

Rational Design of Hetero-ring-Expanded Guanine Analogs with Enhanced Properties for Modified DNA Building Blocks

Jinmei Zhang,[†] Robert I. Cukier,[‡] and Yuxiang Bu^{*,†,‡}

Institute of Theoretical Chemistry, Shandong University, Jinan, 250100, P.R. China, and Department of Chemistry, Michigan State University, East Lansing, Michigan 48824

Received: May 10, 2007

The properties and modes of recognition of physiological DNAs associated with the four natural nucleobases might be extended, in principle, by the design of non-natural nucleobase derivatives. The goal is an expansion of the genetic alphabet, with the possible outcome of producing new DNAs with improved physical or biological properties. In this work, a new series of hetero-ring-expanded guanine analogs are proposed, and their relevant structural characteristics and electronic properties are determined by density functional theory. The stabilities of the decamer DNA duplexes $(\mathbf{dn}\cdot\mathbf{dC})_{10}$ (where \mathbf{n} represents the corresponding expanded guanine analog designed here) are also examined, using molecular dynamics. The simulations show that the designed motifs can form stable DNA-like structures. We determined the pairing energies for the Watson–Crick (WC) hydrogen-bonded dimers between the expanded **G**-analogs and the natural **C**, and found that the pairing energies are close to those of the natural **GC** pair. The calculated adiabatic ionization potentials (IPs) of the size-expanded guanine analogs and their base pairs, and the corresponding vertical ionization potentials, show that some are distinctly smaller than the corresponding natural versions. The HOMO–LUMO energy gaps for most of the size-expanded guanine analogs and their WC base pairs are considerably lower than those of the corresponding natural base and base pairs. Thus, the expanded **G** bases may be considered as DNA genetic motifs, and they may serve as building blocks for potential biological applications and the development of molecular electronic devices.

1. Introduction

Creating unnatural nucleobases has gained increasing attention, primarily motivated by their potential applications in biotechnology and medicinal chemistry, as well as in material science. To date, most efforts have been devoted to the development of modified bases, including the use of nonstandard hydrogen-bond donor/acceptor patterns,¹ van der Waals and hydrophobic interactions,² metal coordination,³ and other alterations.⁴ Clearly, modifying the chemical and physical properties of DNA without affecting its complementary base pairing may play an important role in future applications. For example, increasing the electrical conductivity of DNA by chemical modification of the base pairs may allow for the self-assembly of complex molecular-scale electrical devices.⁵ On the other hand, although the specific Watson–Crick (WC) hydrogen bond in natural DNA plays an important role in the storage and transfer of bio-information, it does not exhaust the possibilities of establishing additional hydrogen bonds. In fact, the specific hydrogen-bond donors and acceptors in the grooves facilitate the recognition of the DNA sequence; the binding to proteins, or small-molecule ligands, can induce changes in both the DNA and protein or small-molecule conformation.⁶

Recent progress indicates that the recognition pattern of physiological DNAs associated with the four natural nucleobases could, in principle, be enlarged by using unnatural nucleobase derivatives, which might be used to expand the genetic alphabet

and to produce new DNAs with improved physical or biological properties.^{2c,7} One interesting development is the synthesis of a new set of “stretched-out” bases (termed **x**-bases, synthesized by Kool’s group⁸), in which the fusion of a benzo ring increases their size by 2.4 Å (see **xG** in Figure 1). These **x**-bases were found to be able to form stable DNA-like structures and to have smaller HOMO–LUMO gaps than the natural one, suggesting that they could enlarge the genetic alphabet and also have potential application in biotechnologies.^{7a,9}

Inspired by these studies, and with the aim of retaining the desired base-pairing properties, we have designed a series of new size-expanded **G**-analogs (shown in Figure 1) that are hetero-(5/6)-membered ring-expanded. There are four groups: **S**₆₁, **S**₆₂, **S**₅₁, and **S**₅₂, simply classified by the types of the spacer rings. In outline, for purposes of classification, the spacer ring in group **S**₆₁ has a π_6^6 aromatic conjugated bond system, and in group **S**₆₂ it has a nonaromatic π_6^8 conjugated bond system. The difference between the two groups, **S**₅₁ and **S**₅₂, lies in the different orientation of the spacer ring. In both groups, the hetero-ring spacers are aromatic and the homo-ring spacers are not π -conjugated and thus nonaromatic. In fact, the ring-expansion strategy has been used to expand nucleobases with varied spacer rings,¹⁰ already including pyridine, pyrazolo, and other heterocyclic rings. However, such modified bases have been used mainly in medicinal chemistry, rational drug design, and investigations of enzyme–substrate interactions. In addition, it is noteworthy that the hetero-ring expansion strategy was also used to modify the pyrimidine to form new base-pairing motifs (tricyclic nucleosides) consisting of four hydrogen bonds. But, its aim was only to enlarge the WC face from three hydrogen

* Author to whom correspondence should be addressed. E-mail: byx@sdu.edu.cn.

[†] Shandong University.

[‡] Michigan State University.

