JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Communication
pubs.acs.org/JACS

Topology of a DNA G-Quadruplex Structure Formed in the HIV-1 Promoter: A Potential Target for Anti-HIV Drug Development

Samir Amrane,*[,†,‡] Abdelaziz Kerkour,[†,‡] Amina Bedrat,[†,‡] Brune Vialet,[†,‡] Marie-Line Andreola,[†,§] and Jean-Louis Mergny[†,‡]

[†]Université de Bordeaux, 33000 Bordeaux, France
[‡]INSERM, U869, IECB, ARNA laboratory, 2 Rue Robert Escarpit 33600 Pessac, France
[§]CNRS UMR 5234, 146 Rue Leo Saignat, 33076 Bordeaux, France

Supporting Information

ABSTRACT: Nucleic acid sequences containing guanine tracts are able to adopt noncanonical four-stranded nucleic acid structures called G-quadruplexes (G4s). These structures are based on the stacking of two or more G-tetrads; each tetrad is a planar association of four guanines held together by eight hydrogen bonds. In this study, we analyzed a conserved G-rich region from HIV-1 promoter that is known to regulate the transcription of the HIV-1 provirus. Strikingly, our analysis of an alignment of 1684 HIV-1 sequences from this region showed a high conservation of the ability to form G4 structures despite a lower conservation of the nucleotide primary sequence. Using NMR spectroscopy, we determined the G4 topology adopted by a DNA sequence from this region (HIV-PRO1: 5′ T**GG**CCT**GGG**C**GGG**ACT**GGG** 3′). This DNA fragment formed a stable two G-tetrad antiparallel G4 with an additional Watson−Crick CG base pair. This hybrid structure may be critical for HIV-1 gene expression and is potentially a novel target for anti-HIV-1 drug development.

DNA or RNA sequences containing guanine tracts are able to adopt noncanonical four-stranded structures called G-quadruplexes (G4s). The inner core of the G4 is based on the stacking of two or more G-tetrads. Each tetrad is a planar association of four guanines held together by eight hydrogen bonds and coordinated with a central Na^+ or K^+ cation.[1] The four G-tracts forming the core delimit four negatively charged grooves linked together by three types of loops[2] (lateral, diagonal, or propeller). Unlike the canonical duplex, these degrees of freedom confer a high level of plasticity to this family of globularly shaped nucleic acid structures. Under near-physiological conditions, intramolecular G4s easily form within milliseconds[3] and can be thermally stable with melting temperatures typically above 50 °C.[4,5] Polymorphism, robustness, and fast folding are altogether intrinsic features of these structures that suggest biological functions. Genome scale bioinformatics analysis showed a significant enrichment of these sequences in various key elements of the human genome such as telomeres, oncogenes, and introns.[5] More recently, *in vivo* studies, using specific G4 probes (antibodies and ligands), support the formation of G4s in cells.[6−8] The implication of G-quadruplexes in virology only begins to be realized with recent investigations in the papilloma,[9] Epstein−Barr,[10] and SARS[11] viruses for instance.

A conserved G-rich region regulates HIV-1 promoter activity. HIV-1 retrovirus infects cells that carry CD4 and one of the chemokine receptors CCR5 or CXCR4. It induces a deficiency of the immune system causing AIDS. Shortly after infection, the two HIV-1 single-stranded RNAs are reverse transcribed by the viral reverse transcriptase into double-stranded DNA. The viral DNA then migrates in the nucleus and is integrated into the genome of the infected cell. At this stage, the integrated provirus is composed of two regulatory regions (5′LTR and 3′LTR) and nine protein-coding genes (Figure 1A). The host cell machinery transcribes the viral genes and produces new viral proteins, and new viruses are assembled.[12] A key step of the viral cycle is the regulation of the transcription of the provirus by a viral promoter located in the 5′ LTR region (Figure 1B). The U3 region of this promoter contains a G-rich sequence 50 nucleotides upstream from the transcription-starting site (TSS), close to the TATA box. This sequence overlaps the so-called minimum promoter composed of three SP1 and two NF-kB binding sites (Figure 1B), which are crucial for the initiation of the transcription.[13,14] The presence of eight blocks of guanines suggests that this region is a good candidate for G4 formation.

