J|A|C|S
ARTICLES
Published on Web 06/22/2004

Significant pK_a Perturbation of Nucleobases Is an Intrinsic Property of the Sequence Context in DNA and RNA

Sandipta Acharya, Jharna Barman, Pradeep Cheruku, Subhrangsu Chatterjee, Parag Acharya, Johan Isaksson, and Jyoti Chattopadhyaya*

Contribution from the Department of Bioorganic Chemistry, Box 581, Biomedical Center, Uppsala University, S-751 23 Uppsala, Sweden

Received March 16, 2004; E-mail: jyoti@boc.uu.se

Abstract: The pH titration and NMR studies (pH 6.6−12.5) in the heptameric isosequential ssDNA and ssRNA molecules, [d/r(5′-CAQ^1***G***Q^{2}AC-3′, with variable Q^1/Q^2)], show that the pK_a of the central ***G*** residue within the heptameric ssDNAs (ΔpK_a = 0.67 ± 0.03) and ssRNAs (ΔpK_a = 0.49 ± 0.02) is sequence-dependent. This variable pK_a of the ***G*** clearly shows that its pseudoaromatic character, hence, its chemical reactivity, is strongly modulated and tuned by its sequence context. In contradistinction to the ssDNAs, the electrostatic transmission of the pK_a of the ***G*** moiety to the neighboring **A** or **C** residues in the heptameric ssRNAs (as observed by the response of the aromatic marker protons of **A**s or **C**s) is found to be uniquely dependent upon the sequence composition. This demonstrates that the neighboring **A**s or **C**s in ssRNAs have variable electrostatic efficiency to interact with the central **G**/***G***$^-$, which is owing to the variable pseudoaromatic characters (giving variable chemical reactivities) of the flanking **A**s or **C**s compared to those of the isosequential ssDNAs. The sequence-dependent variation of pK_a of the central ***G*** and the modulation of its pK_a transmission through the nearest-neighbors by variable electrostatic interaction is owing to the electronically coupled nature of the constituent nucleobases across the single strand, which demonstrates the unique chemical basis of the sequence context specificity of DNA or RNA in dictating the biological interaction, recognition, and function with any specific ligand.

Introduction

The sequence specificity of nucleic acid function[1a,b] can be broadly classified into two types: First involves ligand recognition and interaction by specific nucleic acid sequences such as in protein recognition by both ssDNA[2] and ssRNA[3] and/or specific ligand recognition/binding by RNA[4,5] (as in the aptamer[4] produced by in vitro selection[5b,c] process or in antibiotic binding by ribosomal RNA[5d] or in DNAzyme[6] interaction with different ligands). Many ssDNAs show[2a−h] their functional properties upon binding to specific proteins. These ssDNA binding proteins recognize a particular sequence, as found in transcriptional regulation[2a], in telomere replication,[2b] as well as in recombination (RecA protein),[2c,d] or in replication (RepA protein)[2e]. Similarly, four main sequence specific ssRNA binding proteins[3a] have been recognized as in sex-lethal protein from *D. melanogaster*[3b], N-terminal domain of polyA binding protein (PABP)[3c], Trp RNA-binding attenuation protein (TRAP),[3d] and transcription termination factor Rho protein[3e]. Recent thermodynamic, kinetic, and structural analyses[7] of DNA/RNA interactions with proteins have also shown in general that the specific protein binding is dictated by specific DNA/RNA sequence context, which can be exemplified by specific interaction of a

(1) (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: Berlin, 1988. (b) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic Acids: Structures, Properties and Functions*; University Science Books: Sausalito, CA, 1999. (c) Leninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of Biochemistry*, 2nd ed.; Worth Publishers Inc.: New York, 1993. (d) Cech, T. R.; Golden, B. L. In *The RNA World*, 2nd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: 1999, 321.

(2) (a) Swamynathan, S. K.; Nambiar, A.; Guntaka, R. V. *FASEB J.* **1998**, *12*, 515. (b) Anderson, E. M.; Halsey, W. A.; Wuttke, D. S. *Biochemistry* **2003**, *42*, 3751. (c) Nishinaka, T.; Ito, Y.; Yokoyama, S.; Shibata, T. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 6623. (d) Bar-Ziv, R.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 9068. (e) Bochkareva, E.; Belegu, V.; Korolev, S.; Bochkarev, A. *EMBO J.* **2001**, *20*, 612. (f) Toulmé, J.-J. *NATO ASI Ser., Ser. A: Life Sci.* **1985**, *101 (Chromosomal Protein Gene Expression)*, 263. (g) Weinfeld, M.; Soderlind, K.-J. M.; Buchko, G. W. *Nucleic Acids Res.* **1993**, *21*, 621. (h) Beckingham, J. A.; Cleary, J.; Bobeck, M.; Glick, G. D. *Biochemistry* **2003**, *42*, 4118. (i) Rooman, M.; Liévin, J.; Buisine, E.; Wintjens, R. *J. Mol. Biol.* **2002**, *319*, 67.

(3) (a) Antson, A. A. *Curr. Opin. Struct. Biol.* **2000**, *10*, 87. (b) Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shimura, Y.; Muto, Y.; Yokoyama, S. *Nature* **1999**, *398*, 579. (c) Deo, R. C.; Bonanno, J. B.; Sonenberg, N.; Burley, S. K. *Cell* **1999**, *98*, 835. (d) Bogden, C. E.; Fass, D.; Bergman, N.; Nichols, M. D.; Berger, J. M. *Mol. Cell* **1999**, *3*, 487. (e) Antson, A.; Dodson, E. J.; Dodson, G.; Greaves, R. B.; Chen, X. P.; Gollnick, P. *Nature* **1999**, *401*, 235. (f) Messias, A. C.; Sattler, M. *Acc. Chem. Res.* **2004**, *37*, 279.

(4) (a) Patel, D. J.; Suri, A. K. *Rev. Mol. Biotechnol.* **2000**, *74*, 39. (b) Luzi, E.; Minunni, M.; Tombelli, S.; Mascini, M. *Trends Anal. Chem.* **2003**, *22*, 810.

(5) (a) Jones, S.; Daley, D. T. A.; Luscombe, N. M.; Berman, H. M.; Thornton, J. M. *Nucleic Acid Res.* **2001**, *29*, 943. (b) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818. (c) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611. (d) Ramakrishnan, V. *Cell* **2002**, *69*, 557.

(6) (a) Beaudry, A. A.; Joyce, G. F. *Science* **1992**, *257*, 635. (b) Li, Y. F.; Breaker, R. R. *Curr. Opin. Struct. Biol.* **1999**, *9*, 315. (c) Jaschke, A.; Seelig, B. *Curr. Opin. Chem. Biol.* **2000**, *4*, 257. (d) Feldman, A. R.; Sen, D. *J. Mol. Biol.* **2001**, *313*, 283. (e) Sidorov, A. V.; Grasby, J. A.; Williams, D. M. *Nucleic Acids Res.* **1993**, *32*, 1591.

(7) (a) Yuan, X.; Davydova, N.; Conte, M. R.; Curry, S.; Matthews, S. *Nucleic Acids Res.* **2002**, *30*, 456. (b) Tanner, J. J.; Komissarov, A. A.; Deutscher, S. L. *J. Mol. Biol.* **2001**, *314*, 807. (c) Ding, J.; Hayashi, M. K.; Zhang, Y.; Manche, L.; Krainer, A. R.; Xu, R. *Genes Dev.* **1999**, *13*, 1102. (d) Hudson, B. P.; Martinez-Yamount, M. A.; Dyson, H. J.; Wright, P. E. *Nat. Struct. Mol. Biol.* **2004**, *11*, 257.

10.1021/ja048484c CCC: $27.50

short pyrimidine-rich (UCUUC) sequence[6a] of RNA with polypyrimidine tract binding protein (PTB),an antigen-binding fragment[6b] bound to ssDNA, heterogeneous nuclear ribonucleoprotein (hnRNP) A1 recognition[6c] by ssDNA [d(TTAGGG)$_n$ repeats], and binding of an AU-rich element in the 3′ untranslated region of target mRNA[6d] by tandem zinc finger domain of the protein TIS11d. Stacking between aromatic amino acids and nucleic acid bases (mainly enthalpic in nature) plays an important role[2g,i] in the enzyme specificity with nucleic acid substrate.