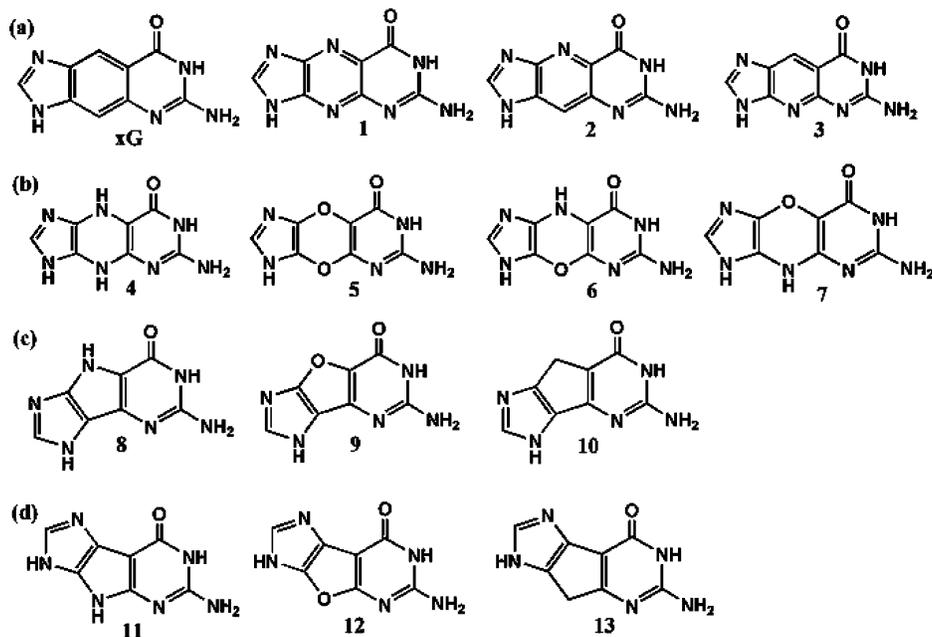


Figure 1. Structures of the hetero-ring-expanded guanine analogs designed here and **xG**. According to the spacer ring, they are classified into four groups: (a) **S**₆₁ (1–3) in which the spacer ring has normal π_6 -conjugated π -bonding; (b) **S**₆₂ (4–7) in which the spacer ring has π_8 -conjugated bonding; (c) **S**₅₁ (8–10); (d) **S**₅₂ (11–13). The spacer rings in both **10** and **13** contain the $>CH_2$ unit in which a C center adopts sp^3 hybridization, being different from the other C members (sp^2 hybridization) in the rings. Thus, they may also be viewed as hetero-rings. In the latter two groups, the hetero-ring spacers are aromatic, but the homo-ring spacers are not π -conjugated and thus nonaromatic.

bonds to four.^{1c} In contrast, the designed **G**-analogs here are expected to have both structural characteristics and biological functionality similar to those of their natural counterpart.

This ring-expansion scheme presents a novel approach to enriching the gene alphabet and to extending DNA-based applications. For promising additional gene motifs, the prospective reference object for the designs should mimic their natural analogues and function in the context of the natural genetic system. Although the stacking interactions between the aromatic DNA bases play a dominant role in duplex stability, and they appear to be driven largely by van der Waals forces and, for hydrophobic molecules, by solvophobic effects,⁸ the necessary hydrophilic sites in the designed bases should be included to maintain normal functionality as for the natural base. Furthermore, maintenance of the necessary hydrophilic sites in the designed DNA motifs without decrease in their overall hydrophobicity may favor their property modulation.

The designed DNA motifs must have comparable or enhanced electronic properties relative to the natural motif. They must also be stable in solution. Thus, we investigate the expanded bases and their base-pairing properties with a combination of quantum chemical density functional theory (DFT) calculations for the electronic properties and molecular dynamics (MD) simulations for the solution stability of the duplexes. We demonstrate here that these hetero-ring-expanded **G**-analogs can form stable duplexes and have electronic properties that are similar to or improved on those of natural **G** and its base pair.

2. Methodology

DFT Calculations. Calculations were performed using the hybrid three-parameter B3LYP density functional approach, as implemented in the GAUSSIAN 03 set of programs.¹¹ The geometries of the size-expanded guanine analogs were optimized at the 6-311+G* level. Frequency calculations were performed as well to ensure that the species represent true minima on the potential energy surfaces. At the same level, the structures of

the Watson–Crick (WC) hydrogen-bonded dimers between the expanded bases and their natural counterpart were obtained, and all the pairing energies were corrected for basis set superposition error (BSSE) using the counterpoise correction. Owing to the large number of mismatches examined, we obtained the initial optimized structures of the mismatches at the 6-31+G* level. Then, the relative energies were estimated from single-point calculations at the 6-311+G* level, and these mismatching energies were also corrected for the effect of the basis set extension. The ionization potentials, and HOMO and LUMO orbital energies, were all computed at the 6-311+G* level.

