Supplementary data

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

Eric Largy[1,2] and Jean-Louis Mergny[1,2,*]

[1] Univ. Bordeaux, Bordeaux, 33000, France
[2] INSERM, U869, IECB, ARNA Laboratory, Pessac, 33600, France

* To whom correspondence should be addressed. Tel: +33 (0)5 4000 3022; Email: jean-louis.mergny@inserm.fr

QUADRUPLEX POLYMORPHISM

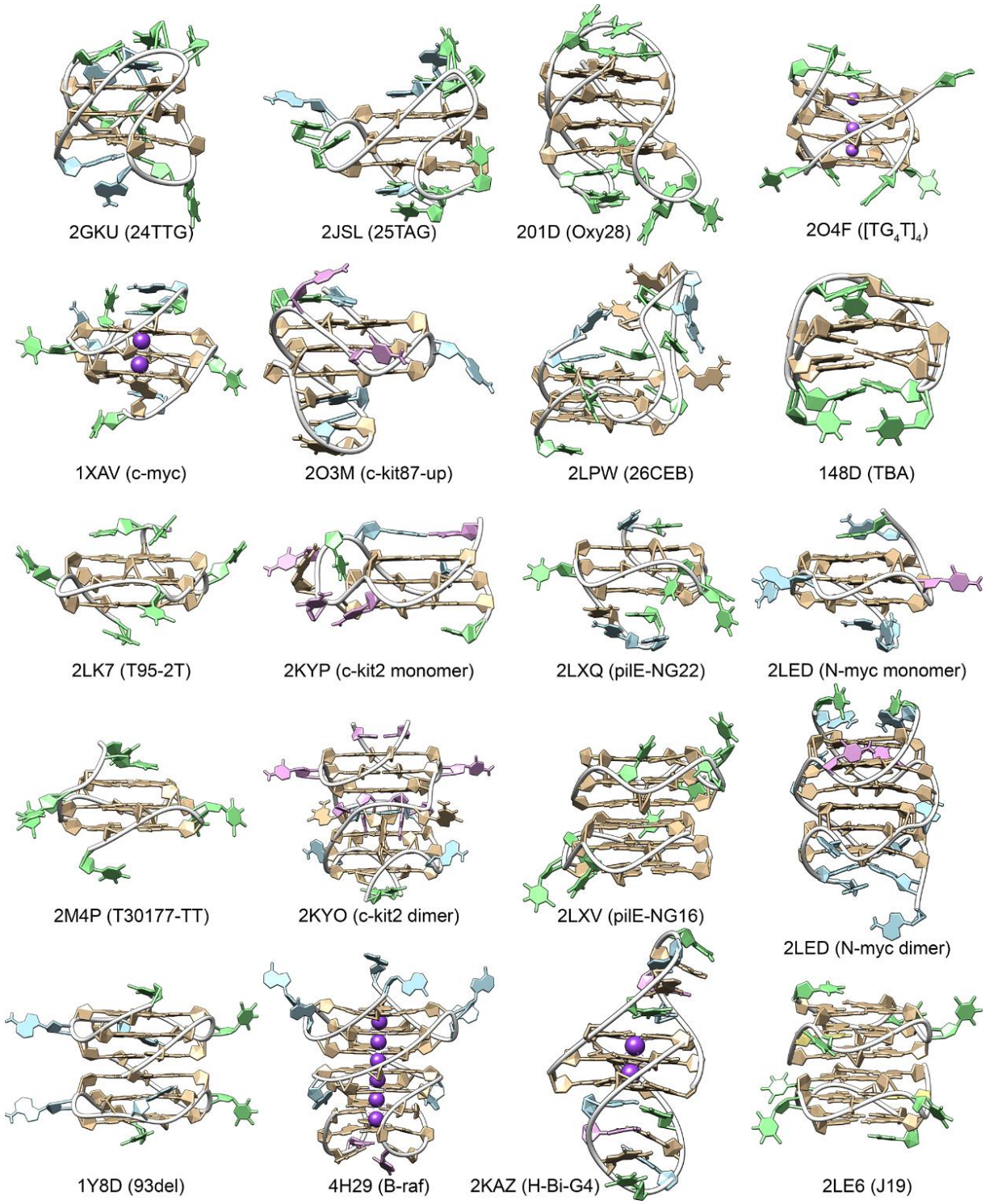

Figure S1. Examples of quadruplex structures deposited in the PDB that were studied by SE-HPLC: 2GKU (1), 2JSL (2), 201D (3), 2O4F (4), 1XAV (5), 2O3M (6), 2LPW (7), 148D (8), 2LK7 (9), 2KYP and 2KYO (10), 2LXQ and 2LXV (11), 2LED and 2LEE (12), 2M4P (13), 1Y8D (14), 4H29 (15), 2KAZ (16), and 2LE6 (17). Guanosines are depicted in brown, inosine in yellow, adenosines in blue, thymidines in green, cytidines in pink, the phosphate backbones as white ribbons, and potassiums as purple spheres where available. This figure was prepared using UCSF Chimera (18).

PEAK FITTING

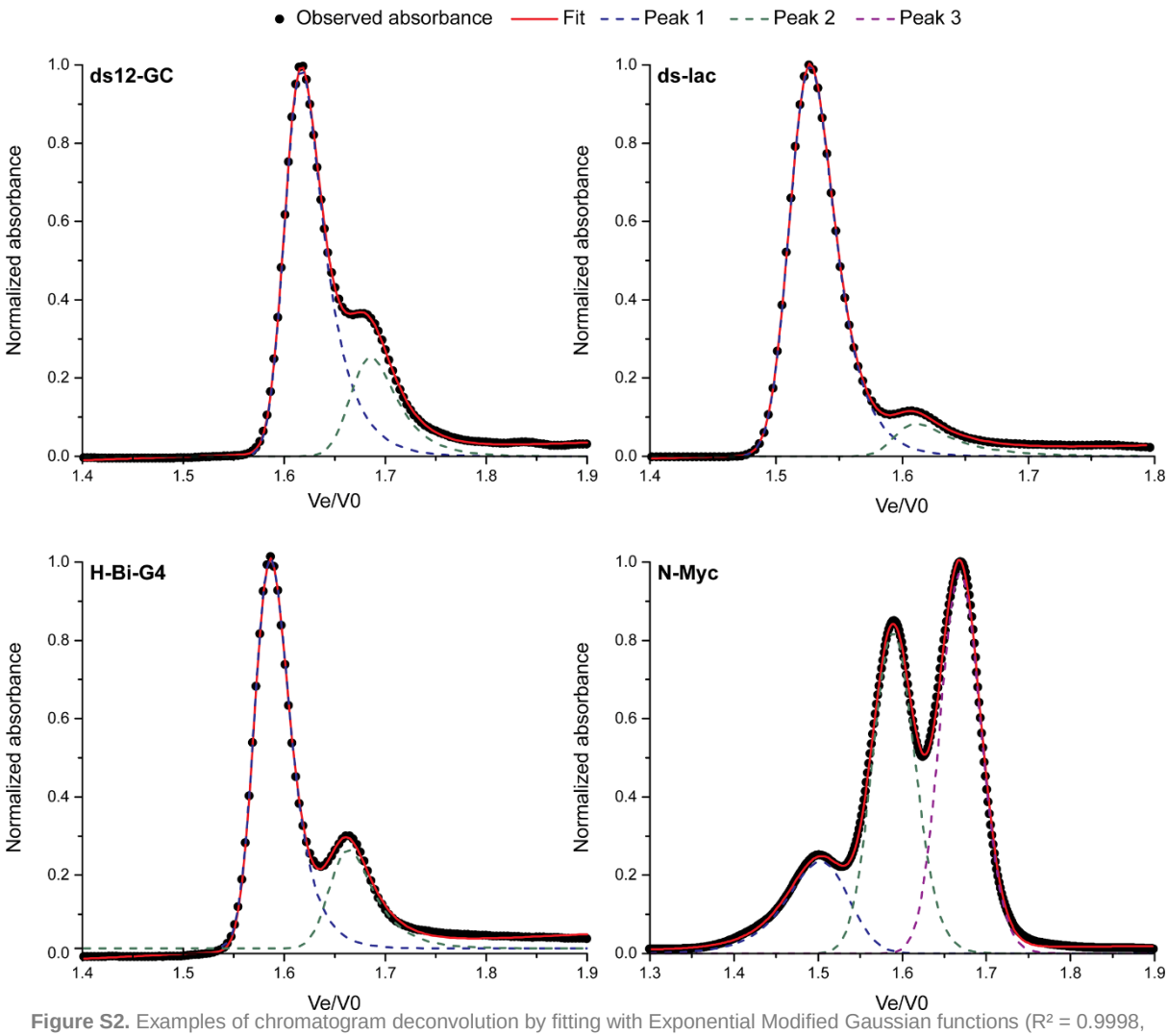

Figure S2. Examples of chromatogram deconvolution by fitting with Exponential Modified Gaussian functions (R^2 = 0.9998, 0.9999, 0.9994, 0.9994, from left to right, top to bottom).

CIRCULAR DICHROISM

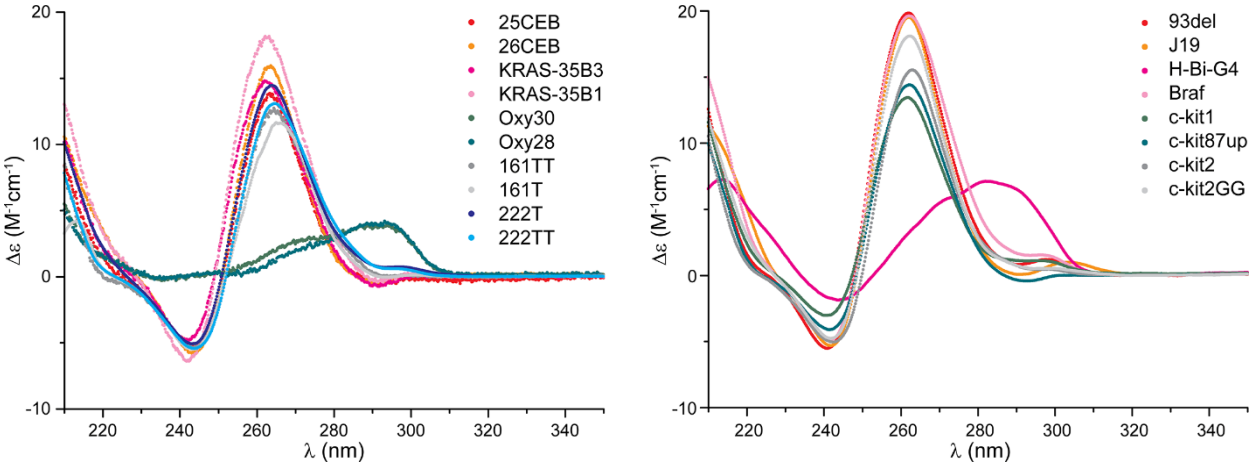

Figure S3. Circular dichroism spectra of selected sequences given in molar dichroic absorption ($\Delta\varepsilon$) based on nucleoside concentration.

