Accepted Article

Title: Structural and Energetic Impact of Non-natural 7-Deaza-8-Azaguanine, 7-Deaza-8-Azaisoguanine and their 7-Substituted Derivatives on H-bonding Pairing with Cytosine and Isocytosine

Authors: Mohit Chawla, Yury Minenkov, Khanh B. Vu, Romina Oliva, and Luigi Cavallo

This manuscript has been accepted after peer review and appears as an Accepted Article online prior to editing, proofing, and formal publication of the final Version of Record (VoR). This work is currently citable by using the Digital Object Identifier (DOI) given below. The VoR will be published online in Early View as soon as possible and may be different to this Accepted Article as a result of editing. Readers should obtain the VoR from the journal website shown below when it is published to ensure accuracy of information. The authors are responsible for the content of this Accepted Article.

To be cited as: ChemBioChem 10.1002/cbic.201900245

Link to VoR: http://dx.doi.org/10.1002/cbic.201900245
Structural and Energetic Impact of Non-natural 7-Deaza-8-Azaguanine, 7-Deaza-8-Azaisoguanine and their 7-Substituted Derivatives on H-bonding Pairing with Cytosine and Isoctyosine

Mohit Chawla,1,* Yury Minenkov,2 Khanh B. Vu,3 Romina Oliva,4,* Luigi Cavallo1,*

1King Abdullah University of Science and Technology (KAUST), Physical Sciences and Engineering Division, Kaust Catalysis Center, Thuwal 23955-6900, Saudi Arabia. 2Moscow Institute of Physics and Technology, Institutskiy Pereulok 9, Dolgoprudny, Moscow Region 141700, Russia. 3NTT Hi-Tech Institute, Nguyen Tat Thanh University, 298-300A Nguyen Tat Thanh Street, Ho Chi Minh City, Viet Nam. 4Department of Sciences and Technologies, University Parthenope of Naples, Centro Direzionale Isola C4, I-80143, Naples, Italy

Email: mohit.chawla@kaust.edu.sa; oliva@uniparthenope.it; luigi.cavallo@kaust.edu.sa

Abstract

We theoretically characterized the impact that the 7-deaza-8-azaguaine (DAG) and 7-deaza-8-azaisoguanine (DAiG) modifications have on the geometry and stability of the G:C Watson-Crick (cWW) base pair and of the G:iC and iG:C reverse Watson-Crick (tWW) base pairs. In addition, we investigated the effect on the same base pairs of seven C7-substituted DAG and DAiG, some of which have been previously experimentally characterized. Our calculations indicate that all these modifications have a negligible impact on the geometry of the above base pairs, and that the modification of the heterocycle skeleton has small impact on the base pair interaction energies. Instead, base pair interaction energies are dependent on the nature of the C7 substituent. For the 7-substituted DAG-C cWW systems we found a linear correlation between the base pair interaction energy and the Hammett constant of the 7-substituent, with higher interaction energies corresponding to more electron-withdrawing substituents. Therefore, the explored modifications are expected to be accommodated in both parallel and antiparallel nucleic acid duplexes without perturbing their
geometry, while the strength of a base pair (and duplex) featuring a DAG modification can in principle be tuned by incorporating different substituents at the C7 position.

**Introduction**

The four canonical nucleotides, adenine (A), uracil/thymine (U/T), guanine (G) and cytosine (C), can be naturally modified both in RNA and in DNA molecules. In RNA, over 110 natural post-transcriptional modifications reported to date greatly enhance its chemical information and functionality.\(^\text{[1]}\) In DNA, nucleotide epigenetic modifications regulate the genes transcription, thus affecting a variety of processes, including human health and disease.\(^\text{[2]}\) In addition, using biotechnology techniques several non-natural (synthetic) nucleotides have been introduced in RNA and DNA for targeted applications.\(^\text{[3]}\) Modification of the nucleobases\(^\text{[4]}\) and of the ribose/deoxyphosphate backbone\(^\text{[5]}\) can impact the structural stability, kinetics, and resistance to enzymatic degradation of nucleic acid molecules.\(^\text{[6]}\) Therefore, designed non-natural nucleotides are being used in probing RNA structure and function,\(^\text{[7]}\) exploring the interaction of RNA molecules with proteins, imparting favorable properties to small interfering RNAs (siRNAs),\(^\text{[3e, 8]}\) expanding the genetic code,\(^\text{[9]}\) and even creating a semi-synthetic organism with increased potential for information storage and retrieval.\(^\text{[10]}\)