The second category of sequence-specific interaction[8] involves base pairing through tertiary interaction involving folding and scaffold building with certain complementary nucleobases (with metal ion as cofactors) through intra- or intermolecular interactions as in the sequence specific cleavage activity for group I[8a,b,i,*o*,p] and group II intron ribozyme,[8d,e,i,p] RNase P RNA,[8c] HDV ribozyme,[8m,n] hammerhead ribozyme,[8f,i,k,q,r] as well as in the substrate sequence specificity surrounding the cleavage site for hairpin ribozyme[8i,j,s] or in the RNA−RNA interaction and recognition (as in kissing hairpins[9a−d] or codon−anticodon recognition[1c] in the ribosome), and unnatural allosteric ribozyme.[9e]

Various modes of stacking (offset, edge-to-face, face-to-face)[10a−c] stabilizing donor−acceptor interactions are electrostatic in nature beside van der Waal and dispersion forces,[10d] which are very difficult to dissect experimentally.[10a,b] The electrostatic forces (Coulombic repulsive or attractive terms) however are shown to play a dominant role in dictating the strength of both stacking[10a−c,11,12] and hydrogen bonding.[13] The free energy of stabilization of both stacking and hydrogen bonding is however dependent upon the relative electronic character (partial ionic charges owing to polarization), as well as by the resulting relative acid−base properties of the donor and the acceptor[13a,e,f,14] which is steered inter alia by the hydrophobic character of the microenvironment.[15] While it is clear that the sequence context of the single-stranded DNA and RNA plays a key role in their interaction with protein (or any other ligand as in the aptamer), it is not yet however understood how this sequence context dictates such a central role in modulating the chemical character of the constituting nucleobases in order to negotiate the biological recognition, interaction, and function. No unambiguous proof has hitherto been available that allows us to understand this in a straightforward manner or how this actually works in physicochemical terms.

It is however clear for some time now that the nearest-neighbor stacking interaction[16] plays an important role in the dangling oligonucleotides, which can actually control the duplex stability. Some more recent evidences[11,12] also suggest that the stacked neighboring nucleobases in oligonucleotides constitute an electronically coupled system, mutually modulating each other's chemical nature and reactivities. Here, we present unambiguous pK_a evidence showing that the chemical nature (pseudoaromaticity) of a specific nucleobase (N) in a given DNA or RNA sequence depends on the chemical nature of the nearest neighbors.

Results and Discussion

We have employed here the pH titration of the aromatic protons in the isosequential heptameric ssDNAs (**8a**−**11a**) and ssRNAs (**8b**−**11b**) by 1D NMR (pH 6.7−12.5) and performed their Hill plot analysis (Figures S1 and S2 in the Supporting Information) for the purpose of understanding the sequence-dependent pK_a modulation of the central ***G*** moiety in the following eight DNA/RNA sequences shown in Figure 1: [d/r(5′-CAQ^1***G***Q^{2}AC-3′): Q^1 = Q^2 = A (**8a**/**8b**) or C (**11a**/**11b**), Q^1 = A, Q^2 = C (**9a**/**9b**), Q^1 = C, Q^2 = A (**10a**/**10b**)]. The trimeric ssDNA/ssRNA molecules, [d/r(A***G***A) (**2a**/**2b**), d/r(A***G***C) (**3a**/**3b**), d/r(C***G***A) (**4a**/**4b**), d/r(C***G***C) (**5a**/**5b**)], constituting the central trinucleotidyl part of the corresponding heptamer, have been used as internal reference compounds. In the isosequential heptameric ssDNAs (**8a**−**11a**) and ssRNAs (**8b**−**11b**), the central trimer sequences (**2**−**5**) are extended both at the 3′ and 5′ ends by the **AC** residues. These ssDNA/ssRNA molecules are designed in such a way that, among all the aromatic residues, only a single anionic species at the N^1 of the ***G*** moiety in the middle of the above sequences can be produced in the alkaline pH. The ***G*** is situated in the middle of these ssDNA/ssRNA sequences (Figure 1) with 5′-purine(**A**)-***G***-purine(**A**)-3′ or 5′-purine(**A**)-***G***-pyrimidine(**C**)-3′

(8) (a) Cech, T. R.; Zaug, A. J.; Grabowski, P. J. *Cell*, **1981**, *27*, 487. (b) Cech, T. R. *Annu. Rev. Biochem.* **1990**, *59*, 543. (c) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849. (d) Padgett, R. A.; Poder, M.; Boulanger, S. C.; Perlman, P. S. *Science* **1994**, *266*, 1685. (e) Sontheimer, E. J.; Gordon, P. M.; Piccirilli, J. A. *Genes Dev.* **1999**, *13*, 1729. (f) Symons, R. H. *Annu. Rev. Biochem.* **1992**, *61*, 641. (g) Noller, H. F.; Hoffarth, V.; Zimniak, L. *Science* **1992**, *256*, 1416. (h) Nissen, P.; Hansen, J.; Ban, N.; Moor, P. B.; Steitz, T. *Science* **2000**, *289*, 920. (i) Takagi, Y.; Warashina, M.; Stec, W. J.; Yoshinari, K.; Taira, K. *Nucleic Acids Res.* **2001**, *29*, 1815. (j) Perrotta, A: T.; Been, M. D. *Nature* **1999**, *350*, 434. (k) Stage-Zimmermann, T. K.; Uhlenbeck, O. C. *RNA* **1998**, *4*, 875. (l) Fedor, M. J. *J. Mol. Biol.* **2000**, *297*, 269. (m) Ferré-D'Amaré, A. R.; Zhou, K.; Doudna, J. A. *Nature* **1998**, *395*, 567. (n) Nakano, S.; Chadalavada, D. M.; Bevilacqua, P. C. *Science* **2000**, *287*, 1493. (o) Shan, S.; Herschlag, D. *RNA* **2000**, *6*, 795. (p) Xiang, Q.; Qin, P. Z.; Michels, W. J.; Freeland, K.; Pyle, A. M. *Biochemistry* **1998**, *37*, 3839. (q) Kore, A. R.; Vaish, N. K.; Kutzke, U.; Eckstein, F. *Nucleic Acids Res.* **1998**, *26*, 4116. (r) Ohmichi, T.; Kool, E. C. *Nucleic Acids Res.* **2000**, *28*, 776. (s) DeRose, V. J. *Chem. Biol.* **2002**, *9*, 961.

(9) (a) Simon, R. W.; Kleckner, N. *Annu. Rev. Genet.* **1988**, *22*, 567. (b) Eguchi, Y.; Itoh, T.; Tomizawa, J. I. *Annu. Rev. Biochem.* **1991**, *60*, 631. (c) Eguchi, Y.; Tomizawa, J. I. *Cell* **1990**, *60*, 199. (d) Eguchi, Y.; Tomizawa, J. I. *J. Mol. Biol.* **1991**, *220*, 831. (e) Jaschke, A. *Curr. Opin. Struct. Biol.* **2001**, *11*, 321.

(10) (a) Hunter, C. A.; Sanders, J. K. M. *J. Am. Chem. Soc.* **1990**, *112*, 5525. (b) Hunter, C. A.; Lawson, K. R.; Perkins, J.; Urch, C. J. *J. Chem. Soc., Perkin Trans. 2* **2001**, 651. (c) Meyer, E. A.; Castellano, R. K.; Diederich, F. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210. (d) Ribas, J.; Cubero, E.; Luque, J.; Orozco, M. *J. Org. Chem.* **2002**, *67*, 7057.

(11) We have recently shown[12] that the ubiquitous electrostatic interactions amongst the nearest-neighbor pseudoaromatic nucleobases in ssRNA both in the neutral and in the ionic states, modeling ligand binding to nucleic acid bases, actually give a measure of the modulation of the pseudoaromatic properties of the constituent nucleobases across the ssRNA.

(12) (a) The pH-dependent NMR studies of single-stranded (ss) di-ribonucleotides demonstrated that the pK_a of a single ionized nucleobase could be measured from the aromatic marker protons of the neighboring nucleobase because they constitute an electronically coupled π-system via intramolecular offset stacking. Acharya, S.; Acharya, P.; Földesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2002**, *124*, 13722. (b) The pH-dependent NMR studies of intramolecular offset stacking in the single-stranded triribonucleotides. Acharya, P.; Acharya, S.; Földesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2003**, *125*, 2094. (c) The pH-dependent NMR titration studies of the hexameric ssRNAs (5′-***G***AAAAC-3′) with a single ionizing nucleotide residue, 5′-***G***, showed that the interplay of nearest-neighbor electrostatic interactions across the hexameric ssRNA chain propagated all the way up to the sixth nucleobase residue. Acharya, P.; Acharya, S.; Cheruku, P.; Amirkhanov, N. V.; Földesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2003**, *125*, 9948.

(13) (a) Shan, S.-O.; Herschlag, D. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 14474. (b) Chen, D. L.; McLaughlin, L. W. *J. Org. Chem.* **2000**, *65*, 7468. (c) Chen, J.; McAllister, M. A.; Lee, J. K.; Houk, K. N. *J. Org. Chem.* **1998**, *63*, 4611. (d) Schmideder, H.; Kasende, O.; Merz, H.; Rastogi, P. P.; Zundel, G. *J. Mol. Struct.* **1987**, *161*, 87. (e) *Annu. Rev. Phys. Chem.***1997**, *48*, 511. (f) Frey, P. A. *Magn. Reson. Chem.* **2001**, *39*, S190.

(14) Acharya, P.; Cheruku, P.; Chatterjee, S.; Acharya, S.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2004**, *126*, 2862.