TECHNICAL CONSIDERATIONS

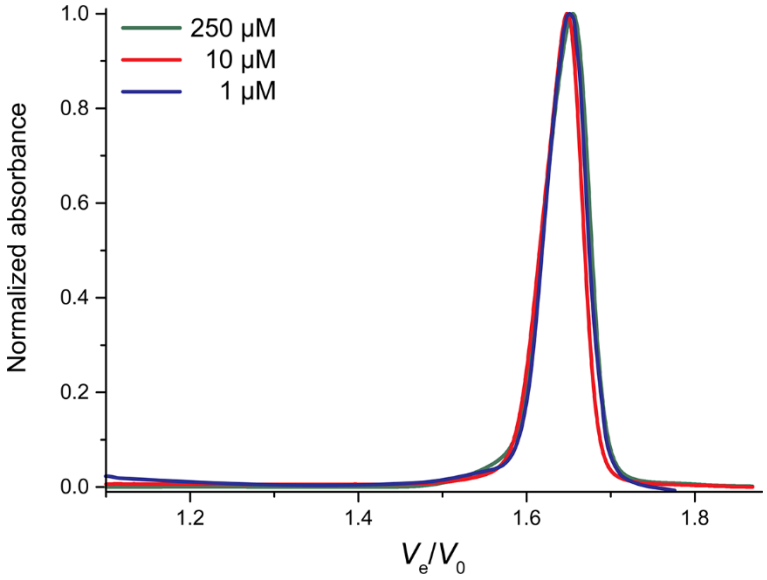

Figure S4. Normalized chromatograms obtained for solutions containing different concentrations (250, 10 and 1 µM) of a 21-mer oligonucleotide (21GG; ε = 215000 M^{-1} cm^{-1}).

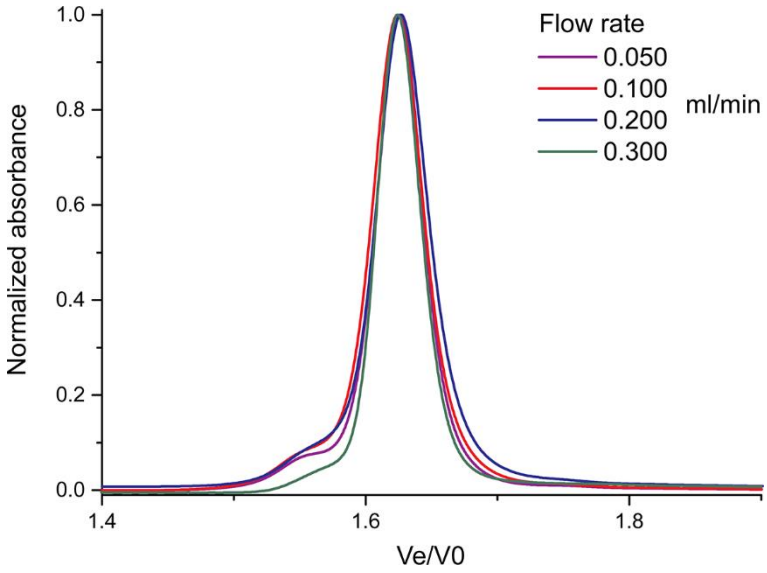

Figure S5. Normalized chromatograms of c-myc obtained at different flow rates.

SINGLE STRANDED DNA

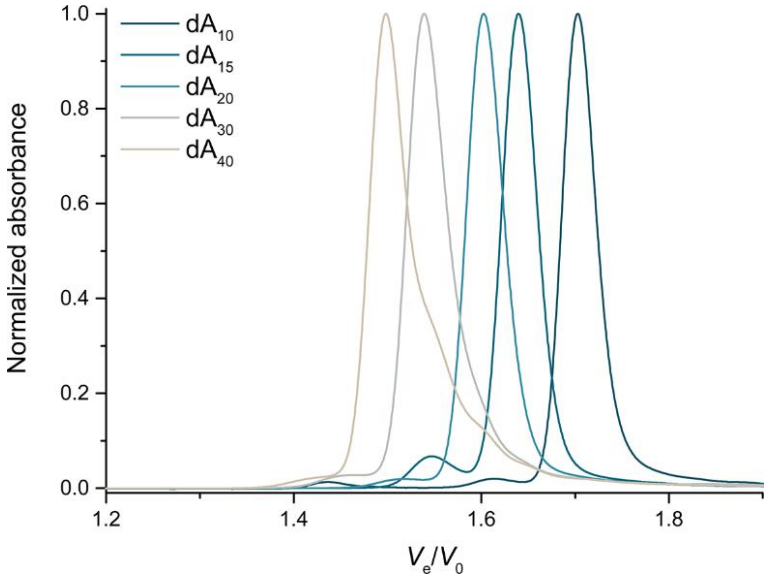

Figure S6. Normalized chromatograms of polypurines d(A_n) oligonucleotides.

HAIRPIN-DUPLEX INTERCONVERSION

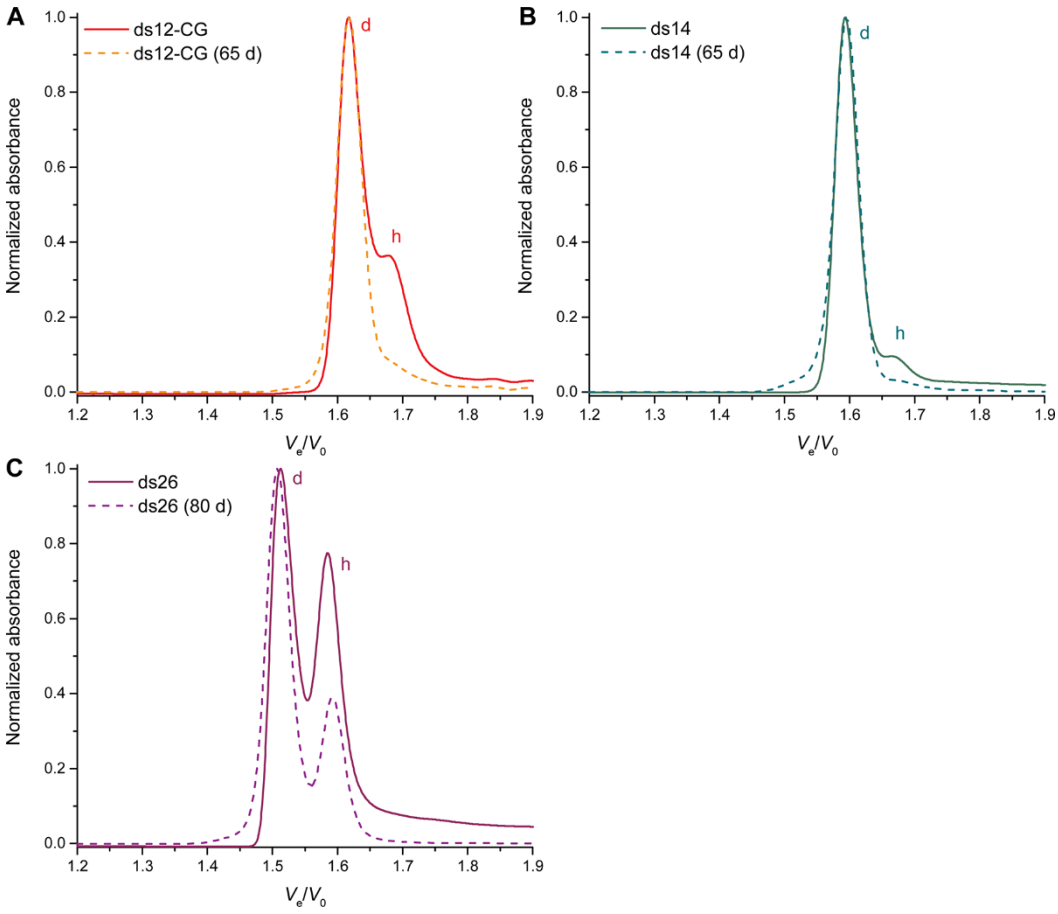

Figure S7. Normalized chromatograms of autocomplementary oligonucleotides within a day from annealing (plain lines), and after a long incubation (65 to 80 days) at room temperature (dashed lines). Monomer is indicated with a *m* and dimer with a *d*.

MISMATCHED DUPLEXES

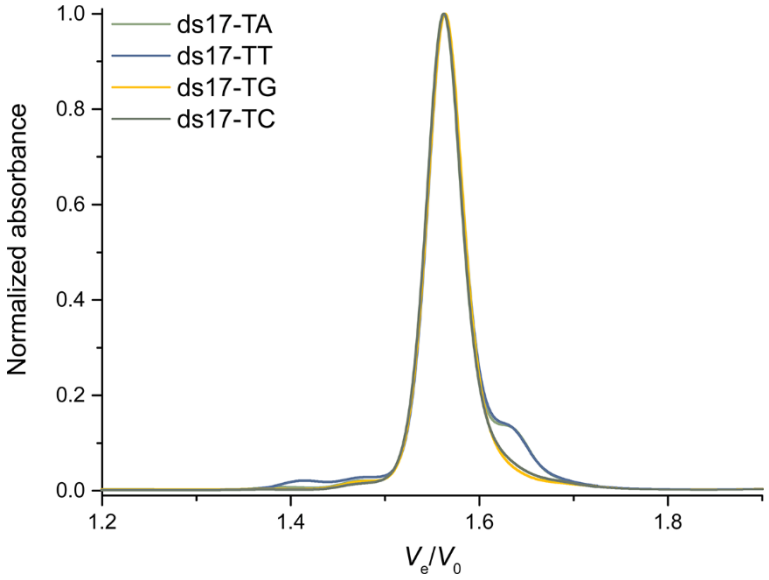

Figure S8. Normalized chromatograms of the heteroduplex ds17-TA, and single-mismatched counterparts ds17-TT, ds17-TG, and ds17-TC.

A-TRACT BENDING

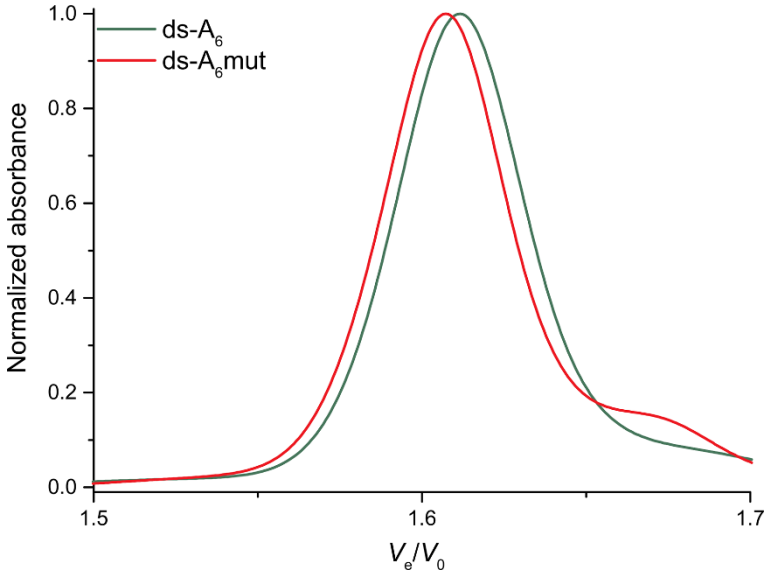

Figure S9. Normalized chromatograms of the heteroduplexes ds-A_6 and ds-A_6mut.