In this scenario, a few 8-azapurine and 7-deaza analogues have been synthesized and characterized,\(^\text{[8, 11]}\) with the latter ones often presenting a variety of 7-substituents.\(^\text{[11a-c,][11b, 11d-f]}\) In the context of RNA, Beal and coworkers focused on exploring modifications of nucleosides that project different substituents either in the minor groove (the sugar edge) or in the major groove (the Hoogsteen edge) of a siRNA duplex.\(^\text{[8, 12-13]}\) Indeed, the synthesis of nucleobase analogues that retain the ‘Watson-Crick’ like pairing and that place the substituents in the major groove is of particular interest, since they preserve the duplex stability and do not alter recognition by the nuclease of the RNA interference (RNAi) pathway.\(^\text{[12]}\) Beal and coworkers also reported the crystallographic structures of two 16-base pair RNA duplexes incorporating the 7-ethynyl-7-deaza-8-aza adenine and 7-triazole-7-deaza-8-aza adenine modifications, and concluded that these non-natural bases can canonically pair to natural uracil and are well-accommodated within an A-form helix.\(^\text{[12]}\) More importantly, the two non-natural nucleosides are read as adenosine by avian myoblastosis virus reverse transcriptase (AMV-RT).\(^\text{[12]}\) Our subsequent quantum mechanics analysis confirmed that the modified 7-deaza-8-
azaadenine (DAA) and its 7-substituted derivatives have a negligible impact on the geometry and interaction energy of the base pairs they make with uracil, both in antiparallel and parallel double-helix strands, and provided a theoretical explanation for that.\[14\]

Similarly, several 7-deaza-8-aza analogues of guanine have also been synthesized and used in different applications in the last couple of decades, especially in the context of DNA.\[15\] A variety of 7-substituted 7-deaza-8-azaguanine (abbreviated here as DAG) residues, including 7-halogen derivatives, have been shown to stabilize the classical antiparallel DNA duplex,\[16\] while 7-deaza-8-azaisoguanine (abbreviated here as DAiG) and its 7-halogen derivatives have been shown to stabilize the parallel DNA duplex when paired to self-complementary strands.\[11g\] DAG has also been used as a scaffold to which complex 7-substituents are added to obtain modified DNAs potentially useful in biotechnological applications, either by introducing a strong fluorescence emission\[17\] or by building supramolecular DNA assemblies (nanostructures).\[18\] More recently, the DAG modification has been introduced in a triplex-forming oligonucleotide (TFO), aimed to target a gene segment to induce genomic modifications, with the ultimate goal of repairing genetic defects. Remarkably, the DAG-containing TFO was shown to have improved binding efficiency towards the target sequence in the presence of potassium and to induce the expected genomic modifications.\[19\] Finally, in the context of RNA, DAG has been used as an analog to map out the interactions of the exogenous guanosine molecule in the Tetrahymena group I ribozyme, and shown to compromise neither its binding nor its reactivity.\[20\]

![Figure 1](https://example.com/fig1.png)

**Figure 1.** Structure of (a) guanine (G); (b) 7-substituted-7-deaza-8-azaguanine (7-X-DAG); (c) isoguanine (iG); (d) 7-substituted-7-deaza-8-aza-isoguanine (7-X-DAiG) with conventional numbering of purines. Values of the Hammett constant characterizing the various substituents (X) are also listed. The sugar-phosphate backbone of the nucleotides has been truncated at C1’ (‘R’ in the Figure).
To complement these experimental studies and rationalize the apparently negligible impact of the above complex modifications on the geometry of nucleic acid duplex structures, often associated to a thermodynamic stabilization, we have undertaken a quantum mechanics study. In particular, herein we investigated a series of different G modifications, starting from the translocation of the N7 to C8 atoms, resulting in 7-deaza-8-aza-guanine (DAG) modification, and continuing with 7 other models where different functional groups are inserted at the C7 position of the DAG residue (see Figure 1). Our focus is on determining the impact of the modifications on the G H-bonding potential with its complementary C base. To this aim, we have investigated the effect of the modifications on the geometry and energetics of the base pair they are involved in, and compared them with those of the corresponding unmodified pair. This is an approach we and others have been applying previously to evaluate the impact of natural and non-natural modifications on various H-bonded base pair systems.\textsuperscript{14, 21} First, we analyzed the impact of the N7 and C8 atoms translocation in the G skeleton, leading to the DAG modification. Next, we analyzed the impact of the ethynyl and triazole substituents on the C7 atom of DAG, as they are widely used for the chemical labeling of nucleic acids (especially using click chemistry reactivity), allowing to image and retrieve them even in whole organisms, thus monitoring their dynamics in living systems.\textsuperscript{12, 22} Furthermore, to better understand the response of the H-bonding capability of the modified base to the nature of the substituent on the C7 position, we also analyzed modified bases bearing either the strongly electron withdrawing NO\textsubscript{2} group, or the strongly electron donating NH\textsubscript{2} group. These two groups are at the extremes of the Hammet scale,\textsuperscript{23} which measures the electron donor and withdrawing capability of substituents, and are known to modify remarkably the properties of the aromatic ring they are bonded to. Finally, the analysis is completed by considering as substituents F, Cl and Br, since halogens as 7-substituents of DAG have been experimentally studied. For all these systems we considered the classic Watson-Crick (cWW) and the reverse Watson-Crick (tWW) geometries, characteristic of antiparallel and parallel duplexes respectively, both in gas phase and in water. For the base pairs in the tWW geometry, the isoguanine–cytosine (iG:C) and guanine–isocytosine (G:iC) pairs have been used as reference systems, as the iG:C/G:iC pairs have been shown to stabilize parallel-stranded nucleic acid duplexes,\textsuperscript{24} while the classical tWW G:C pair has been shown to destabilize them.\textsuperscript{21a, 24a, 25} Therefore, DAiG and all its modifications with the above substituents on the C7 atom have also been modeled (Figure 1).
Models and Computational Details

