in Fig. 2). \(^{19}\) Our strategy is based on the use of regioselectively addressable cyclopeptide platforms for directing the intramolecular assembly of the attached oligonucleotides into the desired tetraplex topology. By site-specifically fixing DNA and/or RNA G4-forming sequences onto cyclopeptide scaffolds, we have been able to trigger the

---

**Fig. 1** (A) G-quartet with Hoogsteen hydrogen bonded guanine and schematic drawing of a G4 structure and (B) hemi-protonated C–C+ base pairs and schematic representation of an *i*-motif structure.

**Fig. 2** Use of the cyclopeptide scaffold for the design of stabilized tetraplex DNA structures: (A) parallel G4-DNA, \(^{19}\) (B) *i*-motif DNA structure without loops, \(^{22}\) (C) antiparallel G4 structure from the HIV virus sequence, \(^{23}\) (D) telomeric *i*-motif mimic 1a and (E) telomeric *i*-motif mimic 1b. For the sake of clarity, the positive charges shared by the two cytosines involved in the C–C base pairs of structures B, D and E are not displayed. Dotted lines represent the base pairing.
formation of topologically controlled and thermally stabilized G4 architectures such as parallel tetramolecular DNA and RNA G4, antiparallel dimeric DNA G4 and DNA-RNA hybrid G4 structures.\textsuperscript{19a,20,21} For the preparation of the parallel DNA (or RNA) tetrameric systems (Fig. 2A), four oxime ligation reactions were simultaneously performed between an aminooxy-containing cyclopeptide platform and DNA (or RNA) strands that bear aldehyde complementary functions, whereas the controlled anchoring of nucleic acid strands to form antiparallel structures such as the \textit{i}-motif structure (Fig. 2B) required two sequential ligations and for that purpose oxime ligation (OL) and copper-catalyzed azide–alkyne cycloaddition (CuAAC) reactions were performed between an aminooxy- and azide-containing cyclopeptide and DNA strands that bore either aldehyde or azide functions.\textsuperscript{22} More recently, a more sophisticated system consisting of a biologically relevant viral G4 (Fig. 2C) has been elaborated with the site-specific attachment of two distinct oligonucleotides in three successive ligations involving the OL, the CuAAC and a third compatible ligation, namely the thiol-iodoacetamide (TC) SN2 reaction. The latter ligation was found to fulfill the criteria of chemical orthogonality associated with the synthesis of such a bioconjugate and the resulting thioether linkage had no perturbative effects on the folding of the G4 structure.\textsuperscript{23} All the resulting tetraplex mimics obtained so far displayed increased stabilities and reduced structural polymorphism. Owing to these properties, the different G4 motifs have found applications in the selection and characterization of structure-specific ligands.\textsuperscript{24}

To further expand this approach and to access more sophisticated biomolecular systems mimicking nucleic acid structures of high biological interest, we explored strategies to fully control the assembly of DNA–peptide conjugates \textit{via} four successive ligations steps. As a proof of concept, we report herein on our unsuccessful (1a, Fig. 2D) and successful (1b, Fig. 2E) optimization efforts to assemble and stabilize a biologically relevant \textit{i}-motif structure derived from the human telomeric sequence (CCCTAA)\textsubscript{4}. Unlike the \textit{i}-motif structure in Fig. 2B, which consisted of four tracts of four cytosines forming a tetramolecular \textit{i}-motif structure without biological relevance, the \textit{i}-motif structure within 1b was designed to form \textit{via} the association of four tracts of three cytosines forming a bimolecular \textit{i}-motif structure displaying two \textsuperscript{5}TAA\textsuperscript{3} loops. This new conjugate therefore precisely mimics the biologically relevant structure of the human telomeric \textit{i}-motif and in particular the loop organization, a feature that is crucial for the subsequent selection and characterization of structure-specific ligands.

To assemble the \textit{i}-motif conjugates 1a and 1b \textit{via} four successive orthogonal reactions, we have chosen to make use of the three previously exploited ligations \textit{(vide supra)}, \textit{i.e.} the oxime ligation (OL), the copper-catalyzed azide–alkyne cycloaddition (CuAAC) and the thiol-iodoacetamide SN2 reaction (TC), with the objective of using one of them twice. Oxime ligation has already been used for the successive incorporation of reporters for the preparation of ODN bis-conjugates.\textsuperscript{25} However both partners for the OL \textit{(i.e.} aldehyde and aminooxy functions) are not very stable and thus, if one would want to perform two OL stepwise for the synthesis of sophisticated systems such as 1a/b, non-trivial protection strategies should be implemented \textit{(i.e.} use of different protecting groups for the aldehyde or aminooxy groups). In our hand, TC reactions have been found to be more sluggish than OL and CuAAC during the formation of DNA–peptide conjugates and therefore repeating twice this reaction in the course of the assembly of these conjugates may prove problematic. On the other hand, CuAAC reactions have been repeatedly used for the successive incorporation of reporters at internal positions or at 3\textsuperscript{\prime},5\textsuperscript{\prime} extremities of nucleic acids.\textsuperscript{26} Furthermore, protected alkynes can be used along with a free alkyne to allow for sequential CuAAC reactions after mild deprotections.\textsuperscript{26a} CuAAC can also be performed in the presence of free amines which can later be converted into azides \textit{via} a mild diazotransfer reaction for subsequent CuAAC reactions.\textsuperscript{27} Finally, both the alkyne and the azide are stable under a wide range of non-reductive experimental conditions. Based on the robustness and flexibility of the CuAAC reaction, we decided to explore the formation of the \textit{i}-motif structural mimic \textit{via} the stepwise assembly of peptide–DNA conjugates through four successive ligations with one OL, one TC and two CuAAC reactions.

