Stabilization of the Noncomplementary Guanine−Adenine Base Pairs by Zn(II) Ions. An ab Initio SCF-MI Study

Federico Moroni,[†] Antonino Famulari,[†] Mario Raimondi,*[,†] and Michal Sabat*[,‡]

Dipartimento di Chimica Fisica ed Elettrochimica and Centro CNR-CSRSRC, Università degli Studi di Milano, via Golgi 19, 20133 Milano, Italy, and Department of Chemistry, University of Virginia, McCormick Road, Charlottesville, Virginia 22904

Received: June 17, 2002; In Final Form: January 9, 2003

The structure and energy of the guanine (G)−adenine (A) base pairs in the G(anti)−A(anti) and G(anti)−A(syn) conformations as well as their complexes with hydrated zinc(II) ions have been investigated by the self-consistent field for molecular interaction (SCF-MI) ab initio method. The formation of covalent bonds between the Zn(II) ions and the N7 sites of the guanine bases results in an increase of the binding energy of the G−A pairs. In addition, the SCF-MI calculations have also been performed for the guanine (G)−guanine (G)−cytosine (C) (G−GC) base triplet and for its complex with a hydrated Zn(II) ion bound to the N7 site of the third-strand guanine. The Zn(II) binding causes a significant stabilization of the G−GC triplet. Possible implications for the conformational changes in alternating d(GA)$_n$ sequences are discussed.

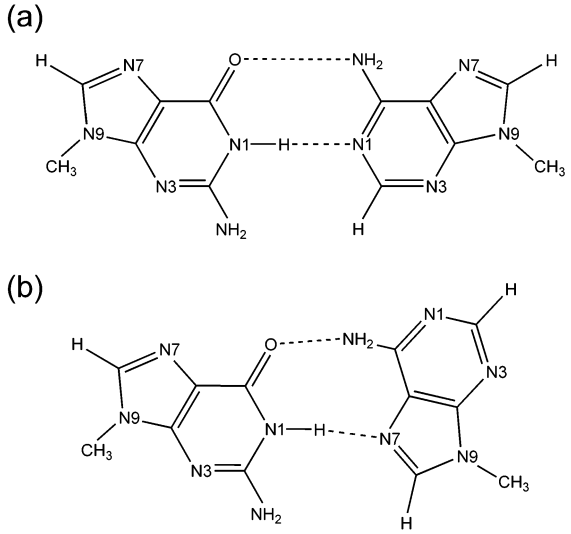
Introduction

The significance of the guanine (G)−cytosine (C) and adenine (A)−thymine (T) Watson−Crick (canonical) base pairs for nucleic acid structure and function has long been established.[1] However, in addition to the Watson−Crick (WC) pairs, there exist a number of noncanonical base pairs that play an important role in several biological processes.[2] For example, non-Watson−Crick base pairs such as AA, GG, GA, CA, and others constitute DNA mismatches that result from erroneous incorporation of bases during DNA replication or recombination.[3,4] Noncanonical base pairs also occur in several unusual nucleic acid systems, including hairpins,[5] triplexes,[6] and quadruplexes.[7]

The efficiency with which a DNA mismatch is repaired depends on the type of base pairs (purine−purine or purine−pyrimidine) involved in this mismatch. Some earlier studies have indicated[8] that the repair of purine−purine G−A mismatches is less efficient than that of all other mismatches in DNA. G−A mismatches have been encountered in several DNA sequences studied by X-ray crystallography[9−12] and NMR spectroscopy.[13−15] The G−A pair can exist in several forms.[16] The most common forms are the G(anti)−A(anti) arrangement, in which the two bases are associated by the N1(G)−H···N1(A) and N6(A)−H···O6(G) hydrogen bonds, and the G(anti)−A(syn) form, stabilized by the N1(G)−H···N7(A) and N6(A)−H···O6(G) hydrogen bonds (Chart 1). In addition, a number of sheared G−A mismatches have been found.[17]

The G−A base pairs also occur in the alternating d(GA)$_n$ DNA sequences, known to adopt a variety of unusual conformations, including parallel[18] and antiparallel[19] self-paired duplexes, as well as purine−purine−pyrimidine triplexes.[20] In a series of reports, Azorin et al.[19−23] indicated that the stability of various structural forms of the d(GA)$_n$ sequences depends on the presence of divalent metal ions such as Zn(II), Cd(II), Co(II), and Mn(II), with Zn(II) being the most effective. In particular, experiments with the d(GA)$_{15}$ and d(GA)$_{20}$ sequences in a variety of environments[24,25] suggested that these sequences form antiparallel homoduplexes, consisting of the alternating G(anti)−A(anti) and G(anti)−A(syn) base pairs and are stabilized by direct Zn(II) binding to the N7 sites of guanines.

CHART 1: (a) G(Anti)−A(Anti) and (b) G(Anti)−A(Syn) Base Pairs



Quantum chemistry has recently emerged as an important tool in the study of the effects exerted by metal ions on nucleic acid bases and nucleotides.[26,27] In particular, a growing number of ab initio investigations show the influence of metal ion binding on the stability and molecular properties of base pairs[28−35] and triplets.[36,37] In this report, we present the results of our ab initio calculations on the main forms of the G−A base pairs. Earlier attempts to estimate the binding energy of these systems included molecular mechanics calculations on DNA sequences incorporating the G−A mismatches[38] and atom−atom potential studies of the mismatched base pairs,[39] as well as ab initio calculations employing geometry optimization in the Hartree−Fock (HF) approximation and energy evaluation with the inclusion of electron correlation using the second-order Moller−

* Corresponding author. Fax: (434)-924-3710. E-mail: ms5c@virginia.edu.
[†] Università degli Studi di Milano.
[‡] University of Virginia.

10.1021/jp026315d CCC: $25.00
Published on Web 04/04/2003

Plesset (MP2) method.[40,41] We applied the SCF-MI (self-consistent field for molecular interactions) theory,[42−44] which has been used successfully to predict the structure and energetics of complexes between the GC Watson−Crick base pair and various metal cations.[33,34,45] To verify the recent findings on zinc-induced conformational changes in $d(GA)_n$ sequences,[19−25] we also carried out calculations on several complexes between the G−A pairs and hydrated Zn(II) ions. The conformational changes occurring in alternating $d(GA)_n$ sequences include the transition to a G−GC intramolecular triplex, stabilized by the Zn(II) ions bound to the N7 sites of the third-strand guanine residues.[20] To elucidate the transition, we performed SCF-MI calculations on the G−GC base triplet and its complex with a hydrated Zn(II) ion.

