

Shape matters: size-exclusion HPLC for the study of nucleic acid structural polymorphism

Eric Largy^{1,2} and Jean-Louis Mergny^{1,2,*}

¹ARNA Laboratory, University of Bordeaux, Bordeaux 33000, France and ²INSERM, U869, IECB, Pessac 33600, France

Received June 25, 2014; Revised August 04, 2014; Accepted August 4, 2014

ABSTRACT

In recent years, an increasing number of reports have been focused on the structure and biological role of non-canonical nucleic acid secondary structures. Many of these studies involve the use of oligonucleotides that can often adopt a variety of structures depending on the experimental conditions, and hence change the outcome of an assay. The knowledge of the structure(s) formed by oligonucleotides is thus critical to correctly interpret the results, and gain insight into the biological role of these particular sequences. Herein we demonstrate that size-exclusion HPLC (SE-HPLC) is a simple yet surprisingly powerful tool to quickly and effortlessly assess the secondary structure(s) formed by oligonucleotides. For the first time, an extensive calibration and validation of the use of SE-HPLC to confidently detect the presence of different species displaying various structure and/or molecularity, involving >110 oligonucleotides forming a variety of secondary structures (antiparallel, parallel, A-tract bent and mismatched duplexes, triplexes, G-quadruplexes and i-motifs, RNA stem loops), is performed. Moreover, we introduce simple metrics that allow the use of SE-HPLC without the need for a tedious calibration work. We show that the remarkable versatility of the method allows to quickly establish the influence of a number of experimental parameters on nucleic acid structuration and to operate on a wide range of oligonucleotide concentrations. Case studies are provided to clearly illustrate the *all-terrain* capabilities of SE-HPLC for oligonucleotide secondary structure analysis. Finally, this manuscript features a number of important observations contributing to a better understanding of nucleic acid structural polymorphism.

INTRODUCTION

Despite being generally depicted as a canonical right-handed B-helix, nucleic acids are predisposed to adopt a variety of secondary structures including, but not limited to, other double-helices (A, Z, A-tract bent, parallel-stranded), triple-helices (triplex), single-stranded hairpins, i-motifs and G-quadruplexes (G4) (1,2). For instance, runs of adenines found in regulatory regions of many organisms are well known to lead to significant bending of the double helix (3), while parallel-stranded duplexes (ps-ds) can be formed through reverse Watson–Crick hydrogen bonding, which are particularly stable for A•T base pairs (4). Binding of a third strand into the major groove of an antiparallel duplex via Hoogsteen or reverse Hoogsteen base-pairing leads to a triplex that can contain different triplets of bases following the motifs purine–purine–pyrimidine (R_xR•Y) or pyrimidine–purine–pyrimidine (Y_xR•Y) (5). The RNA hairpin secondary structure is another notable example of deviation from the norm with its mismatches, bulges and loops, and it is recognized that these structures are of high importance for a wide range of biological phenomena (6). Some G-rich nucleic acid sequences can fold into quadruplex structures through π -stacking of guanine quartets held together by an extensive Hoogsteen H-bonding network and coordination of monovalent metal cations (mostly K⁺ in a cellular context) sandwiched between the quartets (Figure 1 and Supplementary Figure S1) (7–9).

The G4 nucleic acids recently became the epitome of this structural diversity since they constitute a highly polymorphic family that deviates largely from the double-helix, and have been the object of growing scrutiny, in particular with regards to their putative biological roles (10,11). Quadruplex structures contain a number of common variables (loop length, geometry and sequence, number of quartets, relative orientation of strands, glycosidic bond angles) and less common features (strand bulge and snapback, base-pairing in loops, alternative quartets, non-canonical base pairing) (Figure 1 and Supplementary Figure S1). These variables are usually influenced by the primary sequence and the concentration of the nucleic acid strand, as well as by the environment (most notably the identity and

*To whom correspondence should be addressed. Tel: +33 5 4000 3022; Fax: +33 5 5757 3655; Email: jean-louis.mergny@inserm.fr

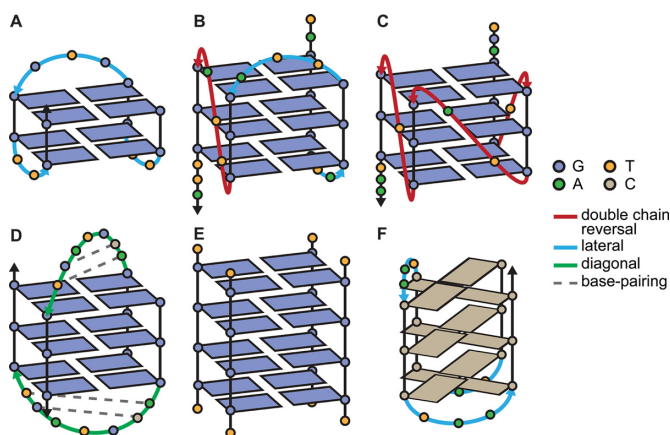


Figure 1. Schematic secondary structures of some of the quadruplex-forming sequences used in this study: TBA (A), 25TAG (B), c-myc (C), H-Bi-G4 (D), $[TG_4T]_4$ (E) and the i-motif 21CC (F). Intra- and bi-molecular structures exhibit different loop lengths (2–7) and conformations (see color legend), number of quartets (2–4), relative strand orientation (A, D: antiparallel; C, E: parallel; B: hybrid 3+1) and molecularity (A, B, C: monomolecular, D: bimolecular, E: tetramolecular). H-Bi-G4 displays very particular loops that form short duplex hairpins, while the human telomeric sequence (25TAG) is one example of polymorphism dependent on termini sequence changes and experimental conditions. Cations and glycosidic bond angles are not shown for the sake of clarity. A more accurate depiction of these structures, and others, can be found in Supplementary Figure S1.

the concentration of monovalent cations) (12–15). In addition, quadruplex structures can be obtained not only by the folding of a single strand but also by hybridization of up to four strands to give polymolecular structures (16,17). Quadruplex-forming oligonucleotides also have the potential to oligomerize or multimerize, by non-covalent stacking of the external quartets, strand interlocking and base-pairing of flanking or loop nucleotides (Supplementary Figure S1: c-kit2, B-raf, 93del, J19, N-myc, Pile-NG16). In summary, a single sequence may fold/hybridize into a variety of structures, sometimes in equilibrium with each other, depending on the experimental conditions. Moreover, the complementary C-rich sequences may fold into an i-motif. This structure is typically observed at slightly acidic pH, where the protonation of cytidine is possible, hence allowing the formation of $C\bullet CH^+$ base pairs arranged in two parallel duplexes associated head-to-tail via base-pair intercalation (Figure 1F) (18).

The understanding of the marked polymorphism of nucleic acids is essential for the development of fields of study such as synthetic biology (19), drug design (e.g. the selective targeting of biologically relevant sequences) (11,20) and the engineering of nanomaterials (9). This has prompted several groups to launch research programs aimed at solving non-canonical structures by nuclear magnetic resonance (NMR) or X-ray spectroscopy (21–23), complemented by low-resolution techniques such as circular dichroism (CD) (24,25). However, these techniques are limited by the need for high sample concentrations or crystal packing forces, and are not fully adapted to the study of structure mixtures, which are often monitored by polyacrylamide gel electrophoresis (PAGE). PAGE suffers from a number of drawbacks such as the duration of the migration, the imprecise

quantification of the species, the buffer conditions, a certain lack of versatility (buffer, temperature) and a possible poor resolution for some polymorphic and oligomeric species (smearing, no entry in the gel). More importantly, the migration depends simultaneously on the molecular weight, the charge (including the phosphate counter-ions) and the shape of the oligonucleotide (possible charge screening effects), which can be tricky to deconvolute.

Size-exclusion chromatography (SEC), also called gel filtration when performed with an aqueous eluent, is commonly used to measure the size of proteins or polymers (26), although its separation principle is not yet fully understood (27–29). DNA being essentially a polymer, SEC intuitively appears to be a method that could be used to assess the hydrodynamic volume of oligonucleotides in native conditions (30). There is no comprehensive study of its general use in the study of DNA secondary structures, but promising seminal studies have paved the way in recent years. The most notable examples are the resolution of polymorphic telomeric and oncogenic quadruplex sequences by Trent *et al.* (31–33), and the study of tetramolecular quadruplexes and i-motifs by Bardin and Leroy (16,34). We decided to explore the possibility of using a UHPLC apparatus equipped with a standard SEC column to study a wide array of secondary structures. Hereafter, > 110 oligonucleotides were analyzed in order to calibrate and validate the assay. Its potential for the quick and confident determination of secondary structures is shown for the first time. The versatility of the method is confirmed using a variety of buffers of different pH, cations and temperatures, and operating on a wide range of nucleic acid concentrations and structures. Throughout the manuscript, a number of observations contributing to a better understanding of nucleic acid structural polymorphism are provided, notably through practical case studies.

MATERIALS AND METHODS

Where possible, the experiments were performed with material free of DNA, RNA, DNase and RNase contaminants. Oligonucleotides were obtained from Eurogentec (Seraing, Belgium) as reverse-phase purified lyophilizates, except SL2, SL2_{mut} and R06 that were provided by Dr Carmelo Di Primo (Inserm U869, Pessac, France). The oligonucleotides were named according to their first appearance in the literature or generally used name unless insufficiently descriptive or ambiguous, in which case a new name was given. Sequence information is given in Tables 1 and 2 (35–71). All salts and solutions were obtained at the molecular biology grade from Sigma-Aldrich (Saint-Quentin Fallavier, France), except Tris(hydroxymethyl)aminoethane (Tris) from Biosolve Chimie (Dieuze, France). Buffer and saline solutions were prepared from ultrapure water (18.2 Ω cm⁻¹ resistivity) and filtered through disposable 0.22- μ m polyethersulfone membrane ‘Stericup-GP’, while samples were filtered through disposable 0.45- μ m PVDF ‘Durapore’ membranes for syringes (4-mm diameter, Millex), both obtained from Merck Millipore (Darmstadt, Germany). Unless otherwise stated, oligonucleotides were prepared in the elution buffer, heated at 90°C for 2 min, allowed to cool down to room temperature, then left overnight at 4°C prior to injection. The samples (typically, 10 μ l at 250

μM in elution buffer, or otherwise mentioned) were injected on a UltiMate 3000 UHPLC system (Thermo Scientific Dionex, Sunnyvale, CA, USA), equipped with an autosampler, a diode array detector and a Thermo Acclaim SEC-300 column (4.6×300 mm; $5\text{-}\mu\text{m}$ hydrophilic polymethacrylate resin spherical particles, 300 \AA pore size). Unless otherwise stated, elution of unfolded, ds- and G4-forming oligonucleotides was performed at 0.150 ml/min in Tris•HCl 50 mM, pH 7.5 , supplemented with KCl (100 mM), with a column temperature of 20.0°C . Parallel-stranded duplexes were eluted with the same buffer supplemented with MgCl_2 (10 mM). Sodium phosphate (20 mM), pH 7.2 , supplemented with NaCl (50 mM) and MgCl_2 (3 mM), was used for the elution of RNA hairpins. I-motif-forming oligonucleotides were analyzed in a MES buffer (50 mM), pH 6.0 , supplemented with KCl (100 mM). Triplexes were eluted with the same buffer with MgCl_2 (10 mM) substituting KCl. Ammonium acetate (100 mM) was used to elute the ESI-MS samples.