Two strategies were explored and involved the use of two diversely functionalized cyclopeptides and three distinct bifunctional C-rich oligonucleotides. The first strategy (Route A, Scheme 1) consists of starting with the oxime ligation followed by a first CuAAC reaction to attach bifunctional oligonucleotide 2. Subsequent thiol-iodoacetamide SN2 (TC) and CuAAC reactions are performed to incorporate the second bifunctional oligonucleotide 3 toward the final conjugate (OL → CuAAC → TC → CuAAC). Unfortunately, Route A was found to be inefficient at affording the expected target 1a (Fig. 2D). A second route (Route B, Scheme 2) was thus explored consisting of starting the assembly with the least yield-efficient TC chemistry using oligonucleotide 10 and achieving the assembly by the most efficient OL reaction. The following synthetic pathway was therefore used: TC → CuAAC → CuAAC → OL. This route allowed us to access conjugate 1b (Fig. 2E). The resulting structural mimic was found to fold into a stable \textit{i}-motif structure at physiological pH and room temperature paving the way to functional applications.

**Results and discussion**

**Route A (OL → CuAAC → TC → CuAAC)**

This strategy requires the attachment of oligonucleotides 2 (\textsuperscript{5}Y-ACCCCTAACCCT-X\textsuperscript{3}) and 3 (\textsuperscript{5}Y-ACCCCTAACCCT-Z\textsuperscript{3}) in which \(X\) represents a 3\textsuperscript{\prime}-aldehyde linker, \(Y\) a 5\textsuperscript{\prime}-alkynyl linker and \(Z\) a 3\textsuperscript{\prime}-thiol linker through OL, CuAAC and TC reactions onto the corresponding peptide scaffold 4a bearing the complementary functional groups (Scheme 1).

Cyclopeptide scaffold 4a was prepared using standard fluorenlymethoxycarbonyl (Fmoc)/IBu solid-phase peptide synthesis (SPPS) on a 2-chlorotriyl resin followed by a cyclization reaction in solution according to the reported procedures.
Briefly, the biotin, aminooxy and azido moieties were incorporated during SPPS by using an Nε-biotin-functionalized lysine building block, an Nε-ethoxyethylidene (Eei) protected aminooxy functionalized lysine building block and the L-azidonorleucine building block, respectively. To equip the cyclopeptide with the chloroacetamide function and the second azido group, two other orthogonally protected lysines (i.e. Nε-Boc and Nε-Dde protected lysine) were introduced during SPPS. After the cleavage from the resin under slightly acidic conditions, to prevent the cleavage of the ethoxy-
ethylidene (Eei) protecting group, and cyclization of the linear peptide under dilute conditions, the chloroacetamide group was anchored on the lysine side chain using chloroacetic anhydride after selective removal of the bis-N-[1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl] (Dde) lysine protecting group with hydrazine. Final treatment under acidic conditions allowed the removal of tert-butoxycarbonyl (Boc) as well as the ethoxyethylidene protections and afforded the desired cyclopeptide scaffold 4a. To synthesize oligonucleotides 2 and 3, solid-phase oligonucleotide syntheses (SPOS), using phosphoramidite chemistry, were performed on a commercially available glyceryl CPG solid support and a 3′-thiol modifier C3−S−SCPG solid support, respectively. The alkynyl function was introduced at the 5′-terminus of oligonucleotides 2 and 3 by using commercially available 5′-hexynyl (β-cyanoethyl) phosphoramidite. Following the cleavage from the solid support and detritylation steps, oligonucleotide 2 was obtained after oxidative cleavage of the 3′-diol group. To obtain oligonucleotide 3, a 100 mM DTT treatment was performed to reduce the disulfide linkage generated upon the cleavage from the solid support.

Toward the incorporation of the first oligonucleotide, 2 was grafted on cyclopeptide scaffold 4a through oxime bond formation to generate intermediate 5. The coupling reaction was carried out in ammonium acetate buffer (pH 4.5) at 55 °C using a slight excess of aminooxy-containing peptide 4a. The reaction was monitored by RP-HPLC and proceeded to com-
completion within 45 min. Purification through RP-HPLC followed by desalting using size-exclusion chromatography (SEC) afforded pure conjugate 5 with an isolated yield of 65%. A first intramolecular CuAAC reaction was then performed in HEPES buffer (pH 7.4) in the presence of CuSO₄, tris(benzyltriazolylmethyl)amine (THPTA) ligand and sodium ascorbate at 37 °C for 2 h to afford, following desalting using size-exclusion chromatography (SEC) and RP-HPLC purification, conjugate 6 in 72% yield. Conjugates 5 and 6 were characterized by ESI MS analysis, which showed an excellent agreement between the experimentally determined molecular weight and the calculated value (Fig. S15 and S17†).