The potential to form a G4 structure in the promoter is highly conserved. We analyzed the conservation of this G-rich sequence over 1684 aligned 5′LTR sequences provided by the HIV-1 database. The LOGO representation[15] generated from this alignment shows a high level of conservation of the eight blocks of guanines (Figures 1B, S1). The conservation of these G-rich blocks contrasts with the poor conservation of the primary sequence overall. Table 1 presents the five most frequent G4 motifs derived from the central part of this G-rich region, between position −57 and −78. It spans the three SP1 binding sites, from the 5′ extremity of SP1-3 to the 3′ extremity of SP1-1. The four blocks of G's are conserved, with slight differences in the number of G's (ranging from two to four) from the two central blocks. The first and last interblock linkers are heterogeneous in length (two to five nucleotides) and base composition. Interestingly, the common feature shared by all

Received: February 12, 2014
Published: March 20, 2014

   dx.doi.org/10.1021/ja501500c | J. Am. Chem. Soc. 2014, 136, 5249−5252

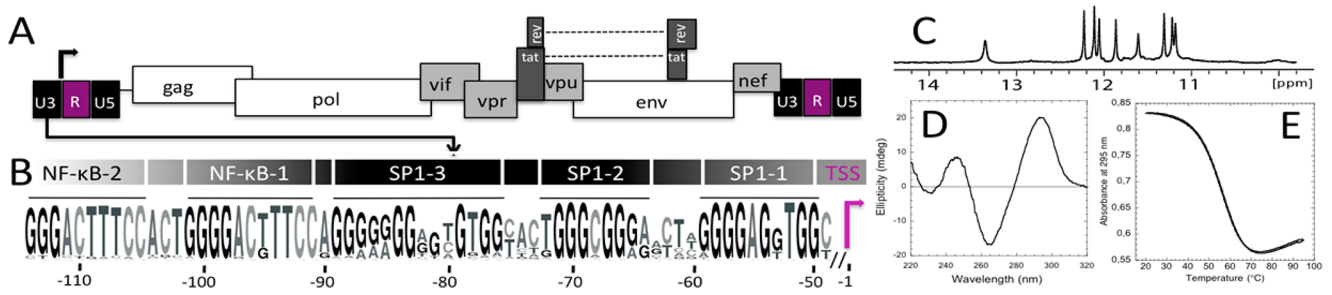

Figure 1. (A) Genomic structure of HIV-1 provirus and (B) LOGO representation of the G-rich region of HIV-1 promoter generated by the weblogo software[15] and based on an alignment of 1684 HIV-1 sequences from the HIV-1 database (www.hiv.lanl.gov). (C) Imino proton NMR spectrum recorded at 25 °C at a concentration of 140 μM. (D) Circular dichroism spectrum recorded at 25 °C at a concentration of 5 μM. (E) UV-melting profile recorded at 295 nm at a strand concentration of 140 μM. All experiments in this study were performed at 25 °C in a buffer composed of 20 mM potassium phosphate pH 6.9 supplemented with 70 mM KCl.

Table 1. G4 Motifs from the HIV-1 Promoter Central Region

N°	Sequence motifs [a]	Occurrence[b]	T_m(°C)[c]
1	N-GG-NNN-GGGCGGG-NNN-GGG	556	-
	T-GG-CCT-GGGCGGG-ACT-GGG	149	56
2	N-GG-NNN-GGGCGGGNNNN-GGG	161	-
	T-GG-TCT-GGGCGGGACTA-GGG	55	44
3	N-GG-NNN-GGGCGGGG-NN-GGG	37	-
	T-GG-CCT-GGGCGGGG-TT-GGG	14	61
4	N-GG-NNN-GGGCGG-NGNN-GGG	163	-
	T-GG-TTT-GGGCGG-AGTT-GGG	146	30
5	N-GG-NN-GGGGCGG-NGNN-GGG	135	-
	T-GG-CC-GGGGCGG-AGTT-GGG	96	35