Thorough equilibration (with at least three column volumes; i.e. 15 ml with our experimental setting) was performed before the first injection. Eluted species were monitored by absorbance measurements ($210\text{--}600$ nm; 1-nm bandwidth, 5-Hz data collection rate, 2-s response time), and hereafter the chromatograms were plotted from the absorbance at 260 nm (normalized to $[0,1]$), using the relative elution volume V_e/V_0 as x-axis, where V_e is the elution volume and V_0 is the dead volume (typically 1.86 ml). Retention time and full width at half-maximum (FWHM) of the peaks were determined with Chromeleon 6.80 (Dionex, Sunnyvale, CA, USA) and verified with PeakFit 4.11 (Systat Software, Inc., San Jose, CA, USA) using an exponent modified Gaussian model (Equation (1)) where Abs_{norm} is the normalized absorbance, h is the amplitude of the peak, w is the width of the peak, s is the distortion of the peak, z is the center of the peak and erf is the error function (72). Examples of peak deconvolution are given in Supplementary Figure S2.

$$\text{Abs}_{\text{norm}} = \frac{hw}{s} \sqrt{\frac{\pi}{2}} \exp\left(\frac{w^2}{2s^2} - \frac{V_e/V_0 - z}{s}\right) \left\{ 1 - \text{erf}\left[\frac{1}{\sqrt{2}}\left(\frac{w}{s} - \frac{V_e/V_0 - z}{w}\right)\right]\right\} \quad (1)$$

Data analysis was performed with QtiPlot 0.9.8.7 (ProIndep Serv S.r.l., Romania) and OriginPro 9.1 (OriginLab, Northampton, MA, USA). For each peak, the nature of the species was confirmed by visual inspection of the characteristic absorbance spectrum (i.e. the presence of a local maximum around 260 nm). Isothermal difference spectra (IDS) were calculated from the absorbance spectra extracted from the areas of the peaks containing pure species, normalized for intensity at 252.3 nm for quadruplexes and 275.2 nm for i-motifs. 22AG, ds26 and dT₁₀ were used as standards to check for potential drifting of the retention times during the course of the study. The complete retention volume (for molecules whose molecular weight is below the lower molecular weight cut-off) has been determined with thymidine (4.79 ml), and the dead volume (MW > 50 kDa) with Blue Dextran (average MW = 2000 kDa; $V_0 = 1.86$ ml).

'Prediction' and 'compactness' plots show the relationship between the logarithm of the molecular weight of, respectively, the oligonucleotides ($\text{MW}_{\text{strand}}$; an average is used for heteroduplexes) or the whole structures ($\text{MW}_{\text{structure}}$; for bi-, tri- and tetra-molecular or higher molecularity species), as a function of the relative elution volume V_e/V_0 . The peak center separation (PCS) is defined as the relative elution volume ratio between two species, where 2 is a bulkier structure than 1, and hence $(V_e/V_0)_2 < (V_e/V_0)_1$ (Equation (2)).

$$\text{PCS} = \frac{(V_e/V_0)_1}{(V_e/V_0)_2} \quad (2)$$

The structure index (SI) of a given peak is an arbitrary score, calculated using Equation (3).

$$\text{SI} = V_e/V_0 \times \log_{10}(\text{MW}_{\text{strand}}) \quad (3)$$

Cohen's effect size d was determined for SI distributions, following Equation (4) using a pooled standard deviation S (Equation 5), with n_i the sample size and S_i the corresponding standard deviation (73).

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S} \quad (4)$$

$$S = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2}} \quad (5)$$

Significance of the observed distribution differences was assessed with non-parametric tests, i.e. the Mann-Whitney 'U test' and the Wilcoxon signed-rank test, the latter is only for pairs of values lying in different distributions arising from a single oligonucleotide. Full parameters and results are provided in Supplementary data.

'Radii of gyration' (R_g) were calculated with UCSF Chimera 1.8.1 (74), from structures deposited in the PDB (the results were averaged when multiple models are available), using a Python script written by Dr Eric Pettersen from a Fortran program by Dr Elaine C. Meng, reported in Supplementary data.

'Native mass spectrometry' samples were prepared by annealing oligonucleotides at a $10\text{-}\mu\text{M}$ strand concentration in 100 mM ammonium acetate, pH 6.9 . Native ESI-MS were obtained using an LCT Premier mass spectrometer (Waters, Manchester, UK). The ESI source voltage was set to 2.2 kV with a desolvation temperature of 60°C , and the sample cone voltage was successively set at 100 , 150 and 200 V. The source pressure was adjusted at 35 mbar and monitored with a Center Two probe (Oerlikon Leybold Vacuum, Cologne, Germany). Injection was performed at $200 \mu\text{l h}^{-1}$.

UV-melting. Melting temperatures were examined by measuring the changes in absorbance at 295 nm as a function of the temperature, using a SAFAS UVmc2 double-beam spectrophotometer (Monte Carlo, Monaco) equipped with a high-performance Peltier temperature controller and a thermostatable 10-cell holder. After heating at 95°C of the samples containing 10 or $5 \mu\text{M}$ oligonucleotide in cacodylate buffer (20 mM $\text{LiAsMe}_2\text{O}_2$, 100 mM KCl, pH 7.2) the absorbance was monitored at 240 , 260 , 273 , 295 and 330 nm on a cycle composed of a cooling down to 2°C

Table 1. Duplex- and triplex-forming oligonucleotides used in this study

Name	Sequence (5' to 3')	Description	Structure	PDB ID
dsIac	GA ₂ T ₂ GTGA(GC) ₂ TCACA ₂ T ₂ C	Lac operon sequence	B-DNA	
ds26	CA ₂ TCG ₂ ATCGA ₂ T ₂ CGATC ₂ GAT ₂ G	Synthetic construct	B-DNA	
ds14	(CG) ₂ A ₃ T ₃ (CG) ₂	Synthetic construct	B-DNA (35)	2DAU
ds10	CGCA ₂ T ₂ GCG	Synthetic construct	B-DNA (36)	1S23
ds12-CG	CGT ₄ A ₄ CG	Synthetic construct	B-DNA (37)	1RVI
ds12-GC	GCT ₄ A ₄ GC	Synthetic construct	B-DNA (37)	1RVH
ds15-h	(CG) ₂ A ₂ GCA ₂ (CG) ₂	Synthetic construct	B-DNA hairpin (38)	2M8Y
ds17-TA	C ₂ AGT ₂ CGTAGTA ₂ C ₃	M.TaqI binding site ^b	B-DNA (39)	
	G ₃ T ₂ ACTACGA ₂ CTG ₂			
ds17-TT ^a	C ₂ AGT ₂ CGTAGTA ₂ C ₃	M.TaqI binding site ^b	Mismatched B-DNA (39)	
	G ₃ T ₂ ACTTCGA ₂ CTG ₂			
ds17-TC ^a	C ₂ AGT ₂ CGTAGTA ₂ C ₃	M.TaqI binding site ^b	Mismatched B-DNA (39)	
	G ₃ T ₂ ACTCCGA ₂ CTG ₂			
ds17-TG ^a	C ₂ AGT ₂ CGTAGTA ₂ C ₃	M.TaqI binding site ^b	Mismatched B-DNA (39)	
	G ₃ T ₂ ACTGCGA ₂ CTG ₂			
ds-A ₆	G ₂ CA ₆ CG ₂	Synthetic construct	A-tract bent B-DNA (40)	1FZX
	C ₂ GT ₆ GC ₂			
ds-A ₆ mut	G ₂ CA ₂ GA ₃ CG ₂	Synthetic construct	A-tract bent B-DNA (40)	1G14
	C ₂ GT ₃ CT ₂ GC ₂			
ps-ds15	A ₅ TA ₂ T ₃ ATAT	Synthetic construct	Parallel duplex (41)	
	T ₅ AT ₂ A ₃ TATA			
ps-ds20	T ₆ A ₂ T ₂ A ₃ TA ₂ TA ₂ T	Synthetic construct	Parallel duplex (41)	
	A ₆ T ₂ A ₂ T ₃ AT ₂ AT ₂ A			
ps-ds24	T ₁₀ AT ₂ A ₄ T ₃ ATA ₂	Synthetic construct	Parallel duplex (41)	
	A ₁₀ TA ₂ T ₄ A ₃ TAT ₂			
PyW1	TCT ₂ CTCT ₃ CT	Synthetic construct	Parallel triplex (42)	
PuC1	AGA ₃ GAGA ₂ GA			
PyH1	TCT ₃ CTCT ₂ CT			

^aThe bases in italics are mismatched.

^bThe sequences contain the 4-bp recognition sequence of the DNA methyltransferase M.TaqI (5'-TCGY-3', where Y = A for the natural substrate).

at a rate of 0.2°C min⁻¹, then heating back up at 95°C at the same rate. Melting temperatures were determined using the baseline method (75), and assuming that the transition equilibrium involves only two states (associated and dissociated). Hence, two baselines were determined by linear regression of the pre- and post-transition plots. The fraction of associated strands as a function of the temperature θ_T was calculated from Equation (6) where L_T are the baselines (0 and 1 superscripts stand for the dissociated and the associated species, respectively), and A_T is the absorbance.

$$\theta_T = (L_{T^0} - A_T)/(L_{T^0} - L_{T^1}) \quad (6)$$

Melting temperatures were extracted by nonlinear fitting (least-squares method) of the $\theta_T = f(T)$ plot using a sigmoidal-type Boltzmann function (Equation (7)), and then solving T for $\theta_T = 0.5$, where A_1 and A_2 are the initial and final values, respectively, T_0 is the center (i.e. the temperature for which $\theta = (A_1 + A_2)/2$) and c is a constant.

$$\theta_T = (A_1 - A_2)/(1 + \exp[(T - T_0)/c]) + A_2 \quad (7)$$

CD experiments were performed with a JASCO J-815 spectropolarimeter equipped with a JASCO CDF-426S Peltier temperature controller, using quartz cells of 10 mm path length. The scans were recorded at 20°C from 210 to