PARALLEL-STRANDED DUPLEXES

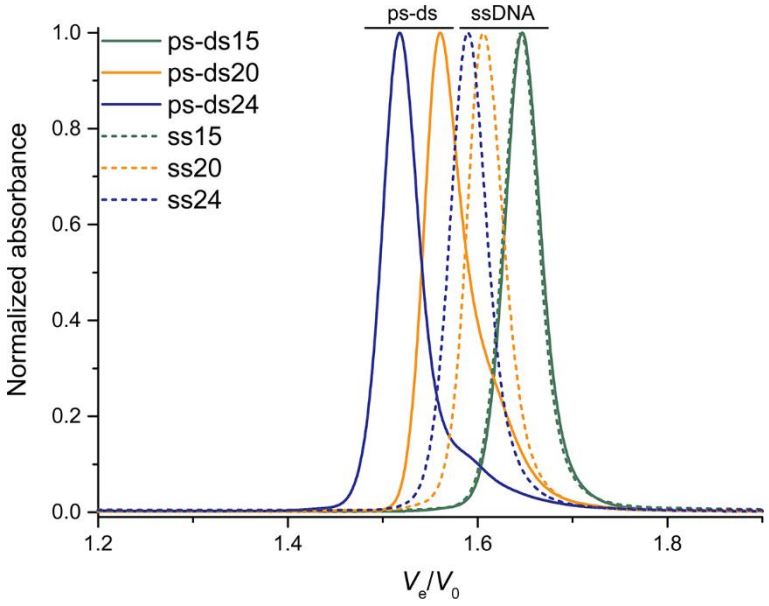

Figure S10. Normalized chromatograms of parallel-stranded heteroduplex forming oligonucleotides (plain lines), and single-stranded sequences prepared in absence of complementary sequence (dashed lines), at a 100-µM strand concentration. *ps-ds* and *ssDNA* labels indicate the successful formation of parallel-stranded duplex and the presence of unstructured single strand, respectively.

TRIPLEX

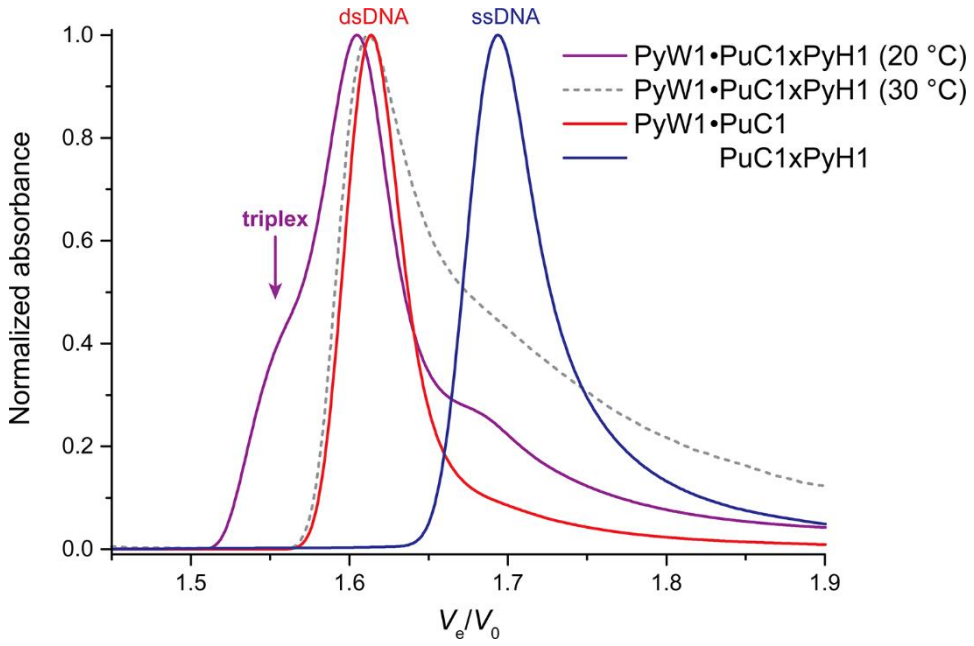

Figure S11. Normalized chromatograms of the triplex structure PyW1•PuC1xPyH1 (at 20 and 30 °C), the Watson-Crick duplex PyW1•PuC1, and a mixture of PuC1 and PyH1 (the Hoogsteen duplex is not formed under these conditions), in MES 50 mM, pH 6.0, supplemented with $MgCl_2$ (10 mM).

HUMAN TELOMERIC MUTANTS

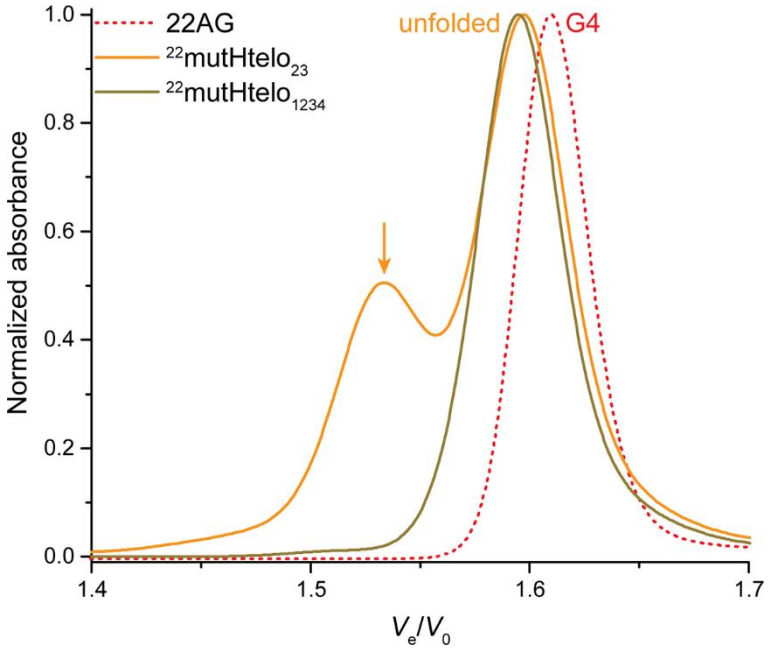

Figure S12. Normalized chromatograms of 22-mer mutants of the human telomeric sequence (plain lines), compared to wild-type sequences (dashed lines), at a 100-µM strand concentration. The arrow indicates the presence of a multimolecular quadruplex.

QUADRUPLEX CHROMATOGRAMS

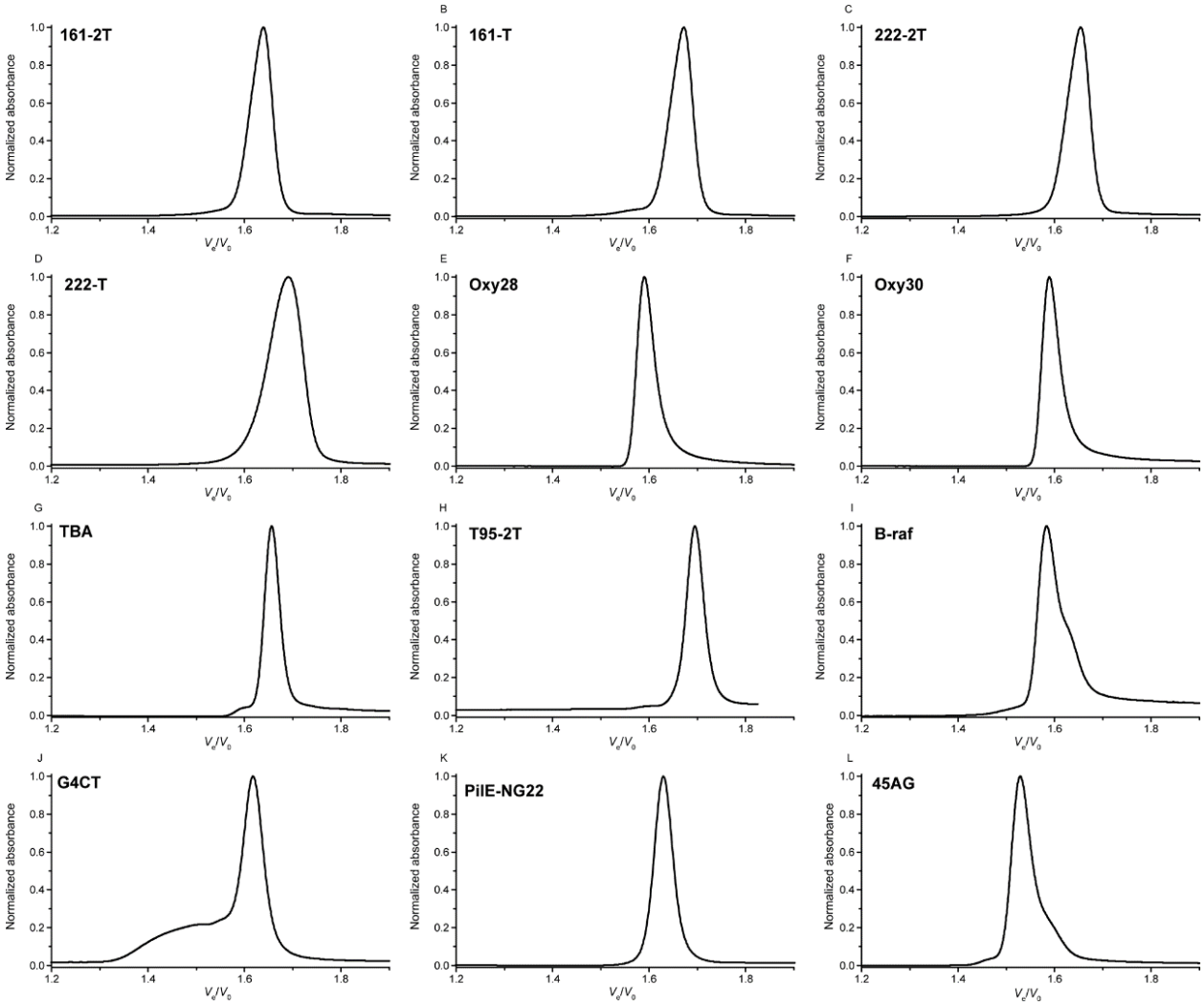

Figure S13. Normalized chromatogram of quadruplex-forming oligonucleotides (250 µM-strand concentration, 100 mM KCl, pH 7.5)

SI DISTRIBUTION STATISTICAL ANALYSIS

In order to confirm that the sample size is large enough to significantly discuss about distribution differences, the Cohen's effect size was calculated using the pooled standard deviation *S*:

d = 3.79, with S = 0.063432

The Mann–Whitney U test was performed on all possible pairs of distributions (null hypothesis: distributions are not different).

Distribution pairs	n	U	Z	exact P	asymptotic P	decision at the 0.01 level
Monomer-dimer	28; 12	336	4.94361	3.58E-10	7.67E-7	significantly different
Tetramer-dimer	10; 12	120	3.92331	3.09E-6	8.73E-5	significantly different
Monomer-tetramer	28; 10	280	4.62438	4.23E-9	3.76E-6	significantly different

Additionally, the Wilcoxon signed-rank test was conducted on data points originating from oligonucleotides giving both a monomer and a dimer peak (null hypothesis: distributions are not different).