Theoretical Background

One of the major drawbacks in the study of molecular clusters employing the supermolecule approach is the introduction of basis set superposition error (BSSE), an error that can be of the same order of magnitude as the interaction energy itself. BSSE is due to the use of truncated basis sets located on the atomic centers of each monomer to compute the wave functions of the complex. In this way, when the molecules approach one another, the monomer-centered basis functions of one fragment can contribute to the description of the other fragments. This involves a more complete functional space at shorter intermolecular distances, causing a bias that gives rise to BSSE.

The most common procedure to correct a posteriori for BSSE is the counterpoise (CP) approach.[46] However, the addition of the partner's functions to counterbalance BSSE, although reasonable, does not provide a definite solution to the problem. The incorrect evaluation of the fragment relaxation contribution[47] and the introduction of "secondary superposition error"[48] can contribute to altering the shape of the PES and the resulting physical picture.

The SCF-MI (self-consistent field for molecular interactions) was proposed as an ab initio variational method that avoids the onset of BSSE in an a priori fashion. In this section, a brief introduction to the most relevant elements of the SCF-MI theory is provided; for a more detailed account of the method, see refs 42−44.

The SCF-MI method, based on modifications of Roothaan equations, consists of the partitioning of the total basis set

$$\chi = (\chi_1|\chi_2|\cdots\chi_K) \tag{1}$$

so that the MOs of different fragments are expanded in different subsets ($M = M_1 + \cdots + M_K$ is the basis set size) and are free to overlap. The resulting total ($M \times N$) matrix of the partitioned molecular orbital coefficients **T** has a block diagonal form

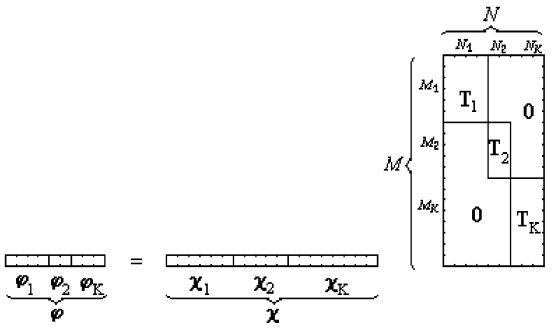
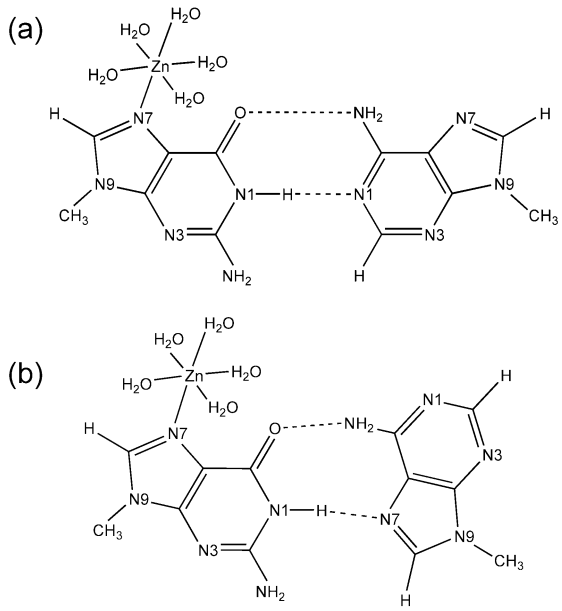


CHART 2: $[Zn(H_2O)_5]^{2+}$ Complexes with the (a) G(Anti)−A(Anti) and (b) G(Anti)−A(Syn) Base Pairs



The appearance of BSSE is avoided by assuming and maintaining the orbital coefficient variation matrix in a block diagonal form. The stationary condition $\delta E = 0$ is equivalent to K secular problems

$$\begin{cases} \mathbf{F}_k'\mathbf{T}_k' = \mathbf{S}_k'\mathbf{T}_k'\mathbf{L}_k \\ \mathbf{T}_k'^{\dagger}\mathbf{S}_k'\mathbf{T}_k' = \mathbf{1}^{N_k} \end{cases} \tag{2}$$

in terms of effective Fock and overlap matrices $\mathbf{F}_k'$ and $\mathbf{S}_k'$. These matrices are Hermitian and exhibit the correct asymptotic behavior: in the limit of infinite separation of the fragments, $\mathbf{F}_k'$ and $\mathbf{S}_k'$ become the Fock and overlap matrices of the individual systems. As a consequence, the SCF-MI binding energy can be simply expressed as

$$\Delta E_{\mathrm{SCF-MI}} = E_{\mathrm{SCF-MI}} - \sum_{k=1}^{K} E_{\mathrm{SCF}}^{k} \tag{3}$$

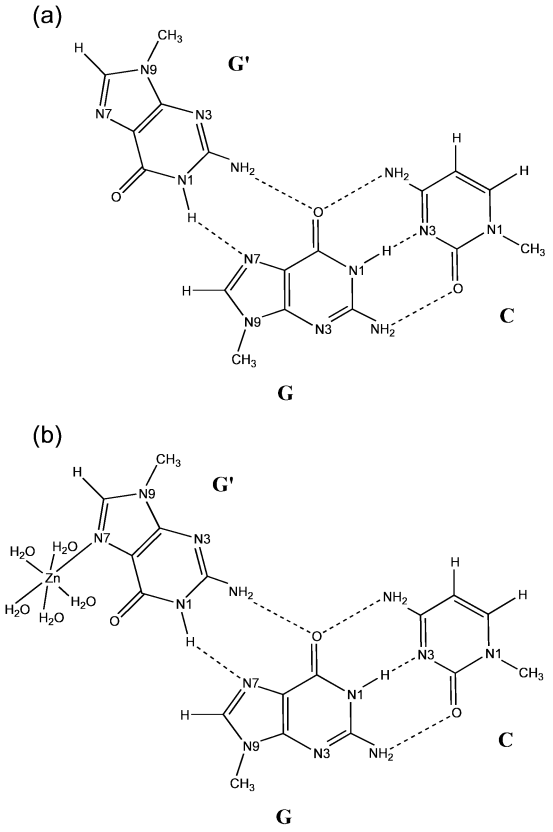
naturally taking into account geometry relaxation effects. The SCF-MI algorithm was implemented[44] into the GAMESS-US suite of programs[49] and also in the particularly efficient PC GAMESS version.[50]

It should be noted that the SCF-MI method directly includes geometry relaxation effects without computational penalty. A particular advantage of the a priori exclusion of BSSE is the possibility of adopting small basis sets and yet obtaining accurate results. The suitability of the 3-21G basis for SCF-MI calculations has been tested on systems of biological interest, such as the nucleic acid base pairs[51] and the complex of cisplatin and other metal cations with the Watson−Crick GC pair,[33−34] with obvious computational advantage.