Table 2. Selected G-quadruplex and i-motif-forming sequences used in this study

Name	Sequence (5' to 3')	Description	Structure ^a	PDB ID
21GG	G ₃ (TTAG ₃) ₃	Human telomere	Polymorphic (43–45) ^b	
22AG	AG ₃ (TTAG ₃) ₃	Human telomere	Hybrid 3+1 (45–47)	
23AG	AG ₃ (TTAG ₃) ₃ T	Human telomere	Hybrid 3+1 (45–47)	2KKA ^c
24TTG	TTG ₃ (TTAG ₃) ₃ A	Modified human telomere	Hybrid 3+1 (43)	2GKU
25TAG	TAG ₃ (TTAG ₃) ₃ ATT	Human telomere	Hybrid 3+1 (48)	2JSL
45AG	G ₃ (TTAG ₃) ₇	Human telomere	Hybrid 3+1 (49)	
Oxy30	T(G ₄ T ₄) ₃ G ₄ T	<i>Oxytricha</i> telomere	Hybrid 3+1 ^d	
Oxy28	(G ₄ T ₄) ₃ G ₄	<i>Oxytricha</i> telomere	Antiparallel (50)/ hybrid 3+1 ^{d,e}	201D
c-myc	TGAG ₃ TG ₃ TAG ₃ TG ₃ TA ₂	c-myc promoter	Parallel (51)	1XAV
c-kit1	G ₃ AG ₃ CGCTG ₃ AG ₂ AG ₃	c-kit promoter	Parallel (52,53) ^d	
c-kit87-up	AG ₃ AG ₃ CGCTG ₃ AG ₂ AG ₃	c-kit promoter	Parallel with a snapback feature ^d (54)	2O3M
c-kit2	CG ₃ CG ₃ CGCTAG ₃ AG ₃ T	c-kit promoter	Monomeric parallel / dimeric parallel ^d (55)	2KYP/2KYO
c-kit2GG	G ₃ CG ₃ CGCTAG ₃ AG ₃	c-kit promoter	Parallel ^d (52)	
c-kit*	G ₂ CGAG ₂ AG ₄ CGTG ₂ C ₂ G ₂ C	c-kit promoter (SP1 binding site)	Antiparallel (56)	
B-raf	G ₃ CG ₄ AG ₅ A ₂ G ₃ A	B-raf promoter	Parallel intertwined dimer (57) ^d	4H29
K-ras 35B3	AG ₃ CG ₂ TGTG ₃ A ₂ GAG ₃ A ₂ GAG ₅ AG ₂	K-ras promoter	Parallel ^d	
K-ras 35B1	AG ₃ CG ₂ TGTG ₃ A ₂ GAG ₃ A ₂ GAG ₅ AG ₂ CAG	K-ras promoter	Parallel ^d	
N-myc	TAG ₃ CG ₃ AG ₃ AG ₃ A ₂	N-myc intron	Monomeric parallel / dimeric parallel (58)	2LED/2LEE
pilE-NG16	G ₃ TG ₃ T ₂ G ₃ TG ₃	<i>N. gonorrhoeae</i> pilin expression locus	Parallel monomer / stacked dimer (59)	2LXV
pilE-NG22	TAG ₃ TG ₃ T ₂ G ₃ TG ₄ A ₂ T	<i>N. gonorrhoeae</i> pilin expression locus	Parallel (59)	2LXQ
G4CT	(G ₄ CT) ₃ G ₄	<i>Treponema pallidum</i> genome	Mainly antiparallel ^f (60)	
T30177-TT	T ₂ GTG ₂ TG ₃ TG ₃ TG ₃ T	Aptamer (HIV-1 integrase inhibitor)	Parallel, with bulge (61)	2M4P
T95–2T	T ₂ G ₃ TG ₃ TG ₃ TG ₃ T	Aptamer (HIV-1 integrase inhibitor)	Parallel monomer / stacked dimer (62)	2LK7
93del	G ₄ TG ₃ AG ₂ AG ₃ T	Aptamer (HIV-1 integrase inhibitor)	Interlocked bimolecular dimeric parallel ^d (63)	1Y8D
J19	GIGTG ₃ TG ₃ TG ₃ T	Aptamer (HIV-1 integrase inhibitor)	Parallel monomer / stacked dimer ^d (64)	2LE6
25CEB	AG ₃ TG ₃ TGTA ₂ GTGTG ₃ TG ₃ T	25CEB human minisatellite locus	Parallel ^d	
26CEB	A ₂ G ₃ TG ₃ TGTA ₂ GTGTG ₃ TG ₃ T	25CEB human minisatellite locus	Parallel ^d (65)	2LPW
TBA	G ₂ T ₂ G ₂ TGTG ₂ T ₂ G ₂	Aptamer (thrombin)	Antiparallel (66)	148D
H-Bi-G4	G ₃ ACGTAGTG ₃	Synthetic construct	Bimolecular, antiparallel, with hairpin loops ^d (67)	2KAZ
161T	TG ₃ TG ₃ T ₆ G ₃ TG ₃ T	Synthetic construct	Parallel ^d	
161TT	T ₂ G ₃ TG ₃ T ₆ G ₃ TG ₃ T ₂	Synthetic construct	Parallel ^d	
222T	TG ₃ T ₂ G ₃ T ₂ G ₃ T ₂ G ₃ T	Synthetic construct	Parallel ^d	
222TT	T ₂ G ₃ T ₂ G ₃ T ₂ G ₃ T ₂ G ₃ T ₂	Synthetic construct	Parallel ^d	
[TG ₄ T] ₄	TG ₄ T	Synthetic construct	Tetramolecular parallel (68)	2O4F
[TG ₅ T] ₄	TG ₅ T	Synthetic construct	Tetramolecular parallel	
[AG ₄ T] ₄	AG ₄ T	Synthetic construct	Tetramolecular parallel	
[AG ₅ T] ₄	AG ₅ T	Synthetic construct	Tetramolecular parallel	
Mito9	TG ₆ TGTCT ₃ G ₄ T ₃ (G ₂ T ₂) ₂ CG ₄ TATG ₄ T	Human mitochondrial DNA	Unknown	
Mito86	TGT ₂ AG ₄ TCATG ₃ CTG ₃ T	Human mitochondrial DNA	Unknown	
21CC	C ₃ (TA ₂ C ₃) ₃	Human telomere	i-motif (69)	
22Py	A ₂ TC ₃ AC ₄ TC ₃ AC ₄ T ₂	c-myc promoter	i-motif (70)	
24mutHtelo	(TTACCG) ₄	Mutated human telomere	Unfolded	
22mutHtelo ₁₂₃₄	AGTG(T ₂ AGTG) ₃	Mutated human telomere	Unfolded	
22mutHtelo ₂₃	AG ₃ (T ₂ AGTG) ₂ T ₂ AG ₃	Mutated human telomere	Unfolded (71)	

^aReferences link to a structure determination by either NMR, X-ray or CD spectroscopy.

^bHybrid 3+1 at high strand concentration and antiparallel at low (3 μM) strand concentration. Addition of flanking bases promotes the hybrid 3+1 conformation (e.g. 22AG, 23AG, 24TTG, 25TAG).

^c2KKA was obtained following a dG₁₄ to dI₁₄ mutation of 23AG, the corresponding 2-quartet structure will thus not be discussed for the calibration.

^dRelative strand orientation observed by CD: see experimental section and Supplementary Figure S3.

^eA hybrid 3+1 conformation is observed by CD, which could be caused by a difference in cation (Na⁺ in the PDB entry).

^fMainly monomolecular in our conditions (100 mM KCl). A tetramolecular structure is formed with increasing KCl concentration (60).

350 nm with the following parameters: 0.2 nm data pitch, 2 nm bandwidth, 0.5 s response, 50 nm min⁻¹ scanning speed, and are the result of three accumulations. Solutions were prepared by diluting the annealed oligonucleotide samples at 10 μM in cacodylate buffer (20 mM LiAsMe₂O₂, 100 mM KCl, pH 7.2). The CD data are blank-subtracted and normalized to molar dichroic absorption ($\Delta\epsilon$) based on nucleoside concentration using Equation (8), with θ the ellipticity in millidegrees, c the nucleoside concentration in mol l⁻¹ and l the path length in cm (Supplementary Figure S3).

$$\Delta\epsilon = \theta / (32980 \times c \times l) \quad (8)$$

RESULTS AND DISCUSSION

Technical considerations

Throughout the manuscript, relatively high strand concentrations are used (typically, 250 μM) in order to confidently compare the results with structures solved by NMR. However, working at high oligonucleotide concentration promotes the formation of multimolecular species. For this reason it can be tricky to compare results obtained with 'low-concentration' techniques (e.g. UV-melting, FRET-melting, CD, PAGE), to 'high-concentration' methods such as NMR. One of the salient advantage of SE-HPLC is its ability to handle a wide range of concentrations. Hence, by tuning the diode array detector, solutions containing as low as 1–10 μM in strand can be analyzed (Supplementary Figure S4), and lower concentrations can be considered depending on the oligonucleotide molar extinction coefficient. Note that, to detect low nucleic acid concentrations, the chromatographic system must be free of unnecessary dead volumes. The use of thin capillaries of appropriate length between the injector and the column, and between the column and the detector is recommended, as well as the use of analytical columns (diameters of 4.6 mm rather than 7.8 mm). SE-HPLC can thus operate as a bridge between the above-mentioned techniques, by simply comparing the solutions at high and low strand concentration, in their proper buffer. A practical example of concentration dependence study is given in the case study section. Moreover, we verified that the oligonucleotides do not interact with the stationary phase by changing the flow rate (0.05–0.30 ml/min): the relative elution volume is unchanged and the obtained chromatograms are superimposable (Supplementary Figure S5).

ssDNA calibration

The experimental setup is first calibrated with 10 d(T_n) sequences of various lengths ($n = 5-70$), which do not form any defined secondary structure. As expected, the relative elution volume V_e/V_0 decreased with increasing oligonucleotide length, ranging from 1.79 to 1.41 (Figure 2A). A near-perfect linear relationship between the decimal logarithm of the molecular weight and V_e/V_0 was observed ($R^2 = 0.9996$), as commonly found for other type of linear polymers (Figure 2B). This plot could be considered as a compactness baseline since the d(T_n) oligonucleotides are unstructured. To ensure that there is no bias caused by the use of polypyrimidine sequences, it is verified that polypurine

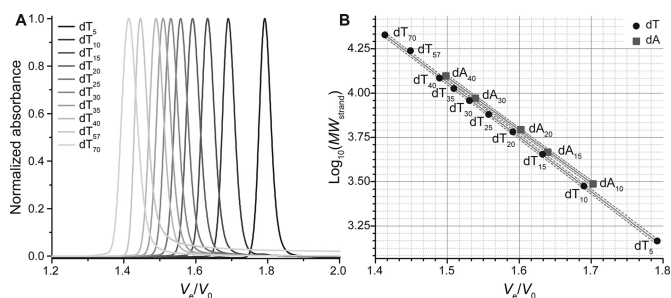


Figure 2. (A) Normalized chromatograms of unstructured polypyrimidine d(T_n) oligonucleotides. (B) Plot of the molecular weight decimal logarithm against the relative elution volume for unstructured polythymidilate (d(T_n), circles) and polyadenylate (d(A_n), squares) tracts. The two sets are linearly fitted independently (solid lines; dashed lines: 95% confidence bands).