(15) (a) Legault, P.; Pardi, A. *J. Am. Chem. Soc.* **1994**, *116*, 8390. (b) Narlikar, G. J.; Herschlag, D. *Annu. Rev. Biochem.* **1997**, *66*, 19. (c) Fersht, A. *Enzyme Structure and Mechanism*; W. H. Freeman: New York, 1984.

(16) (a) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290*, 967. (b) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28*, 1929. (c) Ohmichi, T.; Nakano, S-i.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2002**, *124*, 10367.

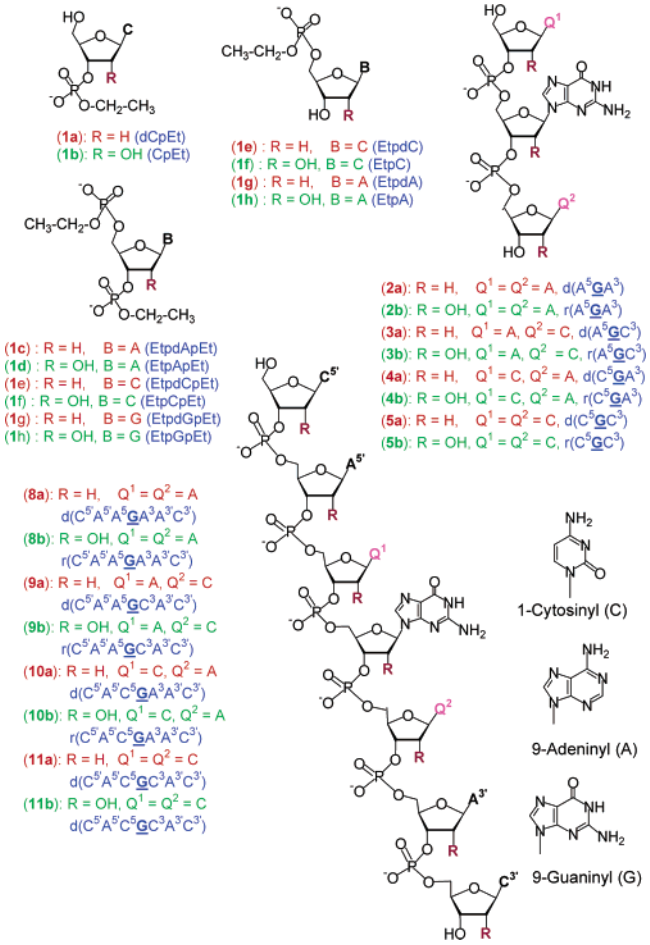

Figure 1.

or 5′-pyrimidine(**C**)-*G*-purine(**A**)-3′ or 5′-pyrimidine(**C**)-***G***-pyrimidine(**C**)-3′ nucleobases. We reasoned that the comparison of the physicochemical properties of these isosequential ssDNAs and ssRNAs within the set of the trimeric or the heptameric units by simple pH-dependent ^{1}H NMR titration studies (^{1}H at 500 or 600 MHz) as well as by chemical shift comparison in the neutral (N) and deprotonated (D) states (i.e., relative stacking/destacking) should tell us about the sequence-dependent electronic environment around ***G*** (apparent pK_{a1}),[17] it should also show the relative strength of the electrostatic propagation[11,12] of the ***G***$^-$ through the neighboring nucleobases (apparent pK_{a2})[17] across the strand of various DNA or RNA sequences. Clearly, the efficiency of this electrostatic propagation of the ***G***$^-$ will be dictated by the chemical nature of the electronically coupled neighboring nucleobases within a sequence context. We also argued that the comparison of the trimers with the heptamers should shed light on how the electrostatic effect of the ***G***$^-$ in the trimeric sequences is modulated when it is inserted in to a larger oligomer with an altered sequence context (Figure 1).

(A) Accuracy of the pH-Dependent NMR Titration Studies. The pK_a's reported (Table 1) for the N^1 center of ***G*** (obtained from δH8***G*** as well as from other marker protons of the neighboring residues, Table 1 and Figure S1 in the Supporting Information) have been obtained by the Hill plot analysis (Figure S2 in the Supporting Information). The error (see Experimental Section for details) in the chemical shift is ±0.001 ppm at 298 K, and the corresponding error in pK_a determination is ±0.01 to ±0.04 except for δH6 of **C**3′ in **9b** (error ±0.05). A negligibly small salt effect on the pK_a shift (±0.003)[18] (a total of 12 mM NaOD was added in small portions to shift the pH from 6.7 to 12.2 through 20−30 pH points. Initial salt concentration at pH 6.6 is zero, see Experimental Section for positive control) and a very small error in the pH readings before

(17) (a) Intrinsic. pK_a of the nucleobase at a specific site in a monomer is the one arising directly from a single ionizable group within a specific pH range with the intrinsic electrostatic interference[15b,c] of the constituent 3′ or/and 5′-phosphates (which is already ionized, pK_a = 1.5[22]), but in the absence of any nearest-neighbor, as we see in the pK_a of N^1 of ***G*** in the monomeric Etp(d/r***G***)pEt (**1g** [pK_a = 9.59 ± 0.01]/**1h** [pK_a = 9.29 ± 0.01]). The presence of electrostatic interaction of ***G*** with additional neighboring electronic groups *such as nucleobases, phosphates, and the pentose-sugar units across the single-stranded* nucleic acids causes the modulation/perturbation of the intrinsic pK_a of the ***G*** (apparent pK_{a1}), as found in the trimeric and heptameric ssDNAs (**2a**−**5a** and **8a**−**11a**) /ssRNAs (**2b**−**5b** and **8b**−**11b**). This can be quantified by comparison with that of the respective monomer unit, Etp(d/r***G***)pEt (**1g**/**1h**). Ideally, such a comparison of the apparent pK_{a1} of the ***G*** is best done *within* the subset of the trimers **2a**−**5a** or the heptamers **8a**−**11a**, thereby ensuring that the electrostatic effect due to the number of phosphates and sugars remains the same within the respective group. In our ssDNAs/ssRNAs, the marker proton from the neighboring nucleobases (**A** or **C**), as well as the ionic phosphates, which are nontitrable in the pH range 6.6−12.5, also shows the pK_a of ***G*** because of the variable electrostatic interaction of ***G***/***G***$^-$ with the neighboring electronic groups [apparent pK_{a2}]. Three distinct cases of variation of apparent pK_{a2} (with respect to apparent pK_{a1}), within each group of the trimers or the heptamers, have been observed in this work: (i) when apparent pK_{a2} > apparent pK_{a1}, it implies an additional electrostatic energy input from the electronic character of the neighboring nucleobase (Type 2 effect);[12c] (ii) when apparent pK_{a2} < apparent pK_{a1}, it suggests an electrostatic screening by the solvent; (iii) when apparent pK_{a1} = apparent pK_{a2}, it means no additional electrostatic energy input from the electronic character of the neighboring nucleobase (Type 1 effect).[12c] The apparent pK_{a2} (Type 2 effect) of the ***G*** observed from other neighboring nucleobases gives the relative increase of the strength of its cross-modulation by a specific neighboring nucleobase in a *sequence-context-dependent* manner. This is because of the differential electrostatic interaction of the neighboring nucleobases (through their variable electronic nature, pseudoaromaticity), owing to their distinctive microenvironments (which dictate the local dielectrics). This is consistent with the fact that the pK_a of the COOH group of, for example, 4-methoxy salicylic acid decreases (pK_a of 3.2) in water (ϵ = 78) (i.e., more acidic) compared to that (pK_a of 7.1) in DMSO (ϵ = 48),[13a] owing to a greater increase of H-bond strength accompanying charge rearrangements (indicated by the pK_a values of the donor/acceptor) in a relatively weaker dielectric medium, as in DMSO, compared to water. This means that when apparent pK_{a2} > apparent pK_{a1}, we have the electrostatic energy input from the corresponding reporter nucleobase owing to its nearest-neighbor influence and/or its location in a more hydrophobic microenvironment. Both the apparent pK_{a1} and apparent pK_{a2} are site-specific and depend on the relative difference in the microenvironments between the ionization site (at ***G*** to ***G***$^-$ in our case) and the position of the neighboring nucleobases in the sequence. This can be easily assessed[12] by the NMR giving the local thermodynamics of the variation of the strength of tandem electrostatic interaction. (b) The ionization of ***G*** itself is its apparent pK_{a1}, and the pK_a of ***G*** obtained from the neighboring nucleobases in the proximity is its apparent pK_{a2}. Both are the result of the two-state protonation D deprotonation equilibrium. The occurrence of apparent pK_{a2} at a distal site is a result of the electrostatic relay of the apparent pK_{a1} from the guanine ionization site. The actual transmission of pK_a is however modulated by the microenvironment and the resulting electronic nature of the nucleobase at the distal site. If the microenvironment around any of the distal nucleobases in a sequence is different, then the specific local hydrophobic environment around that nucleobase will (in comparison with the aqueous environment) have altered charge density and, as a result, will show different pK_{a2}'s in a sequence-specific manner. Thus, for example, three **A**s or three **C**s in DNA **9a** or RNA **9b** sequence can have, in theory, a maximum of six different microenvironments because of nonidentical nearest-neighbor interactions. Thus the two-state protonation D deprotonation equilibrium constant (*K*) for G D ***G***$^-$ is modulated differently by different nucleobases (Q^1/Q^2 or $A^{5'}$/$A^{3'}$ or $C^{5'}$/$C^{3'}$, Figure 1) at different sites to give a set of modulated readings of the same equilibrium constants for ***G***$^-$ with site Q^1 (K_1), with site Q^2 (K_2), with site $A^{5'}$ (K_3), with site $A^{3'}$ (K_4), etc., which are the apparent pK_{a2} readings. Thus, each of these modulated pK_{a2} readings of the equilibrium constants, K_1, K_2, K_3, or K_4, is proportional to the respective electrostatic potential energy (E) at each site [$E = q_1q_2/4\pi\epsilon_o r$, where q_1 = ***G***$^-$ and q_2 = nucleobases, Q^1 or Q^2 or $A^{5'}$ or $A^{3'}$ or $C^{5'}$ or $C^{3'}$ in d/r(5′-$C^{5'}A^{5'}Q^1$**N**$Q^2A^{3'}C^{3'}$-3′ shown in Figure 1); ϵ_o = permittivity factor]. Note that the individual charge density of each nucleobase within the DNA/RNA sequence depends on the sequence context with variable ϵ_o depending upon the relative nature of hydrophobicity or hydrophilicity around respective nucelobases (see above). Thus, $K_1 \neq K_2 \neq K_3 \neq K_4$ within the same sequence when the microenvironment around each nucleobase is different, and this can only be assessed when there is a single ionization point as ***G***$^-$.