Computational Details

The systems under investigation included the base pairs formed by 9-methyladenine (9-MeA) and 9-methylguanine (9-MeGH) (Chart 1) as well as their complexes with hydrated Zn-(II) ions covalently bound to the N7 sites of the bases (Chart 2). The possibility of Zn(II) binding at these sites was indicated by gel electrophoresis experiments carried out on the hairpin-forming $d(GA)_{15}$ sequences modified with dimethyl sulfate

CHART 3: (a) G−GC Base Triplet and (b) Its Complex with the $[Zn(H_2O)_5]^{2+}$ Ion



(DMS) and diethylpyrocarbonate (DEPC).[19] A 9-MeGH−9-MeGH−1-MeC base triplet and its complex with Zn(II) bound to the N7 of the third-strand guanine were also used in the study (Chart 3). The base triplet is composed of three fragments, namely, the two Watson−Crick bases (9-MeGH and 1-MeC) and the third-strand 9-MeGH, forming a complex with the hydrated zinc ion. The total binding energy of this system can be expressed either as a difference between the total energy of the system and the energy of the contributing optimized monomers (eq 3) or as a sum of two-body (pairwise) interaction energies and a three-body term ΔE_3

$$\Delta E^{ABC}_{SCF-MI} = \Delta E^{AB} + \Delta E^{BC} + \Delta E^{AC} + \Delta E_3 \qquad (4)$$

the generic pairwise contribution is defined as the difference

$$\Delta E^{AB} = E^{AB\#}_{SCF-MI} - E^{A}_{SCF} - E^{B}_{SCF} \qquad (5)$$

where # indicates the energy of the AB system calculated at the geometry of the trimer. It should be emphasized that, because the energies of the monomers are determined at their optimized geometries, the pairwise and three-body terms also incorporate the geometry deformation contribution.

All ab initio calculations were performed by applying the SCF-MI procedure implemented in the GAMESS-US package.[49−51] The standard split-valence 3-21G basis set was employed for all atoms other than zinc. The zinc core electrons were described by an effective core potential, as reported by Stevens et al.,[52] with its corresponding triple-ζ basis set for the remaining valence electrons. The use of the 3-21G basis set was instigated by the encouraging results of our previous calculations on various nucleic acid base pairs and their complexes with hydrated metal ions.[33,34] Several satisfactory comparisons, including one with the HF/6-31G* and MP2/6-31G*(0.25)//HF/6-31G** calculations of Šponer et al.,[40] confirm the validity of the 3-21G set. Furthermore, our earlier report on the $[Pt(NH_3)_2]^{2+}$ complex with the guanine−cytosine pair[45] showed very good agreement between the results of the 3-21G SCF-MI calculations and those from the investigations applying a larger SCF-MI basis set, as well as the HF/6-31G**, MP2, and DFT calculations. In addition, an analysis of the electrostatic potential presented in that paper indicated that the flexibility of the 3-21G basis set is sufficient to properly describe the polarization induced by the $[Pt(NH_3)_2]^{2+}$ ion.

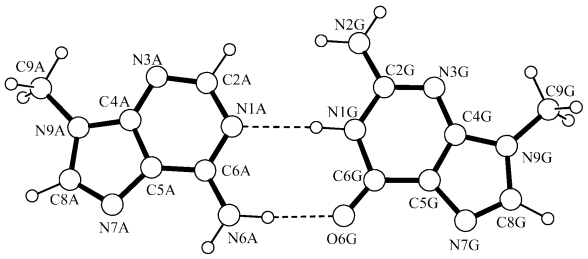


Figure 1. Optimized structure of the 9-MeGH(anti)−9-MeA(anti) base pair.

To investigate geometrical arrangements significant in the framework of the DNA structure, several geometry constraints were imposed on the complexes under study. However, to verify the validity of this approach, the systems were also fully gradient-optimized without constraints, resulting in highly distorted nonplanar structures.

Results and Discussion

9-MeGH(anti)—9-MeA(anti) Base Pair. The optimized structure of this base pair is shown in Figure 1. As mentioned earlier, all atoms of the system, with the exception of the methyl-group hydrogens, were constrained to be coplanar. However, the geometry constraints do not affect the final energy in any significant manner, and all of the torsional components of the energy gradient (both intra- and intermolecular) are reasonably small. The binding energy of the planar system is 10.67 kcal/mol, whereas that of the unconstrained base pair is 10.71 kcal/mol. The geometry of the hydrogen-bond system connecting the bases is identical to that determined by the HF calculations.[40] Thus, the N6(A)···O6(G) and N1(G)···N1(A) donor−acceptor distances are 2.96 and 3.20 Å, respectively, whereas those obtained by the HF approach are 2.95 and 3.19 Å.