Distribution pairs	n	W	Z	exact P	asymptotic P	decision at the 0.01 level
Monomer-dimer[a]	10; 10	55	2.75209	0.00195	0.00592	significantly different

[a] Only for oligonucleotides giving both a monomer and dimer peak.

G4-DNA COMES IN A VARIETY OF SHAPES

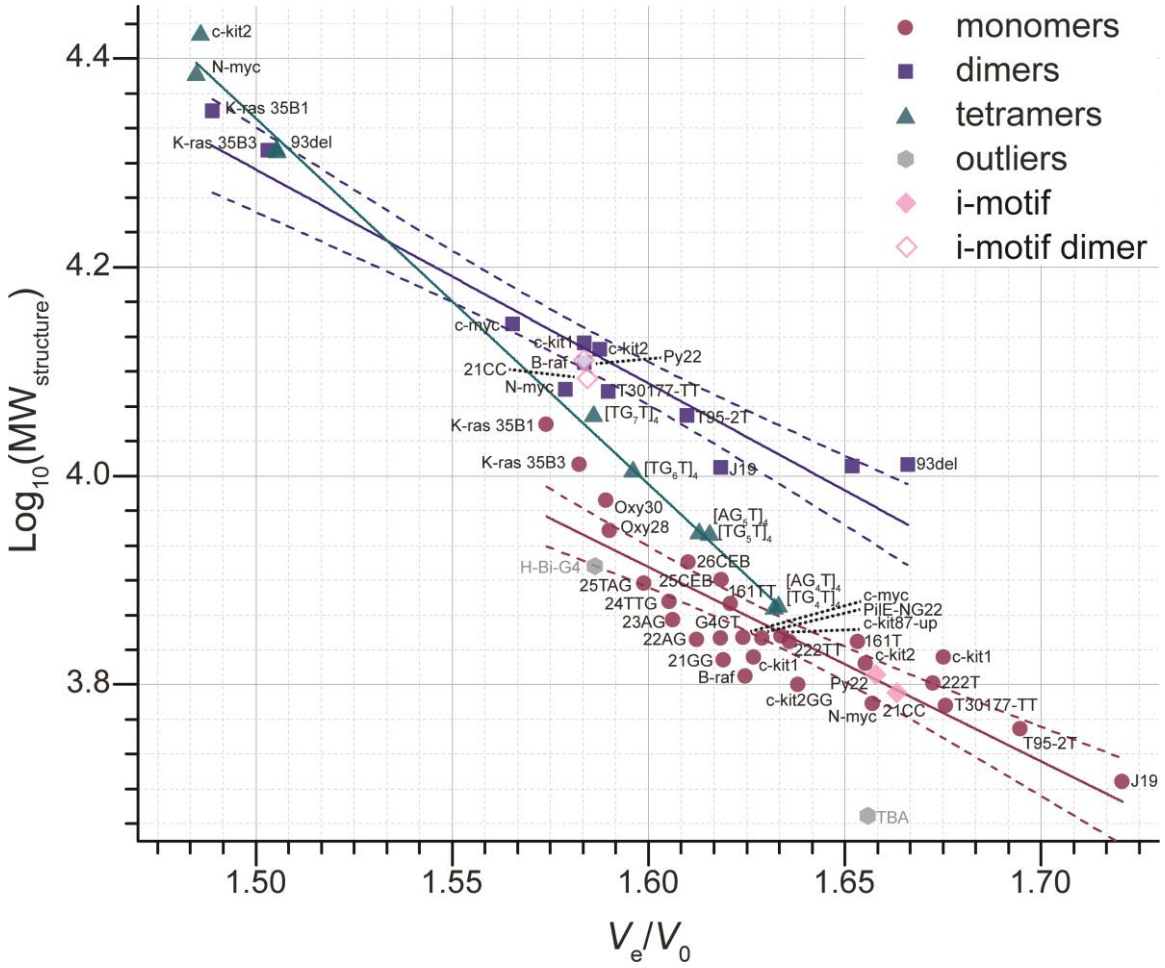

Figure S14. Compactness plot of the $\log_{10}$(MW) against relative elution volume for monomeric (purple), dimeric (blue), and tetrameric (green) quadruplex-forming oligonucleotides, with independent linear fitting (solid line; dashed lines: 95% confidence bands). TBA (only 2 G-quartets) and H-Bi-G4 (a bimolecular G4 with only 3 quartets, hence behaving like a monomer) are trivial outliers. I-motifs are depicted as pink diamonds.

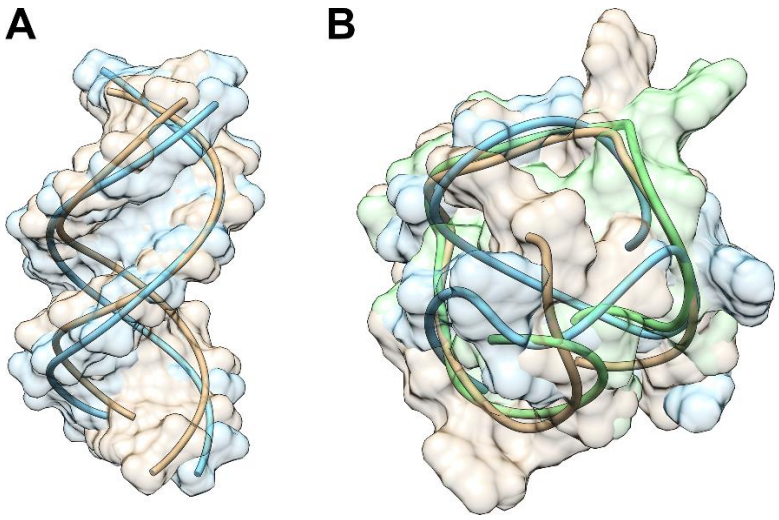

Figure S15. A) Superimposition of two 12-bp ds-DNA structures (1RVI and 1RVH, in blue and tan respectively), B) superimposition of three 22-mer G4-DNA structures (tan: 1XAV, blue: 2O3M, green: 2LXQ). Only backbones and surfaces are depicted.

RELATIONSHIP BETWEEN THE RADIUS OF GYRATION AND THE ELUTION VOLUME

The script used in Chimera to determine radii of gyration is given below:

```
from chimera import Point, openModels, Molecule, sqdistance
from math import sqrt
for m in openModels.list(modelTypes=[Molecule]):
    masses = [a.element.mass for a in m.atoms]
    com = Point([a.coord() for a in m.atoms], masses)
    sum_sq = 0.0
    for a in m.atoms:
        sum_sq += a.element.mass * sqdistance(a.coord(), com)
    print str(m), "center of mass: %s, radius of gyration: %.3f" % (str(com),
        sqrt(sum_sq / sum(masses)))
```

The radii of gyration R_g, calculated by summing the distances of each atom to the center of mass of the structure, are determined from the structures deposited in the PDB, and plotted against V_e/V_0. Both the R_g values and the linear trend slope absolute value of ds-DNA are higher than for G4-DNA. This is expected since ds-DNA is generally more linear than G4-DNA, in analogy with what is found with linear and branched synthetic polymers (19). Dimeric G4 are intuitively more linear than monomeric G4s, since they contain more consecutive stacked G-quartets (≥ 6), while TBA is particularly globular as it only has two quartets. A slightly higher linearity can also be expected for G4 displaying long loops such as H-Bi-G4 and 26Ceb. Both predictions are confirmed by the R_g plot.

We have previously seen that although the logarithm of the molecular weight allows a good first approximation of the volume, and hence of the expected relative elution volume, the globularity of the structure influences significantly its capability to enter the pores of the column beads. In that regard, the radius of gyration is not a much better descriptor of nucleic acid volumes: linear trends between R_g and V_e/V_0 can be observed, but with fairly low R^2 values, and only within a given secondary structure type. This is not very surprising since such results were also found with synthetic polymers (19). Furthermore, the dynamic nature of nucleic acid structures is merely taken into account by this variable, calculated at best by averaging a few NMR models. Other theoretical models (extended CABS, Flory–Huggins) might give better results but were not assayed (19,20).

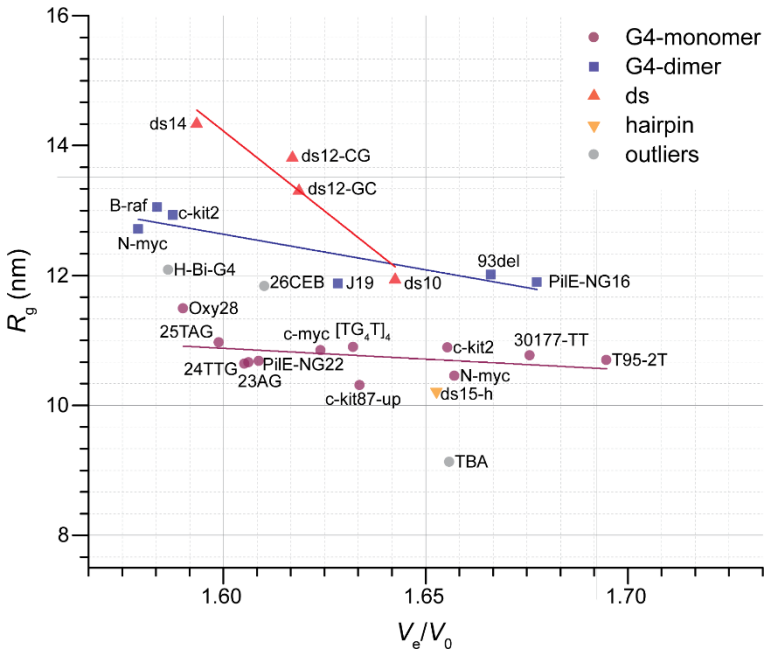

Figure S16. Plot of the radii of gyration (R_g), calculated for PDB-deposited structures, against the relative elution volume obtained for these sequences. Linear trends for homo-duplex, dimeric and monomeric G4 structures are shown as plain lines. H-Bi-G4 is considered as a monomer.

