the FEBS Journal

Solution structure of a truncated anti-MUC1 DNA aptamer determined by mesoscale modeling and NMR

Meriem Baouendi[1], Jean A. H. Cognet[1], Catia S. M. Ferreira[2,3], Sotiris Missailidis[3], Jérome Coutant[4], Martial Piotto[4], Edith Hantz[5] and Catherine Hervé du Penhoat[5]

1 Laboratoire Acides Nucléiques et Biophotonique, Université Pierre et Marie Curie Paris 6, Paris, France
2 Ontario Cancer Institute, University Health Network, Toronto, ON, Canada
3 Chemistry Department, The Open University, Milton Keynes, UK
4 Institut de Chimie, Strasbourg University, Bruker Biospin, Wissembourg, France
5 Laboratoire Chimie, Structure et Propriétés de Biomatériaux et d'Agents Thérapeutiques, Université Paris 13, Bobigny, France

Keywords
anti-MUC1 aptamer; biopolymer chain elasticity; DNA structure; molecular dynamics; NMR solution structure

Correspondence
C. Hervé du Penhoat, Laboratoire Chimie, Structure et Propriétés de Biomatériaux et d'Agents Thérapeutiques, UMR 7244 CNRS, Université Paris 13, UFR SMBH, 74 rue Marcel Cachin, Bobigny 93017, France
Fax: +33 14 83 87 77
Tel: +33 14 83 87 391
E-mail: penhoat8@aol.com

(Received 25 July 2011, revised 13 November 2011, accepted 21 November 2011)

doi:10.1111/j.1742-4658.2011.08440.x

Mucin 1 is a well-established target for the early diagnosis of epithelial cancers. The nucleotides of the S1.3/S2.2 DNA aptamer involved in binding to variable number tandem repeat mucin 1 peptides have been identified using footprinting experiments. The majority of these binding nucleotides are located in the 25-nucleotide variable region of the total aptamer. Imino proton and 2D NMR spectra of truncated and total aptamers in supercooled water reveal common hydrogen-bonding networks and point to a similar secondary structure for this 25-mer sequence alone or embedded within the total aptamer. NMR titration experiments confirm that the TTT triloop structure is the primary binding site and show that the initial structure of the truncated aptamers is conserved upon interaction with variable number tandem repeat peptides. The thermal dependence of the NMR chemical shift data shows that the base-paired nucleotides melt cooperatively at 47 ± 4 °C. The structure of the 25-mer oligonucleotide was determined using a new combined mesoscale molecular modeling, molecular dynamics and NMR spectroscopy investigation. It contains three Watson–Crick pairs, three consecutive mispairs and four Watson–Crick pairs capped by a TTT triloop motif. The 3D model structures (PDB 2L5K) and biopolymer chain elasticity molecular models are consistent with both NMR and long unconstrained molecular dynamics (10 ns) in explicit water, respectively.

Database
Structural data are available in the Protein Data Bank and BioMagResBank databases under the accession numbers 2L5K and 17129, respectively

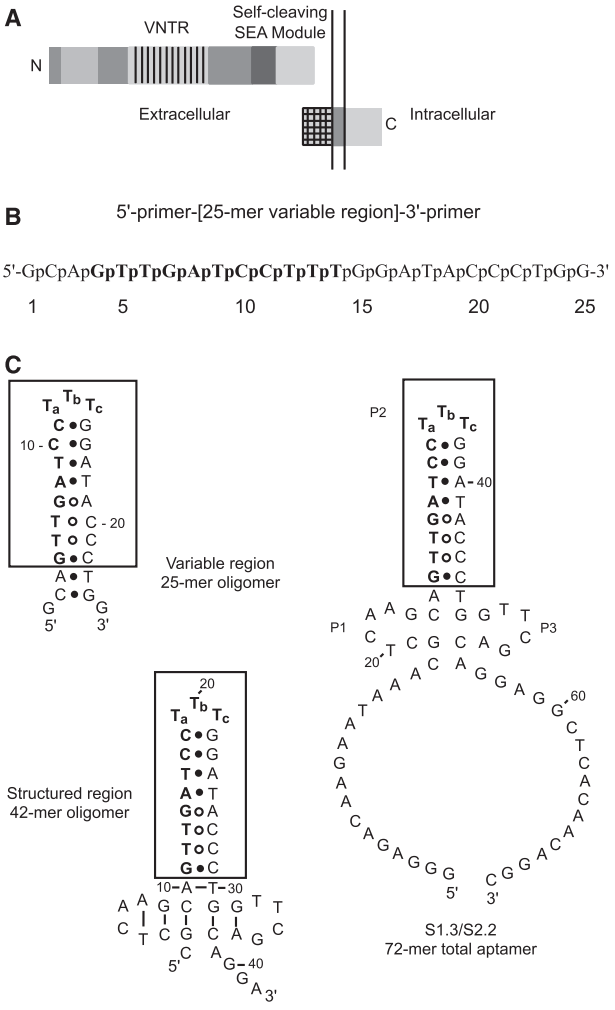
Introduction

Mucin 1 (MUC1), a cell-surface-associated glycoprotein, is overexpressed on the surface of cancer cells [1]. It consists of a high molecular mass glycosylated extracellular domain that includes a variously amplified 20-amino-acid repeat sequence or variable number tandem repeats (VNTR) region and a hydrophobic 31-amino-acid transmembrane domain with a 69-amino-acid cytoplasmic tail (Scheme 1A) [2]. In cancer cells, the VNTR domain is under-glycosylated exposing the VNTR peptides to interactions with molecules in the

Abbreviations
BCE, biopolymer chain elasticity; MD, molecular dynamics; MUC1, mucin 1; ODN, oligodeoxyribonucleotide; SEA, Sperm Enterokinase Agrin (from initial identification in these proteins); VNTR, variable number tandem repeats.

extracellular medium. MUC1 also contains a self-cleaving SEA module (name from initial identification in a Sperm protein, in Enterokinase, and in Agrin) in the extracellular domain. Upon self-cleavage, the resulting extracellular component undergoes internalization, making it a very promising target for therapeutic and diagnostic purposes.

Development of the systematic evolution of ligands by the exponential enrichment process [3,4] leads to the isolation of oligonucleotide sequences with the capacity to recognize target molecules such as peptides or proteins with high affinity and specificity. These oligonucleotide sequences, referred to as 'aptamers', are beginning to emerge as an important class of molecules for both therapeutics and diagnostics [5].



Scheme 1. (A) Domain structure of the mucin 1 glycoprotein, MUC1 [5]. (B) Nucleotide sequence and numbering of the 25-mer DNA oligomer corresponding to the variable region of the 72-mer S1.3/S2.2 total DNA aptamer in the systematic evolution of ligands by the exponential enrichment protocol. (C) NMR-defined secondary structure of the 25-mer and 42-mer truncated aptamers. Canonical and noncanonical base pairs are labeled by filled (•) and unfilled (○) circles, respectively. The putative additional base-pairing (number and type of imino protons in the low-temperature spectra based on chemical shift criteria) for the 42-mer oligodeoxyribonucleotide (ODN) with respect to the 25-mer are indicated with bars.

DNA aptamers that bind to unglycosylated VNTR peptides (a 60-mer peptide corresponding to three repeats and a 9-mer peptide corresponding to the immunogenic region of MUC1) have been reported by Missailidis and co-workers [6]. The aptamer affinity (S1.3/S2.2 K_D = 0.135 nM) was characterized with competition ELISA and SPR spectroscopy. Fluorescence-labeling yielded aptamers that successfully imaged MCF-7 breast cancer cells [6], whereas radiolabeling yielded aptamers with imaging potential [7,8]. A quantitative detection protocol for MUC1 using a three-component DNA hydridization system with quantum dot-labeling based on the S1.3/S2.2 aptamers has also been described recently [9].