9-MeGH(anti)—9-MeA(anti)—Zn(N7G) Complex. The system is formed by the planar 9-MeGH(anti)−9-MeA(anti) base pair and the unconstrained $[Zn(H_2O)_5]^{2+}$ unit with zinc covalently bound to the N7 atom of guanine (Figure 2). The binding energy of this system is 13.86 kcal/mol, about 4 kcal/mol lower than that of the fully unconstrained assembly. However, full optimization resulted in a highly distorted geometry with the nucleobases almost perpendicular to each other. Similar nonplanar conformation has previously been observed in the complexes formed by the CG pair with *cis*-Pt-$[NH_3]_2^{2+}$ and with the free and hydrated magnesium cations.[33−34]

Compared to the metal-free system, the geometry of the biologically relevant planar complex (Figure 2) shows several changes, apparently induced by the zinc binding. The N1(G)···N1(A) distance is shortened to 3.06 Å, whereas at the same time, there is a substantial lengthening of the N6(A)···O6(G) separation to 3.13 Å. In addition, the separation between the H2(A) and H2(G) hydrogen atoms is reduced by 0.33 Å, resulting in a stronger repulsion of the two atoms. The Zn−

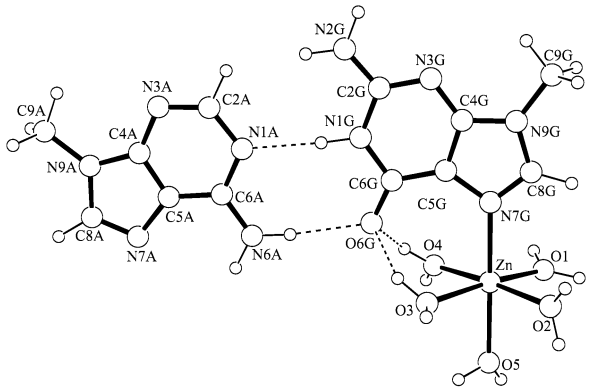


Figure 2. Optimized structure of the 9-MeGH(anti)−9-MeA(anti)−Zn(N7G) complex.

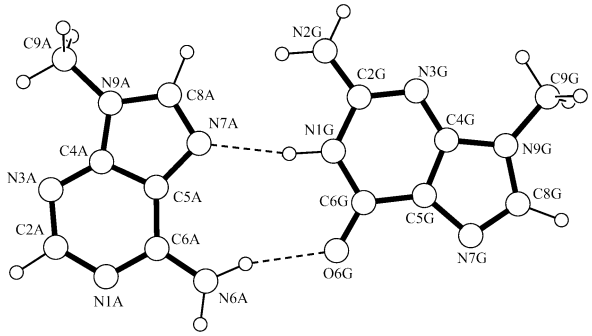


Figure 3. Optimized structure of the 9-MeGH(anti)−9-MeA(syn) base pair.

N7(G) bond length of 2.15 Å in the planar system compares well with that of 2.14(1) Å found in the X-ray structure of [Zn-$(H_2O)_4$(Me-5′-GMP)$_2$] (where Me-5′-GMP = 5′-methylguanosine-5′-monophosphate).[53] Also the Zn−O(water) distances, ranging from 2.07 to 2.11 Å, are similar to those observed in the crystal structure. For all practical purposes, there are no significant alterations in the geometry of the anti-disposed 9-methyladenine base of the Zn complex. On the other hand, changes in the metric parameters of 9-methylguanine show some interesting trends. The most pronounced variations occur in the exocyclic oxo O6 group and its surroundings. The O6 site acts as an acceptor in three hydrogen bonds, two involving the coordinated water molecules and another from the amino group of 9-methyladenine. The C6−O6 bond length increases significantly from 1.22 Å in the metal-free system to 1.27 Å in the Zn complex. The increase is accompanied by reductions of the N1−C6 distance from 1.42 to 1.37 Å and of the C5−C6 distance from 1.42 to 1.40 Å.

The binding energy of the Zn complex is about 3.2 kcal/mol higher than that of the 9-MeGH−9-MeA base pair, indicating additional stabilization of the pair induced by the covalent interaction with the hydrated zinc ion.

9-MeGH(anti)—9-MeA(syn) Base Pair. This assembly is stabilized by two hydrogen bonds (Figure 3). One of them employs the amino group of 9-MeA as a donor and the O6 site of 9-MeGH as an acceptor; the other involves the N1(G) site as a donor and N7(A) site as an acceptor.

The distances characterizing the two hydrogen bonds, 2.92 and 3.20 Å for the N6(A)···O6(G) and N1(G)···N7(A) separations, respectively, are quite similar to those present in the anti−anti arrangement of the two bases. However, the orientation of the adenine amino group in the anti−syn pair is less favorable for a strong hydrogen bond. For example, the H−N6(A)−O6-(G) angle shows a greater deviation (7.6° vs 2.9° in the anti−anti conformation) from the optimal linear arrangement. These differences, along with the diverse orientation of base dipole moments, could be related to the smaller binding energy of the planar anti−syn base system (9.88 kcal/mol) with respect to that of the anti−anti base pair.

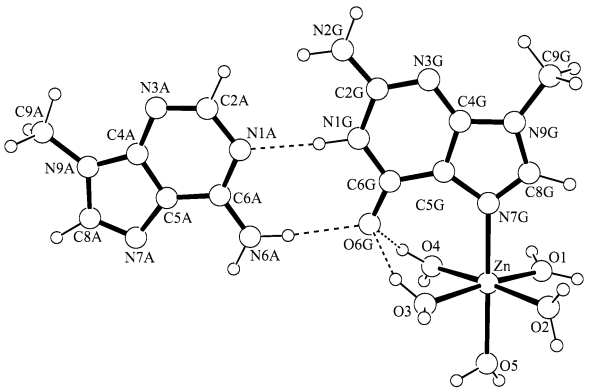


Figure 4. Optimized structure of the 9-MeGH(anti)−9-MeA(syn)−Zn(N7G) Complex.

9-MeGH(anti)—9-MeA(syn)—Zn(N7G) Complex. The optimized geometry of the planar 9-MeGH(anti)−9-MeA(syn)−Zn(N7G) complex is shown in Figure 4. The general features of the system are similar to those of the Zn complex with the anti−anti base pair. There are no noticeable changes in the geometry of 9-methyladenine. On the other hand, the type and numerical values of the variations taking place in the 9-methylguanine moieties are identical to those found in the anti−anti system. Compared to the metal-free system, the geometry of the hydrogen bonds within the complex shows two major modifications. The N6(A)−H···O6(G) separation increases from 2.92 to 3.08 Å, indicating a weakening of this hydrogen bond, whereas the N1(G)−H···N7(A) bond is apparently strengthened, as suggested by the shortening of the distance between its donor and acceptor atoms from 3.20 to 3.09 Å. However, the zinc-induced enhancement of the stability of this complex is less pronounced in comparison to that found for the anti−anti base pair. Furthermore, the binding energy of this complex (10.94 kcal/mol) is only 1 kcal/mol greater than that of the metal-free system.