MULTIMOLECULAR QUADRUPLEXES

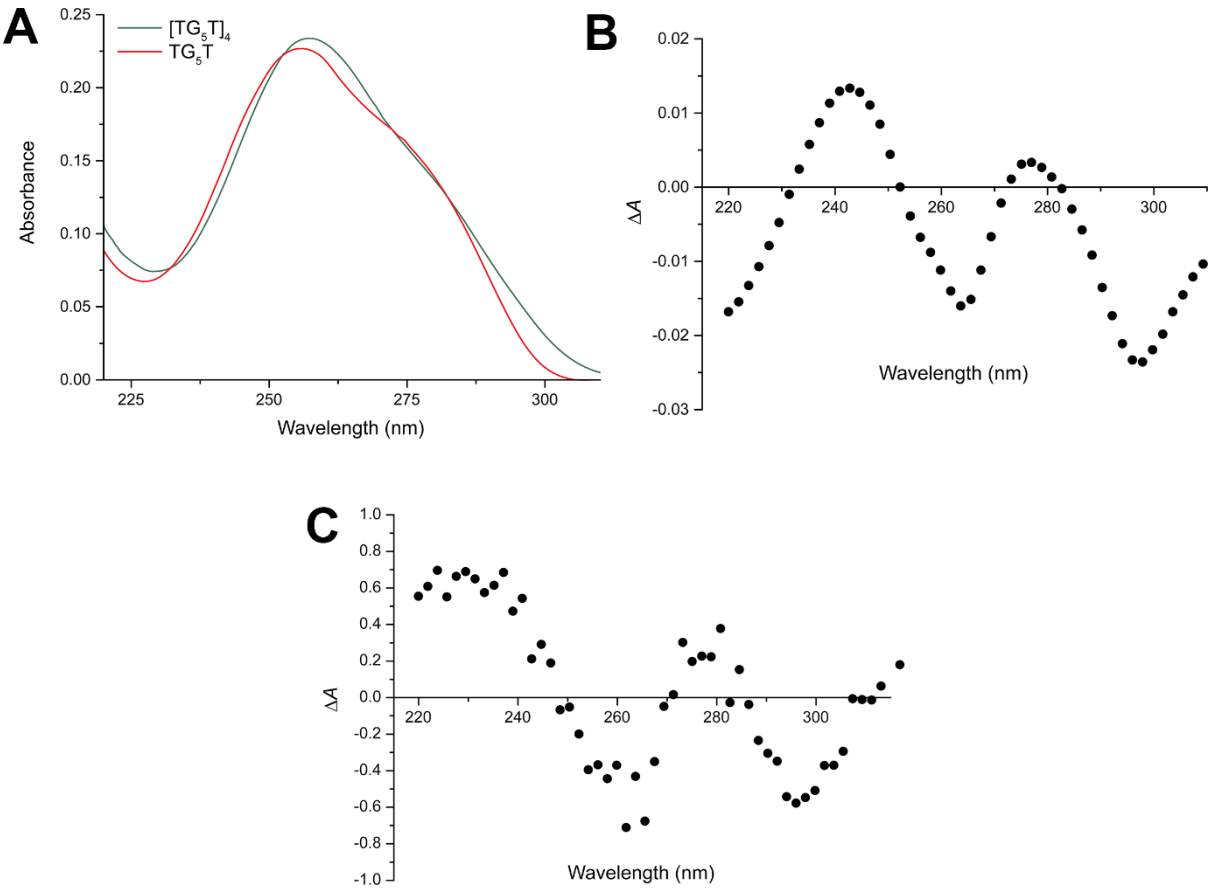

Figure S17. A) Absorbance spectra of TG_5T, integrated under the quadruplex and monomer peaks. B) IDS plotted from A). C) IDS of the monomer/dimer H-Bi-G4 couple. In both cases, a typical quadruplex signature was obtained meaning that the monomolecular species is unfolded while the accelerated peak is a G4, in accordance with their elution volumes.

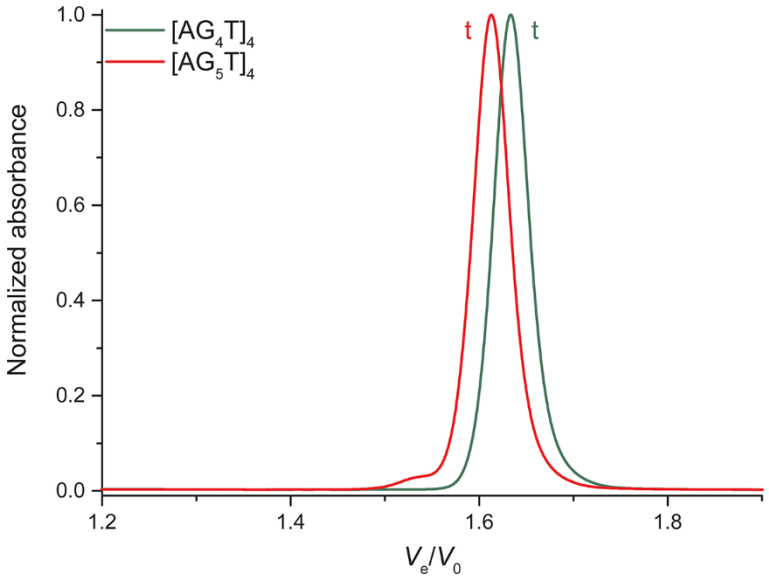

Figure S18. Normalized chromatograms of tetramolecular quadruplex-forming sequences $[AG_4T]_4$ and $[AG_5T]_5$ at a 250-µM strand concentration. Tetramer species are indicated with a *t*.

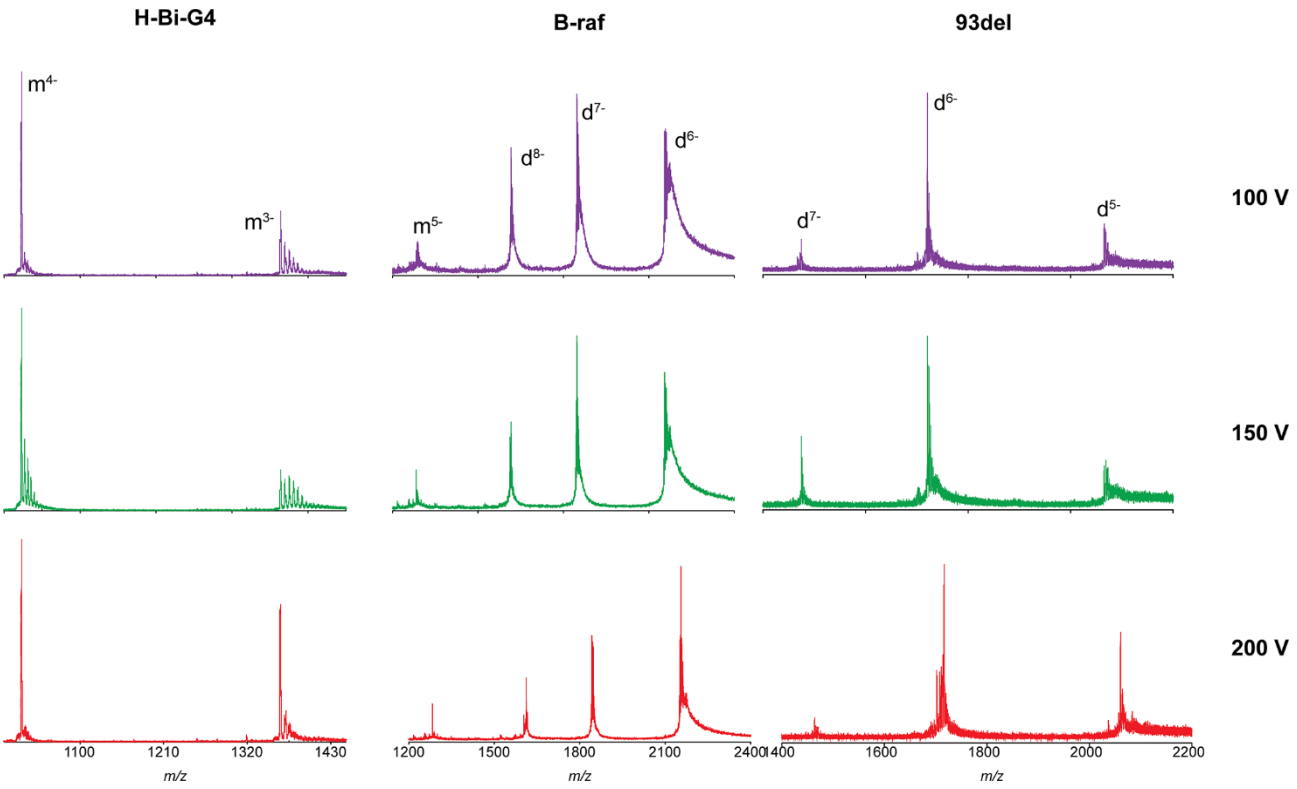

Figure S19. ESI-MS results for some bimolecular-quadruplex-forming sequences (H-Bi-G4, B-raf, and 93del), at various sample cone voltages. *m* stands for monomer, and *d* for dimer; the charge is indicated in superscript.

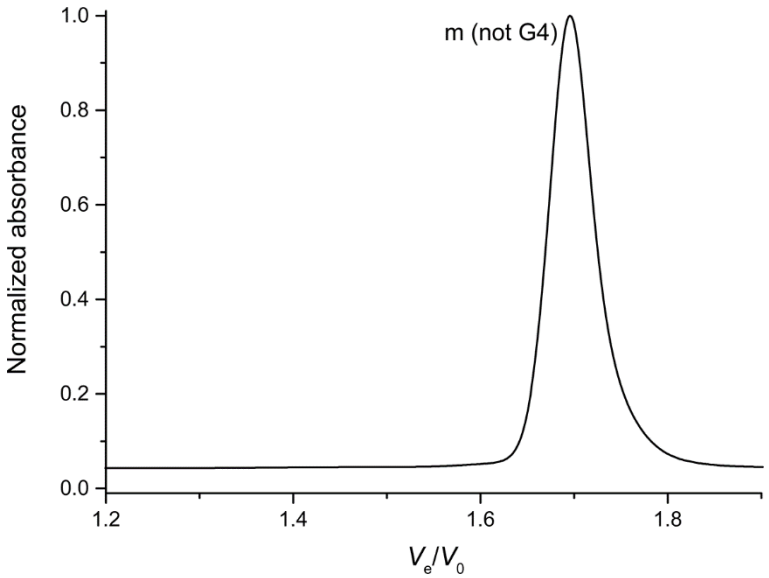

Figure S20. Normalized chromatograms of H-Bi-G4 in 100 mM ammonium acetate.

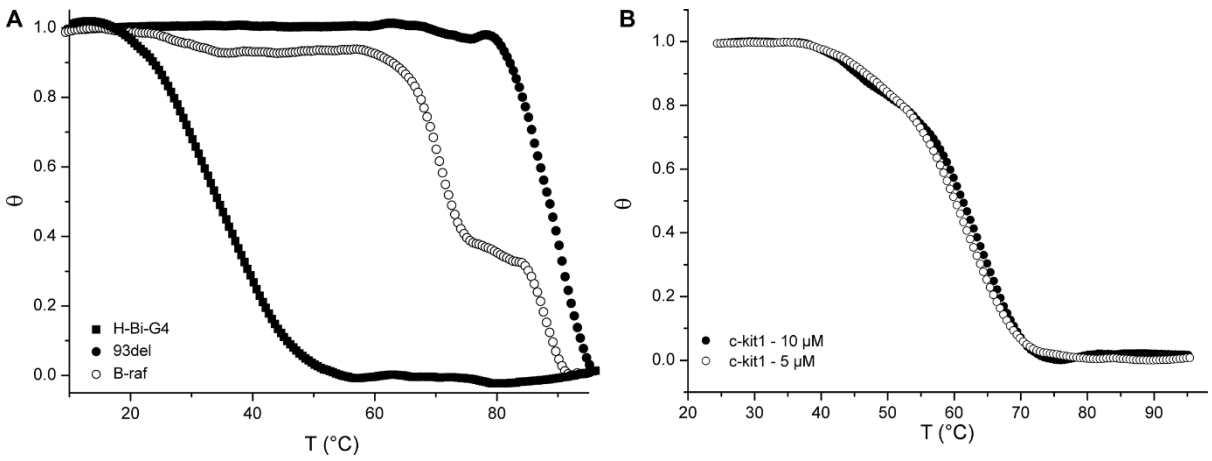

Figure S21. A) UV-melting of bimolecular quadruplexes at 10 μM strand concentration, in 100 mM KCl: while H-Bi-G4 is relatively unstable, both 93del and B-raf display high T_m. The latter is a mixture of a monomer and an interlocked dimer quadruplexes, hence the two clear transitions observed. B) UV-melting of monomolecular quadruplexes formed by c-kit1 at 10 and 5 μM strand concentration.