Chemical modification of these MUC1 aptamers might lead to many other applications if a better understanding of the aptamer/peptides interactions was achieved. The following questions need to be answered. What are the 3D structures and thermodynamical stabilities of the aptamer, the peptide and the aptamer/peptide complex? The answers define which nucleotides are necessary for folding and for binding and therefore those that can be modified without provoking loss of affinity for the target peptide.

In this study, our main goal was to obtain a detailed description of the atomic interactions that are important for the S1.3/S2.2 MUC1 aptamer folding and binding to the VNTR peptides. Footprinting studies were used to establish the nucleotides important for binding. Secondary structure was obtained from imino proton NMR in supercooled water which reveals the hydrogen-bonding patterns of all species present in solution. Indeed, small interconverting nucleic acid structures (< 100 bp) exchange slowly enough on the NMR time scale (> ms) to lead to distinct sets of signals [10,11]. This strategy was applied to three nucleic acid sequences, namely those of the variable region (nucleic acid sequence given in Scheme 1B), the structured region (sequence encompassing all possible Watson–Crick pairs) and the total aptamer (a 25-mer ODN, a 42-mer ODN and a 72-mer ODN, respectively) as outlined in Scheme 1C. NMR titration experiments with the VNTR peptides confirmed the binding site of the truncated MUC1 aptamers. The combined NMR and mesoscopic molecular modeling approach established previously [12] provided excellent 3D models for the aptamer-binding motif (nucleotides in bold in Scheme 1C).

Results

MUC1 aptamer/VNTR peptide interactions

Footprinting assays have been used to characterize and help elucidate the structural determinants of the interactions between DNA or RNA single-stranded aptamers and targets of interest. This technique takes advantage of the enzymatic cleavage properties of DNAse I (used in our study) to form a unique ladder of digested fragments that identifies the contact point of interactions between the two moieties.

To help elucidate the secondary structure of the 72-mer aptamer, the later was labeled at the 5′-end with ^{32}P and refolded. The aptamer was then allowed to interact with a nonglycosylated five tandem repeat MUC1 peptide for 1 h at 37 °C. The mixture was subsequently subjected to limiting digestion using DNAse I and was analyzed by denaturing PAGE (Fig. 1). The results were analyzed in conjunction with the experimentally determined structures shown in Scheme 1C in order to propose a substructure that encompasses the binding nucleotides. Digestion analysis revealed that most of the nucleotides indicated in bold in Fig. 1 and Scheme 1C (encased in rectangles) are indispensable for aptamer binding and possibly crucial for its high affinity. It should be noted that the nearby 5′-end nucleotides may also be involved in binding.

Imino proton spectra of the truncated 25- and 42-mer sequences, alone and in the presence of MUC1 peptides, were recorded to monitor changes in secondary structure (Fig. 2). The nonexchangeable ^{1}H, ^{13}C and ^{31}P chemical shifts of the unlabeled 25-mer ODN sample at 25 °C were assigned using classical 2D NMR methods [13,14], as described previously [12], and deposited in the BMRB (accession number 17129). The canonical Watson–Crick base-paired imino proton signals of the 25-mer (indicated by filled circles, 12–15 p.p.m. in Fig. 2A) were already visible at moderately low temperature (2 °C) but additional imino proton signals (open circles, and triangles) were much more intense in supercooled water [15] where chemical exchange of labile protons with the solvent was much slower (Fig. 2B, −8.5 °C). All exchangeable proton signals were assigned on the basis of NOESY (τ_m 50 ms) data at −8.5 °C (Fig. S1). The broad signal at 11.05 p.p.m. in Fig. 2B was assigned to the thymine H3 protons in the $-T_aT_bT_c-$ triloop through relayed (spin-diffusion) NOEs to H3′(T_a) and/or H3′(T_c). The chemical shift of the H1(G7) imino signal was characteristic of an imino G•A base pair where pairing occurs through Watson–Crick edges [16] (type VIII) [17] as opposed to a sheared G•A base pair [18] (type XI) [17]; the H1(G7)/H2(A19) NOE corroborated this assignment.

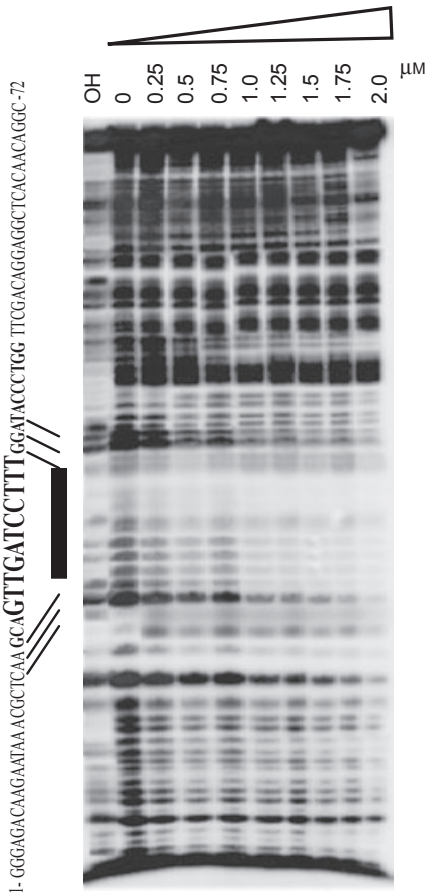


Fig. 1. Autoradiogram of a 18% polyacrylamide/8 M urea gel, showing digestion products of 5′-end-labeled S1.3/S2.2 with DNase I in the presence (0.25–2 μM) or absence (0) of unglycosylated peptide MUC1 5TR. The major protected area is shown by a vertical line. The highlighted DNA bases that were found to interact with the unglycosylated MUC1 peptide containing five tandem repeats are located mostly in the P2 region of S1.3/S2.2 labeled in Scheme 1C.

The remaining two imino signals of the 25-mer ODN in Fig. 2B were tentatively assigned to water-mediated T5•C21w and T6•C20w mispairs (11.98 and 11.82 p.p.m., respectively, labeled with an open circle) because they lie between the range reported for unpaired thymine H3 (10.0–12.0 p.p.m.) [11] and that of thymine H3 engaged in a T–C mismatch (12.66 p.p.m. at 10 °C) [19]. A similar chemical shift value (11.5 p.p.m. at 10 °C) has been reported previously [20] for a water-mediated T–C mismatch. Weak spin-diffusion NOESY cross-peaks in Fig. S1 (detailed in the legend) corroborated these assignments. Moreover, comparison of the normalized intensities of the cross-peaks of the exchangeable protons (ω_2) at the