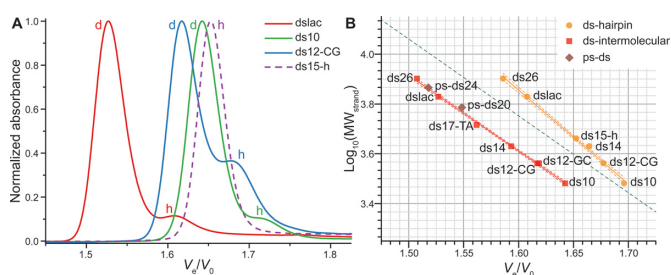


Figure 3. (A) Normalized chromatograms of some auto-complementary oligonucleotides (plain lines), and the hairpin-forming ds15-h (dashed line). Dimeric species are indicated with a *d*, and hairpins with a *h*. Examples of data treatment are given in Supplementary Figure S2. Other chromatograms can be found in Supplementary Figures S7 and S8. (B) Predictive plot of the log₁₀(MW) against V_e/V_0 for duplex-forming oligonucleotides structured in intermolecular antiparallel duplex (red squares) and hairpin (orange circles). The two sets are linearly fitted independently (solid lines; dashed lines: 80% prediction bands). Parallel duplexes are shown as brown diamonds. The linear fit obtained for d(T_n) oligonucleotides is depicted as a green dotted line.

sequences give similar results. Five dA_n sequences are injected ($n = 10-40$), and the plot of log₁₀(MW) against V_e/V_0 of the main peak is linearly fitted ($R^2 = 0.9994$) with a similar slope (-2.98 versus -3.08). However, a small portion of the strands fold into non-canonical secondary structures via A•A base-pairing, resulting in secondary peaks (Supplementary Figure S6).

dsDNA calibration

Next, duplex-forming oligonucleotides, varying in length (10–26 nt) and base composition, are analyzed (Figure 3A). Auto-complementary sequences were chosen as they can be structured in the homo-duplex and/or the hairpin-duplex form, depending on the annealing conditions (nucleic acid concentration, cation concentration, cooling rate). Additionally to the widely studied ds26 and dslac (lac operon) sequences, four sequences are based on the so-called Drew-Dickerson model (ds10, ds14 and the mirror-image sequences ds12-GC and ds12-CG) (35–37). Finally, in order to obtain a reference peak for hairpin structures, the synthetic construct ds15-h, which is not fully auto-complementary, is also injected (Table 1). Indeed, the three

central nucleotides forming the loop in the hairpin mismatch in the case of the homo-duplex, rendering the latter significantly less stable than the former (38).

Relative elution volumes of homo-duplexes and hairpins are plotted as previously described (Figure 3B). The exact elution volume for each peak was determined after proper peak deconvolution as shown in Supplementary Figure S2. The structure (homo-duplex or hairpin) of each peak is assigned: (i) according to its V_e/V_0 , the homo-duplex being trivially twice larger than the corresponding hairpin and (ii) using ds15-h as a reference for hairpins (*vide infra*). In a first approximation, both secondary structures can be considered as linear polymers and it came as no surprise to also find a linear relationship between $\log_{10}(\text{MW}_{\text{strand}})$ and retention time ($R^2 = 0.9974$ and 0.9950 , respectively). The data points are slightly less aligned than previously observed for d(T_n) and d(A_n) tracts, probably owing to small structural differences arising from variable helix bending, different GC content and hairpin loop length and flexibility. Noteworthy, the mirror-image oligonucleotides ds12-GC and ds12-CG gave quasi identical relative elution volumes (1.619 and 1.617, respectively), which is another argument for the method reliability. Another interesting point is the exclusive formation of a hairpin structure for ds15-h, as expected from its sequence, and in accordance with recent NMR spectroscopy experiments (38).

The structure(s) formed by a duplex-forming sequence can be readily identified from the predictive plot since the hairpin and intermolecular duplex areas, defined by the 80% prediction band of the linear fit, are entirely separated on the studied MW range. However, in order to avoid the tedious calibration work, the use of one to two reference sequences of known structures, carefully selected in function of the unknown sample characteristic (analogous MW, same type of secondary structure), might prove sufficient. To help in the achievement of this approach, a simple arbitrary metric can be used, the PCS, which is simply the ratio of V_e/V_0 values from the slowest peak (here, the hairpin) with the fastest peak (here, the intermolecular duplex). The mean PCS found in our conditions is 1.04 ± 0.01 , but the value is dependent on the MW of the oligonucleotide (longer sequences give higher PCS). For this reason it is of the utmost importance to choose reference sequences lying in the same MW range to perform the PCS comparison. Examples of this approach are given below (e.g. the RNA stem loop case studies).

It is interesting to note that the simultaneous detection of the intermolecular duplex and hairpin species also enables the study of the interconversion processes between these two secondary structures. Under the experimental conditions, high salt and strand concentrations, a conversion from the hairpin to the duplex form is observed over time (Supplementary Figure S7). The stem of ds26 is much longer than the stem of ds12 and ds14, which renders its hairpin form relatively more stable than the hairpin of the latter, leading to a higher hairpin/duplex ratio after a long incubation period.

Non-canonical double helices can also be monitored by SE-HPLC. Duplexes presenting a single mismatch (ds17-TT, ds17-TG and ds17-TC) led to chromatograms superimposed with the fully matched counterpart (ds17-TA; Sup-

plementary Figure S8). However, the presence of an A-tract can be detected, because A-tracts induce a significant bending of the double helix as compared to a control duplex where the same A-tract is disrupted (Supplementary Figure S9) (40). Several parallel-stranded duplex forming oligonucleotides are also assayed in a 100-mM KCl, 10-mM MgCl₂ containing buffer, with a strand concentration of 100 μ M (Supplementary Figure S10). While the 24- and 20-mer (ps-ds24 and ps-ds20) indeed form a ps-ds structure, with more unassociated single strand detected in the latter case, the shorter 15-mer (ps-ds15) eluted entirely as unstructured single strands. These results are perfectly consistent with the expected increase in ps-ds stability with a greater number of base pairs.

The study of triplexes by SE-HPLC is also considered. The binding of a third strand in the groove of a DNA duplex should result in a bulkier structure that elutes faster. A 12-mer parallel pyrimidine DNA triplex PyW1•PuC1xPyH1 is analysed at pH 6.0, in a Mg²⁺-containing buffer, and compared to the corresponding Watson–Crick duplex (PyW1•PuC1) (Supplementary Figure S11). As expected, the chromatogram of PyW1•PuC1xPyH1 acquired at 20°C exhibits three peaks corresponding to the triplex, the duplex and the unbound PyH1 strand. The parallel Hoogsteen-duplex part of the triplex can be selectively melted in the column because it is less stable than the antiparallel Watson–Crick one. At 30°C, the duplex peak is conserved, but the triplex peak completely disappears and the proportion of single strand rises as a result.

G4-DNA calibration

A wide range of quadruplex-forming sequences from genomic (telomeres, oncogene promoters, minisatellites) or synthetic (aptamers, human-designed) origin are subsequently analyzed (Table 2). These sequences are selected on the basis of their various structural characteristics, including possible polymorphism and multimerization (through stacking or interlocking), the availability of structural data, the frequency of their use in the field and their biological relevance. The four-repeat human telomeric sequence, represented by four sequences with varying length (22 nucleotides for 22AG to 25 nucleotides for 25TAG), is a well-known example of structural heterogeneity among the G4 field (76). Interestingly, single base modifications can be differentiated with SE-HPLC, albeit with small differences in relative elution volumes, ranging from $V_e/V_0 = 1.599$ to 1.619, with identical full width at half-maximum (FWHM = 0.038 ± 0.001) (Figure 4A). The relative linear alignment of their V_e/V_0 as a function of the MW logarithm ($R^2 = 0.9541$) also tends to indicate an important structural similarity, and homogeneity in case of mixtures. Similarly, 25CEB and 26CEB, which differ by the presence of an extra 5' base for the latter, can be discriminated. However, a difference of a few bases, which increases the molecular weight of an oligonucleotide, does not necessarily impact significantly its hydrodynamic volume. Clearly, the interest of the method does not lie in the discrimination of oligonucleotides based on small length differences, in particular not for quadruplex structures. Gel electrophoresis and capillary

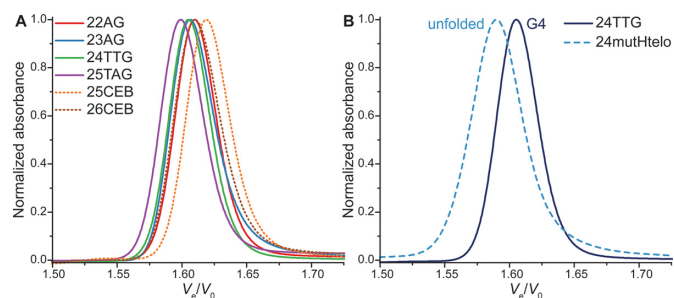


Figure 4. Normalized chromatograms of (A) some monomeric quadruplexes and (B) a 24-mer mutant of the human telomeric sequence (dashed line), compared to the quadruplex-forming sequence 24TTG (plain line).

electrophoresis are two methods more suited for this type of analysis.

To ensure that quadruplexes can be discriminated from unfolded strands, mutants of the 22- and 24-mer human telomeric sequence are analyzed. 22mutHelo₁₂₃₄ and 24mutHelo have all four guanine tracts mutated, which prevents the formation of quadruplex. As a result, it elutes faster than the corresponding more compact quadruplex (22AG and 24TTG, respectively; Figure 4B and Supplementary Figure S12). However, only two G-tracts are mutated in 22mutHelo₂₃ (Table 2), a control used in a recent study (71). While the expected unfolded species is clearly visible and elutes faster than 22AG, a secondary accelerated peak is also detected. This species is likely a multimolecular quadruplex formed by association of the non-mutated guanine tracts.

Prediction plot. Relative elution volumes of sequences forming unambiguously monomeric structures (for which structures have been reported to be monomeric, that gave a single peak by SE-HPLC, and whose names are italicized in Table 2) are used to create a first reference linear trend plot, which is fitted ($R^2 = 0.8041$) in order to start the assignment of other species (not shown). The peaks are subsequently assigned to monomeric, dimeric, another higher-order quadruplex structures or an unstructured strand, based on (i) the position of the datum point compared to this first trend, (ii) the available structural data, (iii) the sequence of the oligonucleotide, (iv) native ESI-MS experiments and (v) UV-melting experiments. Eventually, enough data points are collected to allow the construction of a predictive plot that replaces step (i) (Figure 5A). A number of examples will be given in the sections below, and all other chromatograms are presented in Supplementary Figure S13. The model feeds itself because the more points are plotted, the higher the level of confidence is for the identification of unknown species. Interestingly, despite the significant polymorphism of G4 nucleic acids, monomers ($R^2 = 0.7415$), dimers ($R^2 = 0.9059$) and tetramers ($R^2 = 0.9954$) can be discriminated with a high confidence level: areas defined by the 80% prediction bands of the linear fitting do not overlap (on the MW range studied herein), and the assignment can be readily done by plotting the unknown point. An arbitrary metric, the SI, is calculated for each peak by multiplying $\log_{10}(\text{MW}_{\text{strand}})$ to V_e/V_0 (see Material and methods), and values are grouped by structure type (Fig-