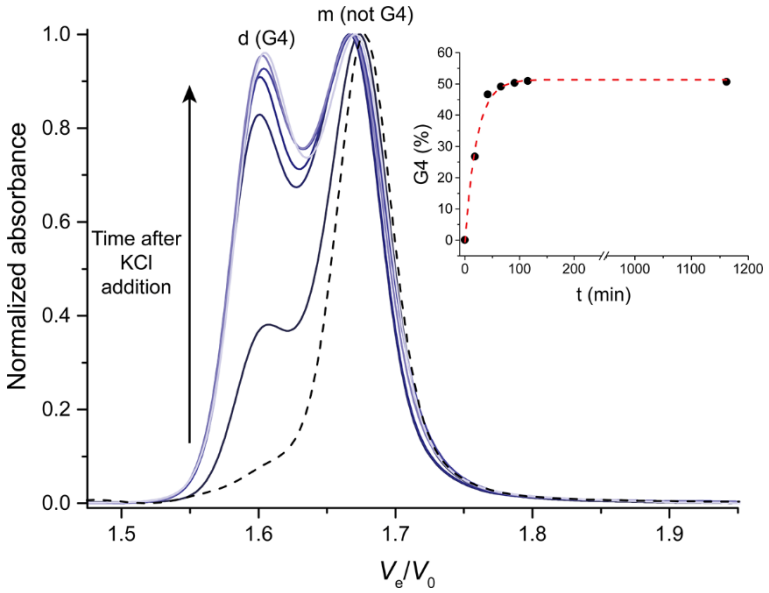

Figure S22. Time-dependent normalized chromatograms of H-Bi-G4 after addition of potassium chloride (plain blue lines), at a 100-μM strand concentration. There is little additional structuration during the elution (dashed black line: H-Bi-G4 prepared in absence of potassium, eluted in a 100-mM KCl containing buffer). Insert: formation of the quadruplex as a function of time.

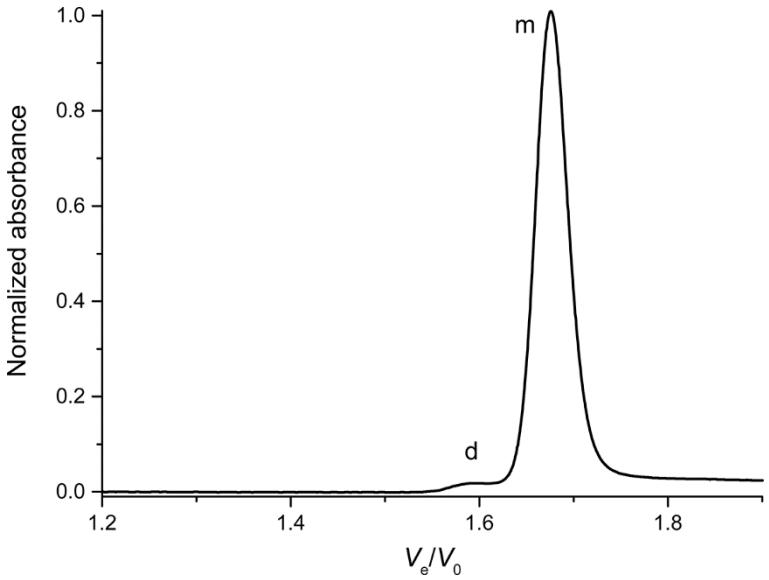

Figure S23. Normalized chromatograms of T30177-TT. Monomer is indicated with a *m* and dimer with a *d*.

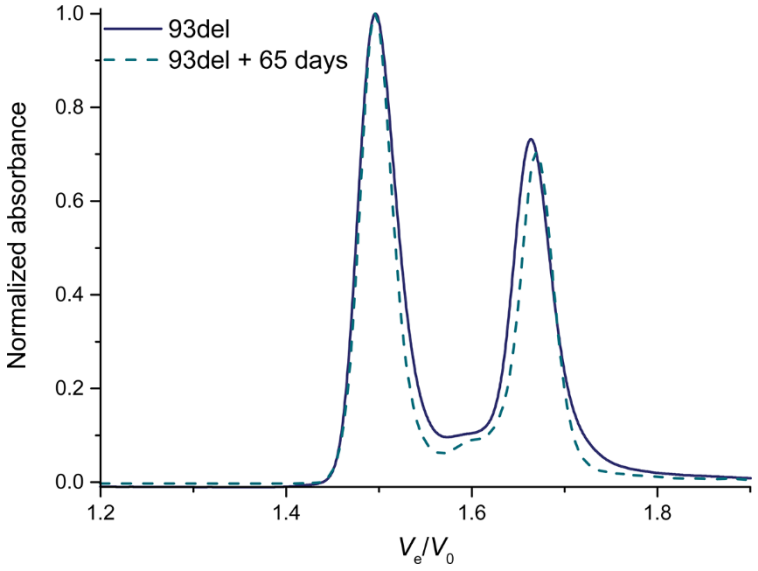

Figure S24. Normalized chromatograms of 93del after annealing (plain line), and after a long incubation at room temperature (dashed line).

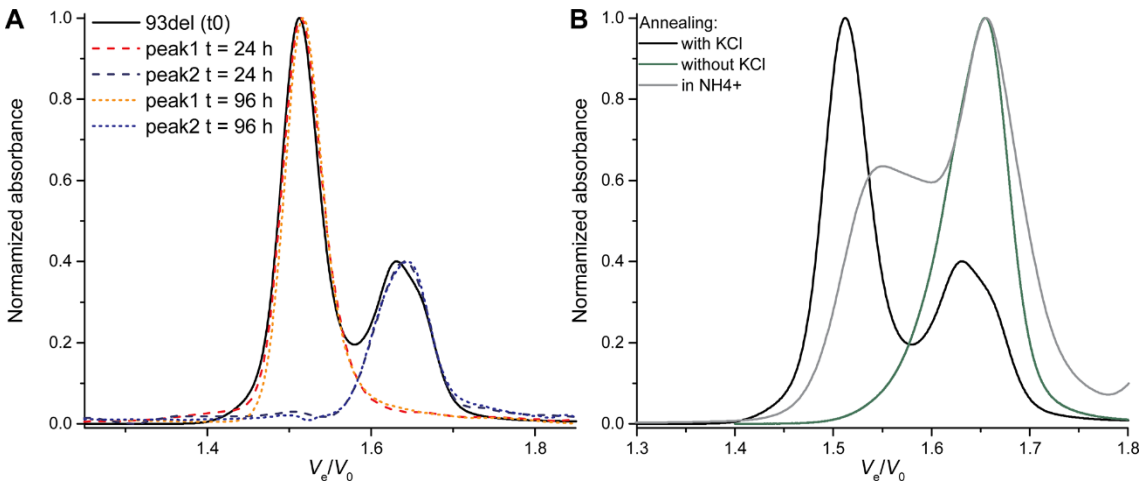

Figure S25. Normalized chromatogram of 93del after annealing in presence of potassium (black line) compared to A) isolated structures incubated 24 and 96 h at r.t after purification. (dashed lines), and B) Annealing condition effects. The tetramer is not formed when the potassium is added only after annealing (green line), while the use of ammonium acetate lead to a different signature (grey line).

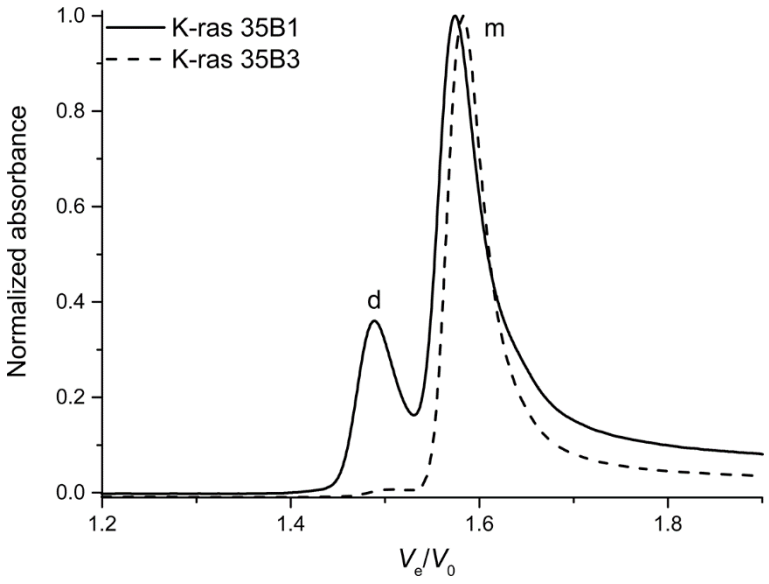

Figure S26. Sequence dependence: normalized chromatograms of sequences from the K-ras oncogene promoters. Monomer species are indicated with a *m* and dimers with a *d*.

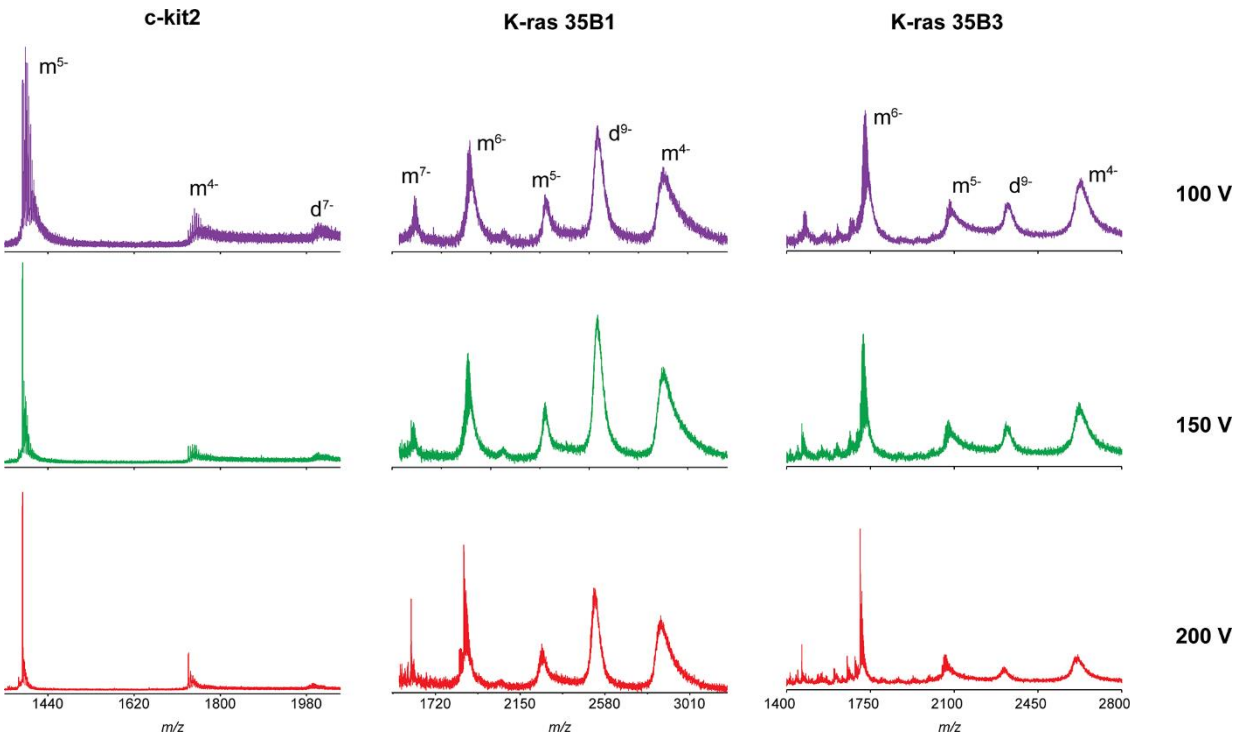

Figure S27. ESI-MS results for some dimeric-quadruplex-forming sequences (c-kit2, K-ras 35B1, and K-ras 35B3), at various sample cone voltages. *m* stands for monomer, and *d* for dimer; the charge is indicated in superscript.