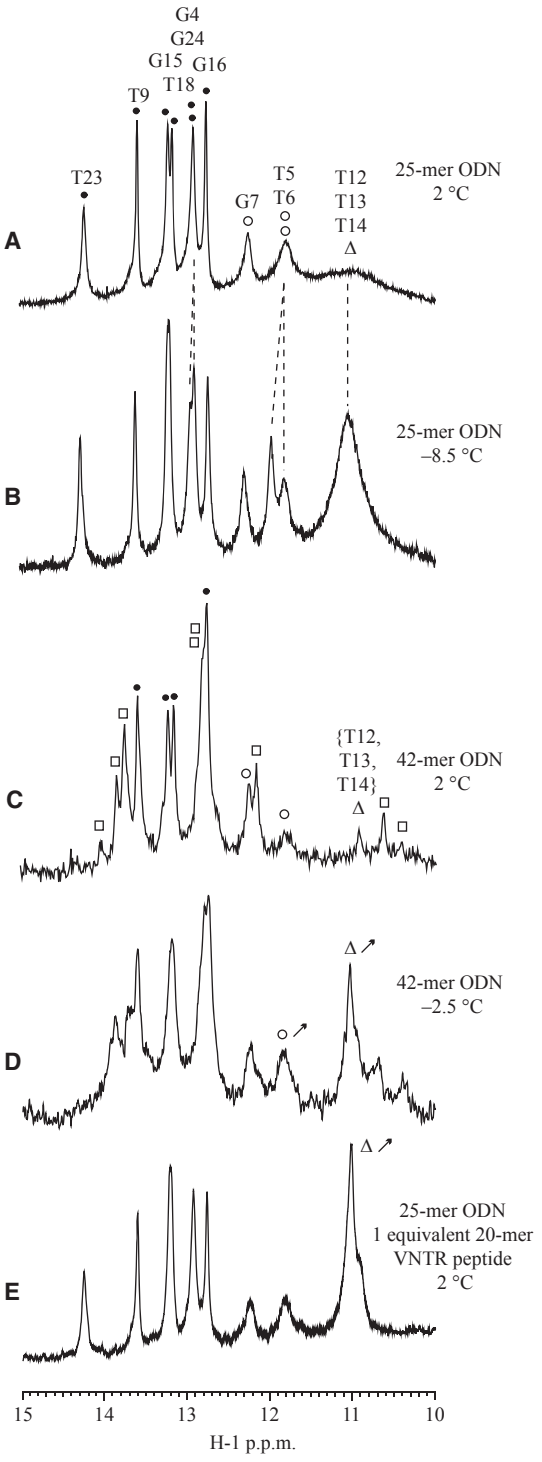


Fig. 2. Imino proton 1D NMR spectra of truncated anti-MUC1 aptamers at low temperature (> 0 °C) and in supercooled water (< 0 °C): the 25-mer at 2 °C (A) and −8.5 °C (B), the 42-mer at 2 °C (C) and −2.5 °C (D), and the 25-mer in the presence of 1 equivalent of the 20-mer repeating unit of the VNTR domain at 2 °C (E). The assignments of the imino proton signals of the 25-mer are indicated in (A). The following symbols are used: canonical base pairs (•), assigned mispairs (○), TTT triloop H3 (Δ) and additional unassigned base pairs (□). Increased signal intensity at lower temperature (C and D) or in the presence of VNTR peptide (A and E) is indicated by an upward arrow.

water frequency [$\omega_1(H_2O)$] in the 2D NOESY spectrum [21,22] of the 25-mer ODN in supercooled water (Fig. S1) showed some protection of H3(T5) and H3(T6) from exchange with the solvent, as would be expected in the case of weak hydrogen bonding. It was concluded that the 25-mer ODN formed seven canonical and three noncanonical base pairs labeled, respectively, by filled (•) and open (○) circles in Scheme 1C.

Comparison of the chemical shifts of the imino signals of the 42-mer ODN in Fig. 2C with those of the 25-mer ODN in Fig. 2A reveals a strong correlation for the triloop–stem region (R = 0.9997) that appears to extend to at least two more noncanonical base pairs (indicated with an open circle at 12.3 and 11.8 p.p.m. in Fig. 2C). This includes almost all of the S1.3/S2.2 aptamer nucleotides important for MUC1 peptide affinity as revealed by the footprinting study (bold and encased in rectangles in Scheme 1C). Several new signals were also detected in Fig. 2C (labeled with an open square). Their chemical shift analysis points to the 42-mer base-pairing that has been indicated with bars in Scheme 1C. As was observed with the 25-mer ODN, the intensity of the signals at 11.8 and 11.0 p.p.m. increased significantly at lower temperature (Fig. 2D).

The signals in the imino proton spectra of the 72-mer S1.3/2.2 aptamer (the 600 MHz spectrum at −9.5 °C is given in Fig. S2) were very broad. However, the most intense signal was centered at 11.0 p.p.m., suggesting that the 72-mer aptamer contained a similar TTT triloop structure to that of the truncated 25- and 42-mer aptamers.

During titration of the 25-mer ODN with the 20-mer VNTR peptide the intensity of the –TTT– triloop H3 signals increased dramatically pointing to considerable protection of these labile protons from exchange with the solvent (Fig. 2A,E). It could be surmised that the 25-mer ODN/20-mer VNTR peptide interaction site involved the P2 stem–loop region (Scheme 1C) in agreement with the footprinting results. Chemical shift mapping of the 25- and the 42-mer proton signals upon titration with the VNTR peptides did not reveal significant variations (i.e. $\Delta\delta < 0.025$ p.p.m.).

Conformational analysis of the 25-mer MUC1 aptamer with NMR

The majority of the NMR data (^{13}C [23] and ^{31}P [24] chemical shifts, intra-residue and sequential NOEs [13], proton–proton and phosphorous–proton coupling constants [14]) pointed to B-DNA conformation. Characteristic sequential NOEs are presented in Scheme S1. Southern (C1′-exo or C2′-endo) sugar pucker, the anti orientation of the χ torsion angle, and both *t* and g^+ orientations of β and γ (inferred from moderately intense H4′(n)/P(n) cross-peaks) [25] could be deduced for most nucleotides. Very broad P(G15)/H5″(G15) and strong H4′(G15)/H5″(G15) cross-peaks in the [^{31}P,^{1}H]-COSY and DQCOSY spectra, respectively, indicated that the γ(G15) adopted the *t* orientation. ^{31}P chemical shifts allowed exclusion of the (*t*) domain of many of the ζ and α torsion angles, as such conformations are associated with downfield phosphorus chemical shifts (such downfield shifts were detected for the C2, T12, and T18 ^{31}P signals) [24]. Finally, hydrogen-bond constraints were included for the Watson–Crick and imino G•A mismatch base pairs based on interstrand NOESY cross-peaks (Fig. S1).

Model structures were obtained by simulated annealing under the NMR constraints with DYANA [26] and the one that best satisfies the NMR constraints has been given in Fig. 3A (model in black). The statistics for the final DYANA simulations and the corresponding average torsion angle values of the NMR-defined model structures are available in Tables S1 and S2. The most striking features of the DYANA structures can be summarized as follows: (a) T12 is folded into the minor groove and the O2(T12)–H22(G15) and O4(T12)–H22(G16) distances are < 3.2 and 4.1 Å, respectively, in all the models; (b) T13 is stacked on the C11–G15 base pair; (c) T14 is partially stacked on T13; (d) G7 and A19 form an unusual imino G•A pair; (e) T5/C21 and T6/C20 are well-positioned to form mispairs inspite of the lack of T•C hydrogen-bond constraints in the DYANA simulations; and (f) all of the thymine methyl groups lie above the aromatic ring of the base of the preceding nucleotide with the exception of CH_3(T12). It is to be noted that such favorable thymidine methyl/π interactions have been found in numerous crystallographic structures [28].