We also explored the electronic effect of the sugar-phosphate moiety on the newly designed guanine analogs and found a very minor effect of the ribose moiety on the structural parameters and energy quantities such as the ionization potentials and HOMO–LUMO energy gaps. These results are in good agreement with observations reported recently by Guerra et al.¹²

Molecular Dynamics Simulations. The natural duplex (d**G**·d**C**)₁₀ was built in canonical B-form using the NUCGEN module in AMBER 8.0,¹³ and starting structures of the modified duplexes were manipulated to change the base pairs. Atom-centered partial charges for the size-expanded guanine analogs were derived by using the AMBER antechamber program (RESP methodology¹⁴). Each structure was immersed in a periodic box of about 3000 TIP3P¹⁵ water molecules, and 18 sodium ions were added to neutralize the charge. After energy minimization to relax strains and heating to 300 K over 20 ps using Langevin dynamics,¹⁶ 4 ns of unrestrained constant temperature and pressure simulations were performed at 300 K and 1 atm with the SANDER module, using the AMBER-99 force field.^{17,18} The Particle Mesh Ewald algorithm¹⁹ was used with a 10 Å cutoff. All bond lengths were constrained using SHAKE,²⁰ allowing a 2 fs integration time step. Since (d**10**·d**C**)₁₀ showed a larger root-mean-square deviation (RMSD) relative to the other simulated duplexes after 2 ns of unrestrained simulation, we extended the time to 6 ns. While the average

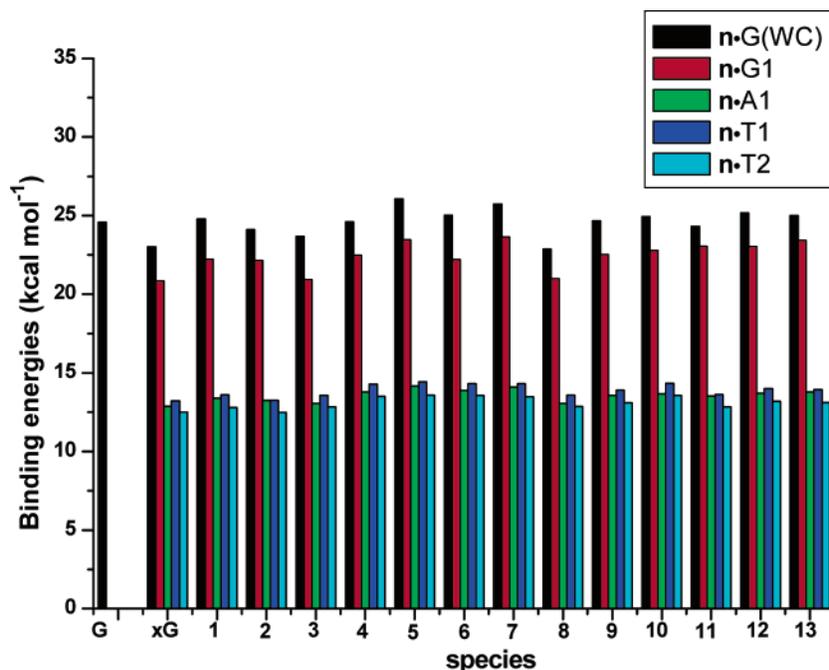


Figure 2. Binding energies for all the modified base pairs in structure $n\text{-C(WC)}$, $n\text{-G1}$, $n\text{-A1}$, $n\text{-T1}$, and $n\text{-T2}$.

structure of $(d10\text{-dC})_{10}$ was obtained from the last 2 ns of its trajectory, the other average structures were generated from the last 3ns of each trajectory. These structures were then analyzed using the program 3DNA.²¹

3. Results and Discussion

The calculations reveal that, aside from the spacer rings, most of the size-expanded G -analogs (except **4**, **6**, and **7**) show slight nonplanarity, similar to the natural one, caused by the inherent pyramidization of the amino groups and the nonplanarity of the aromatic rings.²² But, the nonplanarity of **4**, **6**, and **7** is somewhat different from the others. In the fused rings of **4**, **6**, and **7**, the nitrogen atoms also acquire a pyramidal configuration, similar to the amino group, leading to the large nonplanarity of the fused rings, and the enhancement of the nonplanarity of the pyrimidine and imidazole rings in the base. The size expansions of these size-altered bases are given in Table S1 of Supporting Information as indexed by two distances termed C4C and C5C (C4 and C5 denote the carbons in the pyrimidine ring, and C represents the corresponding carbon in the imidazole ring).

Since these tricyclic analogs still retain the main features of the parent compound, especially the WC faces, they are expected to retain the necessary hydrogen bonds to form the canonical WC pairs, while having enlarged base stacking interactions. Although the major-groove acceptor site, N7, shifts outward by some distance, the minor-groove acceptor atom N3 is preserved, such that the minor-groove hydrogen bonds between DNA and an enzyme may be retained.²³ Additionally, the heteroatomic sites in the spacer rings may increase to provide new active sites. From these features, we believe that these designed bases could be inserted efficiently and extended successfully during DNA synthesis and form stable duplex pairings with their natural counterpart C , as does Kool's-designed xG .^{8d} To verify our hypothesis and to compare these hetero-ring-expanded bases with xG and G , we determined the pairing energies for the WC hydrogen-bonded dimers between the expanded G -analogs (n) and the natural C (termed as $n\text{-C}$) and found that the pairing energies are close to those of the natural GC pair ($24.58 \text{ kcal mol}^{-1}$), with minor differences of

only $1.7\text{--}2.5 \text{ kcal mol}^{-1}$ (Figure 2 and Table S2 of Supporting Information). It is quite a surprise that the pairing stability of almost all designed G -analogs exceeds that of xG , with the sole exception of **8**, and even exceeds that of the natural G (except for **2**, **3**, **8**, and **11**). This observation indicates that insertion of a hetero-ring spacer between the pyrimidine and imidazole yields a rather small effect on the binding property of the bases (even smaller than the benzo insertion), and these G -analogs can still pair with the natural complementary partner C to generate the size-altered stable base pairs.