[a]General motif definition (top) and one representative sequence for each motif (bottom). [b]Number of times each motif appears in the alignment. The five motifs represent a total of 1052 sequences out of the 1684 sequences from the alignment. [c]Melting temperatures determined by UV-melting experiments recorded at 295 nm.

these sequences is the ability to form stable G4 structures *in vitro* as shown by the UV-melting profiles (midpoints of melting transitions, T_m's, range from 30 to 61 °C) recorded for one representative sequence of each G4 motif (Tables 1, S1 and Figure S2). We analyzed the topology of the most represented motif in all HIV-1 subspecies (motif 1, Table 1) using the sequence we call HIV-PRO1 (Table 2).

HIV-PRO1 forms a two-tetrad intramolecular G4. The imino proton NMR spectrum of the HIV-PRO1 sequence presents eight well-resolved imino peaks in the 10−12 ppm region (Figure 1C). These G4 characteristic resonances indicate the formation of a unique two G-tetrad structure

Table 2. HIV-PRO1 Mutated Sequences

Name	Sequence	T_m (°C)
HIV-PRO1	TGGCCTGGGCGGGACTGGG	56
m4T	TGGTCTGGGCGGGACTGGG	46
m7T	TGGCCTTGGCGGGACTGGG	39
M4A-7A	TGGACTAGGCGGGACTGGG	39
m5T	TGGCTTGGGCGGGACTGGG	57
m10T	TGGCCTGGGTGGGACTGGG	55
m15T	TGGCCTGGGCGGGATTGGG	58
m19A	TGGCCTGGGCGGGACTGGA	57

(four imino proton peaks for each G-tetrad). The lower field peak at 13.4 ppm suggests the presence of an additional Watson−Crick base pair. This quadruplex presents an antiparallel type circular dichroism signature with a positive peak at 295 nm and a negative peak at 265 nm (Figure 1D). The UV-melting profile reveals a concentration independent melting temperature of 56 °C (Figures 1E and S3). These results along with the fast electrophoretic mobility on native PAGE (Figure S3) are indicative of an intramolecular folding.

NMR assignment of guanines and cytosines of HIV-PRO1 sequence. Guanine imino (H1) and aromatic (H8) protons of HIV-PRO1 were unambiguously assigned by site specific 5%-enrichment ^{15}N and ^{13}C guanine labeling (Figure 2A, 2B). The ^{15}N edited experiments showed that the imino proton of G7 is involved in the Watson−Crick lower field peak, whereas the imino protons from G2, G3, G8, G9, G12, G13, and G17, but not G11, are involved in the G-tetrads hydrogen bond network. G19 is not involved in the tetrads since the G19 to A19 mutated sequence (m19A) had very similar ^{1}H-1D NMR signatures (Figure 2C) and T_m's (Table 2, Figure S4) as compared to the wild type sequence. Thus, the remaining imino resonance at 12.1 ppm belongs to G18. This was also confirmed by through-bond correlations {^{13}C−^{1}H}-HMBC (Figure S5). The C to T mutations at positions 5, 10, and 15 did not affect the NMR signature (Figure 2C) or the T_m (Table 2, Figure S4) ; all were comparable to those of HIV-PRO1. We took advantage of these mutations to assign the cytosine resonances using their specific H5/H6 TOCSY correlation (Figure 2D). In each case, the disappearance of the TOCSY signal allowed us to determine the H5 and H6 chemical shifts of the mutated C. The C4 to T4 mutation (m4T) did disrupt the structure (Figure 2C) suggesting that C4 might be involved in the Watson−Crick base pair with G7.