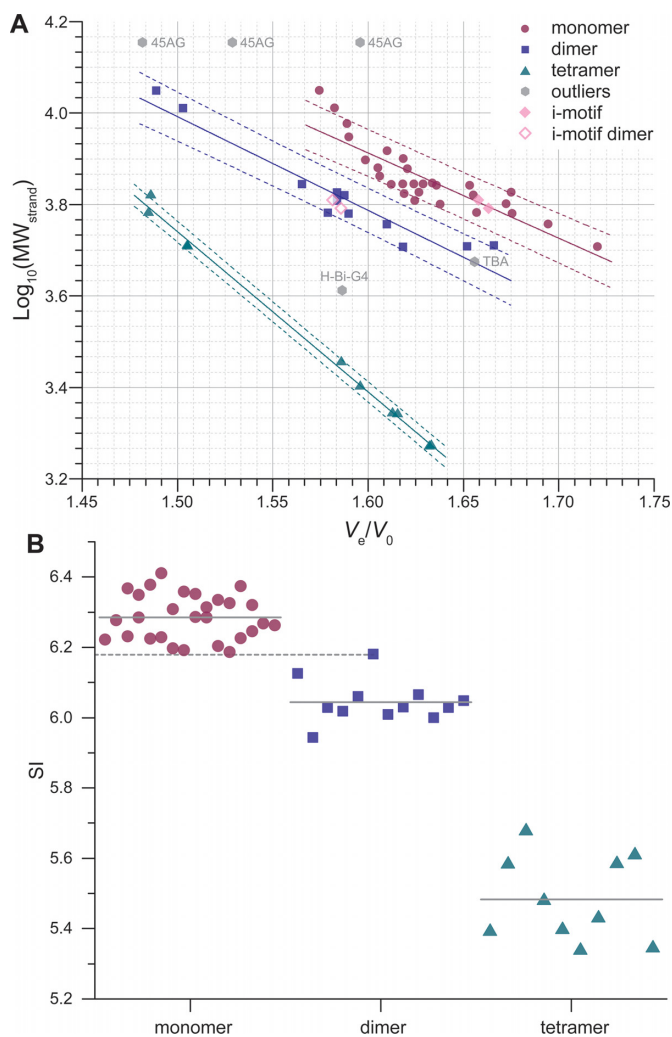


Figure 5. (A) Predictive plot: $\log_{10}(\text{MW}_{\text{strand}})$ against V_e/V_0 for quadruplex-forming oligonucleotides, with linear fitting (solid line; dashed lines: 80% prediction bands). Prediction bands define areas for each category of structures. Outliers are oligonucleotides whose particular structure (e.g. only two tetrads for TBA) leads to a non-alignment of its V_e/V_0 with the ones of its structural category (e.g. other monomers). (B) SI chart obtained from the product of the $\text{MW}_{\text{strand}}$ by V_e/V_0 ; plain gray lines: mean values (monomers: 6.29 ± 0.06 , dimers: 6.04 ± 0.06 , tetramers: 5.48 ± 0.12), dashed gray line: maximum SI value observed for dimers (93del: 6.18).

ure 5B). The resulting chart also allows the discrimination between monomers, dimers and tetramers because there is no overlap between the three groups. The aptamer 93del, which folds into a very compact dimer (*vide infra*), is the only data point close to another structure group, and not lying within one standard deviation unit from the group average. The significant difference between the monomer and dimer SI distributions was further confirmed with statistical tests at high significance levels, additionally to the large Cohen's effect size ($d = 3.79$) (see the experimental section and Supplementary data).

Although it seems appealing to perform such direct assignments, one must be careful when doing so because quadruplexes exhibit a very high degree of polymorphism. Complementary methods may thus be used as a mean of

comparison. For instance, some oligonucleotides folding in unique structures (referred to as outliers in the plots), such as TBA (only two G-quartets) or H-Bi-G4 (a dimer with only three G-quartets, thus eluting like a monomer), do not abide by this calibration. Another tricky case is the 45-mer human telomeric sequence 45AG that gives a chromatogram featuring a major peak flanked by two minor peaks (Supplementary Figure S13). 45AG can theoretically fold into two distinct quadruplex units linked by a TTA trinucleotide. There is no consensus on the topology of arrangement of successive units for long human telomeric sequences as several models have been suggested (beads-on-a-string, same direction stacking, alternate direction stacking) (8). Additionally, one unit may not be folded, yielding a less compact structure that would account for the small accelerated peak. Furthermore, Phan *et al.* have also shown that a seven-repeat human telomeric sequence can fold into a single intramolecular quadruplex unit containing a long loop (49). Finally, the formation of multimolecular assemblies cannot be excluded (77,78). The data points are somewhat out of the calibration range and it is thus risky to interpret the chromatogram that however clearly shows that the folding of long G-rich sequences is not trivial.

Overall, the good separation between structure groups mentioned above avoids a comprehensive, tedious calibration via the injection of a couple of reference sequences and calculation of the peak center ratio (PCS), as shown above for dsDNA. It is defined as the ratio of the V_e/V_0 value from the slowest peak (here, the monomer) with the one of the fastest peak (dimer or tetramer). This value is consistent for sequences giving two peaks resulting from a dimer/monomer mixture (1.048 ± 0.013), and is significantly different from the tetramer/monomer one (1.113 ± 0.005 ; effect size: $d = 6.0$). This can be used to determine whether a secondary peak of a given sample is likely to result from the formation of a dimer or a tetramer, by simply comparing its PCS with standard PCS values inferred from reference oligonucleotides. These results may be linked to other analytical methods since SE-HPLC does not give absolute results.

Exploring quadruplex structural polymorphism

Compactness. From the compactness plot obtained for all identified monomeric quadruplexes, it is clear that the data points are far less aligned (Supplementary Figure S14; $R^2 = 0.7432$) than for unstructured or duplex-structured oligonucleotides, which certainly accounts for the well-known polymorphism of quadruplex nucleic acids (Figure 1 and Supplementary Figure S1). G4-forming oligonucleotides of similar molecular weight can display a significantly different hydrodynamic volume depending on the structural characteristic of the quadruplex they form (length and geometry of loops, number of quartets, bulges). In short, two quadruplex structures are less similar to one another than two duplex helices, for a given molecular weight, which likely leads to the scattered data points. While ssDNA and dsDNA oligonucleotides can be assimilated to superimposable linear polymers, G4-DNA comes in a variety of shapes (Supplementary Figure S15). This distribution can be taken advantage of to evaluate the relative compactness of a quadru-

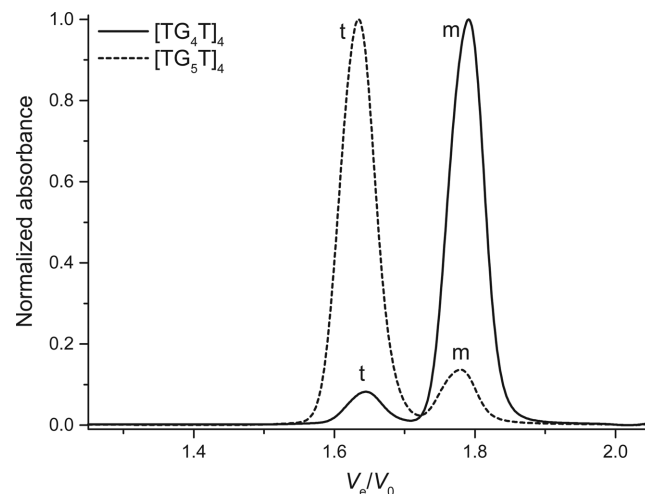


Figure 6. Normalized chromatograms of tetramolecular quadruplex-forming sequences (250 μ M strand concentration) acquired within 30 min after KCl addition (TG₄T) or after a two-day equilibration (TG₅T). Monomer species are indicated with a *m* and tetramers with a *t*.

plex structure. For a given molecular weight, an oligonucleotide with a higher retention time has a smaller hydrodynamic volume and can therefore be considered as more compact. The most compact species are thus located in the right-hand part of the compactness plot (Supplementary Figure S14; e.g. c-myc is more compact than c-kit2), which, unlike the prediction plots, is drawn from the molecular weight of the whole structure. The TBA structure is the least compact of all the assayed monomeric G4 (highest deviation from the trend line), most probably because it only has two G-quartets. The trend line can thus be used as the mean compactness of the studied set of sequences, the ones being *above* being schematically more compact than the average. Incidentally, it can be noted that the dimeric structures are located above the monomeric G4 scatter plot (Supplementary Figure S14; $R^2 = 0.9066$), suggesting that higher-order structures are more compact. Additionally, an analysis of the relationship observed between the radius of gyration, calculated for structures deposited in the PDB, and the elution volume is given in Supplementary data (Supplementary Figure S16). Overall, SE-HPLC highlights the important polymorphism of G4 and allows to discriminate particular structures such as TBA or dimers.

Multimeric quadruplexes. A salient advantage of SE-HPLC is the possibility to monitor the formation of intermolecular structures. Indeed, the simultaneous detection of the monomer and quadruplex peaks allows the quantification of the extent and the kinetics of association, similarly to the hairpin/duplex equilibrium examples given above. In Figure 6, the delayed peaks correspond to the unfolded monomers, whereas the accelerated peaks are tetramolecular ([TG₄T]₄ and [TG₅T]₄) species. Interestingly, IDS can be plotted from the absorbance spectra integrated under the quadruplex and monomer peaks, yielding yet an additional proof of quadruplex formation (Supplementary Figure S17A and B) (79). Finally, [AG₄T]₄ and [AG₅T]₅, which are known to associate at higher rates (17), present chro-

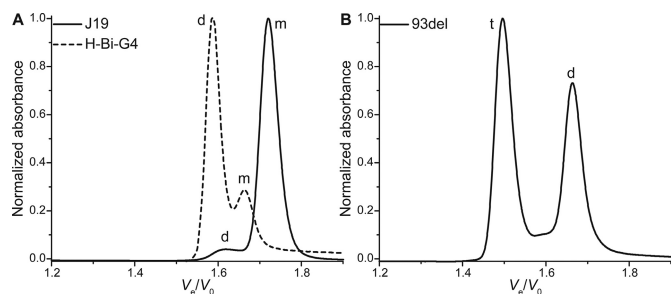


Figure 7. Normalized chromatograms of bimolecular quadruplex-forming sequences: (A) hairpin-looped H-Bi-G4 and stacked-dimer J19, and (B) the stable interlocked dimers 93del. Monomer species are indicated with a *m*, dimers with a *d* and tetramers with a *t*.

matograms devoid of monomer after eight hours at 4°C (Supplementary Figure S18).

Similarly, H-Bi-G4 forms a bimolecular quadruplex in equilibrium with a monomer theoretically unfolded since it only contains two runs of guanines (Figure 7A). This is further verified by plotting the IDS of the monomer/dimer couple, calculated from the absorbance spectra under the pure areas of the peaks (see experimental section and Supplementary Figure S17C). Using ESI-MS at lower strand concentration (10 μM), the bimolecular quadruplex cannot be detected using 100–200 V capillary tensions, which suggests that this structure is relatively unstable, at least in ammonium conditions (Supplementary Figures S19 and S20). This is subsequently confirmed in the presence of potassium by UV-melting experiments ($T_m^{10\ \mu\text{M}} = 34.5^\circ\text{C}$; Supplementary Figure S21A), which explains the large percentage of monomer detected by SE-HPLC. This is an interesting result as it shows that SE-HPLC can be a tool to measure relative stabilities. Additionally, UV-melting experiments further confirmed the intermolecular nature of H-Bi-G4 quadruplex because the melting temperature is concentration dependent ($T_m^{5\ \mu\text{M}} = 30.5^\circ\text{C}$). Finally, since H-Bi-G4 forms a bimolecular structure, a decrease in strand concentration (from 250 to 100 μM) leads to the formation of a lower proportion of quadruplex, as expected from the law of mass action (Supplementary Figure S22). The kinetics of formation is also fairly slow, at room temperature, compared to a typical intramolecular folding.