Thermodynamic analysis of the 25-mer folding by NMR

NMR chemical shifts can be used to monitor the hairpin to unfolded single-strand transition if these species

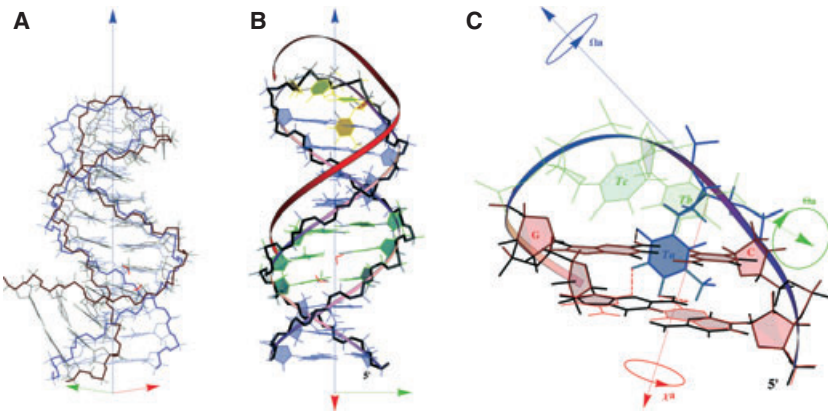


Fig. 3. (A) Superposition of the AMBER-minimized DYANA model (black) that best satisfies the NMR constraints with the BCEmin model (blue) and its two water molecules (red). The two models (23-mers) are superposed on the central segment d(TTGATC).d(GATACC) to show the small bending of the loop into the minor groove, and the larger bending that occurs at the level of the T•C mispairs. (B) Representation of the BCEmin model with the three thymines in the loop (yellow), the Watson–Crick base pairs (blue), the three mispairs (green) and the two bridging water molecules (red) in the minor groove of the T•C mispairs. The reference frame (*O*, *x*, *y*, *z*) was defined for the loop according to the Cambridge conventions [27]. The x and y vectors (red and green, respectively) are 10 Å long, and the z vector (blue) is 50 Å. The loop is viewed from its major groove ($x > 0$), and Ta is in the background (minor groove, i.e. $x < 0$), whereas the mispairs are viewed from their minor grooves. The ribbon (pink) is computed from the theoretical optimal elastic line. Note the similar widths of Watson–Crick and T•C mispairs and the small enlargement due to the G•A mispair. The thick red ribbon indicates the T5 to T14 binding region of the aptamer determined by footprinting experiments. (C) Stabilization of Ta in the minor groove by three hydrogen bonds as a function of the sequence of the two canonical helical B-DNA base pairs (gray cycles) below the loop WC-TaTbTc-GZ: the molecular model features base pair W•Z as G•C (red) or as C•G (plain black), the BCE optimal loop backbone and elastic line (inside red and outside blue), thymine Ta (blue, in the minor groove), and thymines Tb and Tc (green wire, in the major groove). Note that H bonds (dashed red) H3(Ta)/O2 (where Z is C) [23] coincides here with H3(Ta)/N3 (where Z is G16).

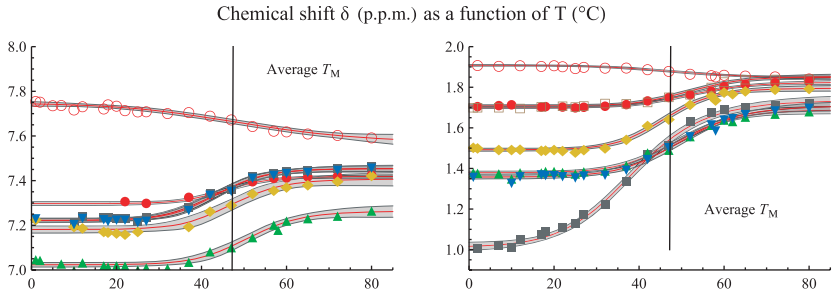


Fig. 4. Thermal behavior of thymine H6 (left) and CH_3 (right) chemical shifts of the 25-mer aptamer based on 1D proton and 2D TOCSY spectra. Data points are plotted with the fits as well as with 95% confidence levels. The melting temperature of the hairpin to single-strand transition, T_M, is indicated by a vertical line. T5, dark gray filled square; T6, yellow filled diamond; T9, blue filled downward triangle; T12, red empty circle; T13, brown empty square; T18, green filled upward triangle; T23, red filled circle. The melting temperature was 47 (±4) °C for all thymine H6 protons. Because T12 is not stacked in the loop, the chemical shift of this aromatic proton varies the least as a function of temperature.

undergo fast exchange over the entire temperature range surrounding the transition. The condition for fast exchange is that the exchange rate be much larger than the difference in the NMR frequencies of the exchanging species [$k_{ex} >> 2\pi(\nu_A - \nu_B)$] [29,30]. The fast exchange regime has been demonstrated for the T6 methyl signal in the mispair, the T12 methyl signal in the loop, and the T18 H6 resonance in the stem (Fig. S3). The H6 (left) and CH_3 (right) chemical shifts of all thymines in the 25-mer aptamer over the 0–80 °C range are displayed in Fig. 4. It has been shown that the latter signals are very direct markers of stacking in nucleic acids because their shift is affected by the 5′-neighboring base [31].

Standard analysis of the UV data as a function of temperature provides the melting temperature, T_M, directly from the numerical derivative (here 38 °C for the 25-mer) (see Annex S1 in the Supporting information) [32,33]. NMR data cannot provide as extensive data as UV: 22 selected data points were recorded by NMR over the 80 °C range. However, within the van't Hoff two-state model assumption, chemical shift data should not vary with temperature in the absence of conformational or chemical exchange.

Within the estimated uncertainty T_M is 47 ± 4 °C for all thymine protons and the corresponding thermodynamic data are collected in Table S3. The most important feature obtained from the data in the different plots in Fig. 4 is that ΔH^0_{vH} is approximately constant throughout the entire oligonucleotide. T5, where ΔH^0_{vH}, ΔS^0_{vH} and T_M, are the smallest, is an exception and it is therefore the most unstable base pair of the entire oligonucleotide. The similar transition midpoint for almost all of the thymines implied that the whole aptamer melts mostly cooperatively with a single transition point, thereby justifying in part the two-state van't Hoff hypothesis.

Stability of the 25-mer folding as probed by BCE, molecular mechanics and molecular dynamics

Construction of the 25-mer oligonucleotide with the biopolymer chain elasticity (BCE) approach [34,35], Fig. 3B, attains three goals difficult to accomplish with other methods: (a) the global minimum of the loop with ideal B-DNA end-point conditions for the stem-loop junction (BCEori), (b) mesoscopic orientations of nucleotides in the loop reference frame through independent angles (Ω, Δχ and θ) to match optimally the best NMR-defined DYANA model (BCEopt_nmr), and (c) the loop with the set of least deformed B-DNA torsion angles in accordance with the NMR-defined ones that also correspond to a local minimum (BCEmin, blue in Fig. 3A) in the AMBER force field [36]. The BCE construction of the $-T_aT_bT_c-$ triloop has already been described elsewhere [34] and was generalized to all known sequences during the course of this study (Tables S4 and S5). The reference frame for the three rotation angles used to construct nucleic acid sequences, Ω_i, χ_i and θ_i with the BCE approach is given in Fig. 3C. The value of the Ω_i rotation angle determines nucleotide position with respect to the global nucleic acid structure. When a nucleotide is in the minor groove, $\Omega < 0$, and when it is in the major groove, $\Omega > 0$. In the case of T_a where $\Omega < 0$, the rotation with angle θ corresponds to a positive swinging angle of the nucleotide in the minor groove from strand II to strand I as shown in Fig. 3C. Below, only the results for the construction of the three consecutive mispairs will be described in detail.