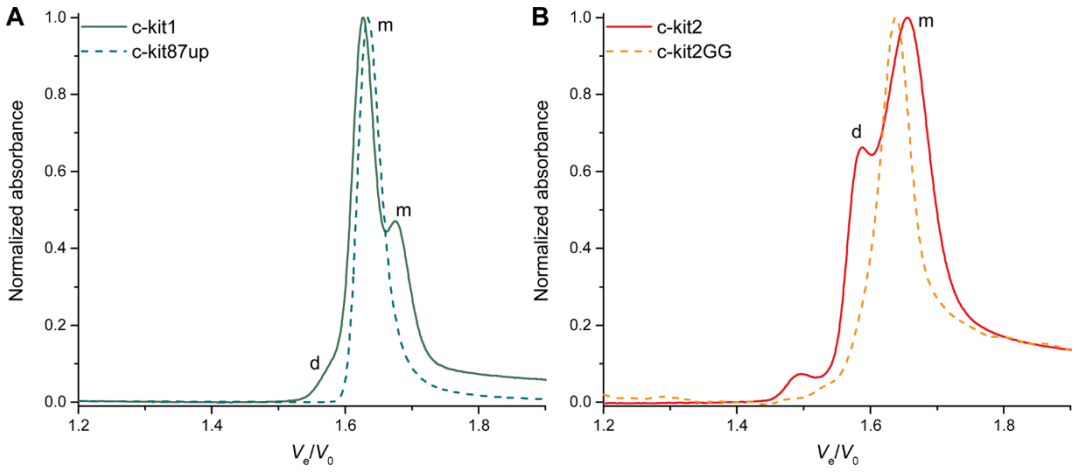

Figure S28. Sequence dependence: normalized chromatograms of sequences from the c-kit oncogene promoters. Monomer species are indicated with a *m* and dimers with a *d*.

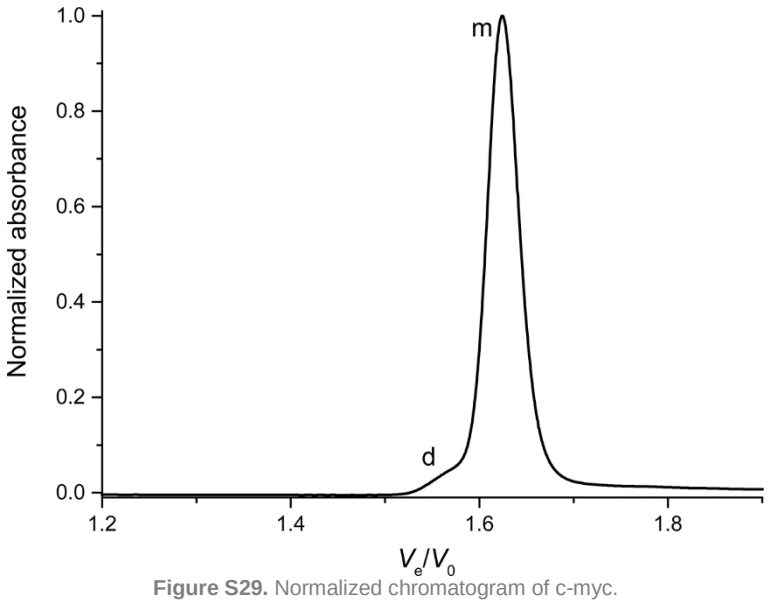

Figure S29. Normalized chromatogram of c-myc.

Mito86 is a 22-mer DNA oligonucleotide prone to form higher-order structures, which can be clearly seen from the comparison with the same-length oligonucleotide 22AG. *Mito86* can theoretically fold into a bulge-containing intramolecular quadruplex but intermolecular structures are very predominantly formed, most likely because it contains only three tracts of guanines. In the same vein, when a sequence is particularly rich in guanines, such as *Mito9*, a wide range of structures particularly difficult to disentangle can be formed, here ranging from the tetramer to the monomer.

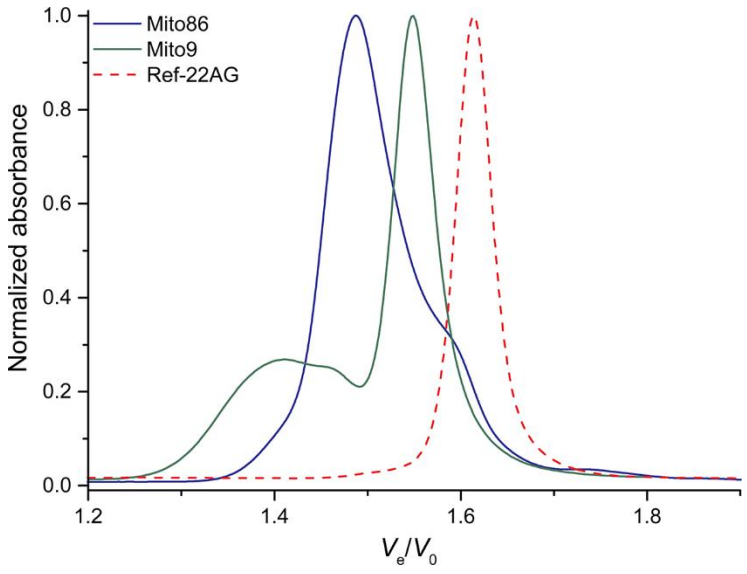

Figure S30. Normalized chromatograms of *Mito9* and *Mito86* , at a 80-µM strand concentration, as compared to the same-length, monomeric oligonucleotide 22AG.

I-MOTIF

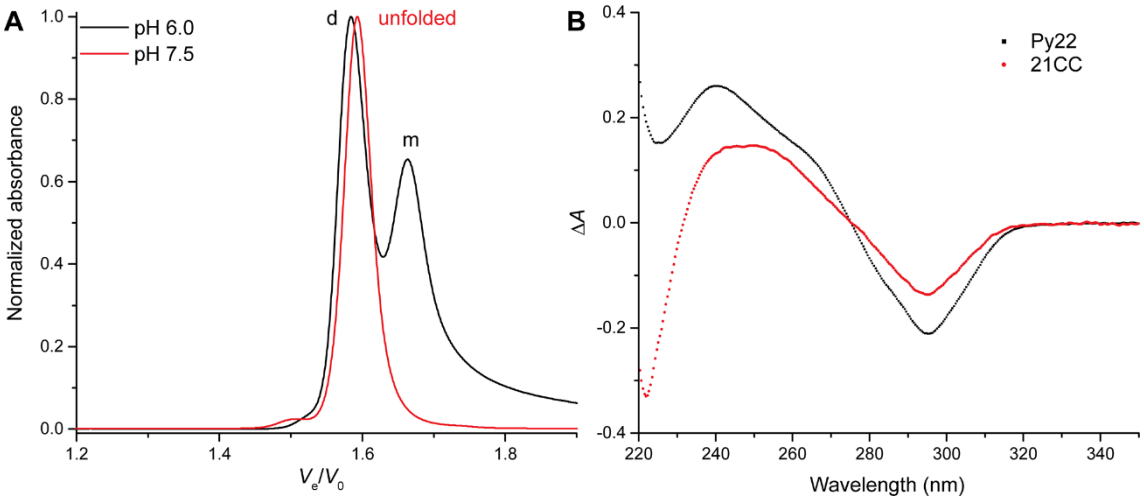

Figure S31. A) pH dependence: normalized chromatograms of the C-rich sequence from the human telomere (21CC); B) IDS signature for dimer/unfolded peaks. Monomer species are indicated with a *m* and dimers with a *d*.

CASE STUDIES

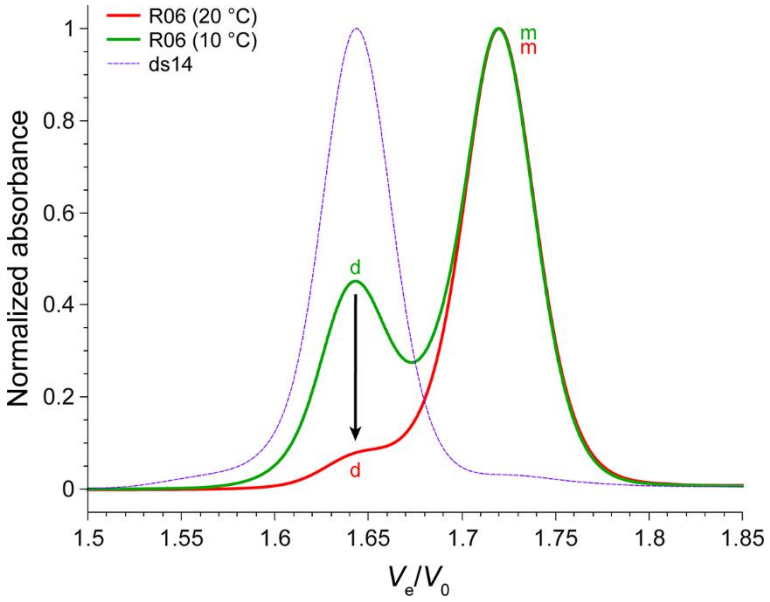

Figure S32. Normalized chromatograms of R06 at 20 and 10 °C (plain lines). Duplex-DNA ds14 shown for reference (dashed line). Monomer species are indicated with a *m* and dimers with a *d*.