(18) (a) Li, Yi.; Breaker, R. R. *J. Am. Chem. Soc.* **1999**, *121*, 5364. (b) Kao, Y. H.; Fitch, C. A.; Bhattacharya, S.; Sarkisian, C. J.; Lecomte, J. T. J.; García-Moreno, B. E. *Biophys. J.* **2000**, *79*, 1637.

Table 1. Comparison of pK_a[a] of the ***G*** in ssDNAs **2a**−**11a** and ssRNAs **2b**−**11b** as well as their monomeric counterparts EtpdGpEt (**1g**) and EtpGpEt (**1h**)

Oligo-ssDNA		^{1}H	pK_a	Oligo-ssRNA		^{1}H	pK_a
EtdGpEt (**1g**)		H8G	9.59 (± 0.01)	EtrGpEt (**1h**)		H8G	9.29 (± 0.01)
d(A^{5}GA^3) (**2a**)	dA^5	H8A	10.33 (±0.02)	r(A^{5}GA^3) (**2b**)	rA^5	H8A H2A	10.04 (±0.01) 9.98 (±0.01)
	dG	H8G	10.24 (±0.01)		rG	H8G	10.03 (±0.01)
	dA^3	H8A	10.45 (± 0.03)		rA^3	H8A H2A	9.97 (±0.01) 10.03 (±0.02)
d(A^{5}GC^3) (**3a**)	dA^5	H8A H2A	10.08 (±0.01) 10.14 (± 0.06)	r(A^{5}GC^3) (**3b**)	rA^5	H8A H2A	10.08 (±0.01) 10.04 (±0.01)
	dG	H8G	10.10 (± 0.01)		rG	**	**
	dC^3	H5C H6C	10.07 (±0.01) 10.11 (± 0.01)		rC^3	H5C	10.05 (±0.01)
d(C^{5}GA^3) (**4a**)	dC^5	H5C H6C	9.91 (± 0.02) 9.95 (± 0.01)	r(C^{5}GA^3) (**4b**)	rC^5	H5C H6C	10.07 (±0.01) 10.18 (±0.01)
	dG	H8G	10.01 (± 0.01)		rG	H8G	10.25 (±0.01)
	dA^3	H2A	9.98 (± 0.01)		rA^3	H8A	10.19 (±0.01)
d(C^{5}GC^3) (**5a**)	dC^5	H5C H6C	9.42 (± 0.04) 9.50 (± 0.01)	r(C^{5}GC^3) (**5b**)	rC^5	H5C H6C	10.20 (±0.01) 10.26 (±0.02)
	dG	H8G	9.49 (± 0.01)		rG	H8G	10.13 (±0.01)
	dC^3	H5C H6C	9.37 (± 0.01) 9.37 (± 0.01)		rC^3	H5C H6C	10.11 (±0.01) 10.16 (±0.01)
d(C$^{5'}$A^{5}A^{5}GA^{3}A$^{3'}$C$^{3'}$) (**8a**)	dC$^{5'}$	-	-	r(C$^{5'}$A$^{5'}$A^{5}GA^{3}A$^{3'}$C$^{3'}$) (**8b**)	rC$^{5'}$	H6C	10.41 (±0.01)
	dA$^{5'}$	H2A	11.19 (±0.02)		rA$^{5'}$	H2A	10.40 (±0.01)
	dA^5	H2A	11.13 (±0.01)		rA^5	H8A H2A	10.58 (±0.01) 10.35 (±0.01)
	dG	H8G	11.06 (±0.01)		rG	H8G	10.58 (±0.01)
	dA^3	H8A H2A	10.86 (±0.02) 11.21 (±0.02)		rA^3	H2A	10.39 (±0.01)
	dA$^{3'}$	H8A	11.15 (±0.01)		rA$^{3'}$	H8A	10.46 (±0.02)
	dC$^{3'}$	H5C H6C	11.13 (±0.02) 11.01 (±0.01)		rC$^{3'}$	-	-
d(C$^{5'}$A$^{5'}$A^{5}GC^{3}A$^{3'}$C$^{3'}$) (**9a**)	dC$^{5'}$	-	-	r(C$^{5'}$A$^{5'}$A^{5}GC^{3}A$^{3'}$C$^{3'}$) (**9b**)	rC$^{5'}$	H6C	10.53 (±0.02)
	dA$^{5'}$	H2A	10.88 (±0.02)		rA$^{5'}$	H2A	10.46 (±0.03)
	dA^5	H8A H2A	10.67 (±0.03) 10.76 (±0.02)		rA^5	H8A	10.80 (±0.04)
	dG	H8G	10.74 (±0.02)		rG	**	**
	dC^3	H5C	10.80 (±0.02)		rC^3	H6C	10.31 (±0.02)
	dA$^{3'}$	H8A H2A	10.61 (±0.03) 10.83 (±0.02)		rA$^{3'}$	H8A	10.34 (±0.03)
	dC$^{3'}$	H5C H6C	10.65 (±0.03) 10.71 (±0.02)		rC$^{3'}$	H5C H6C	10.31 (±0.03) 10.87 (±0.08)
d(C$^{5'}$A$^{5'}$C^{5}GA^{3}A$^{3'}$C$^{3'}$) (**10a**)	dC$^{5'}$	H5C H6C	10.85 (±0.04) 10.78 (±0.03)	r(C$^{5'}$A$^{5'}$C^{5}GA^{3}A$^{3'}$C$^{3'}$) (**10b**)	rC$^{5'}$	H6C	10.50 (±0.02)
	dA$^{5'}$	-	-		rA$^{5'}$	H8A H2A	10.60 (±0.01) 10.50 (±0.02)
	dC^5	H5C H6C	10.71 (±0.04) 10.64 (±0.04)		rC^5	-	-
	dG	H8G	10.79 (±0.04)		rG	**	**
	dA^3	H8A	10.69 (±0.04)		rA^3	H8A H2A	10.56(±0.01) 10.59(±0.01)
	dA$^{3'}$	-	-		rA$^{3'}$	H8A H2A	10.37(±0.02) 10.17(±0.03)
	dC$^{3'}$	-	-		rC$^{3'}$	-	-
d(C$^{5'}$A$^{5'}$C^{5}GC^{3}A$^{3'}$C$^{3'}$) (**11a**)	dC$^{5'}$	H5C H6C	10.41 (± 0.01) 10.55 (± 0.01)	r(C$^{5'}$A$^{5'}$C^{5}GC^{3}A$^{3'}$C$^{3'}$) (**11b**)	rC$^{5'}$	H6C	10.60 (± 0.02)
	dA$^{5'}$	H2A	10.37 (± 0.02)		rA$^{5'}$	H8A H2A	10.69 (± 0.03) 10.59 (± 0.02)
	dC^5	H5C H6C	10.31 (± 0.01) 10.25 (± 0.01)		rC^5	H5C H6C	9.78 (± 0.02) 9.74 (± 0.02)
	dG	H8G	10.39 (± 0.01)		rG	H8G	10.09 (± 0.02)
	dC$^{3'}$	H5C H6C	10.32 (± 0.01) 10.58 (± 0.01)		rC^3	H5C H6C	10.99(± 0.03) 10.70 (± 0.04)
	dA$^{3'}$	H8A	10.42 (± 0.02)		rA$^{3'}$	H8A	10.47 (± 0.01)
	dC$^{3'}$	H5C H6C	10.36 (± 0.01) 10.45 (± 0.02)		rC$^{3'}$	H5C	10.37 (± 0.01)

[a] All pK_a values and the corresponding errors have been calculated from Hill plot analyses (Figure S2 in the Supporting Information). [b] ** denotes see ref 19 for the explanation of the negligible pH-dependent chemical shift change for H8G.

and after each NMR titration point (±0.025) were observed; hence, no buffer was used for our study (see Experimental Section). A sample concentration of 1 mM has been used to rule out any self-association.