HIV-PRO1 adopts an antiparallel chair type G4 topology with an additional CG base pair. The G-tetrad alignments were identified from the NOESY spectrum (Figure 3A) based on the specific imino-H8 connectivity patterns defining two G-tetrads: G2-G18-G12-G9 and G3-G8-G13-G17. The hydrogen-bond directionalities around the G-tetrads are opposite (Figure 3B). G2, G8, G12, and G17 adopt *syn* glycosidic conformations as suggested by the strong intraresidue NOE cross-peaks between the H1′ of the sugar and the H8 of the aromatic base (Figure 3C). In contrast, G3, G9, G13, and G18 adopt *anti* conformations (Figure 3C). These 5′-*syn-anti*-3′ steps are also confirmed by the specific rectangular NOE patterns[16] that result from a double sequential H1′/H8 correlation observed for the 5′-G2-G3-3′ and 5′-G12-G13-3′ steps. Hence, each

dx.doi.org/10.1021/ja501500c | J. Am. Chem. Soc. 2014, 136, 5249−5252

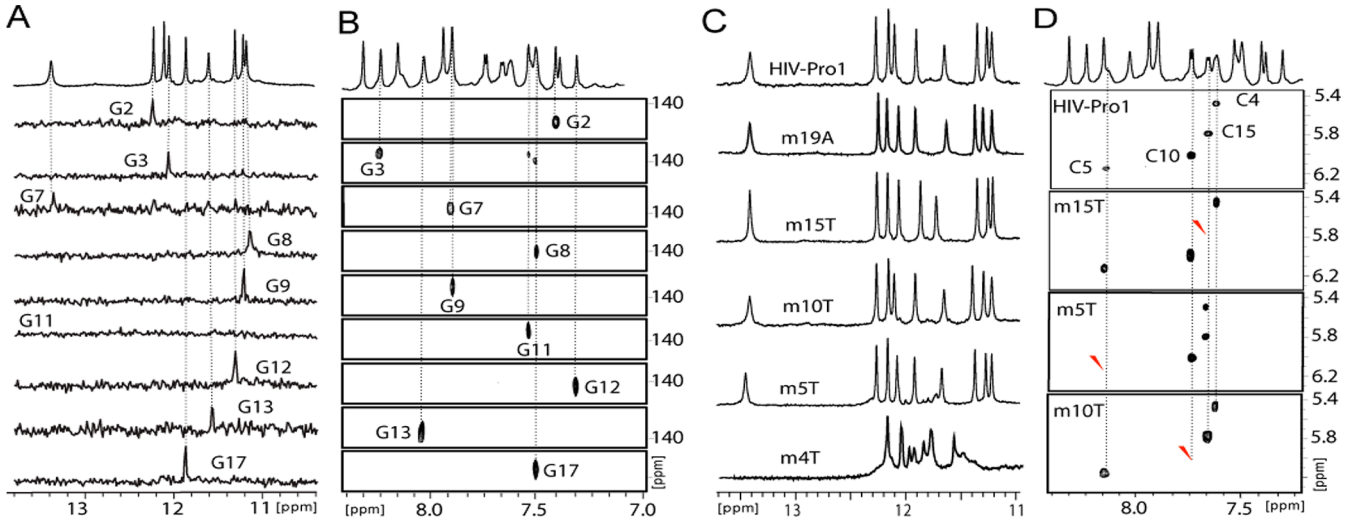

Figure 2. (A) Imino proton (H1) assignments from ^{15}N-filtered experiments and (B) aromatic proton (H8) assignment from ^{13}C-HSQC experiments using 5% ^{15}N−^{13}C enriched guanines at the indicated positions. (C) Imino proton NMR spectrum and (D) aromatic region TOCSY correlations of HIV-PRO1 mutated sequences. A red arrow shows the disappearance of the TOCSY correlation for the mutated cytosine. Oligonucleotides were dissolved in a 90 mM KCl/Kpi buffer at around 100−200 μM strand concentration.