From intermediate 6, we envisioned to perform first the TC reaction and to complete the assembly with a CuAAC reaction. This decision was mainly guided by the instability of the thiol function during the CuAAC reaction. Toward 1a from 6, two alternative pathways have been investigated that consist of performing the diazo transfer reaction either after (a) or before (b) the TC reaction (Scheme 1). Previous studies have shown a poor reactivity for the TC reaction when the first oligonucleotide was already anchored on the cyclopeptide scaffold and consequently an optimization of the TC reaction had to be performed.23 The assembly of conjugate 6 with thiol containing oligonucleotide 3 was thus carried out in water at room temperature in the presence of tris(2-carboxylethyl)phosphine (TCEP) to prevent the formation of disulfide species, potassium iodide (KI) for in situ activation of the reaction through the transient formation of a reactive iodoacetamide group and DIEA to obtain a pH value around 8.5. Furthermore, under these optimized conditions, the addition of KCl is absolutely required for cation shielding to neutralize unfavorable charge repulsions between conjugate 6 and non-complementary oligonucleotide 3. Intermediate 7 was obtained in 74% yield after RP-HPLC purification and then subjected to a diazo transfer reaction with the ISAHC (imidazole-1-sulfonylazide hydrochloride) reagent by using previously reported conditions27 to afford conjugate 8 in 80% yield. The other alternative (way b) has also been explored by performing first the diazo transfer reaction on conjugate 6 to give intermediate 9 and the subsequent TC reaction with oligonucleotide 3 under the same optimized conditions as above to afford conjugate 8 in 30% overall yield. Conjugate 8 obtained by both pathways was characterized by ESI-MS analysis which showed a good agreement between the experimentally determined molecular weight and the calculated value (Fig. S21†). It is noteworthy that the diazo transfer reaction was found to be efficient and compatible with the sophisticated conjugates 6 and 7.

Toward 1a, a second CuAAC reaction had to be achieved through the reaction of the alkylene moiety on the 5'-extremity of the grafted oligonucleotide and the azide function carried by the cyclopeptide. By using the standard and above-mentioned CuAAC conditions as well as in the presence of 500 mM KCl with the aim to neutralize unfavorable charge repulsions between the oligonucleotide strands, no transformation was observed as evidenced by RP-HPLC monitoring (data not shown). We only observed the recovery of the starting material 8. Variation of reactant equivalents and temperatures did not prove to be beneficial for promoting the formation of 1a. As i-motifs are known to form at acidic pH, we thus envisioned to perform the CuAAC reaction under slightly acidic conditions compatible with CuAAC. At slightly acidic pH, the DNA strands of conjugate 8 were expected to adopt an antiparallel topology (pre-organization to an i-motif structure) and thus to hold the azide and alkyn groups close to one another thereby lowering the entropic cost of the CuAAC reaction. CD analysis of the starting conjugate 8 indeed revealed the formation of the anti-parallel topology at pH 4, 5 and 6 whereas at pH 7 no structuring was noticed (Fig. S39†). CuAAC reactions were thus carried out at pH 4, 5, 6 and 7 under the above conditions (with and without KCl) and again no conversion toward 1a could be observed by RP-HPLC monitoring regardless of the conditions: only the starting material 8 was observed (data not shown). The absence of conversion could be explained by the ineffectiveness of the CuAAC reaction under acidic conditions.

Although Route A was found to be unsuitable to afford the desired constrained i-motif 1a it provided precious information to design the second strategy (Route B). Indeed, a major drawback with Route A is that during the last synthetic step, i.e. CuAAC reaction, the molecular weights of the starting material 8 and final product 1a are the same thus precluding the use of mass spectrometry to monitor the transformation and identify the final conjugate. Also, the TC reaction was again shown to be the least efficient and hard to carry out when an oligonucleotide is already grafted on the molecular scaffold. We thus decided for Route B to use the least efficient TC reaction at an early stage of the synthetic process while the most efficient OL reaction would be used in the last step. Furthermore, the intramolecular OL reaction results in the formation of a water by-product leading to a final product whose molecular weight is distinct from that of its precursor, thus allowing the use of MS for the monitoring of the reaction and identification of the final product 1b. Moreover, the OL reaction is known to be more efficient under acidic conditions and the formation of the i-motif is also known to be favored at acidic pH. We thus anticipated a pre-organization of the i-motif leading to a spatial proximity of the two reactive functions (i.e. aldehyde and aminooxy moieties) to promote the last intramolecular reaction.