(B) Modulation of pK_a of *G* by Electrostatic Interaction through ssDNA and ssRNA Strand is Sequence-Dependent. It has been found that the central ***G*** residue in the isosequential heptameric DNA and RNA molecules (Figure 1) shows a sequence-dependent variation in its pK_a (Table 1) owing to its variable electrostatic interactions[17] with the neighboring nucleobases. It is also observed that the electrostatic effect of the G^- formation is transmitted differently through the neighboring

nucleobases across the strand depending upon the sequence context.[17] The modulation of the pK_a of ***G***, as observed from each marker proton of the neighboring **A** and **C** moieties (Table 1, Figure 1), is a result of stepwise (~3.5 Å) nearest-neighbor electrostatic propagation in the electronically coupled ssDNAs or ssRNAs. The apparent pK_{a1} of ***G*** in the group of four trimeric ssDNAs **2a**−**5a** varies from 9.49 ± 0.01 to 10.24 ± 0.01 (ΔpK_{a1} = 0.75), whereas in the group of four heptameric ssDNAs **8a**−**11a**, the apparent pK_{a1}[17] of ***G*** varies from 10.39 ± 0.01 to 11.06 ± 0.01 (ΔpK_{a1} = 0.67). On the other hand, the apparent pK_{a1}[17] of ***G*** in the four trimeric ssRNAs **2b**−**5b** varies from 10.03 ± 0.01 to 10.25 ± 0.01 (ΔpK_{a1} = 0.22), while in a group of four heptameric ssRNAs **8b**−**11b** the apparent pK_{a1}[17] variation is from 10.09 ± 0.02 to 10.58 ± 0.01 (ΔpK_{a1} = 0.49). It is noteworthy that all the above ΔpK_{a1} within the group of trimers or group of heptamers are well above the pK_a estimation error (see Experimental Section for details) of ±0.01 to ±0.04. Thus, the apparent pK_{a1}[17] variation of the ***G*** observed within the set of the trimeric or the heptameric ssDNAs/ssRNAs is a measure of how different/variable is its pseudoaromatic character within each group as a result of variation of the respective sequence context, in which the modulation by the phosphate charges or/and the pentose sugar units within each group of trimers or heptamers remains the same. On the other hand, when the apparent pK_{a1} variation of the ***G*** in ssDNA trimers are compared with the corresponding ssDNA heptamers and, similarly, the ssRNA trimers with the corresponding ssRNA heptamers (in both sets of comparisons the trimeric sequence has been placed in the middle of the heptamer sequence), one can clearly observe the effect of different microenvironments owing to their respective nearest-neighbors depending upon the chain length, phosphate charges or/and the pentose sugar units, as well as the sequence-context owing to variable stacking.

The cross-modulation of apparent pK_{a1} of ***G*** in the neighboring nucleobases, owing to tandem nearest-neighbor electrostatic interactions, is also evident from the variation of the apparent pK_{a2}[17] obtained from each marker proton of its neighboring nucleobase moieties, **A** and **C** in trimers **2a**−**5a** and **2b**−**5b** and heptamers **8a**−**11a** and **8b**−**11b**. The apparent pK_{a2} (Type 1 effect)[17] is the most predominant type of electrostatic interactions among the coupled nucleobases in all trimeric ssDNAs and ssRNAs and heptameric ssDNA sequences.

In contradistinction, both Type 1 and Type 2 effects[17] were observed in the apparent pK_{a2} for the heptameric ssRNAs **8b**−**11b**, depending upon their respective sequence context (Table 1). Heptameric ssRNA sequence r(5′-CAA***G***AAC-3′) (**8b**) (***G*** flanked by 5′-**A** and 3′-**A**) showed comparatively poorer (ΔpK_{a2} = 0.23) cross-modulation of pK_a of ***G*** by the neighboring nucleobases [i.e., apparent pK_{a2} (Type 1 effect)[17]], while the sequences r(5′-CAA***G***CAC-3′) (**9b**) (ΔpK_{a2} 0.56, ***G*** is flanked by 5′-**A** and 3′-**C**), r(5′-CAC***G***AAC-3′) (**10b**) (ΔpK_{a2} = 0.43, ***G*** is flanked by 5′-**C** and 3′-**A**), and r(5′-CAC***G***CAC-3′) (**11b**) (ΔpK_{a2} = 1.25, ***G*** is flanked by 5′-**C** and 3′-**C**) show the cross-modulation of pK_a [i.e., ΔpK_{a2}, where $\Delta pK_{a2} = \{(pK_{a1}\text{ of }\boldsymbol{G})_{\text{from }\delta\text{H8G in ssDNA or ssRNA}}\} - \{(pK_{a2}\text{ of }\boldsymbol{G})_{\text{from marker protons (H8A/H2A/H5C/H6C) of ssDNA or ssRNA}}\}$],[17] owing to the variation of the apparent pK_{a2} (Type 2 effect)[17] as a result of sequence-dependent nearest-neighbor effect. *This means that the pK_{a2} perturbation[17] within the sequence is maximum when the neighboring nucleobases of* ***G*** *are pyrimidines, whereas it is minimum when the neighboring nucleobases of* ***G*** *are purines.*

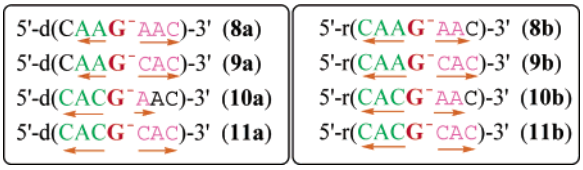

Figure 2. Propagation of the electrostatic interaction at the 3′- and the 5′-ends as a result of $\boldsymbol{G}^-$ formation (propagation (→) at the 3′-end is shown by pink, whereas the propagation at the 5′-end is shown by green; compare apparent pK_{a2} in Table 1).

(C) Electrostatics Modulation Through ssDNA and ssRNA Chain is not Uniform Toward the 3′- and 5′-Direction. The trimeric sequences show similar electrostatic transmission at both the 3′- and 5′-direction of ***G*** for both ssDNA **2a**−**5a** and ssRNA **2b**−**5b**. In contradistinction, the transmission of electrostatics interaction (manifested in apparent pK_{a-2} obtained from the neighbors) of ***G*** in 3′- and 5′-directions in heptameric ssDNAs/ssRNAs is different depending upon the sequence context in the deoxy versus ribo series, which is evident from the pH-dependent chemical shift change [evident from $\Delta\delta_{(N-D)}$, Figure S3 in the Supporting Information] of the marker protons from the neighboring nucleobases. The distance up to which the effect of electrostatics of $\boldsymbol{G}^-$ is transmitted through the neighboring nucleobases (as evident by observation of apparent pK_{a2}) is shown in Figure 2.

In the trimeric ssDNAs **2a**−**5a** and ssRNAs **2b**−**5b** as well as in the heptameric ssDNAs **8a**−**11a** and ssRNA **8b**, the apparent pK_{a2}'s of ***G*** observed from the marker protons of the neighboring nucleobases in both the 3′- and 5′-ends remain almost the same as the apparent pK_{a1} observed from the δH8G itself (Table 1). This is because the pseudoaromatic characters of **A** or **C** in the above sequences have been similarly modulated by the electrostatic interaction of the $\boldsymbol{G}^-$ (apparent pK_{a2}, Type 1 effect),[17] which means that the pseudoaromatic characters of **A**s and **C**s at both 3′- and 5′-ends are uniform.

In contradistinction, the pK_a's of ***G*** in the heptameric ssRNAs **9b**, **10b**, and **11b**, as measured from the neighboring nucleobases at both 3′- and 5′-ends, are further modulated (apparent pK_{a2}, Type 2 effect)[17] from that of the pK_a of ***G*** itself (apparent pK_{a1}).[17] Thus, the variation of the pK_a of ***G*** as measured from the marker protons of **A**s (apparent pK_{a2}, Type 2 effect)[17] shows the modulation of the pseudoaromatic character of **A**s in r(C$^{5'}$A$^{5'}$A^{5}***G***C^{3}A$^{3'}$C$^{3'}$)-3′ (**9b**), r(C$^{5'}$A$^{5'}$C^{5}***G***A^{3}A$^{3'}$C$^{3'}$)-3′ (**10b**), and r(C$^{5'}$A$^{5'}$C^{5}***G***C^{3}A$^{3'}$C$^{3'}$)-3′ (**11b**): A$^{3'}$ (10.34 ± 0.03 from H8A), A$^{5'}$ (10.46 ± 0.03 from H2A), and A^{5} (10.80 ± 0.04 from H8A) in **9b**; A$^{3'}$ (10.37 ± 0.02 from H8A, 10.17 ± 0.03 from H2A), A^{3} (10.56 ± 0.01 from H8A, 10.59 ± 0.01 from H2A), and A$^{5'}$ (10.60 ± 0.01 from H8A, 10.50 ± 0.02 from H2A) in **10b**; and A$^{3'}$ (10.47 ± 0.01 from H8A), and A$^{5'}$ (10.69 ± 0.03 from H8A, 10.59 ± 0.02 from H2A) in **11b**.