9-MeGH(anti)—9-MeA(anti)—Zn(N7A) Complex. This system was optimized with the same geometrical constraints as those in the other Zn complexes, with one exception: no torsional constraints were imposed on the hydrogens of the amino group of adenine, allowing this group to optimize the interaction with the hydrated zinc moiety by pyramidalization.[31] It should be noted that the loss of planarity of the amino group is entirely due to the out-of-plane deviation of the hydrogen on the major groove side. Because the other hydrogen atom remains coplanar with the base, the pyramidalization does not prevent a priori the formation of a strong hydrogen-bond network within the base pair.

However, the presence of zinc bound to the N7 site of adenine has a profound effect on the base pair regardless of the type of geometry optimization performed. Figure 5 shows the structure of the complex after several steps of the constrained geometry optimization. As can be seen, the N1(G)−H···N1(A) hydrogen bond is broken, and the guanine, pivoting on the remaining hydrogen bond, orients its N7/O6 side toward the hydrated cation. Even though the geometry at that particular stage has not reached full convergence, the system exhibits a remarkably high binding energy of 27.78 kcal/mol. Removing the copla-

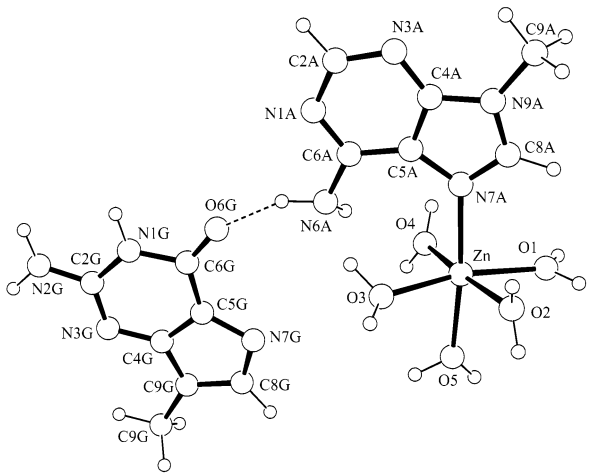


Figure 5. Structure of the 9-MeGH(anti)−9-MeA(anti)−Zn(N7A) complex after several cycles of the constrained geometry optimization.

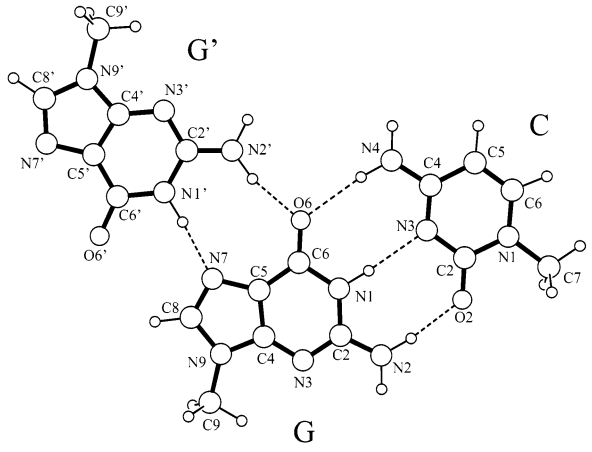


Figure 6. Optimized structure of the 9-MeGH−9-MeGH−1-MeC base triplet.

narity constraints of the two bases results in an even more peculiar structure. The N6(A)−H···O6(G) hydrogen bond breaks, and the O6 and N7 sites of guanine act as hydrogen-bond acceptors for two different water molecules of the zinc solvation shell. In this case, the binding energy increases to more than 47 kcal/mol. Yet another consequence of the zinc complexation to the N7 site of adenine is the out-of-plane arrangement of the cation. The zinc atom is located ca. 0.5 Å above the plane of the base pair, whereas the one bound to guanine lies in the plane of the base.

9-MeGH—9-MeGH—1-MeC Base Triplet. The G′−GC base triplet is composed of the canonical WC GC pair and the third-strand guanine G′ bound in a reverse Hoogsteen fashion to the WC guanine base (Figure 6). Even though no constraints were imposed on the geometry optimization of the triplet, the structure retained almost perfectly planar arrangements of the bases. The total binding energy for this system is 38.72 kcal/mol. The decomposition of this value into the pairwise contributions together with the hydrogen-bond distances characterizing the triplet is reported in Table 1. The presence of the third-strand guanine does not seem to entail any significant influence on the geometry and conformation of the GC pair. The O2(C)···N2(G), N3(C)···N1(G), and N4(C)···O6(G) distances shown in Table 1 are almost identical to the values of 2.99, 3.05, and 2.94 Å, respectively, found in the isolated GC base pair.[54] Furthermore, a comparison of the binding energy for the isolated GC base pair (22.45 kcal/mol)[54] with the GC pairwise contribution to the total interaction energy of the triplet (22.10 kcal/mol) implies only negligible perturbation of the system.

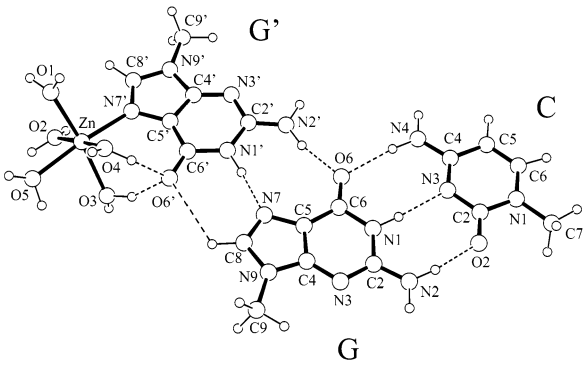


Figure 7. Optimized structure of the 9-MeGH−9-MeGH−1-MeC−Zn(N7G) System.