In order to explore the stability of the modified base pairs under study, we have modeled them and calculated their interaction energy by density functional theory and post-HF methods.

Modeling the interaction system. We first focused on studying the impact on H-bonding of a heterocyclic ring with a simple translocation of the N7-C8 atoms in a guanine heterocycle. For this, we took into account 7-deaza-8-azaguanine (DAG) H-bonded to cytosine (C) forming a DAG:C cWW base pair. The base pair system was modeled by preserving the H-bonding pattern as in the G:C cWW pair. Next, in order to look at the effect of varying electronic properties, we modeled different C7 substituents of DAG H-bonded to cytosine. The studied substituents are ethynyl, 1,2,3-triazole, –NO2, –NH2, –F, –Cl, –Br (see Figure 1), giving rise to the 7-E-DAG:C, 7-T-DAG:C, 7-NO2-DAG:C, 7-NH2-DAG:C, 7-F-DAG:C, 7-Cl-DAG:C, 7-Br-DAG:C base pairs (see Figure 2). The coordinates of the modeled geometries of the modified base pairs were built starting from the optimized geometries of the canonical G:C cWW base pair. Therefore, in total we modeled 8 different geometries of the non-natural modified base pair systems presenting the cWW geometry. The glycosidic bonds of these base pairs are oriented in ‘cis’, which corresponds to the geometries pertinent to the canonical antiparallel stranded RNA structure.

More base pair combinations have been studied, which correspond to the reverse (trans) Watson-Crick geometry (abbreviated as tWW), characteristic of parallel stranded duplex structures. For the base pairing involving guanine and cytosine residues, the G:iC and iG:C base pairs have been considered as references rather than the classical G:C tWW pair. Thus, a total of 8 modified base pairing geometries were modeled corresponding to the G:iC pair, and another set of 8 base pairs were modeled corresponding to iG:C geometries, including all the substituents at the C7 position shown in Figure 1. For all the model systems described above, the base pairing geometries is truncated at the C1’ atom of the ribose. This is the standard approach used in literature.[21b, c, 25a, b]

QM calculations and Electron Density Analysis.

These methods are adopted from previous work [14, 21, 25a, b, 26] and the details of which can be found in Supplementary Information.
**Results and Discussion**

Eight different modifications of the guanine residue on the ‘Hoogsteen edge’, all of them being derivatives of 7-deaza-8-azaguanine (DAG), have been modeled in canonical cWW base pairs with C (see Figure 2). The same 8 modified bases and 8 additional non-natural bases, introducing analogous modifications on isoguanine, have also been modeled in non-canonical tWW base pairs with isocytosine and cytosine, respectively (Figures 3 and 4). All the above 24 base pairs have been geometrically and energetically investigated and compared to the corresponding unmodified systems. Table 1 summarizes the calculated interaction energies for the investigated base pairs. Optimal geometries and H-bonding distances in gas and in water are shown in Figure 2 for the *cis* Watson-Crick base pairs and in Figure 3 and 4 for the *trans* base pairs. To shed light on the possible decreased/increased stability consequent to the modifications, we also calculated and compared the electron densities of the modified versus the unmodified systems. They are shown in Figures 5, 6 and S1, and discussed in the following when appropriate.