Figure 7 features other interesting examples of quadruplex dimers, which differ by the dimerization mode (stacking, interlocking). Purely stacked quadruplex dimers such as J19 can be detected (Figure 7A), which highlights the fact that SE-HPLC allows to operate in *native* conditions. Only a very small amount of stacked dimers is also detected for the T30177-TT aptamer (Supplementary Figure S23); the two thymines in 5' have been shown to inhibit dimer formation (61). Some sequences do not fold at all into monomeric G4 despite being theoretically able to do so. Hence, the smallest structure formed by 93del ($V_e/V_0 = 1.666$; Figure 7B) is indeed attributed to a very stable interlocked dimer ($T_m^{10\ \mu\text{M}} > 88^\circ\text{C}$; Supplementary Figure S21A), which is further confirmed by mass spectrometry (Supplementary Figure S19), and is consistent with the NMR structure published by Phan *et al.* (63). Another bulkier structure characterized by an accelerated peak is detected, and the SE-HPLC signa-

ture does not change after two-month incubation at room temperature (Supplementary Figure S24). This species is likely a tetramer, and the relative area of the peak diminishes when the strand concentration is lowered to 10 μM (54–31%), consistent with a multimolecular structure. However, the corresponding *m/z* is not detected in the conditions of mass spectrometry (10 μM strand concentration, 100 mM ammonium acetate; Supplementary Figure S19). Both the dimer and the tetramer can be isolated by SE-HPLC and are stable over a few days (Supplementary Figure S25A), which indicates that the tetramer is likely not formed by a simple end-stacking of two dimers. Interestingly, the tetrameric structure is not formed when potassium is added only after annealing (Supplementary Figure S25B).

Similarly, sequence-dependent polymorphism can be easily detected for various quadruplex-forming oncogene promoter sequences. An interesting example comparison to make is between K-ras 35B1 and K-ras 35B3 that only differ by the sequence of their 3'-terminus (Table 2). K-ras 35B1 forms a mixture of monomer and dimer, which is also observed by native mass spectrometry (Supplementary Figures S26 and S27). However, the absence of the trinucleotide CAG in 3' inhibits almost entirely the formation of dimer.

In the same vein, c-kit1 and c-kit87-up differ solely by the absence or presence, respectively, of an adenosine at the 5'-terminus. While c-kit87up displays a unique peak, c-kit1 gives a triple-peak signal (Supplementary Figure S28A). Based on the predictive plot, the two main peaks are attributed to monomer structures. The accelerated shoulder likely is a fairly unstable dimer since it cannot be detected at lower concentration by mass spectrometry, and since the UV-melting profile is almost concentration independent (Supplementary Figure S21B). Note also that the melting is not fully sigmoidal, most probably due to the presence of a mixture of two monomeric structures. Also worthy of a comparison is the c-kit2/c-kit2GG pair, where the latter lacks two flanking bases, effectively inhibiting the formation of a dimer (Supplementary Figures S27 and S28B). Note that, additionally to the monomer and dimer peaks, c-kit2 gives a peak at a lower retention time that suggests the formation of an additional higher-order structure, not reported so far.

Another extensively studied oncogene promoter of well-known polymorphism, c-myc, shows a high sequence dependency by SE-HPLC, and full results will be detailed elsewhere. Note that with the c-myc sequence studied herein, a small percentage of dimer is detected (<4%; Supplementary Figure S29), which was not the case with an electrophoretic gel mobility shift assay performed at much lower concentration (2 μM), published alongside the NMR structure (51).

An important proportion of the sequences analyzed in this study has been selected and sometimes modified by other research groups in order to obtain clean NMR spectra and ultimately to determine their structure(s). Such sequences yield fairly clean chromatograms where monomers are usually prevalent. However, analysis of unbiased oligonucleotides, such as G4-forming sequences identified *in silico* from genomes, may lead to more complex SEC signatures. This is observed with a number of sequences extracted from different genomes; oligonucleotides fold into complex mixtures, and a number of higher-order structures

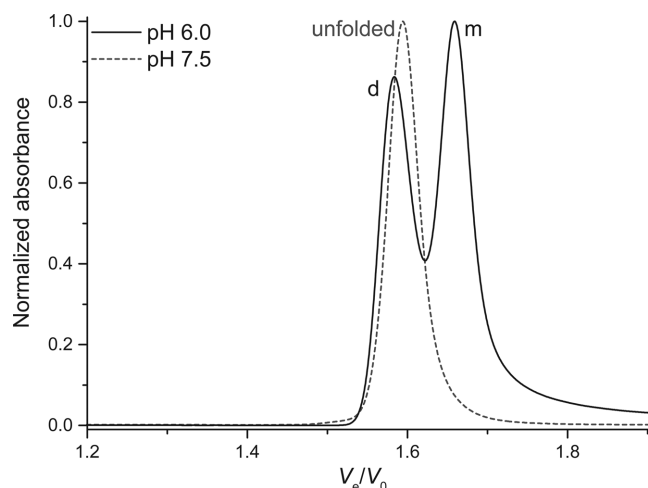


Figure 8. pH dependence: normalized chromatograms of the C-rich sequence from the human c-myc promoter (Py22). Monomer species is indicated with a *m*, dimer with a *d*.

can be detected, notably because they deviate significantly from the normative ‘four repeats of three guanines’ often found in G4-related studies. Two examples taken from the human mitochondrial DNA genome are described in Supplementary data (Supplementary Figure S30). Case studies provide other examples (*vide infra*).

i-motifs

The formation of i-motifs requires a partial protonation of the cytosines. This pH dependence can be easily monitored by SE-HPLC. At near physiological pH (7.5), a single peak is observable for both Py22 and 21CC, corresponding to the unstructured single strands (Figure 8 and Supplementary Figure S31A, respectively). However, at pH 6.0, the strands can fold into a more compact intramolecular i-motif, and the corresponding delayed peak appears. The second peaks observed at pH 6.0 do not strictly elute at the same elution volume than the unfolded controls. The typical i-motif signatures inferred from the IDS analysis of that couple of peaks, together with the elution volumes, reveal that these species are i-motif dimers (Supplementary Figure S31B) (79).

Case studies

RNA stem loops. The Hepatitis C virus genome contains three contiguous stem-loops in its 3'-UTR region (SL1, SL2 and SL3). A recent study suggests that SL2 (r(UCACG₂CUAGCUGUGA₃G₂UC₂GUGA)) might co-exist as a mixture of monomeric and dimeric structures, in rapid exchange, since several bands and smearing were observed by native gel electrophoresis (80). In order to confirm this hypothesis, we studied SL2 using SE-HPLC, making the necessary changes to work in similar conditions: the temperature is set at 10°C and a different eluent is used (see experimental section). Note that the use of lower temperatures leads to higher viscosity of the eluent and in turn to higher pressures. It

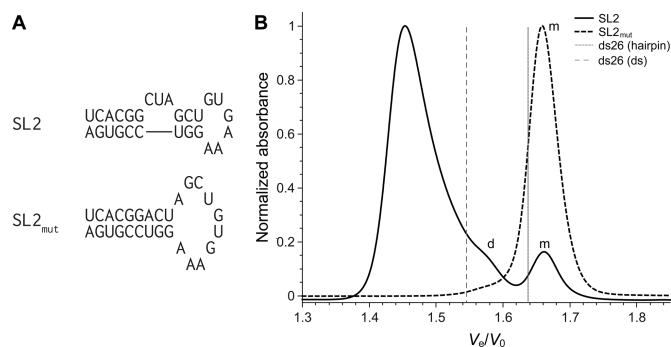


Figure 9. (A) Secondary structures of the RNA stem loops predicted by the Mfold web server, reported by Palau *et al.* (80). (B) Normalized chromatograms of SL2 and SL2_{mut} RNA oligonucleotides (plain lines). Relative elution volume of the duplex and hairpin forms of ds26 are depicted with vertical lines for reference. Monomer species are indicated with a *m* and dimers with a *d*. The major peak for SL2 corresponds to a tetramer or higher-order structure.

might therefore be necessary to lower the flow rate to work at pressures compatible with the stationary phase. A mutant sequence SL2_{mut} that includes an extra adenine to form a perfect hairpin is used as monomeric control (r(UCACG₂ACUAGCUGUGA₃G₂UC₂GUGA); Figure 9A). Clearly, SL2 forms at least three structures ($V_e/V_0 = 1.45, 1.57, 1.66$). The monomeric hairpin form can easily be identified by comparison with SL2_{mut} being the reference monomeric hairpin ($V_e/V_0 = 1.66$) (Figure 9B). The second peak ($V_e/V_0 = 1.57$) gives a PCS value of 1.058, very close to the hairpin/intermolecular value found for ds26 (1.054), and for this reason it is likely that this peak corresponds to a bimolecular species. The FWHM of the fastest SL2 peak is particularly large, which is consistent with a number of unidentified higher-order species co-existing in equilibrium that translates into a smear in native gel electrophoresis.

The RNA hairpin aptamer R06 (r(G₂UCG₂UC₃AGACGAC₂)), specific for the *trans*-activating responsive RNA element of HIV-1 (81), is suspected of forming a dimer but it has not been demonstrated so far. This sequence was therefore examined by SE-HPLC using the same experimental conditions as for SL2. It appears that roughly 30% of the strands associate in a dimeric structure (PCS = 1.046, consistent with the value of the reference ds14: 1.046) (Supplementary Figure S32). It is also observed that when the temperature is raised to 20°C, the proportion of dimer drops dramatically, highlighting once again the influence of experimental conditions on secondary structure formation. It is also interesting to note that there is no shift in relative elution volumes when the temperature is changed, which allows to confidently compare the results.

A bit of concentration is needed. The SP1 binding site in the *c-KIT* promoter, c-kit* (Table 2) was reported as the first non-human telomeric sequence to fold in an antiparallel two-quartet quadruplex (56), despite the fact that three of the four guanine tracts are composed of only two guanines, and the presence of five cytosines that are usually considered as detrimental for quadruplex formation. However, analysis by SE-HPLC reveals that, at high strand concentra-

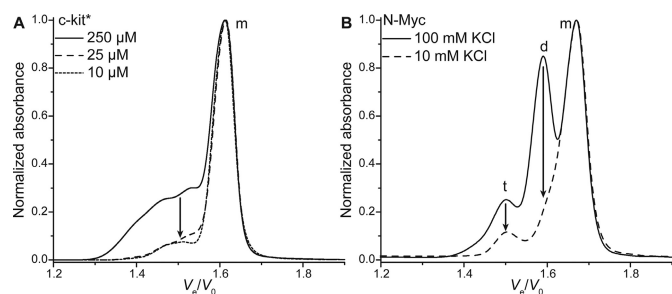


Figure 10. Normalized chromatograms of (A) c-kit* after annealing at various concentrations and (B) N-myc (10 μ M) after annealing with 10 or 100 mM KCl. Monomer species are indicated with a *m*, dimers with a *d* and tetramers with a *t*.

tion (250 μ M), a mixture of bulkier structures co-exist with the expected monomeric quadruplex (Figure 10A). These structures are likely to be multimeric because lower concentrations (25 and 10 μ M) lead to a dramatic decrease of the relative intensity of the related peaks (from 38% to 4.4% and 1.9%, respectively). In order to work with the quadruplex expected to fold in the genome, operating at low concentration is thus advised. More generally, one should keep in mind that the secondary structure(s) of nucleic acids can be altered by their concentration, and SE-HPLC is an efficient method to control the formation of higher-order structures within a relatively large concentration range.