TABLE 1: Energy Decomposition (kcal/mol) and Hydrogen-Bond Distances (Å) in the 9-MeGH−9-MeGH−1-MeC (G′−GC) and 9-MeGH−9-MeGH−1-MeC−Zn (N7G′) (Zn−G′−GC) Triplets

	G′−GC	Zn−G′−GC
ΔE^{G-C}	22.10	20.71
$\Delta E^{G'-G}$	16.02	34.26
$\Delta E^{G'-C}$	−2.14	−1.98
ΔE_3	2.74	6.98
O2(C)···N2(G)	2.96	2.86
N3(C)···N1(G)	3.05	3.06
N4(C)···O6(G)	2.99	3.15
N1(G′)···N7(G)	2.95	2.89
N2(G′)···O6(G)	3.17	2.83

9-MeGH—9-MeGH—1-MeC—Zn(N7G) System. Zinc binding to the N7 position of the third-strand guanine does not affect the planarity of the triplet in any appreciable way. However, the presence of the hydrated metal cation significantly affects the hydrogen-bond network within the base triplet (Table 1 and Figure 7). Both the N1(G′)···N7(G) and N2(G′)···O6(G) donor−acceptor distances are shortened, the latter by more than 0.3 Å. These appreciable changes suggest a strengthening of the two hydrogen bonds and consequently an enhancement of the purine−purine interaction. On the other hand, the O6(G)···N4(C) distance involving O6 of the central guanine increases from 2.99 to 3.15 A. This elongation is accompanied by a shortening of the N2(G)···O2(C) distance, whereas the central hydrogen bond of the GC pair is not affected. The pronounced variations of both the G′···G and G···C intermolecular distances suggest long-range polarization effects due to the presence of the metal cation. This is also indicated by an analysis of the binding energies. The hydrated zinc bound to the third-strand guanine significantly enhances the total binding energy of the system (59.97 kcal/mol) by more than 50%. Most of the enhancement comes from a dramatic increase in the purine−purine pairwise contribution, which has almost doubled. Similar cation-induced base pairing enhancement has previously been observed in other complexes between hydrated zinc ions and base triplets.[36] It should be noted that the interpretation of the binding energy in the 9-MeGH−9-MeGH−1-MeC−Zn(N7G) complex is based on the assumption that hydrated zinc and the third-strand guanine constitute one subsystem. A more detailed analysis has not been performed here; however, assuming that the charge of the hydrated zinc ion creates a region of positive electrostatic potential around the N1−H and N2−H sites of the third-strand guanine similar to that found in the $[Pt(NH_3)_2GH]^{2+}$ complex,[45] it is reasonable to expect a significant strengthening of the

attraction by the negative N7 and O6 sites of the remote guanine, leading to a net increase in the corresponding guanine−guanine pairwise contribution.

The presence of long-range polarization effects is confirmed by an increase of the three-body term by more than 4 kcal/mol. This value offsets to a large extent the small reduction of the central guanine−cytosine pairwise term. Thus, it can be concluded that the presence of the hydrated zinc cation stabilizes the triplet as an entity involving all three bases and not only by enhancing the interaction between the two guanine bases.

The nature of the hydrogen bonding in the hydrated zinc complexes with the G−GC base triplets, as well as in the complexes involving the GA base pairs mentioned earlier, merits a more detailed discussion. This bonding is dominated by electrostatic and polarization interactions, as the charge of the hydrated zinc ion bound to the N7 site of guanine is expected to create a region of positive electrostatic potential around the base, with a subsequent strengthening of the hydrogen bonds, in which guanine acts as a proton donor. On the other hand, a reduction in the strength of hydrogen bonds can be anticipated for the systems in which guanine is a proton acceptor. The covalent contribution to the hydrogen bonding in the systems under investigation is negligible, as indicated by our previous study on the Watson−Crick GC and AT base pairs by the spin-coupled (SC) techniques.[55]

Biological Significance

Noncomplementary G−A base pairs have been found in a variety of DNA sequences and conformations. Both the G(anti)−A(anti) and G(anti)−G(syn) forms have been observed in X-ray structures.[9−12] On the other hand, NMR studies have indicated that single G−A mismatches usually adopt the G(anti)−A(anti) conformation in solutions at neutral pH.[56] All quantum-chemical calculations, including those presented in this paper, show that the G(anti)−A(anti) base pair is slightly more stable than the G(anti)−A(syn) pair.

Generally, DNA duplexes are destabilized by single G−A mismatches. However, sequences with several adjacent (tandem) G−A base pairs can be significantly more stable than those composed of the canonical base pairs only.[57] The stability of tandem G−A mismatches depends on the pH, the salt concentration, and the sequence context.[58] Long runs of the G−A pairs have also been encountered in such forms as the homopurine−homopyrimidine $d(GA)_n$ sequences occurring quite frequently in eukaryotic genomic DNA. Several reports have indicated a remarkable structural polymorphism of these sequences (for a review, see refs 16 and 59). At neutral pH in the presence of Zn(II) ions, they can form [GA(GA−TC)] intramolecular triplexes, which are transformed into (GA−GA) antiparallel-stranded homoduplexes when the Zn(II) ion concentration is above 3 mM.[24] Although several details are already known, the exact role of Zn(II) ions in inducing the structural transitions is still obscure. Gel electrophoresis and UV-melting experiments suggest that the stability of the (GA−GA) homoduplexes could result from the direct covalent binding of Zn(II) ions to the N7 sites of guanines. Our calculations on the 9-MeGH(anti)−9-MeA(anti)−Zn(N7G) model do show that the binding energy of the G(anti)−A(anti) pair increases by over 3 kcal/mol when Zn(II) is covalently bound to the N7 site of the guanine base. On the other hand, the stabilization caused by Zn(II) binding is less significant in the G(anti)−A(syn) pair. On the basis of their experiments with the $d(GA-TC)_{22}$ plasmid inserts[20] as well as synthetic $d(GA)_{15}$[24] and $d(GA)_{20}$[25] sequences, Azorin et al. concluded that the stems of the GA−GA hairpins are composed of alternating G(anti)−A(anti) and G(anti)−A(syn) base pairs. These formations could be stabilized by strong stacking interactions between the adjacent G−A pairs. Zn(II) binding to the N7 sites of the G(anti) residues could provide additional stabilization, although it is not clear whether and to what extent this binding would affect the stacking interactions. Zn(II) binding to the N7 sites of adenines seems much less likely, as indicated by our calculations for the 9-MeGH(anti)−9-MeA(anti)−Zn(N7A) system.