To examine the relative pairing preferences of these modified G -analogs, the pairing energies for the mismatches formed by these size-expanded bases with the natural bases were determined. Since the most favorable mismatches in $G\text{-}G$, $G\text{-}A$, and $G\text{-}T$ structures for G ²⁴ are $G\text{-}G1$, $G\text{-}A1$, and $G\text{-}T1$, which only contain hydrogen bonds at their WC faces, and the pairing energy of $G\text{-}T1$ is close to $G\text{-}T2$ (see Figures S1 and S2 of Supporting Information), only the just-noted four types of mismatches in the $G\text{-}G1$, $G\text{-}A1$, and $G\text{-}T$ pairing patterns (termed hereafter as $G\text{-}G\text{-}$, $G\text{-}A\text{-}$, and $G\text{-}T\text{-}$ type mismatches, respectively) were examined for the size-expanded bases (n). Indeed, we found that the $n\text{-C}$ WC pair is still the most stable structure for the base n , followed by the $n\text{-G}$, $n\text{-A}$, and $n\text{-T}$ mismatches (Figure 2). This is just the case as for the natural G . Thus, it can be concluded that size-expansion of the guanine by the spacer ring has a minimal effect on its hydrogen-bonding ability. To further confirm our conclusion, taking **1** and **4** as examples, we examined all of their possible hetero and homo base mismatches (17 for **1** and 26 for **4**) and found that all these mismatches have smaller pairing energies than those of their corresponding WC $n\text{-C}$ pairs (Tables S3 and S4 of Supporting Information). All these results show that these size-altered G -analogs exhibit relatively high pairing selectivity for C over the other natural bases, and the magnitude of the effect is comparable to that for the natural G . This similarity may be surmised to be partly attributable to the approximate conservation of the natural G structure for the expanded modifications. That is, they still retain the main structural features of the natural G . In summary, we

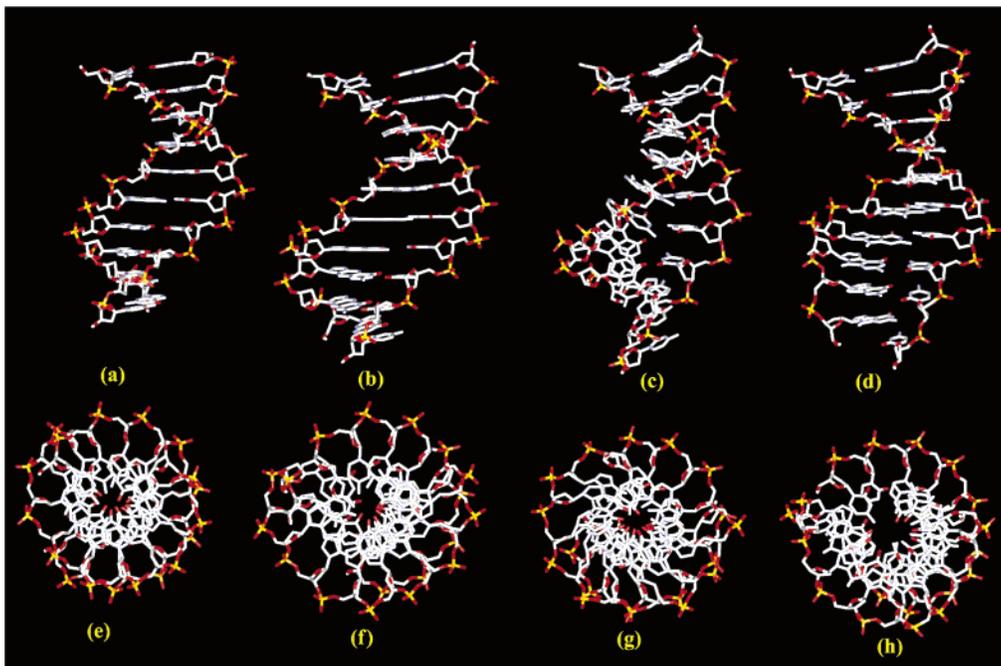


Figure 3. Side and top views of the average structure of the 10-mer duplexes: (dG·dC)₁₀ (a and e), (d1·dC)₁₀ (b and f), (d8·dC)₁₀ (c and g), and (d11·dC)₁₀ (d and h).

conclude that these size-expanded **G**-analogs could selectively pair with **C** in the WC structure, forming size-altered duplexes.

To investigate the dynamical properties of these newly designed bases, when incorporated to form duplexes, and to generate detailed structural models of the size-expanded duplexes, MD simulations were performed for these duplexes (dn·dC)₁₀ (**n** denotes the given bases in Figure 1). All the duplexes in 4 ns unrestrained MD simulations are stable, as monitored by their RMSD relative to the starting structures, and no major helical unfolding or transitions to single-strand structures were found during the simulations (see Figure S3 of Supporting Information for RMSD plots). Their trajectory averaged geometries are depicted in Figure 3 and Figure S4 of Supporting Information. It can be seen that they still possess double-helical structures with helicity and backbone conformation apparently similar to those of natural DNA and, with the exceptions of (d8·dC)₁₀ and (d9·dC)₁₀, the base pairs in the duplexes are practically planar. The base pairs in the duplexes (d8·dC)₁₀ and (d9·dC)₁₀ are slightly nonplanar, but the hydrogen-bonding schemes are still well preserved in all cases.