**Route B (TC → CuAAC → CuAAC → OL)**

This strategy requires the successive attachment of oligonucleotide 10 (5'-Y-ACCCTAACCCT-X') in which X represents a 3'-azide linker and Y a 5'-thiol linker and the above-described oligonucleotide 2 through TC, CuAAC and OL reactions onto the corresponding peptide scaffold 4b bearing the complementary functional groups (Scheme 2). Cyclopeptide 4b thus contains a protected aminooxy group for the OL reaction, an alkylene function for the first CuAAC reaction, a chloroacetamide moiety for the TC reaction and an amino function to be transformed into an azide for the second CuAAC reaction. It was prepared in the same manner as peptide 4a using standard fluorenylmethoxycarbonyl (Fmoc)/tBu solid-phase
peptide synthesis (SPPS) on a 2-chlorotrityl resin and the cyclization reaction in solution (see Scheme S2†). Again the biotin and aminoxy moieties were incorporated during SPPS by using an Nε-biotin-functionalized lysine building block and an Nε-ethoxymethylene (Eei) protected aminox group functionalized lysine building block and two other orthogonally protected lysines (i.e. Nε-Allo and Nε-Dde protected lysine) were introduced during SPPS for the introduction of chloroaacetamide and azide moieties as above. The alkylene group was used as a building block during SPPS by using an Nε-pentyne functionalized lysine building block. After the cleavage from the resin under slightly acidic conditions (i.e. to prevent the removal of the Eei aminox protecting group) and cyclization of the linear peptide, the chloroaacetamide group was anochro on the lysine side chain using chloroaacet anhydride after selective removal of the Dde protecting group with hydrazine. It should be noted that allylic alcohol should be used during the Dde cleavage to prevent the reduction of Alloc protection. The Alloc protecting group was next removed through treatment with Pd(0)(Ph3)4 in the presence of phenysilane and the Alloc protecting group was next removed through treat-
ment as described above to a neutralized lysine building block. After the cleavage from the commercially available 3′-PT-amino-modifier C3 CPG. The thiol function was introduced at the 5′-terminus by using the commercially available 5′-S-trityl modifier C6 phosphoramidite. The introduction of an azide function was achieved through the diazo transfer reaction as described above on the S-trityl protected thiol oligonucleotide followed by thiol deprotection in the presence of silver nitrate as reported.29

With Route B, the first step consists of the use of the TC reaction which was carried out under the above-mentioned conditions, i.e. in the presence of tris(2-carboxyethyl)phosphine (TCP), potassium iodide (KI) and DIEA. In this case, the addition of KCl for cation shielding is not required. It should be noted that due to the presence of an azide group on oligonucleotide 10, TCEP had to be added during the reaction and not prior to the addition of all other components in order to avoid partial reduction of the azide group. Desalting using size-exclusion chromatography (SEC) followed by RP-HPLC purification afforded pure conjugate 11 with an isolated yield of 53%. A first intramolecular CuAAC reaction was then performed in PBS buffer (pH 7.4) in the presence of CuSO4, THPTA ligand and sodium ascorbate at 37 °C for 2 h. As previously observed, the addition of KCl was required in order to neutralize unfavorable charge repulsions. After the addition of EDTA, desalting using size-exclusion chromatography (SEC) and RP-HPLC purification, conjugate 12 was obtained in 40% yield. The free amino function on the peptide moiety was then transformed into an azide by using a diazo transfer reaction which was carried out as described above to afford, in a quantitative yield, conjugate 13 after purification by size-exclusion chromatography. Conjugates 11-13 were characterized by ESI-MS analysis which showed a good agreement between the experimentally determined molecular weight and the calculated value (Fig. S26, S29 and S31†).

From intermediate 13, two pathways can be considered to access the expected product 1b. Indeed, a first OL could be performed to attach oligonucleotide 2 followed by a second CuAAC reaction or vice versa. As mentioned above, the major drawback of the CuAAC reaction is that there is no molecular weight difference between the starting material and the reaction product and the retention time difference is expected to analyze and purify the crude reaction mixture. Furthermore, the CuAAC reaction was found to be inefficient in the last step of the assembly of 1a. For these reasons, we envisioned to perform the CuAAC reaction first followed by the most reactive oxime reaction as the last step. The CuAAC reaction was carried out under the optimized conditions as described above, i.e. in the presence of 500 mM KCl for neutralizing the unfavorable charge repulsions between 13 and oligonucleotide 2. However, a low conversion rate was obtained at room temperature (around 20–30% according to LC-MS monitoring). Increasing the temperature to 60 °C was found to be beneficial for affecting intermediate 14 with a satisfactory 48% yield after desalting using size-exclusion chromatography (SEC) and RP-HPLC purification.

The last step consisted of the one-pot deprotection of the aminox function and the oxime reaction. Treatment of conjugate 14 under mild acidic conditions (1% TFA) allowed the cleavage of the Eei protecting group and the simultaneous oxime bond formation. As anticipated, the monitoring of the reaction by LC-MS analysis showed the loss of the Eei protection as well as the loss of a water molecule thus demonstrating the formation of the oxime bond to afford the i-motif forming conjugate 1b (Fig. 3 and S35–S38†). It should be mentioned that the intermediate-free aminox compound was not observed which emphasizes the high reactivity of the oxime reaction. As expected, the efficiency of the oxime formation might have been promoted by the preorganization of the DNA strands into an i-motif structure at the acidic pH required to perform the oxime reaction.

Characterization of i-motif formation

To confirm the formation of an i-motif structure within 1b and to gauge its stability, CD studies were performed at different pH values at a concentration of 2 μM at 20 °C. Previous investigations have indeed shown that i-motif DNA can be associated with a specific CD signature which is markedly distinct for the CD signature of C-rich DNA fragments not forming the i-motif structure.30 The CD spectrum of the biomolecular assembly 1b recorded at pH 4 displayed a positive peak at 286 nm and a negative peak at 262 nm, characteristic of the formation of the i-motif. For comparison, the CD spectrum of the native telomeric sequence 15 d(5′CCCTAATCCCT3′) was recorded and showed similar signals (Fig. 4A shows the CD spectra of 1a and 15 at pH = 4) that confirms the formation of the i-motif structure in 1b.