Mispair G•A may be one of the widest with a C1′–C1′ distance of 12.8 Å, whereas T•C may be one of the smallest with a C1′–C1′ distance of 9.2 Å [19]. Assembling such structurally different mispairs in a manner that allows for favorable stacking interactions

would be difficult. The mispair T•Cw, which includes a water molecule, has been characterized with a C1′–C1′ distance of 11.2 Å [19] close to Watson–Crick base pairs (10.8 Å). The requirements for minimizing the deformations of the sugar–phosphate backbone and also for optimizing stacking, and our NMR observations that the two T•C mispair H3 exchange rapidly with water at −11.9 °C (Fig. S1), argue in favor of T•Cw mispairs.

For construction of the T•Cw mispairs, it is sufficent to rotate each base, initially cast in helical B-DNA, outwards by 6.0° about an axis parallel to helical axis and passing through the midpoint of C5′ and C3′ of each pyrimidine. These two opposite rotations are sufficient to bring together atoms O4(T)/H42(C) in the major groove to form a hydrogen bond and to open the minor groove in order to form a water-mediated hydrogen bond between H3(T) and N3(C). For the G•A mispair, it is sufficent to translate each base, initially cast in helical B-DNA, outwards by 1.0 Å along a line parallel to N1–C4 of each purine. These opposite translations are sufficient to form bifurcated hydrogen bond H22(G)/N1(A), and both H1(G)/N1(A) and O6(G)/H62(A) hydrogen bonds. These very small rotations or translations from the Watson–Crick base pair geometry, i.e. from small deformation of the backbone, partially account for the unexpected stability of these three mispairs stacked together in helical B-DNA (Fig. 3B).

The molecular dynamics (MD) simulations, initiated with the BCEmin conformation, were performed without constraints. All torsion angles and the orientations of the triloop–stem junction nucleotides (–C10–C11–T12–T13–T14–G15–G16–) were observed to be remarkably stable throughout the 10-ns MD trajectory despite the presence of three consecutive mispairs (Fig. 5; see Fig. S4 for the mispaired nucleotides). The second most surprising feature is the conformational stability of

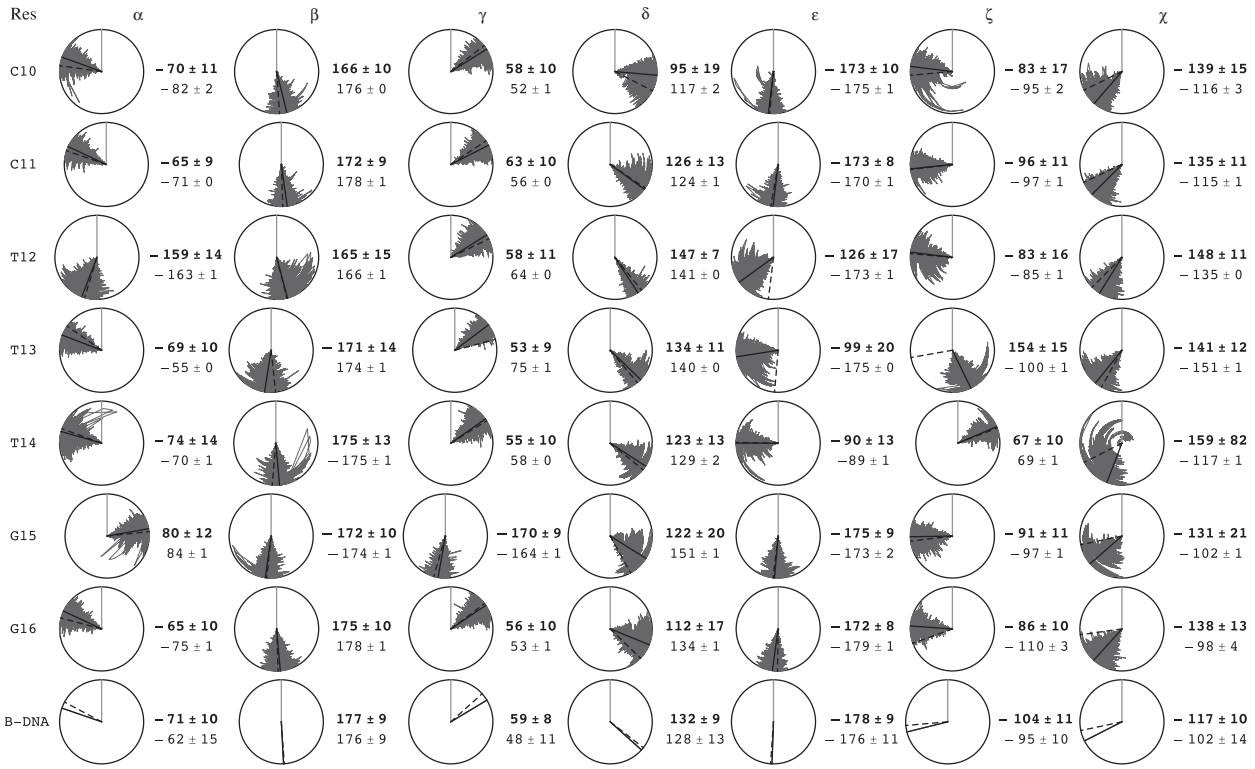


Fig. 5. Torsion angles visited by the stem–loop nucleotides (C10–G16) of the BCE model of the 25-mer, BCEmin, during the 2500 ps production period of free molecular dynamics trajectory simulation at 25 °C in explicit water with the AMBER force field [36]. In the dial frames, backbone and glycosidic torsion angle values of the nucleotides increase clockwise with 0 at the top of the dials. Time increases from the center of the circumference and the detailed trajectories are shown in gray. Average values over the molecular dynamics are depicted by a thick black line and values and standard deviation are in bold, whereas NMR/DYANA-derived values are indicated by a dashed black line and printed below. (Bottom) Continuous lines and values in bold are from molecular dynamics simulations and molecular modeling of NMR-derived data [37], dashed lines are mean values and SD of high-resolution (< 1.9 Å) crystal structures of B-DNA with bimodal distributions BIr and BIIr ($\varepsilon = -114 \pm 15$; $\zeta = 174 \pm 14$) [38].

thymines, T_a (T12) and T_b (T13). T_a is anchored in the minor groove to the two base pairs below the loop by three very stable hydrogen bonds. T_a is laid down very compactly in the minor groove against the two base pairs edges and the 5′ strand (Fig. 3C). Although not hydrogen bonded, T_b remains well stacked on C11. T_c was initially partially stacked on T_b, but transition of $\chi(T_c)$ (Fig. 5) led to unstacking and rotation into the solvent, as observed previously by NMR for another TTT triloop–stem sequence [23].

The global mesoscopic model and the MD trajectories reveal several stabilizing interactions in the TTT triloop region (Fig. 3C). Three hydrogen bonds can be predicted from the model, O2(T_a)–H22(G15), H3(T_a)–N3(G16) and O4(T_a)–H22(G16) are seen to be very stable (2.1 ± 0.2, 2.4 ± 0.3 and 2.0 ± 0.2 Å, respectively) during the entire duration of the 10-ns MD trajectory. These values are very close to the C11•G15 or C10•G16 Watson–Crick hydrogen-bond lengths observed during this MD simulation [O2(C)–H21(G)] (Table S6).

Some additional findings from MD simulations are as follows: (a) the C2pA3 and T18pA19 steps revert from B_I to B_{II} conformation in the AMBER force field [36]; (b) the stability of the mispaired region during the 10-ns MD trajectory is in agreement with cooperative folding as revealed by the chemical shift titrations as a function of temperature; (c) all thymine methyls interact with the closest aromatic ring of the preceding nucleotide as observed by NMR with the exception of T14 (Table S6); and (d) despite apparent g^-,g^+ conformation that is not favored for ε(T14), ζ(T14) (−90°, +68°), the agreement between NMR and the MD conformation is excellent and corroborates the exceptional stability of the fold at the sharp turn [ε(T14), ζ(T14), α(G15), and γ(G15)]. This conformation corresponds to a local minimum, as shown previously [12].