The main helical characteristics of these duplexes were determined from the average structures from the simulations (Table S5, Supporting Information). In natural B DNA, the helix pitch is 34 Å and the twist is 36°, which results from a rise per base pair of 3.4 Å and 10 base pairs per turn. However, we found that the natural duplex (dG·dC)₁₀ has a smaller helix twist (31.4 Å) and rise (3.03 Å) resulting in a 34.8 Å helix pitch, and its helical parameters also show some A-type characteristics, consistent with previous unrestrained simulations.^{25a,b} This may be due to the fact that **G**-tracts, or runs of four or more **GC** base pairs, have special structural properties.^{25c,d} They may favor the A DNA helix conformation and, depending on the local water activity, can induce a B-to-A DNA transition.²⁶ The duplexes constructed by the designed **G**-analogs have a smaller or similar helix twist (27°–31°) to that of the natural one, leading to a greater or similar number of base pairs per turn (11–13), and the inter-base-pair distances are 2.7–3.3 Å. In fact, most of these duplexes have reduced inter-base-pair distances compared to that of the natural duplex (3.03 Å). For

most of these duplexes (except for (d12·dC)₁₀ and (d13·dC)₁₀), the larger number of base pairs per turn and shorter rise per base pair in helices generally result in larger helix pitches. In fact, we found that while the duplexes constructed by members in **S**₅₁ have a diameter similar to that of the natural one (measured by the C1'–C1' distance) and belong to the B family, the other duplexes have larger diameters due to the size expansion of the bases and have conformations intermediate between the A and B DNA families. In the helices constructed by members in **S**₆₁, **S**₆₂, and **S**₅₂, while their glycosidic bond orientations still fall within the B-type, the slides between base pairs and displacements of the phosphates are intermediate in character, and it is especially obvious for (d4·dC)₁₀ and duplexes formed by members in **S**₅₂. In addition, we found that, aside from **S**₅₂, the major and minor grooves in the duplexes are generally wider with respect to the natural one. In contrast, the major grooves are reduced in helices formed by **S**₅₂, but the minor grooves in the helices are still wider. Since most proteins bind DNA via the major groove, the narrower major groove may suggest that they are less biologically active. However, due to the short length and special sequence of these duplexes, the above data are inconclusive.

It is noteworthy that, due to the larger aromatic surface of the bases and shorter inter-base-pair distance, most of these newly designed bases may have enhanced stacking interactions with adjacent bases, which significantly contributes to the stability of the duplexes. The MD simulation results along with the slightly increased binding energies suggest that these size-expanded **G**-analogs can still form stable helices. These findings agree with previous experimental reports for **xG**-containing DNA duplexes^{8d} and are a prediction for the duplexes containing our designed bases, for which no experimental data are available.

Though the physical mechanism of DNA conductivity is still under debate, it has been suggested that the structure of DNA with a π -electron system of four bases stacked on each other can provide a mechanism for electron transfer along DNA that involves hole-hopping from one **G** to the next.²⁷ As known, the HOMO–LUMO gaps and the ionization potentials (IPs) of the bases are two important quantities that reflect the conductiv-

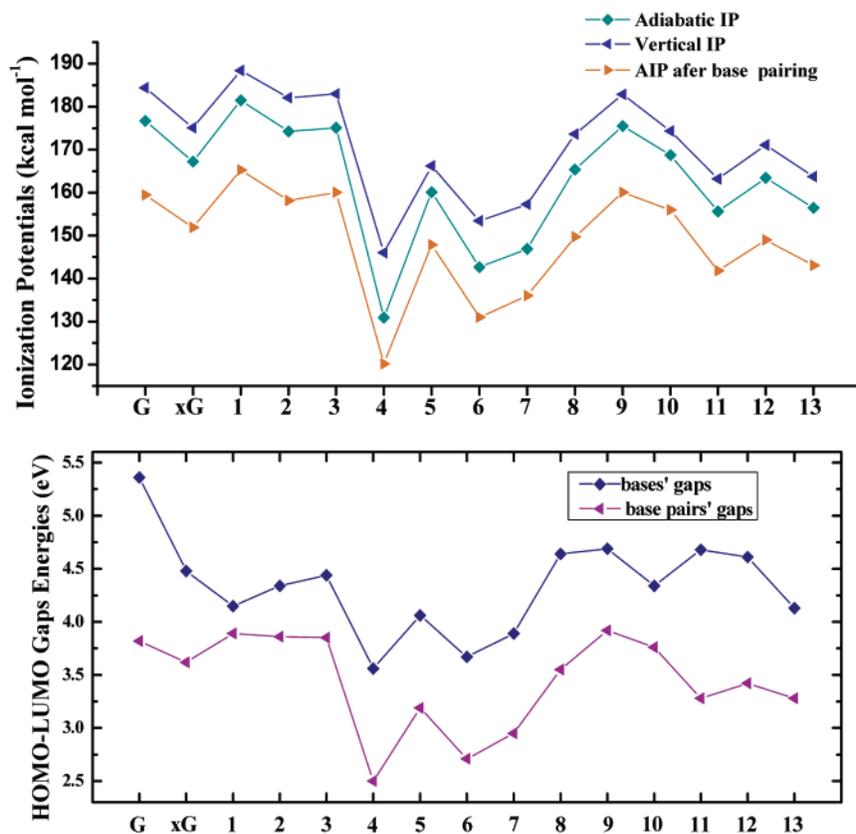


Figure 4. (Upper) Adiabatic ionization potentials (IPs) of the size-expanded guanine analogs and their base pairs together with vertical ionization potentials of the bases (including **G** and **xG**). (Lower) The HOMO–LUMO energy gaps for the size-expanded guanine analogs and their Watson–Crick (WC) base pairs.