It has been postulated that the transition from the $d(GA-TC)_n$ B-DNA duplexes to the [GA(GA−TC)] intramolecular triplexes is facilitated by Zn-induced destabilization of the duplexes and simultaneous stabilization of the base triplets.[59] In fact, the large increase of 21.25 kcal/mol in the G−GC triplet binding energy due to the Zn coordination to the N7 site of guanine indicates a significant stabilization of the base triplet. However, the GA(GA−TC) triplexes are composed of alternating G−GC and A−AT bases triplets.[6] Some earlier calculations[36] suggested that these mixed purine−purine−pyrimidine triplexes might actually be much less stabilized by Zn binding to the N7 atoms of the third-strand adenines than the adjacent G−GC triplets, which coordinate Zn through the third-strand guanine N7 sites. Consequently, the reduced relative stability of the A−AT triplets could play some role in the rearrangement of the triplexes into the GA hairpins. On the other hand, a very recent report[60] indicated that intramolecular parallel triplexes with G−GC and T−AT base triplets can also exist in the presence of Zn(II) ions. Because the thymine bases are usually not involved in metal binding,[61] the stabilization of these difficult to obtain triplexes is most likely due to the zinc interaction with the N7 sites of guanines. Thus, it might be possible that the Zn(II) binding to the third-strand guanines is the decisive stabilizing force in all mixed DNA triplexes containing G−GC triplets.

Summary

Structures and energies of the noncomplementary G−A base pairs and their complexes with a hydrated Zn(II) ion have been studied using the self-consistent field for molecular interaction (SCF-MI) calculations. Alternating G(anti)−A(anti) and G(anti)−A(syn) base pairs are believed to be present in the stems of the GA−GA DNA hairpins. The formation of covalent bonds between the Zn(II) ions and the N7 sites of the guanine bases in the stems increases the binding energy of the G(anti)−A(anti) and G(anti)−A(syn) pairs by ca. 3.2 and 1.1 kcal/mol, respectively. The difference in the binding energies, although small, does indicate a somewhat better stabilization of the G(anti)−A(anti) pair. Zn(II) binding to the N7 sites of adenines in the G(anti)−A(anti) pairs appears less likely, as indicated by our calculations for the 9-MeGH(anti)−9-MeA(anti)−Zn(N7A) system. The geometry optimization of this system resulted in highly distorted structures with no apparent biological significance. On the other hand, Zn(II) binding to the N7 site of the third-strand guanine induces an appreciable stabilization of the G−GC base triplet, which might be an important factor in the formation of mixed DNA triplexes containing this as well as other base triplets in their sequences.

The effect of the solvent on the binding energy was not directly addressed in the present report. However, in a previous study on the hydration of the Watson−Crick GC base pair,[54] we found an enhancement of the binding energy due to the formation of additional hydrogen bonds involving the water molecules of the hydration sphere; a similar effect cannot be excluded here.

References and Notes

(1) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic Acids. Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.

(2) Saenger, W. *Principles of Nucleic Acid Structure.* Springer-Verlag: New York, 1984.

(3) Drake, J. W. *The Molecular Basis of Mutation*; Holden-Day: San Francisco, CA, 1970.

(4) Friedberg, E. C. *DNA Repair*; W. H. Freeman and Co.: New York, 1985.

(5) Hilbers, C. W.; Heus, H. A.; van Dongen, M. J. P.; Wijmenga, S. S. In *Nucleic Acids and Molecular Biology*; Eckstein, F., Lilley, D. M. J., Eds.; Springer-Verlag: Berlin, 1994; Vol. 8, p 56.

(6) Soyfer, V. N.; Potaman, V. N. *Triple-Helical Nucleic Acids*; Springer-Verlag: New York, 1996.

(7) Patel, D. J.; Bouaziz, S.; Kettani, A.; Wang, Y. In *Oxford Handbook of Nucleic Acid Structure*; Neidle, S., Ed.; Oxford University Press: New York, 1999; p 389.

(8) Fersht, A. R.; Knill-Jones, J. W.; Tsui, W.-C. *J. Mol. Biol.* **1982**, *156*, 37.

(9) Hunter, W. N.; Brown, T. In *Oxford Handbook of Nucleic Acid Structure*; Neidle, S., Ed.; Oxford University Press: New York, 1999; p 313.

(10) Prive, G. G.; Heinemann, U.; Chandrasegaran, S.; Kan, L.-S.; Kopka, M. L.; Dickerson, R. E. *Science* **1987**, *238*, 498.

(11) Brown, T.; Hunter, W. N.; Kneale, G. G.; Kennard, O. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 2402.

(12) Brown, T.; Leonard, G. A.; Booth, E. D.; Chambers, J. *J. Mol. Biol.* **1989**, *207*, 455.

(13) Fazakerley, G. V.; Boulard, Y. *Methods Enzymol.* **1995**, *261*, 145.

(14) Greene, K. L.; Jones, R. L.; Li, Y.; Robinson, H.; Wang, A. H.-J.; Zon, G.; Wilson, W. D. *Biochemistry* **1994**, *33*, 1053.

(15) van Dongen, M. J. P.; Mooren, M. W. W.; Willems, E. F. A.; van der Marel, G. A.; van Boom, J. H.; Wijmenga, S. S.; Hilbers, C. W. *Nucleic Acids Res.* **1997**, *25*, 1537.

(16) Chou, S.-H.; Reid, B. R.; In *Oxford Handbook of Nucleic Acid Structure*; Neidle, S., Ed.; Oxford University Press: New York, 1999; p 331.

(17) Chou, S.-H.; Zhou, L.; Reid, B. R. *J. Mol. Biol.* **1997**, *267*, 1055.

(18) Rippe, K.; Fritsch, V.; Westhof, E.; Jovin, T. M. *EMBO J.* **1992**, *11*, 3777.

(19) Huertas, D.; Bellsolell, L.; Casasnovas, J. M.; Coll. M.; Azorin, F. *EMBO J.* **1993**, *12*, 4029.

(20) Beltran, R.; Martinez-Balbas, A.; Bernues, J.; Bowater, R.; Azorin, F. *J. Mol. Biol.* **1993**, *233*, 966.

(21) Bernues, J.; Beltran, R.; Casasnovas, J. M.; Azorin, F. *Nucleic Acids Res.* **1990**, *18*, 4067.

(22) Casasnovas, J. M.; Huertas, D.; Ortiz-Lombardia, M.; Kypr, J.; Azorin, F. *J. Mol. Biol.* **1993**, *233*, 671.

(23) Martinez-Balbas, A.; Azorin, F. *Nucleic Acids Res.* **1993**, *21*, 2557.