The influence of pH on the formation of the i-motif structure was then investigated with both compounds 1b and oligonucleotide 15. By varying the pH from 4 to 7 the shape of the signal changed with the appearance of a positive peak at
around 270 nm at pH = 7 (Fig. 4B) for the native telomeric sequence 15 which can be attributed to the “unstructured” DNA motif in agreement with reported studies.\(^{15a}\) The intermediate conjugate 8 also shows the CD characteristics of unfolded C-rich DNA fragments at pH 7 (Fig. S39†), whereas for constrained system 1b, the characteristic signal of the i-motif structure is still observed at pH 7 (Fig. 4C). These data are of interest as they suggest that interaction studies with potential ligands as well as with biological components (e.g., proteins) might be performed at physiological pH with the mimic structure 1b.

The thermal stability of the i-motif structure formed within 1b was further evaluated by CD denaturation studies at pH 6 and at neutral pH 7 (Fig. S40 and S41†). At pH 6 the melting temperature was 40 °C, a temperature which is on par with the reported melting temperature of the native telomeric i-motif sequence at pH 5.5.\(^{15a}\) On the other hand, we found that the telomeric i-motif structural mimic within 1b possesses a melting temperature of 31 °C at pH 7. This demonstrates a significant stabilization from the native structure, which was shown to unfold into a random coil at pH 6.5 with no determinable \(T_m\) at physiological pH,\(^{15a}\) and further confirms that 1b might be exploited to study the interaction of i-motif-targeting ligands at neutral pH and room temperature.

**Experimental section**

**General remarks**

ESI mass spectra were recorded using an Esquire 3000 spectrometer from Bruker or an Acquity UPLC/MS system from Waters equipped with an SQ Detector 2 and a column heater set at
60 °C. MALDI-TOF mass spectra were recorded using an Autoflex Bruker with hydropicolinic acid (HPA) (45 mg, ammonium citrate (4 mg) in 500 μL H₂O/CH₃CN) as the matrix. Peptides were analyzed in the positive mode and oligonucleotides and conjugates in the negative mode.

Peptide synthesis

General details. The peptide synthesis was performed on a Syro II synthesizer using the Fmoc/tBu strategy on a 2-chlorotrityl resin. The course of reactions was monitored by using a Waters UPLC-MS system including reverse phase chromatography using a Nucleosil C18 column (130 Å, 2.1 × 50 mm, 1.7 μm) and with detection by UV spectrometry at 214 nm and 250 nm and combined electron spray ionization mass spectrometry. RP-HPLC purifications were performed on a Gilson or Waters system with a Nucleosil C18 column (100 Å, 250 × 21 mm, 7 μm) with UV monitoring at 214 nm and 250 nm. A 20 mL min⁻¹ flow linear gradient was applied from 95% solvent A1 (0.1% formic acid in water) and 5% solvent B1 (0.1% formic acid in acetonitrile) to 100% B1 for 20 min.

Synthesis of peptide 4a. The linear peptide was synthesized using 2-chlorotrityl (loading of 0.83 mmol g⁻¹) commercial resin. Fmoc-Gly-OH (3 equiv.) was coupled on the resin in anhydrous DCM in a glass reaction vessel fitted with sintered glass. The pH was adjusted to 8 using DIEA. The mixture was stirred for 1 h at room temperature. A DCM/MeOH/DIEA solution (17/2/1, v/v/v) was then added. The Fmoc protecting group was then removed using three washes with a 20% piperidine solution in DMF (40 mL). The resin loading was assayed by quantification of free dibenzofulvene using UV absorbance at 299 nm (loading of 0.40 mmol g⁻¹, yield: 48%). The elongation was then performed on a Syro II using the Fmoc/tBu strategy on the above-prepared resin using HTBU as an activator. Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Lys(Dde)-OH, Fmoc-Lys(Alloc)-OH and Fmoc-Lys(biotin)-OH were commercially available. Fmoc-azidonorleucine and Fmoc-Lys(Eei-Aoa)-OH were synthesized according to reported protocols. The elongated peptide was then cleaved from the resin by using DCM/TFE/AcOH solution (70/20/10, v/v/v). The resulting solution was evaporated under vacuum and the peptide was precipitated in ether to obtain a white powder. The crude product was used without any further purification in the next step. The cyclization of the linear peptide was performed in DMF at 10⁻³ M concentration in the presence of PyBOP (1.2 equiv.). The pH was adjusted to 8–9 using DIEA and the solution was stirred at room temperature until complete peptide cyclization occurred (UPLC monitoring). The solvent was evaporated under vacuum and then the crude peptide was precipitated in ether and purified by RP-HPLC and freeze-dried to obtain a white powder. ESI-MS (+): m/z calcd for C₇₀H₁₁₄N₂₀O₁₈S: 1591.8, m/z found: 1613.9 [M + Na]⁺, HR-ESI-MS (+): m/z calcd for C₇₀H₁₁₄N₂₀O₁₈S: 1591.8186, m/z found: 1591.8196. The Boc protection on the lysine side chain and ethoxymethylide on the aminooxy moiety were then cleaved through treatment with a TFA/TIS/H₂O (80/16/4, v/v/v) solution and the mixture was stirred for 1 h at room temperature. The solvent was then evaporated under vacuum and the crude peptide was precipitated in ether to afford 4a as a white powder. ESI-MS (+): m/z calcd for C₆₉H₁₀₂N₂₀O₁₅S: 1421.7, m/z found: 1421.9 [M + H]⁺.