ity or the electron transport in DNA, as they are closely related to the DNA band gap and hole-accepting ability. Thus to verify if these **G**-analogs could be used in future molecular electronic devices or as added DNA motifs, we examined their HOMO–LUMO gaps and IPs, in a comparison with those of the natural one. This should provide basic information for the conductance of the duplexes constructed from these bases.

The theoretical adiabatic IPs of the size-expanded **G**-analogs and their base pairs are presented in Figure 4 (upper). As occurs for natural **G**, the base pairings lower the adiabatic IPs of the size-expanded bases, and generally they preserve a similar order for all the bases. Moreover, in general, the members in **S**₆₁ have IPs similar to that of **G**, and are also very close to **xG**, followed by **S**₅₁. However, IPs of the bases in **S**₆₂ and **S**₅₂ are reduced greatly, by more than 10 kcal mol⁻¹, and the lowest one **4** (120.16 kcal mol⁻¹) is reduced by 39.29 kcal mol⁻¹. Interestingly, we found that each member in **S**₅₂ has an IP of ~10 kcal mol⁻¹ lower than the corresponding member (bearing the same spacer ring but different orientation) in **S**₅₁. For example, the adiabatic IP of **11** after pairing is 7.85 kcal mol⁻¹ lower than that of **8**. Moreover, the molecules with the O-containing spacers have relatively higher IPs than those of the other bases in the group, indicating that the IPs of the size-expanded **G**-analogs are related to the size and hetero-atom types in the spacers. As previously mentioned, the aromatic rings for **4**, **6**, and **7** show distinct nonplanarity; thus, we calculated the vertical IPs of these bases and also plot them in Figure 4 (upper). The energy difference between vertical and adiabatic IPs arise from the relaxation energies of the nuclear framework. Indeed, the relaxation energies of **4**, **6**, and **7** (9.76–15.07 kcal mol⁻¹) are larger than those of other molecules (5.58–8.20 kcal mol⁻¹). But, the vertical and adiabatic IPs preserve exactly the same order. These results suggest that the relaxation of the nuclear

framework is not a crucial factor for the IP changes of these molecules. Since size-expansion generally decreases the IPs, and even the largest IP (possessed by **1**) among these expanded bases or base pairs is still lower than those of **A**, **C**, and **T**, we can conclude that the addition of a spacer ring to **G** would not change the fact that the purine moiety (**G** and **G**-analogs) is the most easily oxidized base among DNA nucleobases. Thus the electron-loss center created in a DNA duplex will still localize on **G** and its derivatives. In fact, in comparison with **G**, most of the size-expanded **G**-analogs should be much easier to oxidize due to their lower IPs. This is important, because a previous study²⁸ showed that for the design of hole-transport-based DNA nanowires a low oxidation potential is an important factor.

The HOMO–LUMO gaps of the size-expanded **G**-analogs and their base pairs are listed in Figure 4 (lower). The addition of a spacer to **G** has considerably lowered its HOMO–LUMO energy gap, and the gap is further reduced after pairing. More interesting is that the extent of lowering for these size-expanded bases is not as great as for the natural **G** (lowered by 1.54 eV). In fact, the extent for each size-expanded base is also different, resulting in some of them (**S**₆₁ and **S**₅₁) with HOMO–LUMO gaps close to those of **G**•**C** and also **xG**•**C**. In general, each group has similar energy gaps, especially for the base pairs. In fact, the HOMO–LUMO gaps of the base pairs (**n**•**C**) formed by **1**, **2**, **3**, and **9** (3.89, 3.86, 3.85, and 3.92 eV, respectively) are a little larger than that of **G**•**C** (3.82 eV), but the differences are not large. In contrast, the remaining two groups (**S**₆₂ and **S**₅₂) have much smaller gaps than the natural one, especially for **4**, **6**, and **7** (2.50, 2.71, and 2.95 eV, respectively), and **4** has the smallest gap. This is just the case as for the IPs. Therefore, in summary, the modified duplexes constructed from these bases may have smaller band gaps than the natural duplex.

As a lower band gap facilitates charge migration, these modified duplexes may be considered to have better conductivity, with potential applications in nanotechnology.

The above analyses reveal that among the designed **G**-analogs, most bases have both low IPs and small HOMO–LUMO gaps (even smaller than **xG**), which would lead to a larger hole-injection ability and charge transfer in helices. In fact, all members in **S**₆₂ and **S**₅₂ possess these two characteristics, especially for **4**, **6**, and **7**. Due to their larger aromatic surfaces, these size-expanded **G**-analogs may exhibit fluorescence²⁹ and may have stronger π – π interactions, suggesting that helices constructed from other size-expanded bases may also have smaller gaps relative to the natural one, leading to improved conductivity, and opening up new opportunities for molecular electronic engineering.