Finally, bending at the first mispair T5•C21 is observed in the 10-ns MD at 25 °C, as with the best NMR structure (Fig. 3A). This phenomenon is correlated to extended or water-mediated hydrogen bonds and it is attenuated in the MD trajectory at 0 °C (Table S6). The typical distance for such bonds is 3.7 Å.

Although there are still few systematic thermodynamics studies of TTT DNA hairpins [39] it is well-documented that T_M can attain unusually high values (70 °C). Calculation of T_M for the NMR-defined secondary structure of the 25-mer ODN (Scheme 1C) using thermodynamic parameters established by SantaLucia and Hicks [39] for mispairs within the stem, and for terminal mispair computed as dangling ends and assuming for the TTT loop that $\Delta G^0 = -1.0$, $\Delta H^0 = -1.0$ kcal·mol^{-1} at 37 °C [$\Delta S^0 = -(\Delta G - \Delta H)/T = 0$], leads to a very similar value (48 °C) to those in Table S3. However, the computed, ΔS^0_{vH} and ΔH^0_{vH} values are too large.

Discussion

The 3D structure and thermodynamical stability of the aptamer-binding motif have been successfully characterized using several methods. Footprinting has confirmed that the nucleotides of the 72-mer anti-MUC1 aptamer (P2 in Scheme 1C) that bind to MUC1 VNTR peptides are located in the 25-mer variable sequence optimized during the systematic evolution of ligands by the exponential enrichment process. Additional nucleotides at the 5′-end (P1 in Scheme 1C) may also play a role in binding. Comparative imino proton and homonuclear 2D spectra in supercooled water have shown that the P2 structural motif is common to both truncated and total 72-mer S1.3/2.2 MUC1 aptamers.

Titration of the truncated aptamers with MUC1 peptides indicates that this binding scaffold does not evolve during initial contact with the MUC1 peptides. The stability of the P2 TTT triloop–stem has been demonstrated by thermodynamic analysis. Surprisingly, the NMR titration experiments (i.e. the lack of chemical shift variations) point to much lower dissociation constants (mM–μM range) than those established for the total 72-mer S1.3/2.2 MUC1 aptamer with binding (sub-nM) and biological assays. A similar increase in K_D (μM range) has been reported for other truncated S1.3/2.2 MUC1 aptamers that include the nucleotide sequence of the variable region and a 3′-extension [9]. These discrepancies could be accounted for by a complex sandwich-type of binding mode simultaneously involving more than one interaction. The topologies of both the VNTR domain (up to 100 repeats) and the structured region of the total aptamer (minor and major groove binding sites at the TTT triloop–stem and other sites that could result from tertiary interactions between the various stem–loop motifs) constitute suitable molecular scaffolds for multiple binding sites.

The –$T_aT_bT_c$– hairpin motif binds strongly to other biologically important peptides and proteins such as actinomycin [40] and adenoassociated virus 2, a popular gene therapy vector [23]. The former complex corresponds to minor-groove binding. In the latter case, it has been suggested that the Watson–Crick edges of T_b and T_c that face into the major groove form a suitable scaffold for protein binding.

In this study, 3D models of a 25-mer DNA sequence have been established using a combined BCE approach and NMR (DYANA, PDB 2L5K). The models are built *a priori* using the BCE approach, and then modified at the mesoscopic level and the torsion angle level accord-

ing to NMR observations. This strategy has provided a starting structure that was stable throughout a 10-ns MD trajectory in the absence of any constraints. It has proved to be successful for DNA hairpins with and without mispairs and should apply to all hairpins. Because several model structures of the MUC1 VNTR peptides have been reported this investigation paves the way to the study of MUC1 aptamer/peptide interactions *in silico*.

Materials and methods

Samples

The 25-mer nucleic acid sequence comprising the variable region of the total aptamer with substituents at G1–O5′–R1 (R1 = O2P–O–*n*-hexyl–NH2) and G25–O3′–R2 (R2 = 3′ or inverted thymine) was obtained by solid-phase phosphoramidite synthesis. The molecular mass and purity were verified by MALDI-TOF MS and reverse-phase chromatography, respectively. The 42-mer ODN (Integrated DNA Technologies, Leuven, Belgium; Eurogentec, Liège, Belgium) and the full 72-mer aptamer (Eurogentec, Liège, Belgium; MWG-Eurofins, Ebersberg, Germany) were obtained from commercial suppliers. The 9-mer peptide containing the immunogenic epitope and the repeating unit of the VNTR domain of MUC1, respectively, were prepared by the Oligonucleotide and Peptide Synthesis Unit of the University of Nottingham, Nottingham, UK.

DNase I footprinting study of the total aptamer

Approximately 4 pmol of aptamer was 5′-end-labeled using 5 pmol of [^{32}P]ATP[γP] (6000 Ci·mmol^{-1}; Amersham) with T4 polynucleotide kinase (New England Biolabs). The aptamer was purified with QIAquick Nucleotide removal kit. Approximately 150 000 cpm of the labeled probe was used in each reaction mixture for the footprinting experiment. Between 0.25 and 2 μM of the 5TR-MUC1 was prepared in binding buffer. MUC1–aptamer complexes were incubated in 50 μL of binding buffer (100 mM NaCl, 5 mM $MgCl_2$, 0.1 mM $CaCl_2$, 10 mM KCl) for 30 min at 37 °C. DNase I digestion was carried out by the addition of 1 μL of DNase I (10 U·μL^{-1}) in 1 × binding buffer for 1 min at 37 °C. The reaction was stopped by the samples being extracted once with an equal volume of phenol-chloroform (1 : 1), ethanol precipitated and resuspended in 5 μL of sequencing formamide loading buffer. After denaturation at 80 °C for 5 min, samples were subjected to electrophoresis on 15% polyacrylamide–8 M urea gels at 1100 V and autoradiographed. A partial alkaline hydrolysis ladder of the same labeled DNA probe was run in parallel on the same gel. The gel was dried onto a blotting paper and was further exposed to a storage phosphor imaging screen (GE, Amersham, Salem, Massachusetts, USA) for 1 h, and scanned on a GE Typhoon imager, using the TYPHOON CONTROL Software (Amersham (GE), Salem, Massachusetts, USA).

UV spectroscopy of the 25-mer ODN

Absorbance versus temperature cooling and heating curves were obtained using a UVIKON 942 spectrophotometer from 6 to 80 °C. The temperature of the bath was increased or decreased at a rate of 10 °C·h^{-1}, thus allowing complete thermal equilibrium of the cells. The melting point was determined from the maximum of the temperature dependence of the first derivative.

NMR spectroscopy

Three types of NMR samples were prepared with the DNA oligomers dissolved in 20 mM phosphate buffer at pH 6.5 containing 50 mM NaCl and 5 μM NaN_3: (a) 300 μL of a 2 mM ODN sample in 100% D_2O in a Shigemi tube, (b) 300 μL of a 1 mM sample in a 90/10 mixture of H_2O and D_2O in a Shigemi tube, and (c) 250 μL of samples (3 mM 25-mer, 3 mM 42-mer and 0.3 mM 72-mer, respectively) in a 90/10 mixture of H_2O and D_2O in capillary tubes (Wilmad, 1.7 mm o.d.) [15] placed in a 5 mm NMR tube.