In the same vein, the concentration of monovalent cation can induce significant conformational changes. A guanine-rich sequence from the intron of the N-myc gene has been found to fold into a monomeric and a dimeric quadruplexes (Supplementary Figure S1), in equilibrium, in the presence of physiological concentrations of potassium, but to form preferentially the monomer at lower potassium concentrations (58). Analysis of this equilibrium shifting by SE-HPLC yields comparable results (Figure 10B). When annealed in the presence of 100 mM potassium, two main peaks of comparable intensities corresponding to the monomer and dimer are observed, as well as a minor peak likely to result from the formation of a tetrameric structure. Annealing in a 10-mM potassium containing buffer lead to a dramatic decrease of the dimer peak, consistent with what was found by Trajkovski *et al.* Also, once formed at low potassium concentration, the monomer has a low propensity to convert to the dimer over time, after addition of potassium (Supplementary Figure S33). Although the structural information that can be inferred from SE-HPLC are far less detailed than the ones from NMR, it allows to study a wide range of experimental conditions in a reduced time (\sim 20 min per sample), and at lower nucleic acid concentration. Another example focused on the various structures formed by sequences following the TG_nT pattern ($n = 4$ –20) is given in Supplementary data (Supplementary Figures S34 and S35).

Concluding remarks

SEC is a widely employed analytical method in the synthetic polymer and protein fields but its use in the study of nucleic acid structures remains scarce. The most notable

examples are the resolution of polymorphic telomeric and oncogenic sequences by Miller *et al.* (31,32), and the study of tetramolecular quadruplexes and i-motifs by Bardin and Leroy (16,34). By performing the present study, we have shown that SEC can be used to study nucleic acid structures, with no theoretical limitations regarding their nature. Both DNA and RNA forming a variety of structures are monitored, but one could also imagine studying nucleic acids with other backbones (PNA, LNA, hybrids), modified bases or carrying functionalities. Oligonucleotides carrying both a FAM (6-carboxyfluorescein) and DABCYL (4-((4-(dimethylamino)phenyl)azo)benzoic acid) in 5' and 3', respectively, are analyzed as an example and the dye and quencher are easily detected at low concentration (1 μ M; Supplementary Figure S36).

SE-HPLC allows the easy distinction between unstructured strands, hairpins, intermolecular duplexes and triplexes. Dimerization of hairpins can be readily detected as observed with the case studies on RNA stem loops. Quadruplex multimers can be confidently discriminated from the classical monomers, and SE-HPLC appears to be a powerful tool to study sequence-dependent polymorphism and to disentangle complex multimer mixtures (e.g. the TG_nT case study). The compactness of various structures can be compared, giving global insights into the examined folds. The study of (slow) time-dependent phenomena is also possible as shown with various examples (H-Bi-G4, 93del and N-myc). In the same vein, the association constants of multimolecular quadruplex can be monitored as exemplified by the comparison with the unstable dimeric quadruplex H-Bi-G4 (a significant amount of monomer is detectable), and the exceptionally stable 93del (no monomer is detectable).

Compared to other methods used to study nucleic acid structures, such as ESI-MS, NMR, CD and X-Ray diffraction, SE-HPLC compensates its lack of resolution (structures of close hydrodynamic volumes cannot be discriminated) by its ease and speed of use, and most notably by its versatility. We have shown that a number of experimental parameters can be tuned to work in conditions relevant to a given study. Thus, the eluent (potassium-, sodium- or ammonium-rich buffers for quadruplexes, acidic buffer for i-motifs and triplexes, and magnesium-containing buffers for ps-ds, triplex and RNA stem loops), the temperature and the sample concentration are successfully changed to obtain the desired conditions. This is of the utmost importance since these parameters can dramatically affect the structure of nucleic acids, as inferred from the various case studies. The concentration range available (μ M–mM) is particularly interesting because it allows to easily compare results obtained by techniques operating at high (NMR) and low (ESI-MS, CD, UV-melting) concentrations.

As mentioned previously, SE-HPLC displays a relatively low resolution compared to other techniques such as NMR or native mass spectrometry. The resolution can theoretically be increased by coupling multiple columns in series. However, complete resolution of the peaks will result in non-equilibrium conditions that might in turn lead to structure interconversions and/or dissociations, owing to the low eluent flows typically used. Ion mobility mass spectrometry is a technique that can achieve greater resolution, albeit in

the gas phase (82), and it should be noted that SE-HPLC is theoretically amenable to mass spectrometry coupling. Ultracentrifugation is another alternative that has been shown to be effective for the determination of hydrodynamic properties of macromolecules, as exemplified with the 21-mer human telomeric sequence by Chaires *et al.* (83).

Compared to the widely used PAGE, SE-HPLC (i) does not depend from the m/z ratio of the structures and, as a corollary, is not affected by charge screening effects; (ii) provides a reliable molecular weight of the species; (iii) allows reliable quantifications; (iv) provides a spectroscopic characterization of each species (UV-vis spectrum, IDS); (v) is an easy mean of isolation and purification of the different species (providing that they are not interconverting); (vi) allows a fine and easy tuning of the experimental conditions (buffer, pH, salt, temperature, etc.) and (vii) can handle a wide range of oligonucleotide quantities using adapted columns.

For all the reasons mentioned above, we believe that SE-HPLC can nicely complement NMR, crystallography, CD or ESI-MS experiments. From a broader point of view, this work highlights the intricacies of working with nucleic acids, and most notably hairpin- and quadruplex-forming sequences. Small changes in sequence and/or conditions (buffer pH, cation nature and concentration, strand concentration, temperature, incubation time, etc.) can lead to drastic modifications of the structure(s), which in turn can alter the results of a study. Moreover, available prediction algorithms typically lead to a number of false positive and negative results and fail to predict the formation of multimolecular structures. In a genomic context, such structures are unlikely to appear, except maybe for highly repeated sequences (e.g. telomeres). However, when studying predicted quadruplex-forming sequences using oligonucleotides, such structures might appear and hence skew the results. Therefore, each new sequence structure should be evaluated in the specific conditions of the study it will be conducted in, and we believe that SE-HPLC is a quick and easy method to do so.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

E.L. thanks A. Marchand for his help with native mass spectrometry, W. Palau and Dr C. Di Primo for the kind gift of SL2, SL2_{mut} and R06 oligonucleotides and subsequent discussions, Dr V. Gabelica for the gift of TG_nT oligonucleotides and fruitful discussions, Dr S. Amrane and Dr V. D'Atri for their insights into quadruplex structures and Dr M. Renders for proofreading of the manuscript. J.L.M. dedicates this manuscript to the memory of Jean-Louis Leroy, who pioneered the use of SEC for unusual nucleic acid structures.

FUNDING

Agence Nationale de la Recherche (OligoSwitch [ANR-12-IS07-0001], 'Quarpediem' [ANR-12-BSV8-0008-01], 'VIBBnano' [ANR-10-NANO-04-03]). Funding for open access charge: Inserm.

Conflict of interest statement. None declared.

REFERENCES

- Choi, J. and Majima, T. (2011) Conformational changes of non-B DNA. *Chem. Soc. Rev.*, **40**, 5893–5909.
- Rich, A. (1993) DNA comes in many forms. *Gene*, **135**, 99–109.
- Haran, T.E. and Mohanty, U. (2009) The unique structure of A-tracts and intrinsic DNA bending. *Q. Rev. Biophys.*, **42**, 41–81.
- Ramsing, N.B. and Jovin, T.M. (1988) Parallel stranded duplex DNA. *Nucleic Acids Res.*, **16**, 6659–6676.
- Vasquez, K.M. and Glazer, P.M. (2002) Triplex-forming oligonucleotides: principles and applications. *Q. Rev. Biophys.*, **35**, 89–107.
- Svoboda, P. and Di Cara, A. (2006) Hairpin RNA: a secondary structure of primary importance. *Cell. Mol. Life Sci.*, **63**, 901–908.
- Collie, G.W. and Parkinson, G.N. (2011) The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.*, **40**, 5867–5892.
- Phan, A.T. (2010) Human telomeric G-quadruplex: structures of DNA and RNA sequences. *FEBS J.*, **277**, 1107–1117.
- Davis, J.T. (2004) G-quartets 40 years later: from 5'-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed.*, **43**, 668–698.
- Mergny, J.-L. (2012) Alternative DNA structures: G4 DNA in cells: itae missa est? *Nat. Chem. Biol.*, **8**, 225–226.
- Lipps, H.J. and Rhodes, D. (2009) G-quadruplex structures: in vivo evidence and function. *Trends Cell Biol.*, **19**, 414–422.
- Palacky, J., Vorlickova, M., Kejnovska, I. and Mojzes, P. (2012) Polymorphism of human telomeric quadruplex structure controlled by DNA concentration: a Raman study. *Nucleic Acids Res.*, **41**, 1005–1016.
- Le, H.T., Miller, M.C., Buscaglia, R., Dean, W.L., Holt, P.A., Chaires, J.B. and Trent, J.O. (2012) Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4. *Org. Biomol. Chem.*, **10**, 9393–9404.
- Dai, J., Carver, M. and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172–1183.
- Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.*, **35**, 7429–7455.
- Bardin, C. and Leroy, J.L. (2008) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.*, **36**, 477–488.
- Mergny, J.L., De Cian, A., Ghelab, A., Saccà, B. and Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res.*, **33**, 81–94.
- Guéron, M. and Leroy, J.-L. (2000) The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.*, **10**, 326–331.
- Agarwala, P., Pandey, S. and Maiti, S. (2013) G-quadruplexes as tools for synthetic biology. *ChemBioChem*, **14**, 2077–2081.
- Belmont, P., Constant, J.F. and Demeunynck, M. (2001) Nucleic acid conformation diversity: from structure to function and regulation. *Chem. Soc. Rev.*, **30**, 70–81.
- Adrian, M., Heddi, B. and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11–24.
- Neidle, S. and Parkinson, G.N. (2008) Quadruplex DNA crystal structures and drug design. *Biochimie*, **90**, 1184–1196.
- Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264–277.
- Vorlickova, M., Kejnovska, I., Sagi, J., Renciuik, D., Bednarova, K., Motlova, J. and Kypr, J. (2012) Circular dichroism and guanine quadruplexes. *Methods*, **57**, 64–75.
- Kypr, J., Kejnovska, I., Renciuik, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.*, **37**, 1713–1725.
- Striegel, A.M., Yau, W.W., Kirkland, J.J. and Bly, D.D. (2009) *Modern Size-Exclusion Liquid Chromatography*. John Wiley & Sons, Inc., Hoboken, NJ.
- Wang, Y., Teraoka, I., Hansen, F.Y., Peters, G.H. and Hassager, O. (2010) A theoretical study of the separation principle in size exclusion chromatography. *Macromolecules*, **43**, 1651–1659.