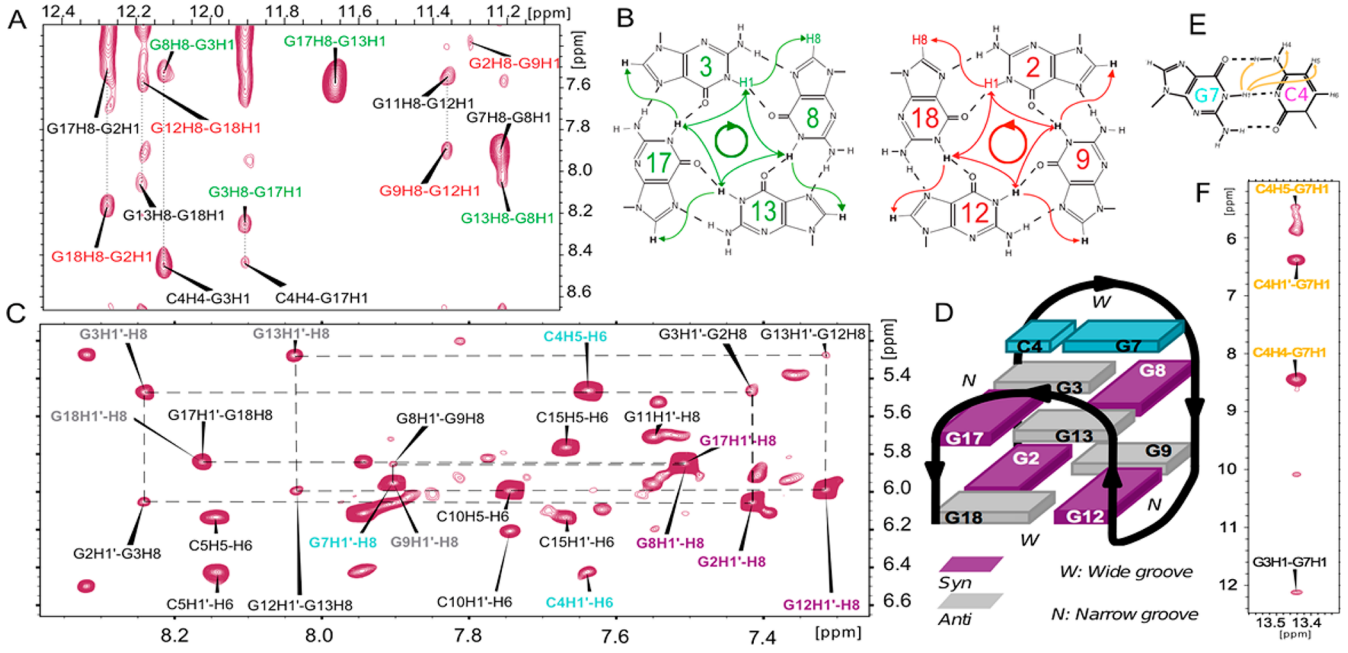

Figure 3. (A,F) 2D ^{1}H NOESY spectrum (300 ms mixing time) in water showing H1/H8 correlations. Green and red labels correspond to intratetrad correlations for the upper and bottom tetrads (B); yellow labels correspond to the GC base pair specific correlations (E). (C) NOESY spectrum in D_2O (350 ms mixing time) showing H1′/H6−H8 and H5−H6 correlations with purple and gray labels for *syn* and *anti* guanines. Spectra were recorded at 25 °C in the 90 mM KCl/Kpi buffer at a concentration of 600 μM. (D) Topology of the hybrid G-quadruplex structure.