NMR measurements (^{1}H, ^{13}C and ^{31}P) were performed using either a Unityplus 500 or Advance 500 or 600 MHz NMR spectrometers. The spectra were processed and analyzed using standard spectrometer software or the programs NMRPIPE [41], PROSA [42] and XEASY [43]. Chemical shifts were indirectly referenced to the methyl signal of 5 mM 4,4-dimethyl-4-silapentane-1-sulfonic acid in D_2O using the IUPAC conversion factors for ^{13}C and ^{31}P available at the BMRB website. Data were acquired at several temperatures (4 °C intervals on average) between −11.9 and 80 °C.

NMR distance and torsion angle constraints for the 25-mer ODN have been deposited in the PDB (accession number 2L5K) Conservative hydrogen-bond restraints were included in the structure calculations for the G•C pairs: a upl of 2.25 Å for the acceptor distance, and of 3.25 Å for the donor distance, respectively. Constraints for the type VIII symmetric heteropurine imino G•A base pair [17] (H1(G7)/N1(A19) and H61(A19)/O6(G7) hydrogen bonds, as above) were also included. The G1 and G25 dangling end nucleotides were omitted in the DYANA [27] simulations.

Thermodynamic analysis of the hairpin to unfolded single-strand transition of the 25-mer ODN

The melting temperature, T_M, and the van't Hoff thermodynamic parameters, ΔH^0_{vH} and ΔS^0_{vH} were determined with

the MATHEMATICA software as follows. Variations of the chemical shifts (δ) of H6 and of the methyl protons of all the thymines as a function of temperature in the range between 1 and 80 °C were directly fitted to the function:

$$\delta[T] = \delta_2 + \frac{\delta_1 - \delta_2}{1 + e^{\frac{-\Delta H^0_{vH}}{R}\left(\frac{1}{T} - \frac{1}{T_M}\right)}} \quad (1)$$

where δ1 and δ2 are the mean values of the plateaux respectively before and after the melting transition, T_M and ΔH^0_{vH} are two parameters provided by the fit to the experimental data, and R is the molar gas constant. The line-widths of these signals were monitored over the entire temperature range to ensure that the fast exchange regime applied.

Molecular modeling with the BCE approach

Construction of the stem–loop DNA sequences by bending canonical B-DNA onto an elastic line with the BCE approach has been described previously [12,34,35] (mesoscopic parameters are given Table S4). Ω is the angle value required to orient each nucleotide about the tangent to the elastic line that models the sugar–phosphate backbone at the global scale in the BCE approach. Δχ is the algebraic glycosidic angle variation with respect to the canonical value (reference value $\chi = -117°$). θ is the rotation angle about the axis defined as follows by a point and a vector. The point is on the elastic line with curvilinear abscissa half way between that of C5′ and O3′. The vector is perpendicular to the tangent to the elastic line and to the glycosidic vector.

The construction of the mispairs was achieved along the lines of a geometric method for which the Object Command Language was developed [44]. We proceeded with the same method, only with the MATHEMATICA language that enables the user to easily construct molecular objects and their associated reference frames, to rotate and translate them with respect to the whole molecule reference frame, or with respect to each other. As this NMR investigation and previous studies suggested that one of the two hydrogen bonds between T and C is mediated by a water molecule in DNA [19,37] hydrated T•Cw mispairs were built in the aptamer BCE model.

AMBER energy refinement

The 20 molecular models of the DYANA ensemble and the BCE model BCE_opt_nmr were energy refined with AMBER [36] without restraints with a distance-dependent dielectric constant until the rms energy gradient was < 0.05 kcal·(mol·Å)$^{-1}$. Models that were refined without constraints have been deposited in the PDB (20 models, PDB 2L5K). BCEopt_nmr was refined with restraints as follows. The force constant was set equal to 900 kcal·(mol rad^2)$^{-1}$. In order to maintain the C2′-*endo* conformation of all sugar puckers, torsion angles, $\delta_{i,0}$, (C5′–C4′–C3′–O3′) were forced to nominal value, $\delta_{i,0} = 144°$. As indicated by NMR, the torsion angles α of T12, the (ε,ζ) of T14 and (α,γ) of G15 of the conformation generated above were modified with AMBER to values (−160°), (−90°, 80°) and (80°, −160°), respectively. Similarly, the ^{31}P NMR data pointed to the B_{II} conformation at the C2pA3 and T18pA19 steps so that (ε, ζ) torsions of these nucleotides were also explored. All restrained structures were subsequently relaxed without restraints. This procedure afforded the energy-minimized BCE model used to initiate the molecular dynamics trajectory (Cartesian coordinates available upon request from jean.cognet@upmc.fr).

AMBER molecular dynamics trajectories

Molecular dynamics simulations were performed as previously described [12] using the AMBER 8 package [36]. The aptamer with the two water molecules bound to the T•Cw mispairs, was placed in a box that contained 3634 TIP3P water molecules (corresponding to a 12-Å hydration shell), 25 K^+ and 3 Cl^- ions (corresponding to a concentration close to 0.25 M of added KCl). The final MD trajectories were implemented without restraints. The equilibrium phase lasted 400 ps after which 10 ns MD trajectories were generated at 25 °C followed by 10 ns at 0 °C. Molecular structures were recorded every 0.5 ps for analysis.

Acknowledgements

EH and CH du P acknowledge access to NMR spectrometers at Paris 5 University and funding from Paris 13 University. Professor Jean Gariepy is kindly acknowledged for the provision of materials and facilities for the footprinting work. SM acknowledges financial support from the Open University.

References

1 Taylor-Papadimitriou J, Burchell J, Miles DW & Dalziel M (1999) MUC1 and cancer. *Biochim Biophys Acta* **1455**, 301–313.
2 Levitin F, Stern O, Weiss M, Gil-Henn C, Ziv R, Prokocimer Z, Smorodinsky NI, Rubinstein DB & Wreschner DH (2005) The MUC1 SEA module is a self-cleaving domain. *J Biol Chem* **280**, 33374–33386.
3 Tuerk C & Gold L (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.
4 Ellington A & Szostak JW (1990) *In vitro* selection of RNA molecules that bind to specific ligands. *Nature* **346**, 818–822.