This is due to the different partial ionic charges within the same nucleobase as well as owing to the differences in their intrinsic pseudoaromatic characters depending upon their nonidentical microenvironments[17]. This also means that the pseudoaromatic characters of A$^{3'}$ ≠ A$^{5'}$ ≠ A^{5} in ***9b****, A$^{3'}$ ≠ A^{3} ≠ A$^{5'}$ in* ***10b****, and A$^{3'}$ ≠ A$^{5'}$ in* ***11b*** *are nonuniform.* The pseudoaromatic characters of **C**s have also been assessed by comparing

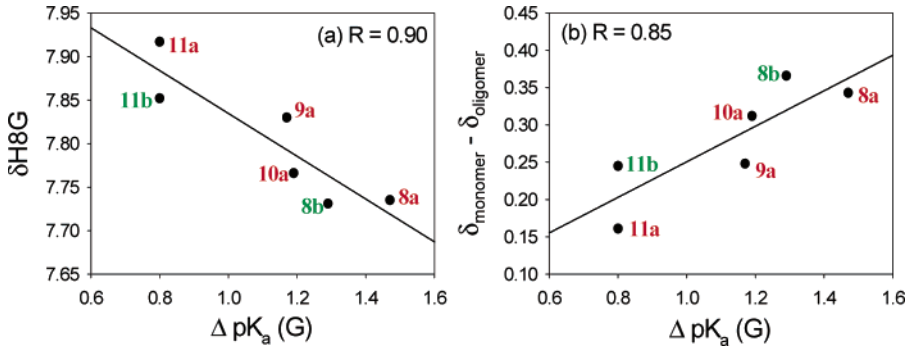

Figure 3. (A) Plot of ΔpK_a [ΔpK_a = (pK_a of G)$_{from\ dH8G\ in\ ssDNA\ or\ ssRNA}$ − (pK_a of G)$_{from\ dH8G\ in\ EtpdGpEt\ or\ EtpGpEt}$] as a function dH8G (Table S3 in the Supporting Information) for the heptameric ssDNAs **8a**−**11a** and ssRNAs **8b** and **11b**, which shows a correlation coefficient of 0.9, based on linear regression analysis. (B) Plot of the oligomerization shift (i.e., the chemical shift difference between the monomer and oligomers, Table S1 and Figure S4 in the Supporting Information) as a function of ΔpK_a in the neutral state for the heptameric ssDNAs **8a**−**11a** and ssRNAs **8b** and **11b**. For ssRNAs **9b** and **10b**, 9-guaninyl did not show any apparent pK_{a1} but showed apparent pK_{a2} (see ref 17). The plot of oligomerization shift versus ΔpK_a shows a straight line, which shows a correlation coefficient *R* of 0.85, based on linear regression analysis.

the pK_a's of ***G*** as measured from the neighboring marker protons of **C**s (apparent pK_{a2}, Type 2 effect):[13] $C^{3'}$(10.31 ± 0.03 from H5C, 10.87 ± 0.08 from H6C), C^3 (10.31 ± 0.02 from H6C), $C^{5'}$ (10.53 ± 0.02 from H6C) in **9b**; $C^{5'}$ (10.50 ± 0.02 from H6C) in **10b**; and $C^{3'}$ (10.37 ± 0.01 from H5C), C^3 (10.99 ± 0.03 from H5C, 10.70 ± 0.04 from H6C), $C^{5'}$ (10.60 ± 0.02 from H6C), and C^5 (9.78 ± 0.02 in H5C, 9.74 ± 0.02 from H6C) in **11b**. *Thus, this also shows that the pseudoaromatic characters of* $C^{3'} \neq C^3 \neq C^{5'}$ *in* ***9b*** *and* $C^{3'} \neq C^{3'} \neq C^{5'} \neq C^5$ *in* ***11b*** *are electronically nonuniform.*

(E) Correlation of ΔpK_a of *G* in the Heptameric ssDNAs and ssRNAs with Their Respective *δ*H8G as well as with the Oligomerization Shift. The apparent pK_{a1} of ***G*** as observed in heptameric ssDNAs **8a**−**11a** and ssRNAs **8b** and **11b** (Table 1) is always more basic by ca. 0.2−0.3 units compared to the pK_a of ***G*** in their monomeric counterparts[14] EtpdGpEt and EtpGpEt, respectively. This oligomerization-promoted modulation of pK_{a1} of ***G*** is due to the different phosphate charges in the oligomers compared to the monomer as well as differential nearest neighbor interactions, thereby changing the overall microenvironment. However, comparing such modulation of pK_a of ***G*** [i.e., ΔpK_a, where ΔpK_a = {(pK_a of ***G***)$_{from\ \delta H8G\ in\ ssDNA\ or\ ssRNA}$} − {($pK_a$ of ***G***)$_{from\ \delta H8G\ in\ EtpdGpEt\ or\ EtpGpEt}$}] as a function of either (i) the chemical shift of ***G*** (Figure 3A) or (ii) the oligomerization shift (Figure 3B), within the series of our heptameric ssDNA and ssRNA sequences (in which the number of phosphate charges remain the same) at the neutral pH, shows the effect of the sequence context promoted modulation of the intrinsic differences in their respective chemical environments. Thus, both the plot of ΔpK_a as a function *δ*H8G and that of the oligomerization shift (Figure S4 in the Supporting Information) as a function of ΔpK_a show a correlation with a high correlation coefficient (0.9 and 0.85, respectively). The chemical shift and the oligomerization shift of ***G*** (*δ*H8) are manifestations of the strength of stacking at ***G*** as well as the sugar−phosphate backbone conformation[19] around ***G*** in the heptameric ssDNAs and ssRNAs (see, for example, *δ*H8G in **9b** and **10b** did not show any pH-dependent chemical shift change[19] because of competing destacking and relative phosphate orientation vis-à-vis G^- formation). *Such high correlation reveals that the* pK_a *perturbation (*ΔpK_a*) of* ***G*** *in any oligomer increases with its increasing stacking interaction with the nearest-neighbors.* It is also clear from the above correlation (Figure 3) that the ones with the middle 5′-purine(A)-***G***-purine(A)-3′ sequences (as in the heptamers **8a** and **8b**) are most stacked, and those having a middle 5′-pyrimidine(C)-***G***-pyrimidine(C)-3′ sequences (as in the heptamers **11a** and **11b**) are least stacked, which are evident from the upfield chemical shifts at the N-state (Table S3 in the Supporting Information) as well as from higher oligomerization shifts (Tables S1, S2 and Figure S4 in the Supporting Information) for the former compared to those of the other sequences. However, in the case of trimeric ssDNA and ssRNA sequences, such correlations could not be obtained, as their structures are presumably more random than those of the heptamers.

Experimental Section

(A) pH-Dependent ^{1}H NMR Measurement. All NMR experiments were performed in Bruker DRX-500 and DRX-600 spectrometers. The NMR samples for compounds **2**−**5** and **8**−**11** (Figure 1) were prepared in D_2O solution (concentration of 1 mM in order to rule out any chemical shift change owing to self-association) with δ_{DSS} = 0.015 ppm as internal standard (DSS = 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt). All pH-dependent NMR measurements have been performed at 298 K. The pH values shown in pH-dependent chemical shift plots in the Figure S1 in the Supporting Information already include the correction for the deuterium effect from pH meter reading [pH = pD − 0.4].[20a] The pH meter is equipped with a calomel microelectrode

(19) In case of ssRNA trimer **3b** and hexamers **9b** and **10b**, there is almost no change in *δ*H8G with ionization of ***G*** ($\Delta\delta_{N-D}$ < 0.01 ppm, Figure S3 in the Supporting Information). In the case of **8b**, *δ*H8***G*** is deshielded ($\Delta\delta_{N-D}$ −0.085 ppm, Figure S3 in the Supporting Information) with ionization of ***G***. This can be due to three reasons: (a) The 5′-phosphate orientation with respect to ***G*** is such that it causes repulsion between the negative charge of the phosphate and that of the *π*-system of the imidazole moiety of ***G***, causing deshielding of *δ*H8G, and, consequently, would not allow the observation of formation of G^-. (b) Change in the chemical shift of H8***G*** with pH is the manifestation of G^- formation (which would cause shielding of *δ*H8G). (c) The destacking (i.e., the deshielding of *δ*H8G) of guanine from the nearest-neighbors is because of G^- formation. In cases where *δ*H8***G*** shows no change with pH, the effect of the proximity of the phosphate and the destacking are of equal strength but opposite in nature to the chemical shift change of H8G owing to G^- formation, and therefore mutually compensating each other's effect. So G^- formation is not observable from the H8G chemical shift (as in **9b** and **10b**). In the case where *δ*H8***G*** is deshielded as a function of pH, the chemical shift change due to destacking is more predominant than that due to G^- formation. This means that the actual ionization of a specific nucleobase as the pH changes may not observable by its own marker proton, H8G ($\Delta\delta_{N-D} \sim 0$, depending upon its nature of nearest-neighbor interaction), but by the pH-dependent chemical shift change of the neighboring marker protons(s). This problem can be circumvented by pH-dependent shifts by either ^{13}C- or ^{15}N-labeled oligonucleotides.[15a,21]

(20) (a) Force, R. K.; James, D. C. *Anal. Chem.* **1974**, *46*, 2049. (b) Acharya, S.; Földesi, A.; Chattopadhyaya, J. *J. Org. Chem.* **2003**, *68*, 1906.