Synthesis of peptide 4b. Peptide 4b was prepared using the same procedure as for peptide 4a. The following protected amino acids were used: Fmoc-Ala-OH, Fmoc-Gly-OH, Fmoc-Lys(Boc)-OH, Fmoc-Pro-OH, Fmoc-Lys(Dde)-OH, Fmoc-Lys(Alloc)-OH and Fmoc-Lys(biotin)-OH which are commercially available and Fmoc-Lys(–N-pentynoic acid)-OH and Fmoc-Lys(Eei-Aoa)-OH building blocks which were prepared as previously reported. After cleavage from the resin with DCM/TFE/AcOH solution (70/20/10, v/v/v) and cyclization as above, the cyclized intermediate peptide was purified by RP-HPLC and freeze-dried to obtain a white powder. ESI-MS (+): m/z calcd for C₆₉H₁₁₄N₁₈O₁₉S: 1716.92, m/z found: 1716.97 [M + H]⁺. Dde removal and introduction of the chloroacetamide group were carried out as above. ESI-MS (+): m/z calcd for C₇₃H₁₁₈N₁₈O₁₉S: 1629.34, m/z found: 1629.95 [M + H]⁺. The peptide was then dissolved in anhydrous CH₃Cl/anhydrous DMF (3/1, v/v) and Pd[Pb]₃(0.2 equiv.) and PhSiH₃ (100 equiv.) were added. The reaction mixture was stirred at room temperature for 30 minutes and the reaction was quenched with MeOH. The solution was evaporated under vacuum and finally purified by RP-HPLC to afford peptide 4b. ESI-MS (+): m/z calcd for C₇₃H₁₁₈N₁₈O₁₉S: 1545.8018, m/z found: 1545.8018, m/z found: 1545.8003 [M⁺].

Oligonucleotide synthesis

General details. The oligonucleotides were prepared using β-cyanoethylphosphoramidite chemistry on a 3400 DNA synthesizer at the 1 μmol scale. RP-HPLC analyses were performed on a Waters HPLC system using a C18 Nucleosil column (Macherey-Nagel, 100 Å, 250 × 4.6 mm, 5 μm) with UV-monitoring at 260 and 280 nm. The RP-HPLC purifications of oligonucleotides were performed on a Gilson system with a Nucleosil C18 column (Macherey-Nagel, 100 Å, 250 × 10 mm, 7 μm) with UV-monitoring at 260 and 280 nm using a 4 mL min⁻¹ flow linear gradient. Solvent A2 (50 mM triethylammonium acetate buffer with 5% acetonitrile) and solvent B2 (acetonitrile with 5% water) were used.
Synthesis of oligonucleotide 2. Oligonucleotide 2 was prepared according to a previously reported procedure.\textsuperscript{22} ESI-MS (−): m/z calcd for C\textsubscript{112}H\textsubscript{148}N\textsubscript{37}O\textsubscript{50}P\textsubscript{12}S\textsubscript{2}: 3502.60, m/z found: 3502.68 [M − H]−.

Synthesis of oligonucleotide 3. The 5′-hexynyl, 3′-S-protected intermediate oligonucleotide precursor of oligonucleotide 3 was obtained from automated synthesis using commercially available 5′-hexynyl (β-cyanoethyl) phosphoramidite and 3′-thiol-modifier C3 S-S CPG resin at the 1 μmol scale. The crude product was purified by RP-HPLC. ESI-MS (−): m/z calcd for C\textsubscript{116}H\textsubscript{156}N\textsubscript{37}O\textsubscript{50}P\textsubscript{12}S\textsubscript{2}: 3624.5, m/z found: 3625.5 [M − H]−. Dithiothreitol (100 equiv.) was then added to a solution of 3′-S-protected oligonucleotide in Tris HCl buffer 1 M, pH 8.5 to obtain a final concentration of 10−3 M. The mixture was stirred for 1 h at room temperature and the crude product was purified by RP-HPLC and then desalted to afford oligonucleotide 3. ESI-MS (−): m/z calcd for C\textsubscript{113}H\textsubscript{151}N\textsubscript{38}O\textsubscript{69}P\textsubscript{12}S: 3534.4, m/z found: 3534.9 [M − H]−.

Synthesis of oligonucleotide 10. The precursor oligonucleotide bearing a 5′-S-trityl group and a 3′-amino group was obtained from automated synthesis using the commercially available 5′-thiol modifier C6 phosphoramidite and the 3′-Pt-amino-modifier C3 CPG resin at the 1 μmol scale. The crude product was purified by RP-HPLC. ESI-MS (−): m/z calcd for C\textsubscript{132}H\textsubscript{171}N\textsubscript{38}O\textsubscript{69}P\textsubscript{12}S: 3797.73, m/z found: 3797.11 [M − H]−. The intermediate oligonucleotide was then suspended in NaHCO\textsubscript{3} buffer (50 mM, pH 8.5)/MeOH (3/1, v/v) in a closed screw-cap plastic vial. CuSO\textsubscript{4} (3 equiv.) and imidazole-1-sulfonyl azide hydrochloride (ISAHC, 30 equiv.) were added. Afterwards, the solution was heated at 60 °C. The reaction was followed by RP-HPLC and UPLC-MS. Finally, the solution was cooled to room temperature and desalted by SEC. ESI-MS (−): m/z calcd for C\textsubscript{113}H\textsubscript{155}N\textsubscript{40}O\textsubscript{69}P\textsubscript{12}S: 3821.5, m/z found: 3821.8 [M − H]−. The azide containing oligonucleotide was next suspended in triethylammonium acetate (TEAA) buffer (0.1 M, pH 6.5). 0.15 volume of 1 M aqueous silver nitrate solution and 0.5 volume of acetic acid were added. The mixture was vortexed and left to react for 45 minutes at room temperature. Afterwards, 0.2 volume of 1 M aqueous DTT solution was added. The mixture was vortexed and left to react for 15 minutes at room temperature. Afterwards, the solution was centrifuged in order to remove the silver–DTT complex and the supernatant was recovered. The precipitate was washed with 3 volumes of TEAA buffer. Finally, the solution was desalted by SEC. ESI-MS (−): m/z calcd for C\textsubscript{113}H\textsubscript{155}N\textsubscript{40}O\textsubscript{69}P\textsubscript{12}S: 3579.63, m/z found: 3579.58 [M − H]−.