4. Conclusions

In summary, we have found that all thirteen of these designed size-expanded **G**-analogs can form DNA-like structures, suggesting that they may form alternative genetic systems. Moreover, while some of these bases have electronic properties similar to those of natural **G**, others have improved electronic properties, leading to possible applications as molecular electronic devices. This design, inspired by Kool's benzo-expansion scheme,⁸ represents a significant extension that is directed toward the goal of the rational design of biologically functional nucleotides with improved properties. For example, the increased major/minor-groove active sites for some of these modified duplexes³⁰ may favor the formation of **G**-quadruplex-like structures and the modulation of DNA function by proteins. Additionally, due to the versatility of nucleosides, their size modification could result in significant biological effects and thus increase their other potential biochemical applications.

Another important aspect reported here is that this hetero-ring expansion scheme, at least the six-membered hetero-ring expansion scheme, may also present a challenge regarding the ring-expansion strategy. Clearly, the six-membered ring-expansion scheme (hetero-ring and benzene) may generate similar hydrogen-bonding topology, and similar expanded duplexes, with the same complementary base. Since the genetic alphabet might be enlarged by these additional base letters, what does need clarification is how to combine them as done for **xG-xC** and **xA-xT** as well as for **xG-C**, **G-xC**, **xA-T**, and **A-xT** versus **G-C** and **A-T**, subject to their different duplex structures. In fact, the increase in hydrophilic sites at the hetero-ring spacer in the major/minor grooves generates an approach to address this and other questions, because of their different binding abilities with proteins, and with small molecules that bind to DNA, compared with their benzo-expanded counterparts.

This theoretical design and our preliminary property analyses may provide helpful information to guide new experimental studies of duplexes containing the hetero-ring-expanded **G** bases. Theoretical studies are underway to explore the details of the conductance mechanism, including the multimer energy gaps, HOMO/LUMO distributions, and other properties.

Acknowledgment. This work was supported by NSFC (20633060, 20573063 to Y.B.), NIH (Grant No. GM62790 to R.I.C.), NCET (to Y.B.), and Virt. Lab for Comput Chem & SCC of CNIC-CAS, MCBILIN at MSU, and HPCC at SDU.

Supporting Information Available: Geometries of the designed bases **1** and **17** and their corresponding various base pairs, MD simulation RMSDs, average structures of the decamer

duplexes, size expansions and relevant energies of the designed base pairs, and the comparison of helix parameters of the modified duplexes and natural A/B DNA. This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) (a) Piccirilli, J. A.; Krauch, T. S.; Moroney, E. S.; Benner, A. *Nature* **1990**, *343*, 33–47. (b) Moser, M. J.; Prudent, J. R. *Nucleic Acids Res.* **2003**, *31*, 5048–5053. (c) Minakawa, N.; Kojima, N.; Hikishima, S.; Sasaki, T.; Kiyosue, A.; Atsumi, N.; Ueno, Y.; Matsuda, A. *J. Am. Chem. Soc.* **2003**, *125*, 9970–9982. (d) Hikishima, S.; Minakawa, N.; Kuramoto, K.; Fujisawa, Y.; Ogawa, M.; Matsuda, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 596–598.
- (2) (a) McMinn, D. L.; Ogawa, A. K.; Wu, Y. Q.; Liu, J. Q.; Schultz, P. G.; Romesberg, F. E. *J. Am. Chem. Soc.* **1999**, *121*, 11585–11586. (b) Kool, E. T.; Morales, J. C.; Guckian, K. M. *Angew. Chem., Int. Ed.* **2000**, *39*, 990–1009. (c) Henry, A. A.; Romesberg, F. E. *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
- (3) (a) Rakitin, A.; Aich, P.; Papadopoulos, C.; Kobzar, Y.; Vedenev, A. S.; Lee, J. S.; Xu, J. M. *Phys. Rev. Lett.* **2001**, *86*, 3670–3673. (b) Tanaka, K.; Tengeji, A.; Kato, T.; Toyama, N.; Shionoya, M. *Science* **2003**, *299*, 1212–1213. (c) Tanaka, K.; Shionoya, M. *J. J. Org. Chem.* **1999**, *64*, 5002–5003. (d) Meggers, E.; Holland, P. L.; Tolman, W. B.; Romesberg, F. E.; Schultz, P. G. *J. Am. Chem. Soc.* **2000**, *122*, 10714–10715. (e) Weizman, H.; Tor, Y. *J. Am. Chem. Soc.* **2001**, *123*, 3375–3376.
- (4) Other strategies include (i) exploring covalently linked base pairs; (ii) exploring compounds without hydrogen-bonding functionalities, which might nevertheless act as substrates for DNA polymerases and self-pair in DNA; and the mixed strategies that exploit several of these key features.
- (5) Okamoto, A.; Tanaka, K.; Saito, I. *J. Am. Chem. Soc.* **2003**, *125*, 5066–5071.
- (6) (a) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984. (b) Geirstanger, B. H.; Wemmer, D. E. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, *24*, 463–493.
- (7) (a) Leconte, A. M.; Romesberg, F. E. *Nature* **2006**, *444*, 553. (b) Kool, E. T. *Acc. Chem. Res.* **2002**, *35*, 936–943. (c) Benner, S. A. *Acc. Chem. Res.* **2004**, *37*, 784–797. (d) Sivakova, S.; Rowan, S. J. *Chem. Soc. Rev.* **2005**, *34*, 9–21.
- (8) (a) Liu, H.; Gao, J.; Lynch, S. R.; Saito, Y. D.; Maynard, L.; Kool, E. T. *Science* **2003**, *302*, 868–871. (b) Gao, J.; Liu, H.; Kool, E. T. *J. Am. Chem. Soc.* **2004**, *126*, 11826–11831. (c) Liu, H.; Gao, J.; Kool, E. T. *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402. (d) Liu, H.; Gao, J.; Kool, E. T. *J. Org. Chem.* **2005**, *70*, 639–647.
- (9) (a) Fuentes-Cabrera, M.; Sumpter, B. G.; Wells, J. C. J. *Phys. Chem. B* **2005**, *109*, 21135–21139. (b) Fuentes-Cabrera, M.; Sumpter, B. G.; Lipkowski, P.; Wells, J. C. J. *Phys. Chem. B* **2006**, *110*, 6379–6384.
- (10) (a) Leonard, N. J.; Morrice, A. G.; Sprecker, M. A. *J. Org. Chem.* **1975**, *40*, 356–366. (b) Leonard, N. J.; Sprecker, M. A.; Morrice, A. G. *J. Am. Chem. Soc.* **1976**, *98*, 3987–3994. (c) Hirota, K.; Maruhashi, K.; Asao, T.; Senda, S. *Chem. Pharm. Bull.* **1982**, *30*, 3377. (d) Seley, K. L.; Januszczak, P.; Hagos, A.; Zhang, L.; Dransfield, D. T. *J. Med. Chem.* **2000**, *43*, 4877–4883.
- (11) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, M.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*; Gaussian, Inc.: Pittsburgh, PA, 2003.
- (12) (a) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *J. Am. Chem. Soc.* **2000**, *122*, 4117–4128. (b) Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. *Chem.–Eur. J.* **1999**, *5*, 3581–3594. (c) Guerra, C. F.; Bickelhaupt, F. M. *Angew. Chem., Int. Ed.* **1999**, *38*, 2942–2945.
- (13) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8*; University of California: San Francisco, 2004.
- (14) Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357–1377.