(21) (a) Legault, P.; Pardi, A. *J. Am. Chem. Soc.* **1996**, *119*, 621. (b) Ravindranathan, S.; Butcher, S. E.; Feigon, J. *Biochemistry* **2000**, *39*, 16026

(in order to measure the pH inside the NMR tube) calibrated with standard buffer solutions (in H_2O) of pH 7 and 10. The pD of the sample has been adjusted by simple addition of microliter volumes of NaOD solutions (0.01 M, 0.1 M, and 0.5 M). The assignments for all compounds [Figure S7 in the Supporting Information] at 298 K have been performed by using ^{1}H NOESY, ^{31}P decoupled ^{1}H COSY, ^{1}H TOCSY, and ^{31}P−^{1}H correlation spectroscopy for compounds **2**−**5** and **8**−**11**. The ^{1}H NOESY spectra were also recorded at 278 K for the compounds **2**−**5** [Figure S7 in the Supporting Information] or at 283 K for the compounds **8**−**11** [Figure S7 in the Supporting Information] in the neutral pH. All ^{1}H spectra have been recorded using 128 K data points and 64 scans. All NOESY spectra for **2**−**5** and **8**−**11** were recorded on 500 and 600 MHz spectrometers with a mixing time (τ_m) of 800 ms. For each FID of NOESY, ^{31}P decoupled ^{1}H DQF−COSY and TOCSY spectra, 64 scans were recorded with a delay of 2 s and the data were zero-filled to 4 × 1 K in the t_1 and t_2 directions, then Fourier transformed, phase adjusted, and baseline corrected in both dimensions using polynomial functions. ^{31}P−^{1}H Correlation spectroscopy was performed in the absolute magnitude mode using 64 scans with a delay of 2 s, and then the spectra were zero-filled to 1 × 1 K data points in the t_1 and t_2 directions, then Fourier transformed, phase adjusted, and baseline corrected in both dimensions using polynomial functions.

(B) pH Titration of Aromatic Protons in 2−5 and 8−11. (a) Accuracy of p*K*$_a$. The pK_a's reported here for the N^1 center of ***G*** (obtained from δH8***G*** as well as from other marker protons of the neighboring residues) have been obtained by the Hill plot analysis of the pH-dependent ^{1}H chemical shifts measured by both 500 and 600 MHz NMR spectrometers using an identical condition (Figure S1 in the Supporting Information). The error in the chemical shift is ±0.001 ppm at 298 K, which represents digital resolution on the same sample with DSS as the internal reference. The error for subsequent NMR measurements on the same sample was also within ±0.001 ppm at 298 K. The error in pK_a determination is ±0.01 to ±0.04 except for δH6 of **C**3′ in **9b** (error ±0.04). All individual errors of respective pK_a values are shown in parentheses in Table 1. These accurate pK_a values allow us to safely attribute the observed pK_a differences larger than ±0.05, respectively, for various nucleobase residues to the differential intramolecular electrostatic interactions experienced by different pseudoaromatic nucleobases along the ssDNA chain. A sample concentration of 1 mM has been used for all NMR experiments in order to rule out any chemical shift change owing to self-association, although no chemical shift change was observed up to 20 mM with the monomeric phosphates. The pH measurements were performed twice inside the NMR tube, both before and after each NMR titration point (30−40 pH points within the pH range of 6.7 to 12.3 for each compound), and the pH readings were found to vary only ±0.025; hence no buffer was used for our study.

(b) Positive Control for the Determination of Salt Effect Induced Chemical Shift. Our pH-dependent positive control studies ($7.02 \leq$ pH ≤ 11.9) with r(ApA), having no ionization site at the 9-adeninyl base in this pH range, but an ionizable 2′-OH group showed negligible $\Delta\delta_{(N-D)}$ for δH8A and δH2A (0.002−0.004 ppm). However, increasing the pH further up to 12.2, the range for $\Delta\delta_{(N-D)}$ becomes 0.004−0.012 ppm without showing any plateau (indicating continued ionization of 2′-OH in such high alkaline pH; pK_a of 2′-OH vicinal to the internucleotidic phosphodiester in r(ApA) is 12.4[20b]). This chemical shift change of the aromatic protons of 9-adeninyl as the 2′-oxyanion is formed because of the change of the electron-withdrawing character of the 2′-substitutent, i.e., 2′-OH versus 2′-oxyanion.[14] Considering the pH titration for all ssRNAs, we find the maximum pH reached is 12.5 for **10b**. The pH-dependent chemical shift of those aromatic marker protons (δH8A$^{5'}$, δH2A$^{5'}$, δH2A^3) of **10b**, having observed titration profile, however, shows a clear plateau at the high pH; i.e., the last three pH points (pH 12.17, 12.33, and 12.5) indicate (Figure S1 in the Supporting Information) constant or almost negligible ($\Delta\delta \approx 0.002$ ppm) chemical shift change. This therefore indicates the negligible effect of 2′-OH ionization and salt effect at 12.5 for **10b** on the chemical shift change of its aromatic marker protons. We can therefore safely attribute any $\Delta\delta_{(N-D)} > 0.012$ ppm to the effect of the nearest-neighbor interactions. A less than 0.02 pK_a shift change was observed for a total 20 mM NaCl[18a] concentration. We therefore observe negligibly small salt effect as we gradually increase the pH from pH 6.7 (neat D_2O without any NaOD) to (∼16 mM NaOD at pH 12.2) on the pK_a shift (±0.003).[18b]

(c) pH Range Studied. The pH titration studies for d(AGA) (**2a**) (pH 6.61−12.45); d(AGC) (**3a**) (pH 6.73−12.15); d(CGA) (**4a**) (pH 6.94−12.11); d(CGC) (**5a**) (pH 7.21−11.31); d(CAAGAAC) (**8a**) (pH 7.09−12.5); d(CAAGCAC) (**9a**) (pH 6.91−12.58); d(CACGAAC) (**10a**) (pH 7.48−12.5); and d(CACGCAC) (**11a**) (pH 7.43−12.33) consist of ∼25−33 data points (see Figure S1 in the Supporting Information). Similarly the pH titration studies for r(AGA) (**2b**) (pH 6.79−11.8); r(AGC) (**3b**) (pH 7.05−11.95); r(CGA) (**4b**) (pH 6.72−12.02); r(CGC) (**5b**) (pH 7.2−12.16); r(CAAGAAC) (**8b**) (pH 7.0−12.14); r(CAAGCAC) (**9b**) (pH 7.23−12.05); r(CACGAAC) (**10b**) (pH 6.96−12.5); and r(CACGCAC) (**11b**) (pH 6.75−11.72) consist of ∼25−33 data points (Figure S1). The corresponding Hill plots for **2a**−**5a** and **8a**−**11a** as well as **2b**−**5b** and **8b**−**11b** are given in the Supporting Information (Figure S2 in the Supporting Information), and the pK_a's shown in Table 1 have been calculated from Hill plot analyses (see section C for details).

(C) p*K*$_a$ Determination. The pH-dependent [over the range of pH 6.6−12.5, with an interval of pH 0.2−0.3] ^{1}H chemical shifts[12c] (δ, with error ±0.001 ppm) for **2**−**5** and **8**−**11** show a sigmoidal behavior [Figure S1 in the Supporting Information]. The pK_a determination is based on the Hill plot analysis[12c,15a] using equation pH = log((1 − α)/α) + pK_a, where α represents the fraction of the protonated species. The value of α is calculated from the change of chemical shift relative to the deprotonated (D) state at a given pH ($\Delta_D = \delta_D - \delta_{obsd}$ for deprotonation, where δ_{obsd} is the experimental chemical shift at a particular pH), divided by the total change in chemical shift between the neutral (N) and deprotonated (D) state (Δ_T). So the Henderson−Hasselbach type equation can then be written as pH = log(($\Delta_T - \Delta_D$)/Δ_D) + pK_a. The pK_a is calculated from the linear regression analysis of the Hill plot [Figure S2 in the Supporting Information].