strand runs into opposite directions with respect to its two neighbors and delimits two wide and two narrow grooves. The wide groove (W) is defined by an *anti*→*syn* step across the groove while the narrow groove (N) is defined by a *syn*→*anti* step (Figure 3D). Each corner of the G-tetrads is connected by three lateral linkers establishing an antiparallel chair-type G4 topology. The C4-C5-T6-G7 and A14-C15-T16 loops span two wide grooves, and the central C10-G11 loop spans a narrow groove. Clear NOE cross-peaks between the H1 of G7 and the H5 and H4 protons of C4 (Figure 3E, 3F) established the formation of a C4-G7 Watson−Crick base pair giving a hairpin-like conformation to the first loop. This base pair spans a wide groove as already observed by Lim and Phan in a different sequence.[17] It stacks on top of the upper tetrad as suggested by ^{1}H NOE cross-peaks with protons from G3, G8, and G17 (Figure 3A, 3C, 3F). The disruption of this base pair (m4T, m7T) decreased the T_m by 10−17 °C (Table 2, Figure S4) explaining the unusually high stability (T_m = 56 °C) observed for this two-layered G4. The further destabilization of m7T suggests that G7 might still be able to stack on the upper tetrad even in the absence of C4 (m4T) while adenine residues are not able to stabilize the G4 in the same fashion (m4A-7A). The *Bombyx mori* telomeric sequence TAGG(TTAGG)$_3$ adopts the same topology without the base pair resulting in a similar drop

dx.doi.org/10.1021/ja501500c | J. Am. Chem. Soc. 2014, 136, 5249−5252

in melting temperature (T_m = 41 °C).[18] Bioinformatics searches have been focused on sequences forming G4s with three tetrads. This hybrid structure should trigger more interest toward the two-layered G4s.

G4 formation potentially regulates HIV-1 promoter activity. Bioinformatic analysis of the human genome have already shown that G4s motifs often overlap or are located in the close neighborhood of zinc-finger transcription factor binding sites (SP1 or MAZ) suggesting a role in the regulation of mammalian genes.[19,20] The G-rich region of HIV-1 promoter overlaps with SP1 and NF-kB binding sites. This conserved "G-richness" might translate into a conservation of the ability to form G4 structures through all HIV-1 subspecies. Hence, G4s should play an important role in HIV-1 promoter regulation. It was recently suggested that the formation of a G4 structure in this region is able to repress the activity of the HIV-1 promoter.[21] The role of this G4 is reminiscent of the regulation of oncogene promoters such as *c-myc*,[22] *PDGFR-β*,[23] and *BCL-2*.[24]

Targeting HIV-1 Virus with G4 Ligands. G-quadruplexes are promising pharmacological targets. Their polymorphism suggests that a high degree of drug specificity can be achieved.[25] This approach is being investigated in anticancer research by targeting G4s located in telomeres or oncogenes.[22] Perrone et al. showed that G4 ligands such as BRACO-19 are able to inhibit HIV-1 infectivity.[21] Despite a relatively weak activity ($IC_{50} \approx 3\ \mu M$), these results suggest that G4s can also be targeted to treat AIDS. This new antiviral strategy presents three advantages: (i) These compounds target viral DNAs and RNAs, the actual source of the disease; (ii) the high conservation of these targets across all HIV-1 subspecies suggests that they are important for HIV-1 fitness and not mutable, and the emergence of resistant strains should be limited; (iii) the HIV-PRO1 sequence cannot be found in the Human genome. We are screening libraries of G4 ligands to identify new antiviral drugs. Our preliminary data showed that ligands such as BRACO-19 can recognize this G4 (Figure S6). Efforts will now be made to identify G4 ligands that preferentially interact with the HIV quadruplex.

■ ASSOCIATED CONTENT

Supporting Information

Additional experimental data for NMR, CD, PAGE, and UV-melting (Figures S1−S6). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

s.amrane@iecb.u-bordeaux.fr

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Prof. Anh Tuân Phan, Dr Brahim Heddi, Dr. Cameron Mackereth and Axelle Grélard for helpful discussions. S.A. is the recipient of an ANRS postdoctoral fellowship. This work was supported by grants from ANR (Quarpdiem, T-Kinet, and Oligoswitch) and Conseil Régional d'Aquitaine ("Chaire d'accueil" to J.L.M. and Aquitaine-Midi Pyrénées call).