(24) Ortiz-Lombardia, M.; Jimenez-Garcia, E.; Garcia-Bassets, I.; Azorin, F. *J. Biomol. Struct. Dyn.* **1998**, *16*, 243.

(25) Ortiz-Lombardia, M.; Eritja, R.; Azorin, F.; Kypr, J.; Tejralova, I.; Vorlickova, M. *Biochemistry* **1995**, *34*, 14408.

(26) Šponer, J.; Leszczynski, J.; Hobza, P. *Biopolymers* **2002**, *61*, 3.

(27) Basch, H.; Krauss, M.; Stevens, W. J. *J. Am. Chem. Soc.* **1985**, *107*, 7267.

(28) Anwander, E. H. S.; Probst, M. M.; Rode, B. M. *Biopolymers* **1990**, *29*, 757.

(29) Burda, J. V.; Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1997**, *101*, 9670.

(30) Šponer, J.; Sabat, M.; Burda, J. V.; Leszczynski, J.; Hobza, P. *J. Phys. Chem. B* **1999**, *103*, 2528.

(31) Gresh, N.; Šponer, J. *J. Phys. Chem. B* **1999**, *103*, 11415.

(32) Šponer, J.; Sabat, M.; Gorb, L.; Leszczynski, J.; Lippert, B.; Hobza, P. *J. Phys. Chem. B* **2000**, *104*, 7535.

(33) Famulari, A.; Moroni, F.; Sironi, M.; Raimondi, M. *Comput. Chem.* **2000**, *24*, 341.

(34) Famulari, A.; Moroni, F.; Sironi, M.; Gianinetti, E.; Raimondi, M. *J. Mol. Struct. (THEOCHEM)* **2000**, *529*, 209.

(35) Munoz, J.; Šponer, J.; Hobza, P.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2001**, *105*, 6051.

(36) Šponer, J.; Sabat, M.; Burda, J. V.; Doody, A. M.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1998**, *16*, 139.

(37) Munoz, J.; Gelpi, J. L.; Soler-Lopez, M.; Subirana, J. A.; Orozco, M.; Luque, F. J. *J. Phys. Chem. B* **2002**, *106*, 8849.

(38) Keepers, J. W.; Schmidt, P.; James, T. L.; Kollman, P. A. *Biopolymers* **1984**, *23*, 2901.

(39) Poltev, V. I.; Shulyupina, N. V. *J. Biomol. Struct. Dyn.* **1986**, *3*, 739.

(40) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Phys. Chem.* **1996**, *100*, 1965.

(41) Šponer, J.; Leszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* **1996**, *14*, 117.

(42) Famulari, A.; Gianinetti, E.; Raimondi, M.; Sironi, M.; Vandoni, I. *Theor. Chem. Acc.* **1998**, *99*, 358.

(43) Gianinetti, E.; Vandoni, I.; Famulari, A.; Raimondi, M. *Adv. Quantum Chem.* **1998**, *31*, 251.

(44) Famulari, A.; Gianinetti, E.; Raimondi, M.; Sironi, M. *Int. J. Quantum Chem.* **1998**, *69*, 151.

(45) Pelmenshchikov, A.; Zilberberg, I. L.; Leszczynski, J.; Famulari, A.; Sironi, M.; Raimondi, M. *Chem. Phys. Lett.* **1999**, *314*, 496. Raimondi, M.; Famulari, A.; Gianinetti, E.; Sironi, M.; Moroni, F. Modification of the Roothaan equations for the ab initio calculation of interactions in large molecular systems in the absence of basis set superposition error. In *Large-Scale Scientific Computations of Engineering and Environmental Problems II*; Griebel, M., Margenov, S., Yalamov, P., Eds.; Vieweg: Wiesbaden, Germany, 2000; p 335.

(46) Boys, S. F.; Bernardi, F. *Mol. Phys.* **1970**, *19*, 553.

(47) Xantheas, S. S. *J. Chem. Phys.* **1996**, *104*, 8821.

(48) Karlström, G.; Sadlej, A. J. *Theor. Chim. Acta* **1982**, *61*, 1.

(49) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347.

(50) Granovsky, A. A. *The PC GAMESS*; Laboratory of Chemical Cybernetics: Moscow State University (MSU): Moscow, Russia, 2001 (available at http://classic.chem.msu.su/gran/gamess/index.html).

(51) Famulari, A.; Raimondi, M.; Sironi, M. Hartree−Fock study of hydrogen bonded systems in the absence of basis set superposition error: The nucleic acid base pairs. In *New Trends in Chemistry and Physics*; Maruani, J., McWeeny, R., Minot, M., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 2001; Vol. 1, p 313.

(52) Stevens, W. J.; Krauss, M.; Basch, H.; Jasien, P. G. *Can. J. Chem.* **1992**, *70*, 612.

(53) Miller, S. K.; VanDerveer, D. G.; Marzilli, L. G. *J. Am. Chem. Soc.* **1985**, *107*, 1048.

(54) Moroni, F.; Famulari, A.; Raimondi, M. *J. Phys. Chem. A* **2001**, *105*, 1169.

(55) Raimondi, M.; Famulari, A.; Gianinetti, E. *Int. J. Quantum Chem.* **1999**, *74*, 259.

(56) Gao, X.; Patel, D. J. *J. Am. Chem. Soc.* **1988**, *110*, 5178.

(57) Li, Y.; Zon, G.; Wilson, W. D. *Biochemistry* **1991**, *30*, 7566.

(58) Cheng, J.-W.; Chou, S.-H.; Reid, B. R. *J. Mol. Biol.* **1992**, *228*, 1037.

(59) Bernues, J.; Azorin, F. In *Nucleic Acids and Molecular Biology*; Eckstein, F., Lilley, D. M. J., Eds.; Springer-Verlag: Berlin, 1995; Vol. 9, p 1.

(60) Khomyakova, E. B.; Gousset, H.; Liqiuer, J.; Huyh-Dinh, T.; Gouyette, C.; Takahashi, M.; Florentiev, V. L.; Taillandier, E. *Nucleic Acids Res.* **2000**, *28*, 3511.

(61) Marzilli, L. G. In *Advances in Inorganic Biochemistry. 3. Metal Ions in Genetic Information Transfer*; Eichhorn, G. L., Marzilli, L. G., Eds.; Elsevier North-Holland: New York, 1981; p. 47.