(D) Calculations of Oligomerization Shift. Oligomerization shifts (Figure S4 as well as Tables S1 and S2 in the Supporting Information) are calculated for the individual nucleotide residues in an oligo-ssDNA (O_{DNA}) **2a**−**5a** and **8a**−**11a** with respect to the monomeric 2′-deoxy-3′-ethylphophates (dNpEt or EtpdNpEt or EtpdN) [i.e., the difference of chemical shift: $\Delta\delta_{(dNpEt/EtpdNpEt/EtpdN\ -\ ODNA)}$, in ppm, at 298 K] at the neutral (N) and deprotonated (D) states. Similarly, oligomerization shifts are calculated for the individual nucleotide residues in an oligo-ssRNA (O_{RNA}) **2b**−**5b** and **8b**−**11b** with respect to the monomeric ribo 3′-ethylphophates (NpEt or EtpNpEt or EtpN) [i.e., the difference of chemical shift: $\Delta\delta_{(NpEt/EtpNpEt/EtpN\ -\ ORNA)}$, in ppm, at 298 K] at the neutral (N) and deprotonated (D) states. See Figure S4 as well as Tables S1 and S2 in the Supporting Information for details.

Conclusions and Implications

(1) The net result of this cross-talk, between two neighboring aglycones, is the creation of a unique set of aglycones in an oligo- or polynucleotide, whose physicochemical property and the pseudoaromatic character are completely dependent upon both the sequence makeup and whether they are stacked or unstacked. Thus the actual physicochemical integrity of **N** in d/r(5′-C$^{5'}$A$^{5'}$Q^1**N**Q^{2}A$^{3'}$C$^{3'}$-3′) is dictated by the variable pseudoaromatic character of both neighboring Q^1 and Q^2 [Q^1 =

(22) (a) Kumler, W. D.; Eiler, J. J. *J. Am. Chem. Soc.* **1943**, *65*, 2355. (b) Cozzone, P. J.; Jardetzky, O. *Biochemistry* **1976**, *15*, 4853. (c) Chamberlin, S.; Merino, E. J.; Weeks, K. M. *PNAS* **2002**, *99*, 14688.

Q^2 = A (**8a**/**8b**) or C (**11a**/**11b**), Q^1 = A, Q^2 = C (**9a**/**9b**), Q^1 = C, Q^2 = A (**10a**/**10b**), Figure 1]. The properties of Q^1 and Q^2 are further tuned by the pseudoaromatic nature of the flanking 5′-$C^{5'}A^{5'}$ and 3′-$A^{3'}C^{3'}$ nucleotides in d/r(5′-$C^{5'}A^{5'}Q^1$**N**$Q^2A^{3'}C^{3'}$-3′). It is noteworthy that the respective apparent pK_{a2} values of ***G*** observed from the marker protons of 5′-$C^{5'}A^{5'}$ and 3′-$A^{3'}C^{3'}$ nucleotides are also found to be different, which means that the electronic characters of A residues or C residues are nonuniform ($A^{5'} \neq A^{3'}$ and $C^{5'} \neq C^{3'}$). This suggests that the pseudoaromatic character of **N** (where N = ***G***) in all our heptameric DNA and RNA sequences (Figure 1) can have at least 2^4 numbers of variations, depending upon the tunable chemical nature of the neighboring Q^1 and Q^2.

(2) Recent studies have shown that the strength of a hydrogen bond (A−H···B) between a donor (A) and acceptor (B) can be assessed on the basis of pK_a difference (ΔpK_a) between the two heteroatoms involved in the hydrogen bond.[13a−d] Thus, those donor−acceptor which have stronger hydrogen bonds are those which have relatively similar pK_a values ("pK_a match"),[13a−d,14] as in the case of a very strong hydrogen bond (where ΔpK_a = 0, $\Delta H°_{[H-bond]} = -24.9$ kJ mol^{-1})[13d] between 2,4,6-trichlorophenol (pK_a 5.99) and butylamine retinal Schiff base (pK_a 5.99). Our recent studies have confirmed this[14] to show that the reason for the stronger base pairing in RNA−RNA duplexes than in the DNA−DNA duplexes is actually based on this fundamental fact that the donor and acceptor nucleobases in the monomeric components of a RNA−RNA duplex have more similar pK_a values (ΔpK_a = 5.53) than those in the DNA counterparts (ΔpK_a = 6.29). This means that the unique electronic characters of the donor and acceptor allow the H-bonded proton in RNA−RNA duplexes to be shared more equally than those in DNA−DNA duplexes. Studies have also shown[13b] the preferential strengthening of base pairing in triplex formation when the participating modified nucleobase (2-aminopyrimidine) having increased pK_a (6.8) relative to dC (4.3) to give better pK_a matching, which gives more effective protonation at the physiological pH, leading to improved hydrogen bonding capability.

It is therefore very likely that the perturbation of pK_a of a particular nucleobase due to sequence-dependent nearest-neighbor electrostatic effect, as found in this work, should modulate the base pairing strength with the complimentary strand, which is also likely to influence the fidelity of any interaction or recognition involving base pairing like replication, transcription, translation, and triplex formation, depending upon the sequence context of DNA or RNA.

Acknowledgment. Generous financial support from the Swedish Natural Science Research Council (Vetenskapsrådet), the Swedish Foundation for Strategic Research (Stiftelsen för Strategisk Forskning), and Philip Morris USA Inc. is gratefully acknowledged.

Supporting Information Available: (Figure S1) Plot of pH-dependent (6.6−12.5) ^{1}H chemical shifts (δH) for different aromatic protons of the trimeric (**2a**−**5a**) and heptameric (**8a**−**11a**) ssDNA as well as trimeric (**2b**−**5b**) and heptameric (**8b**−**11b**) ssRNA. (Figure S2) The Hill plots for ssDNA **2a**−**5a** and **8a**−**11a** as well as ssRNA **2b**−**5b** and **8b**−**11b**. (Figure S3) The difference of chemical shifts [$\Delta\delta_{(N-D)}$, in ppm] of all identical aromatic marker protons of the neighboring nucleobases in the isosequential ssDNA (**2a**−**5a** and **8a**−**11a**) and ssRNA (**2b**−**5b** and **8b**−**11b**) at the neutral (N) and the deprotonated (D) states. (Figure S4) The oligomerization shifts at the neutral state and at the deprotonated state. Oligomerization shifts are calculated for the individual nucleotide residues in an oligo-ssDNA trimers **2a**−**5a** and heptamers **8a**−**11a** as well as in oligo-ssRNA trimers **2b**−**5b** and heptamers **8b**−**11b** with respect the appropriate monomeric reference compounds. (Figure S5) The pairwise subtraction of the chemical shifts [$\delta_{(deoxy-ribo)}$, in ppm] of all identical aromatic marker protons of the nucleobases in the isosequential ssDNA (**2a**−**5a**) and (**8a**−**11a**) and isosequential ssRNA (**2b**−**5b**) and (**8b**−**11b**) at the neutral (N) and the deprotonated (D) states. (Figure S6) The stack plots of the pH-dependent ^{1}H NMR chemical shifts (in D_2O) of the aromatic protons for isosequential ssDNA (**2a**−**5a** and **8a**−**11a**) and ssRNA (**2b**−**5b** and **8b**−**11b**) at 298 K. (Figure S7) The NMR assignments are shown for ssDNA (**2a**−**5a** and **8a**−**11a**) and ssRNA (**2b**−**5b** and **8b**−**11b**) using ^{1}H NOESY (both at 298 K and at 283 or 278 K), DQF-COSY and TOCSY (at 298 K) as well as ^{31}P−^{1}H correlation spectroscopy at the N-state. The connectivity and proton assignments are shown for each spectrum. (Table S1) The oligomerization shift estimated from ^{1}H chemical shift at the neutral (N) state at 298 K for aromatic protons for ssDNA **2a**−**5a** and **8a**−**11a** as well as ssRNA **2b**−**5b** and **8b**−**11b** using appropriate monomeric reference compounds. (Table S2) The oligomerization shift estimated from ^{1}H chemical shift at the deprotonated (D) state at 298 K for aromatic protons of for ssDNA **2a**−**5a** and **8a**−**11a** as well as ssRNA **2b**−**5b** and **8b**−**11b** using appropriate monomeric reference compounds. (Table S3) ^{1}H chemical shifts [δ_H, in ppm] at the neutral (N) and the deprotonated (D) states at 298 K for monomeric compounds **1a**−**1q**, ssDNA trimers **2a**−**5a**, ssDNA heptamers **8a**−**11a** as well as ssRNA trimers **2b**−**5b**, ssRNA heptamers **8b**−**11b**. (Table S4) ^{1}H chemical shift differences [$\Delta\delta_{doexy-ribo}$, in ppm] at the neutral (N) and the deprotonated (D) states at 298 K between the monomeric 2′-deoxy analogues **1d**, **1j**, **1p, 1f, 1h, 1l, 1n** and monomeric ribo analogues **1e**, **1k, 1q, 1g, 1i, 1m**, **1o**, respectively, as well as between the ssDNAs **2a**−**5a**, **8a**−**11a** and the corresponding ssRNAs **2b**−**5b**, **8b**−**11b**, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

JA048484C