Conjugate synthesis

Synthesis of conjugate 5 (oxime ligation). Oligonucleotide 2 (1 equiv., 500 nmol) was dissolved in 0.4 M ammonium acetate buffer (pH 4.5, concentration 10−3 M) and aminooxy peptide 4a (2 equiv., 1 μmol) was added. The solution was stirred at 55 °C for 45 min. The crude mixture was purified by RP-HPLC. Conjugate 5 was desalted using an NAP-25 column and freeze dried (325 nmol, 65% yield). MALDI-TOF-MS (−): m/z calcd for C\textsubscript{173}H\textsubscript{243}N\textsubscript{27}O\textsubscript{68}P\textsubscript{12}S\textsubscript{2}: 4906.3, m/z found: 4906.9 [M − H]−.

Synthesis of conjugate 6 (CuAAC reaction). Conjugate 5 (300 nmol) was dissolved in 100 mM HEPES buffer (pH 7.4, concentration 10−4 M) and CuSO\textsubscript{4} (6 equiv.), THPTA (30 equiv.), and sodium ascorbate (30 equiv.) were added to this solution. The reaction mixture was stirred at 37 °C for 2 h and the reaction was stopped with the addition of 0.5 M ethylenediaminetetraacetic acid (EDTA) solution (50 equiv.). The resulting reaction mixture was desalted using an NAP-25 column and the crude product was purified by RP-HPLC. Conjugate 6 was desalted using an NAP-25 column and freeze dried (216 nmol, 72% yield). ESI-MS (−): m/z calcd for C\textsubscript{173}H\textsubscript{245}N\textsubscript{27}O\textsubscript{68}P\textsubscript{12}SCl: 4906.3, m/z found: 4907.2 [M − H]−.

Synthesis of conjugate 7 (thioether formation). Conjugate 6 (1 equiv., 100 nmol) and oligonucleotide 3 (2 equiv., 200 nmol) were dissolved in a solution of H\textsubscript{2}O/CH\textsubscript{3}CN (9:1, v/v) at 5.10−4 M. TCEP (2 equiv.), 500 mM KCl, DIAE (45 equiv.), and KI (60 equiv.) were added to this solution. The reaction mixture was stirred at room temperature for 3 h and the crude product was purified by RP-HPLC. Conjugate 7 was desalted using an NAP-25 column and freeze dried (74 nmol, 74% yield). ESI-MS (−): m/z calcd for C\textsubscript{286}H\textsubscript{395}N\textsubscript{60}O\textsubscript{153}P\textsubscript{24}S\textsubscript{2}: 8405.2, m/z found: 8403.5 [M − H]−.

Synthesis of conjugate 8 (diazo transfer reaction), way (a). To a solution conjugate 7 (40 nmol, concentration 3.10−4 M) in a solution of 50 mM NaHCO\textsubscript{3}/(MeOH/H\textsubscript{2}O, 1/1) 75/25 at pH 8.5 were added CuSO\textsubscript{4} (2.8 equiv.) and ISAHc (20 equiv.). The reaction mixture was stirred at 60 °C for 30 min and the reaction was quenched with 0.5 M EDTA solution (50 equiv.). The product was purified by RP-HPLC and then desalted and freeze-dried (31.6 nmol,: 79% yield). ESI-MS (−): m/z calcd for C\textsubscript{286}H\textsubscript{393}N\textsubscript{60}O\textsubscript{153}P\textsubscript{24}S\textsubscript{2}: 8430.8, m/z found: 8433.8 [M − H]−.

Synthesis of conjugate 9 (diazo transfer reaction). The same diazo transfer protocol was used from conjugate 6 (100 nmol) to obtain conjugate 9 (50 nmol, 50% yield). MALDI-TOF-MS (−): m/z calcd for C\textsubscript{173}H\textsubscript{241}N\textsubscript{28}O\textsubscript{4}P\textsubscript{12}S\textsubscript{2}Cl: 4932.3, m/z found: 4934.4 [M − H]−.

Synthesis of conjugate 8 (thioether formation), way (b). The same thiol chloroacetamide reaction was carried out from conjugate 9 (40 nmol) and oligonucleotide 3 (80 nmol) to afford conjugate 8 (24 nmol, 60% yield).