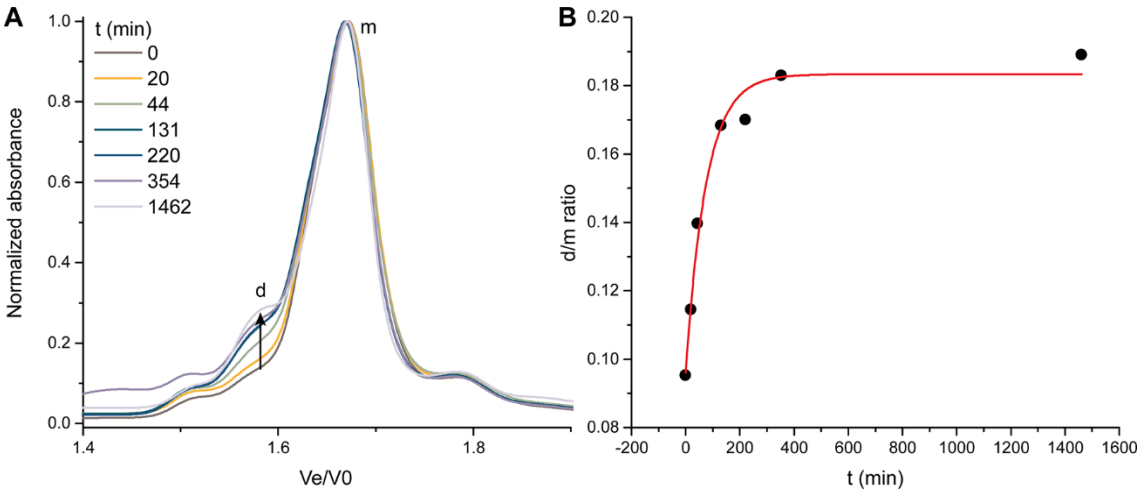

Figure S33. A) Time-dependent normalized chromatograms of N-myc (at a 100-µM strand concentration, and annealed in 10 mM KCl), after addition of potassium (final concentration: 100 mM); and B) plot of the dimer/monomer ratio vs. time. Red line: exponential fit.

TG_nT SEQUENCES.

To show the potential of SE-HPLC for the study of the polymorphism of quadruplex nucleic acids, the various structures formed by sequences following the TG_nT pattern (*n* = 4 to 20; **Table S1**) are studied.

Table S1. TG_nT sequences result summary

n	MW (Da)[a]	V_e/V_0[b]	FWHM[c]
4	1863.3	1.632	0.0679
5	2192.5	1.616	0.0743
6	2521.7	1.596	0.0824
7	2850.9	1.586	0.0808
8	3180.1	1.582	0.0985
9	3509.3	1.590	0.1244
10	3838.5	1.578	0.1292
11	4167.7	1.575	0.1486
12	4496.9	1.628	0.1632
14	5155.3	1.609	0.0985
15	5484.5	1.603	0.1018
16	5813.7	1.586	0.1002
18	6472.1	1.577	0.1325
20	7130.5	1.572	0.1034

[a] Molecular weight of the monomer
[b] Main peak is considered
[c] Full width at half maximum in relative elution volume units

Joly *et al.* recently used mass spectrometry to demonstrate that, although TG_4T and TG_5T oligonucleotides form well-known tetramolecular assemblies, longer sequences are prone to associate in alternative tri- or bimolecular structures (in 150 mM ammonium acetate buffer) (21). The “classical” tetramolecular quadruplex is only observed for *n* = 4 to 7, while *n* = 8—11 and *n* = 12—20 sequences forms trimolecular and bimolecular species, respectively. These sequences (**Table S1**) are annealed in 100 mM K^+-containing buffer, and analyzed after 4 days incubation using SE-HPLC.

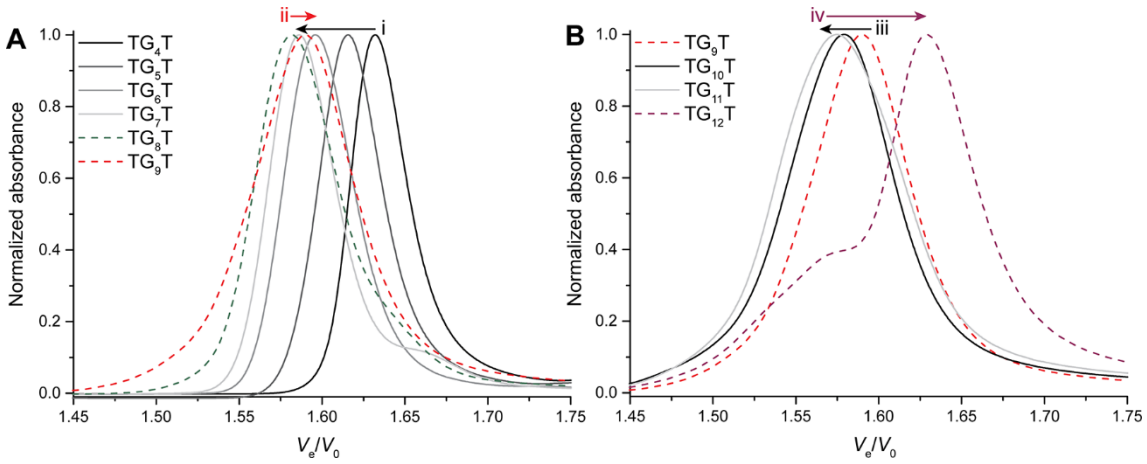

Figure S34. Chromatograms obtained for TG_nT sequences (A: n = 4—9, B: n = 9—12). Arrows indicate: i) an increase of the tetramolecular structure volume (decrease of V_e/V_0), ii) the appearance of smaller (trimolecular) species (increase of V_e/V_0), iii) an increase of the trimolecular structure volume (decrease of V_e/V_0), and iv) the appearance of a bimolecular structure (increase of V_e/V_0).

For n = 4—7, a single peak is present, with a retention time decreasing as these structures of similar shapes grow larger (**Figure S34A**). At n = 8 (dashed green chromatogram), a shoulder appears, which can be quantified by the widening of the peak taken as a whole (FWHM rising from 0.0679 at n= 4 to 0.0985 at n = 8; **Table S1**). Tetramolecular structures whose strands are staggered by one more bases might also contribute to this widening. The new type of dominant structure that is arising is smaller than its tetramolecular counterpart, because its retention time is higher. For n = 9, this alternative structure becomes predominant and following sequences until n = 11 follow this trend with decreasing retention times, suggesting that this is the same alternative structure growing in volume (**Figure S34B**). According to Joly *et al.*, this structure is likely to be trimolecular, which is supported by the expected retention time of these species considering their molecular weight. The FWHM of these peaks is also quite large (0.1244 to 0.1486), which indicates that there is a mixture of species in each case, possibly including structures with staggered strands. At n = 12 (dashed purple chromatogram), the trimolecular peak relative intensity dramatically decreases and another, significantly delayed, peak appears. This peak has been attributed to a bimolecular structure, again on the basis of its retention time, in accordance with the results obtained by Joly *et al.*

Longer sequences (n =13 to 20; **Figure S35**) have decreasing retention times most probably because they keep a similar structure but have increasing hydrodynamic volumes. For n = 16, 18 and 20, accelerated species can also be observed and might result from the presence of some trimolecular species or alternative higher-order assemblies. Despite the presence of long tracts of guanines, no polymerization is observed.

These results are overall very consistent with previously reported data and exemplify the potential use of SE-HPLC to study quadruplex DNA structure polymorphism and molecularity

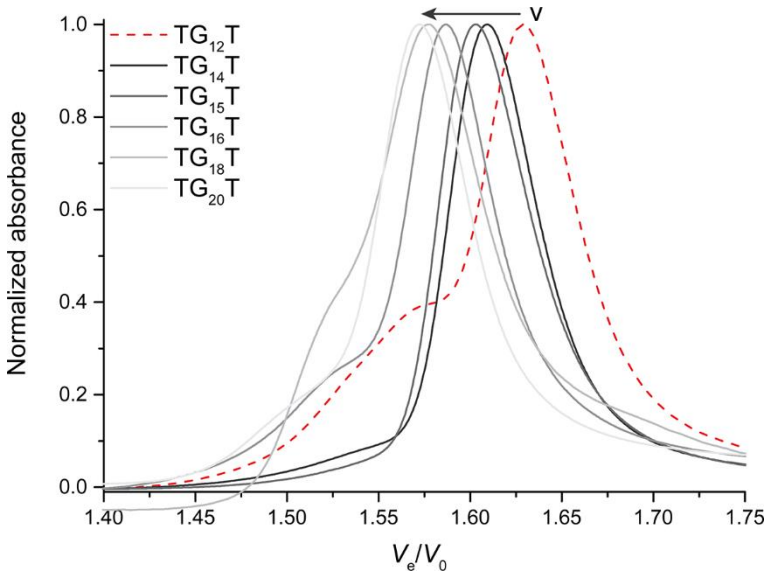

Figure S35. Chromatograms obtained for TG_nT sequences (n = 12—20). Arrows indicate an increase of the tetramolecular structure volume (decrease of V_e/V_0).

FUNCTIONALIZED OLIGONUCLEOTIDES

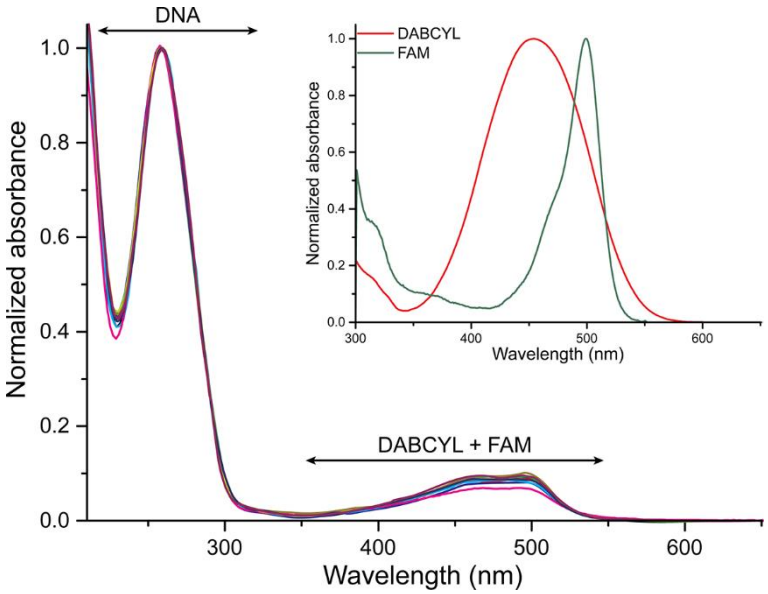

Figure S36. Absorbance spectra obtained by summing of the spectra under the peaks from chromatograms of 1-µM DNA samples. The spectra were zero-corrected at 650 nm, then normalized to 1 at 259.3 nm. The insert shows the absorbance spectra of FAM and DABCYL, in water.

REFERENCES

1. Luu, K.N., Phan, A.T., Kuryavyi, 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.
2. Phan, A.T., Kuryavyi, 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.
3. 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.
4. 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., Sect. D: Biol. Crystallogr.*, **63**, 682-688.
5. 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.
6. Phan, A.T., Kuryavyi, 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.
7. Amrane, S., Adrian, M., Heddi, B., Serero, A., Nicolas, A., Mergny, J.-L. and Phan, A.T. (2012) Formation of pearl-necklace monomorphic G-quadruplexes in the human CEB25 minisatellite. *J. Am. Chem. Soc.*, **134**, 5807-5816.
8. 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.
9. 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. *Chem. --Eur. J.*, **18**, 14752-14759.
10. Kuryavyi, 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.
11. Kuryavyi, 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 (London, England : 1993)*, **20**, 2090-2102.
12. 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.
13. 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.

14. Phan, A.T., Kuryavyi, 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.
15. 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.
16. 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.
17. 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.
18. 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.
19. 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.
20. 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.
21. Joly, L., Rosu, F. and Gabelica, V. (2012) d(TG(n)T) DNA sequences do not necessarily form tetramolecular G-quadruplexes. *Chem. Commun.*, **48**, 8386-8388.