- (15) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–936.
- (16) Kottalam, J.; Case, D. A. *Biopolymers* **1990**, *29*, 1409–1421.
- (17) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
- (18) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
- (19) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (20) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–336.
- (21) Lu, X.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
- (22) (a) Sponer, J.; Hobza, P. *J. Phys. Chem.* **1994**, *98*, 3161–3164. (b) Hobza, P.; Sponer, J. *Chem. Rev.* **1999**, *99*, 3247–3276.
- (23) Morales, J. C.; Kool, E. T. *J. Am. Chem. Soc.* **1999**, *121*, 2323–2324.
- (24) (a) Hobza, P.; Sandorfy, C. *J. Am. Chem. Soc.* **1987**, *109*, 1302–1307. (b) Sponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965–1974.
- (25) (a) Trantírek, L.; Štefl, R.; Vorlíčková, M.; Koča, J.; Sklenář, V.; Kypr, J. *J. Mol. Biol.* **2000**, *297*, 907–922. (b) Cheatham, T. E., III; Srinivasan, J.; Case, D. A.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1998**, *16*, 265–280. (c) Vargason, J. M.; Henderson, K.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7265–7270. (d) Ng, H.-L.; Kopka, M. L.; Dickerson, R. E. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 2035–2039.
- (26) (a) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. *Biopolymers* **1973**, *12*, 89–110. (b) Ivanov, V. I.; Minchenkova, L. E.; Minyat, E. E.; Frank-Kamenetskii, M. D.; Schyolkina, A. K. *J. Mol. Biol.* **1974**, *87*, 817–833. (c) Minchenkova, L. E.; Schyolkina, A. K.; Chernov, B. K.; Ivanov, V. I. *J. Biomol. Struct. Dyn.* **1986**, *4*, 463–476.
- (27) (a) Ratner, M. *Nature* **1999**, *397*, 480–481. (b) Jortner, J.; Bixon, J.; Langenbacher, M.; Michel-Beryerle, T. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12759–12765.
- (28) In general, three important factors that are prerequisites for designing real DNA nanowires may be outlined as (i) a highly ordered π -stacking array, (ii) low oxidation potential, and (iii) suppressed oxidative degradation. See ref 5 and Baik, M.-H.; Silverman, J. S.; Yang, I. V.; Ropp, P. A.; Szalai, V. A.; Yang, W. T.; Thorp, H. H. *J. Phys. Chem. B* **2001**, *105*, 6437–6444.
- (29) Due to the larger aromatic surfaces, some of these size-expanded G-analogs probably yield fluorescence as x-bases;⁸ however, the fluorescence mechanism may be different. The exploration of the detailed fluorescence mechanism is still underway.
- (30) For helices formed by S₅₂, their major grooves are reduced, so quadruplex formation or protein interactions may, in fact, be less favored. However, their minor grooves are enlarged, favoring the interaction of helices with proteins. This is a qualitative prediction, and the details should be explored by combining the actual amino acids, peptides, or proteins. We plan to examine the details of all these bindings, including quadruplex formation, protein interactions, and others.