Synthesis of conjugate 11 (thioether formation). From the starting material oligonucleotide 10 (1 eq., 220 nmol) and peptide 4b (440 nmol, 2 eq.), the same protocol as for the synthesis of conjugate 7 was carried out excepted that TCEP was added during the course of the reaction and no KCl was added (116 nmol, yield: 53%). ESI-MS (−): m/z calcd for C\textsubscript{183}H\textsubscript{268}N\textsubscript{58}O\textsubscript{86}P\textsubscript{24}S\textsubscript{2}: 5089.46, m/z found: 5089.68 [M − H]−.

Synthesis of conjugate 12 (CuAAC reaction). To a solution of conjugate 11 (1 equiv., 150 nmol) in 50 mM PBS buffer (pH 7.4) were added CuSO\textsubscript{4} (3 equiv.), THPTA (30 equiv.), sodium ascorbate (30 equiv.), 500 mM KCl and DMF (10% vol). The reaction mixture was stirred at 37 °C for 2 h and the reaction was quenched with 0.5 M EDTA solution (50 equiv.). The product was desalted using an NAP-25 column and then purified by RP-HPLC and freeze-dried (60 nmol, yield: 40%).
ESI-MS (−) m/z calcd C\textsubscript{182}H\textsubscript{268}N\textsubscript{58}O\textsubscript{86}P\textsubscript{12}S\textsubscript{2}: 5089.46, m/z found: 5089.62 [M − H\textsuperscript{+}].

Synthesis of conjugate 13 (diazo transfer reaction). From conjugate 12 (1 eq., 90 nmol), the diazo transfer reaction was carried out as for conjugate 6 to afford conjugate 13 in a quantitative yield. ESI-MS (−) m/z calcd C\textsubscript{183}H\textsubscript{267}N\textsubscript{60}O\textsubscript{88}P\textsubscript{12}S\textsubscript{2}: 5116.45, m/z found: 5116.21 [M − H\textsuperscript{+}].

Synthesis of conjugate 14 (CuAAC reaction). From the starting conjugate 13 (1 eq., 70 nmol) and oligonucleotide 2 (1.2 eq., 84 nmol), the same protocol as for the preparation of conjugate 12 was applied (42 nmol, yield: 48%). MALDI TOF-MS (−) m/z calcd C\textsubscript{295}H\textsubscript{411}N\textsubscript{57}O\textsubscript{134}P\textsubscript{26}S\textsubscript{2}: 8617.05, m/z found: 8613.61 [M − H\textsuperscript{+}].

Synthesis of conjugate 1b (oxime ligation). Conjugate 14 (1 equiv., 9 nmol) was suspended in water (350 μL) in a closed screw-cap glass vial. A 2% aqueous TFA solution (350 μL) was added in order to obtain a final solution of TFA 1% vol. The reaction mixture was stirred at room temperature for 1 h. The product was desalted by SEC and then freeze-dried. The yield was considered as quantitative. MALDI TOF-MS (−) m/z calcd C\textsubscript{291}H\textsubscript{405}N\textsubscript{97}O\textsubscript{154}P\textsubscript{24}S\textsubscript{2}: 8528.99, m/z found: 8526.68 [M − H\textsuperscript{+}].

Circular dichroism studies

Analyses were performed on a Jasco J-810 spectropolarimeter using a 1 cm length quartz cuvette. Spectra were recorded at 20 °C or every 5 °C in the range of 5 to 90 °C with a wavelength range of 220 to 330 nm. For each temperature, the spectrum was an average of three scans with a 0.5 s response time, a 1 nm data pitch, a 4 nm bandwidth and a 200 nm min\textsuperscript{−}1 scanning speed. Melting temperatures were obtained using Boltzmann fit on Origin software. Each curve fit was only accepted with an $r_{\text{value}}>0.99$.

Conclusion

In summary, we have designed and assembled a peptide–DNA conjugate capable of folding into a stable i-motif structure. The assembly was performed via the stepwise connection of two i-motif forming DNA strands through four consecutive orthogonal ligation reactions. The sequential synthetic pathway was optimized to provide full control in the assembly of such a sophisticated i-motif forming conjugate. This detailed optimization effort sheds some new light on the concrete hurdles associated with the assembly of such advanced molecular objects. For instance, we found that TC and CuACC reactions were less capable than the OL reaction to achieve the challenging formation of intramolecular covalent bonds within bulky molecular conjugates. We also further demonstrate that a mild acidic treatment can be used for the one-pot aminooxy deprotection and oxime bound formation in the presence of the acid-sensitive heteropolymer DNA sequence.

The resulting conjugate was found by CD to fold, at room temperature, into an i-motif structure at acidic and neutral pH. This stabilization of the secondary structure was obtained without major modifications of a large portion of the native structure. The stabilized i-motif mimic may therefore be used to study, at physiologically relevant pH, the interaction of the i-motif with putative i-motif targeting ligands by using SPR. This unique molecular object might also be used to pull down and further characterize proteins that interact with i-motif structures. In both cases, one could take advantage of the biotin moiety built in the i-motif conjugate for surface immobilization. Finally, such a novel i-motif forming a conjugate with shifted pH sensitivity might be exploited to design original DNA nanodevices with new sets of characteristics and performances.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche (ANR-16-CE11-0006-01). The Nanobio-ICMG platform (FR2607) is acknowledged for providing facilities for the synthesis and purification of oligonucleotides (R. Lartia) as well as for mass spectrometry analyses (A. Durand, L. Fort and R. Guéret).

References