5 Nimjee SM, Rusconi CP & Sullenger BA (2005) Aptamers: an emerging class of therapeutics. *Annu Rev Med* **56**, 555–583.
6 Ferreira CSM, Matthews CS & Missailidis S (2006) DNA aptamers that bind to MUC1 tumour marker: design and characterization of MUC1-binding single-stranded DNA aptamers. *Tumor Biol* **27**, 289–301.
7 Borbas KE, Ferreira CSM, Perkins A, Bruce JI & Missailidis S (2007) Design and synthesis of mono- and multimeric targeted radiopharmaceuticals based on novel cyclen ligands coupled to anti-MUC1 aptamers for the diagnostic imaging and targeted radiotherapy of cancer. *J Bioconj Chem* **18**, 1205–1212.
8 DaPieve C, Perkins AC & Missailidis S (2009) Anti-MUC1 aptamers: radiolabelling with ^{99m}Tc and biodistribution in MCF-7 tumour-bearing mice. *Nuclear Med Biol* **36**, 703–710.
9 Cheng AKH, Su H, Wang YA & Yu HZ (2009) Aptamer-based detection of epithelial tumor marker mucin 1 with quantum dot-based fluorescence readout. *Anal Chem* **81**, 6130–6139.
10 Hobartner C & Micura R (2003) Bistable secondary structures of small RNAs and their structural probing by comparative imino proton NMR spectroscopy. *J Mol Biol* **325**, 421–431.
11 Varani G, Aboul-ela F & Allain FHT (1996) NMR investigation of RNA structure. *Progr NMR Spectrosc* **29**, 51–127.
12 Santini G, Cognet JAH, Xu D, Singarapu K & Hervé du Penhoat C (2009) Nucleic acid folding determined by mesoscale modeling and NMR spectroscopy: solution structure of d(GCGAAAGC). *J Phys Chem B* **113**, 6881–6893.
13 Wüthrich K. (1986) *NMR of Proteins and Nucleic Acids*, pp. 203–255. Wiley, New York.
14 Kim SG, Lin LJ & Reid BR (1992) Determination of nucleic acid backbone conformation by ^{1}H NMR. *Biochemistry* **31**, 3564–3574.
15 Skalicky JJ, Sukumaran DK, Mills JL & Szyperski T (2000) Towards structural biology in supercooled water. *J Am Chem Soc* **122**, 3230–3231.
16 Nagaswamy U, Voss N, Zhang Z & Fox GE (2000) Database of non-canonical base pairs found in known RNA structures. *Nucleic Acids Res* **28**, 375–376. http://prion.bchs.uh.edu/bp_type
17 Saenger W (1984) *Principles of Nucleic Acid Structure* (Cantor CR, ed.). Springer, New York.
18 Chou SH, Chin KH & Wang AHJ (1990) Unusual DNA duplex and hairpin motifs. *Nucleic Acids Res* **31**, 2461–2474.
19 Boulard Y, Cognet JAH & Fazakerley GV (1997) Solution structure as a function of pH of two central mismatches C•T and C•C in the 29 to 39 K-*ras* gene sequence by nuclear magnetic resonance and molecular dynamics. *J Mol Biol* **268**, 331–347.
20 Allawi HT & SantaLucia J (1998) Thermodynamics of internal C•T mismatches in DNA. *Nucleic Acids Res* **26**, 2694–2701.
21 Guéron M & Leroy JL (1995) Studies of base pair kinetics by NMR measurement of proton exchange. *Methods Enzymol* **261**, 383–413.
22 Szyperski T, Götte M, Billeter M, Perola E, Cellai L, Heumann H & Wüthrich K (1999) NMR structure of the chimeric hybrid duplex r(gcaguggc)·r(gcca)d(CTGC) comprising the tRNA–DNA junction formed during initiation of HIV-1 reverse transcription. *J. Biomol. NMR* **13**, 343–355.
23 Chou SH, Tseng YY & Chu BY (2000) Natural abundance heteronuclear NMR studies of the T3 mini-loop hairpin in the terminal repeat of the adeno-associated virus 2. *J Biomol NMR* **17**, 1–16.
24 Heddi B, Foloppe N, Bouchemal N, Hantz E & Hartmann B (2006) Quantification of DNA BI/BII backbone states in solution. Implications for DNA overall structure and recognition. *J Am Chem Soc* **128**, 9170–9177.
25 Sarma RH, Mynott RJ, Wood DJ & Hrusha FE (1973) Determination of the preferred conformations constrained along the C4′–C5′ and C5′–O5′ bonds of beta-5′-nucleotides in solution. Four-bond phosphorus-31–proton coupling. *J Am Chem Soc* **95**, 6457–6459.
26 Güntert P, Mumenthaler C & Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* **273**, 283–298.
27 Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J Mol Biol* **313**, 229–237.
28 Umezawa Y & Nishio M (2002) Thymine-methyl/π interaction implicated in the sequence-dependent deformability of DNA. *Nucleic Acids Res* **30**, 2183–2192.
29 Kaplan JI & Fraenkel G (1980) *NMR of Chemically Exchanging Systems*. Academic Press, London.
30 Proctor DJ, Ma H, Kiersek E, Kierzek R, Gruebele M & Bevilacqua PC (2004) Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry* **43**, 14004–14014.
31 Wijmenga SS, Kruithof M & Hilbers CW (1997) Analysis of ^{1}H chemical shifts in DNA: assessment of the reliability of ^{1}H chemical shift calculations for the use of structure refinement. *J Biomol NMR* **10**, 337–350.
32 Puglisi JD & Tinoco I (1989) Absorbance melting curves of RNA. *Methods Enzymol* **180**, 304–325.

33 Marky LA & Breslauer KJ (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* **26**, 1601–1620.
34 Pakleza C & Cognet JAH (2003) Biopolymer chain elasticity: a novel concept and a least deformation energy principle predicts backbone and overall folding of DNA TTT hairpins in agreement with NMR distances. *Nucleic Acids Res* **31**, 1075–1085.
35 Santini GPH, Pakleza C & Cognet JAH (2003) DNA tri- and tetra-loops and RNA tetra-loop hairpins fold as elastic biopolymer chains in agreement with PDB coordinates. *Nucleic Acids Res* **31**, 1086–1096.
36 Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW & Kollman PA (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J Am Chem Soc* **117**, 5179–5197.
37 Cognet JAH, Boulard Y & Fazakerley GV (1995) Helical parameters, fluctuations, alternative hydrogen bonding, and bending in oligonucleotides containing a mismatched base pair by NOESY distance restrained and distance free molecular dynamics. *J Mol Biol* **246**, 209–226.
38 Schneider B, Neidle S & Berman HM (1997) Conformations of the sugar–phosphate backbone in helical DNA crystal structures. *Biopolymers* **42**, 113–124.
39 SantaLucia J & Hicks D (2004) The thermodynamics of DNA structural motifs. *Annu Rev Biophys Biomol Struct* **33**, 415–440.
40 Chin KH, Chen FM & Chou SH (2003) Solution structure of the ActD-5′-CCGTT$_3$GTGG-3′ complex: drug interaction with tandem G•T mismatches and hairpin loop backbone. *Nucleic Acids Res* **31**, 2622–2629.
41 Delaglio F, Grzesiek S, Vuister G, Zhu G, Pfeifer J & Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293.
42 Guntert P, Dotsch V, Wider G & Wüthrich K (1992) Processing of multi-dimensional NMR data with the new software PROSA. *J Biomol NMR* **2**, 619–629.
43 Bartels C, Xia T, Billeter M, Güntert P & Wüthrich K (1995) The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* **6**, 1–10.
44 Gabarro-Arpa J, Cognet JAH & Le Bret M (1992) Object command language: a formalism to build molecular models and to analyze structural parameters in macromolecules, with applications to nucleic acids. *J Mol Graphics* **10**, 166–173.

Supporting information

The following supplementary material is available:
Fig. S1. Imino protons of NOESY cross-peaks in supercooled water.
Fig. S2. The imino proton spectrum of the 72-mer S1.3/2.2 aptamer in supercooled water.
Fig. S3. Temperature dependence of the line-widths of selected aromatic protons and thymine methyls.
Fig. S4. Torsion angles of the mispaired nucleotides (T5 to G7 and A19 to C21) during the MD trajectories.
Scheme S1. Sequential B-DNA NOEs of the 25-mer.
Table S1. Statistics from DYANA simulations of the 25-mer.
Table S2. Average torsion angles of the DYANA ensemble.
Table S3. Thermodynamic data for the 25-mer oligonucleotide.
Table S4. Ω_i, χ_i and θ_i values of the $T_aT_bT_c$ triloop nucleotides used to build all known TTT hairpins.
Table S5. Hydrogen bonds predicted with BCE for –WX–TaTbTc–YZ– sequences.
Table S6. Statistics of putative hydrogen bond and thymine methyl/aromatic ring distances (Å) for the 10 ns MD trajectory.
Annex S1. Thermodynamic equations.

This supplementary material can be found in the online version of this article.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.