■ REFERENCES

(1) Phan, A. T. *FEBS J.* **2010**, *277*, 1107.
(2) Marusic, M.; Sket, P.; Bauer, L.; Viglasky, V.; Plavec, J. *Nucleic Acids Res.* **2012**, *40*, 6946.
(3) Zhang, A. Y.; Balasubramanian, S. *J. Am. Chem. Soc.* **2012**, *134*, 19297.
(4) Guédin, A.; Alberti, P.; Mergny, J. L. *Nucleic Acids Res.* **2009**, *37*, 5559.
(5) Eddy, J.; Maizels, N. *Nucleic Acids Res.* **2008**, *36*, 1321.
(6) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. *Nat. Chem.* **2013**, *5*, 182.
(7) Rodriguez, R.; Miller, K. M.; Forment, J. V.; Bradshaw, C. R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S. P. *Nat. Chem. Biol.* **2012**, *8*, 301.
(8) Henderson, A.; Wu, Y.; Huang, Y. C.; Chavez, E. A.; Platt, J.; Johnson, F. B.; Brosh, R. M.; Sen, D.; Lansdorp, P. M. *Nucleic Acids Res.* **2014**, *42*, 663.
(9) Tluckova, K.; Marusic, M.; Tothova, P.; Bauer, L.; Sket, P.; Plavec, J.; Viglasky, V. *Biochemistry* **2013**, *52*, 7207.
(10) Norseen, J.; Johnson, F. B.; Lieberman, P. M. *J. Virol.* **2009**, *83*, 10336.
(11) Tan, J.; Vonrhein, C.; Smart, O. S.; Bricogne, G.; Bollati, M.; Kusov, Y.; Hansen, G.; Mesters, J. R.; Schmidt, C. L.; Hilgenfeld, R. *PLoS Pathog.* **2009**, *5*, e1000428.
(12) Pomerantz, R. J.; Horn, D. L. *Nat. Med.* **2003**, *9*, 867.
(13) Pereira, L. A.; Bentley, K.; Peeters, A.; Churchill, M. J.; Deacon, N. J. *Nucleic Acids Res.* **2000**, *28*, 663.
(14) Rooebuck, K. A.; Saifuddin, M. *Gene Expr.* **1999**, *8*, 67.
(15) Crooks, G. E.; Hon, G.; Chandonia, J. M.; Brenner, S. E. *Genome Res.* **2004**, *14*, 1188.
(16) Adrian, M.; Heddi, B.; Phan, A. T. *Methods* **2012**, *57*, 11.
(17) Lim, K. W.; Phan, A. T. *Angew. Chem., Int. Ed. Engl.* **2013**, *52*, 8566.
(18) Amrane, S.; Ang, R. W.; Tan, Z. M.; Li, C.; Lim, J. K.; Lim, J. M.; Lim, K. W.; Phan, A. T. *Nucleic Acids Res.* **2009**, *37*, 931.
(19) Kumar, P.; Yadav, V. K.; Baral, A.; Saha, D.; Chowdhury, S. *Nucleic Acids Res.* **2011**, *39*, 8005.
(20) Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2008**, *36*, 2700.
(21) Perrone, R.; Nadai, M.; Frasson, I.; Poe, J. A.; Butovskaya, E.; Smithgall, T. E.; Palumbo, M.; Palu, G.; Richter, S. N. *J. Med. Chem.* **2013**, *56*, 6521.
(22) Balasubramanian, S.; Hurley, L. H.; Neidle, S. *Nat. Rev. Drug Discovery* **2011**, *10*, 261.
(23) Chen, Y.; Agrawal, P.; Brown, R. V.; Hatzakis, E.; Hurley, L.; Yang, D. *J. Am. Chem. Soc.* **2012**, *134*, 13220.
(24) Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. *J. Am. Chem. Soc.* **2014**, *136*, 1750.
(25) Balasubramanian, S.; Neidle, S. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345.