28. Sun, T., Chance, R.R., Graessley, W.W. and Lohse, D.J. (2004) A study of the separation principle in size exclusion chromatography. *Macromolecules*, **37**, 4304–4312.
29. Brooks, D.E., Haynes, C.A., Hritcu, D., Steels, B.M. and Muller, W. (2000) Size exclusion chromatography does not require pores. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 7064–7067.
30. Yoshio, K. and Shigeru, N. (2003) Size exclusion chromatography of nucleic acids. In: Wu, C-S (ed) *Handbook of Size Exclusion Chromatography and Related Techniques: Revised and Expanded*. Vol. 91, CRC Press, Boca Raton, FL, pp. 716.
31. Miller, M.C., Le, H.T., Dean, W.L., Holt, P.A., Chaires, J.B. and Trent, J.O. (2011) Polymorphism and resolution of oncogene promoter quadruplex-forming sequences. *Org. Biomol. Chem.*, **9**, 7633–7637.
32. Miller, M.C. and Trent, J.O. (2011) Resolution of quadruplex polymorphism by size-exclusion chromatography. *Curr. Protoc. Nucleic Acid Chem.*, **45**, 17.3.1–17.3.18.
33. Dailey, M.M., Miller, M.C., Bates, P.J., Lane, A.N. and Trent, J.O. (2010) Resolution and characterization of the structural polymorphism of a single quadruplex-forming sequence. *Nucleic Acids Res.*, **38**, 4877–4888.
34. Leroy, J.-L. (2009) The formation pathway of i-motif tetramers. *Nucleic Acids Res.*, **37**, 4127–4134.
35. Denisov, A.Y., Zamaratski, E.V., Maltseva, T.V., Sandström, A., Bekiroglu, S., Altmann, K.H., Egli, M. and Chattopadhyaya, J. (1998) The solution conformation of a carbocyclic analog of the Dickerson-Drew dodecamer: comparison with its own X-ray structure and that of the NMR structure of the native counterpart. *J. Biomol. Struct. Dyn.*, **16**, 547–568.
36. Valls, N., Wright, G., Steiner, R.A., Murshudov, G.N. and Subirana, J.A. (2004) DNA variability in five crystal structures of d(CGCAATTGCG). *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 680–685.
37. Stefl, R., Wu, H., Ravindranathan, S., Sklenár, V. and Feigon, J. (2004) DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc. Natl Acad. Sci. U. S. A.*, **101**, 1177–1182.
38. Lim, K.W. and Phan, A.T. (2013) Structural basis of DNA quadruplex-duplex junction formation. *Angew. Chem. Int. Ed. Engl.*, **52**, 8566–8569.
39. Granzhan, A., Largy, E., Saettel, N. and Teulade-Fichou, M.-P. (2010) Macrocyclic DNA-mismatch-binding ligands: structural determinants of selectivity. *Chemistry*, **16**, 878–889.
40. Macdonald, D., Herbert, K., Zhang, X., Polgruto, T. and Lu, P. (2001) Solution structure of an A-tract DNA bend. *J. Mol. Biol.*, **306**, 1081–1098.
41. Yatsunyk, L.A., Pietrement, O., Albrecht, D., Tran, P.L.T., Renciuik, D., Sugiyama, H., Arbona, J.M., Aimé, J.P. and Mergny, J.L. (2013) Guided assembly of tetramolecular g-quadruplexes. *ACS Nano*, **7**, 5701–5710.
42. Sugimoto, N., Wu, P., Hara, H. and Kawamoto, Y. (2001) pH and cation effects on the properties of parallel pyrimidine motif DNA triplexes. *Biochemistry*, **40**, 9396–9405.
43. Luu, K.N., Phan, A.T., Kuryavii, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K⁺ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
44. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
45. Renciuik, D., Kejnovska, I., Skolakova, P., Bednarova, K., Motlova, J. and Vorlickova, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K⁺ solutions. *Nucleic Acids Res.*, **37**, 6625–6634.
46. Gray, R.D., Petraccone, L., Trent, J.O. and Chaires, J.B. (2010) Characterization of a K⁺-induced conformational switch in a human telomeric DNA oligonucleotide using 2-aminopurine fluorescence. *Biochemistry*, **49**, 179–194.
47. Miller, M.C., Buscaglia, R., Chaires, J.B., Lane, A.N. and Trent, J.O. (2010) Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.*, **132**, 17105–17107.
48. Phan, A.T., Kuryavii, V., Luu, K.N. and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K⁺ solution. *Nucleic Acids Res.*, **35**, 6517–6525.
49. Yue, D.J., Lim, K.W. and Phan, A.T. (2011) Formation of (3+1) G-quadruplexes with a long loop by human telomeric DNA spanning five or more repeats. *J. Am. Chem. Soc.*, **133**, 11462–11465.
50. Wang, Y. and Patel, D.J. (1995) Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. *J. Mol. Biol.*, **251**, 76–94.
51. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
52. Waller, Z.A., Sewitz, S.A., Hsu, S.T. and Balasubramanian, S. (2009) A small molecule that disrupts G-quadruplex DNA structure and enhances gene expression. *J. Am. Chem. Soc.*, **131**, 12628–12633.
53. Tran, P.L.T., Largy, E., Hamon, F., Teulade-Fichou, M.-P. and Mergny, J.-L. (2011) Fluorescence intercalator displacement assay for screening G4 ligands towards a variety of G-quadruplex structures. *Biochimie*, **93**, 1288–1296.
54. Phan, A.T., Kuryavii, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.*, **129**, 4386–4392.
55. Kuryavii, V., Phan, A.T. and Patel, D.J. (2010) Solution structures of all parallel-stranded monomeric and dimeric G-quadruplex scaffolds of the human c-kit2 promoter. *Nucleic Acids Res.*, **38**, 6757–6773.
56. Raiber, E.A., Kranaster, R., Lam, E., Nikan, M. and Balasubramanian, S. (2012) A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.*, **40**, 1499–1508.
57. Wei, D., Todd, A.K., Zloh, M., Gunaratnam, M., Parkinson, G.N. and Neidle, S. (2013) Crystal structure of a promoter sequence in the B-raf gene reveals an intertwined dimer quadruplex. *J. Am. Chem. Soc.*, **135**, 19319–19329.
58. Trajkovski, M., Webba da Silva, M. and Plavec, J. (2012) Unique structural features of interconverting monomeric and dimeric G-quadruplexes adopted by a sequence from the intron of the N-myc gene. *J. Am. Chem. Soc.*, **134**, 4132–4141.
59. Kuryavii, V., Cahoon, L.A., Seifert, H.S. and Patel, D.J. (2012) RecA-binding pilE G4 sequence essential for pilin antigenic variation forms monomeric and 5' end-stacked dimeric parallel G-quadruplexes. *Structure*, **20**, 2090–2102.
60. Rehm, C., Holder, I.T., Groß, A., Wojciechowski, F., Urban, M., Sinn, M., Drescher, M. and Hartig, J.S. (2014) A bacterial DNA quadruplex with exceptional K⁺ selectivity and unique structural polymorphism. *Chem. Sci.*, **5**, 2809–2809.
61. Mukundan, V.T. and Phan, A.T. (2013) Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.*, **135**, 5017–5028.
62. Do, N.Q. and Phan, A.T. (2012) Monomer-dimer equilibrium for the 5'-5' stacking of propeller-type parallel-stranded G-quadruplexes: NMR structural study. *Chemistry*, **18**, 14752–14759.
63. Phan, A.T., Kuryavii, V., Ma, J.-B., Faure, A., Andréola, M.-L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. *Proc. Natl Acad. Sci. U. S. A.*, **102**, 634–639.
64. Do, N.Q., Lim, K.W., Teo, M.H., Heddi, B. and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res.*, **39**, 9448–9457.
65. Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J.-L. and Phan, A.T. (2012) Formation of pearl-necklace monomeric G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.*, **134**, 5807–5816.
66. Schultze, P., Macaya, R.F. and Feigon, J. (1994) Three-dimensional solution structure of the thrombin-binding DNA aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Biol.*, **235**, 1532–1547.
67. Balkwill, G.D., Garner, T.P., Williams, H.E. and Searle, M.S. (2009) Folding topology of a bimolecular DNA quadruplex containing a stable mini-hairpin motif within the diagonal loop. *J. Mol. Biol.*, **385**, 1600–1615.
68. Creze, C., Rinaldi, B., Haser, R., Bouvet, P. and Gouet, P. (2007) Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li⁺ ions. *Acta Crystallogr. D Biol. Crystallogr.*, **63**, 682–688.
69. Phan, A.T. and Leroy, J.L. (2000) Intramolecular i-motif structures of telomeric DNA. *J. Biomol. Struct. Dyn.*, **17**, 245–251.
70. Dai, J., Hatzakis, E., Hurley, L.H. and Yang, D. (2010) I-motif structures formed in the human c-MYC promoter are highly

- dynamic—insights into sequence redundancy and I-motif stability. *PLoS One*, **5**, e11647–e11647.
71. Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R. and Oyoshi, T. (2013) Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem. Biol.*, **20**, 341–350.
72. Di Marco, V.B. and Bombi, G.G. (2001) Mathematical functions for the representation of chromatographic peaks. *J. Chromatogr. A*, **931**, 1–30.
73. Hartung, J., Knapp, G. and Sinha, B.K. (2008) Various measures of effect size. In *Statistical Meta-Analysis with Applications*. John Wiley & Sons, Inc., Hoboken, NJ, pp. 13–23.
74. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.
75. Mergny, J.L. and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
76. Buscaglia, R., Gray, R.D. and Chaires, J.B. (2013) Thermodynamic characterization of human telomere quadruplex unfolding. *Biopolymers*, **99**, 1006–1018.
77. Xu, Y., Sato, H., Sannohe, Y., Shinohara, K. and Sugiyama, H. (2008) Stable lariat formation based on a G-quadruplex scaffold. *J. Am. Chem. Soc.*, **130**, 16470–16471.
78. Zhang, N., Phan, A.T. and Patel, D.J. (2005) (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.*, **127**, 17277–17285.
79. Mergny, J.L., Li, J., Lacroix, L., Amrane, S. and Chaires, J.B. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
80. Palau, W., Masante, C., Ventura, M. and Di Primo, C. (2013) Direct evidence for RNA-RNA interactions at the 3' end of the Hepatitis C virus genome using surface plasmon resonance. *RNA*, **19**, 982–991.
81. Lebars, I., Legrand, P., Aimé, A., Pinaud, N., Fribourg, S. and Di Primo, C. (2008) Exploring TAR-RNA aptamer loop-loop interaction by X-ray crystallography, UV spectroscopy and surface plasmon resonance. *Nucleic Acids Res.*, **36**, 7146–7156.
82. Ferreira, R., Marchand, A. and Gabelica, V. (2012) Mass spectrometry and ion mobility spectrometry of G-quadruplexes. A study of solvent effects on dimer formation and structural transitions in the telomeric DNA sequence d(TAGGGTTAGGGT). *Methods*, **57**, 56–63.
83. Li, J., Correia, J.J., Wang, L., Trent, J.O. and Chaires, J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res.*, **33**, 4649–4659.