Before starting a detailed characterization of base pairs with the modified bases, we tested the capability of the RIMP2 interaction energies to capture the severe modification of the heterocycle skeleton. As CCSD(T) results are widely accepted as the gold-standard in quantum chemistry in general and in the calculation of nucleobase pairs geometry and energy, in particular,[27] for a subset of systems we also calculated DLPNO-CCSD(T)/CBS energies.[26a, b, 28] Specifically, we calculated three reference base pairs, G:C cWW, G:iC tWW and iG:C tWW, and the corresponding base pairs where G/iG is substituted by a DAG/DAiG nucleobase: DAG:C cWW, DAG:iC tWW and DAiG:C tWW (Table 1, values in parenthesis). The comparison between the DLPNO-CCSD(T)/CBS energy values and those obtained by the RIMP2 approach shows that differences are minor. They range between 0.1 kcal/mol for G:C cWW and 1.3 kcal/mol for DAG:iC. Further, differences in the corresponding $E_{\text{mod}}$ values are smaller, ranging between 0.3 kcal/mol for DAiG:C and 0.9 kcal/mol for DAG:C cWW and DAG:iC tWW. This substantial agreement between the DLPNO-CCSD(T) and RIMP2 values supports and defines the RIMP2 energies discussed in the following. Due to the high computational cost of the DLPNO-CCSD(T) protocol we have chosen, it is impossible to apply it to all the systems and geometries discussed in the following.

| Table 1. Geometry and interaction energies, for the cWW and tWW G:C and modified G:C base pairs. All energies are reported in kcal/mol. $E_{\text{Int}}$ is the interaction energy of the two bases in the base pair. |
pair without inclusion of deformation energy. $E_{\text{Def}}$ is the deformation energy of the two bases calculated as the difference between the energy of the geometry they have in the base pair and the energy of the geometry they have when optimized alone. $E_{\text{Bind}}$ is the sum of interaction and deformation energies, i.e. $E_{\text{Bind}} = E_{\text{Int}} + E_{\text{Def}}$. $E_{\text{mod}}$ is the difference between the interaction energy of the modified base pair and of the reference pair, i.e. G:C cWW for cWW geometries, and G:iC and iG:C tWW for tWW geometries. Negative and positive values of $E_{\text{mod}}$ indicate that the modified base pair is more stable or less stable than the reference G:C base pairs. The energies are calculated at RIMP2/aug-cc-pVTZ level on B3LYP-D3/cc-pVTZ optimized geometries. Reference energy values calculated at DLPNO/CBS are reported in parentheses.

<table>
<thead>
<tr>
<th>System</th>
<th>Geometry</th>
<th>$E_{\text{int}}$</th>
<th>$E_{\text{Def}}$</th>
<th>$E_{\text{Bind}}$ (gas)</th>
<th>$E_{\text{mod}}$ (gas)</th>
<th>$E_{\text{Bind}}$ (water)</th>
<th>$E_{\text{mod}}$ (water)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G:C</td>
<td>cWW</td>
<td>-30.74</td>
<td>2.75</td>
<td>-27.98 (-28.11)</td>
<td>0.00</td>
<td>-12.49</td>
<td>0.00</td>
</tr>
<tr>
<td>DAG:C</td>
<td>cWW</td>
<td>-30.33</td>
<td>2.66</td>
<td>-27.66 (-28.65)</td>
<td>0.32 (-0.54)</td>
<td>-12.46</td>
<td>0.03</td>
</tr>
<tr>
<td>7-E-DAG:C</td>
<td>cWW</td>
<td>-30.82</td>
<td>2.62</td>
<td>-28.19</td>
<td>-0.19</td>
<td>-12.66</td>
<td>-0.17</td>
</tr>
<tr>
<td>7-T-DAG:C</td>
<td>cWW</td>
<td>-29.59</td>
<td>2.48</td>
<td>-27.11</td>
<td>0.87</td>
<td>-12.53</td>
<td>-0.04</td>
</tr>
<tr>
<td>7-F-DAG:C</td>
<td>cWW</td>
<td>-31.02</td>
<td>2.69</td>
<td>-28.33</td>
<td>-0.35</td>
<td>-12.64</td>
<td>-0.15</td>
</tr>
<tr>
<td>7-Cl-DAG:C</td>
<td>cWW</td>
<td>-31.07</td>
<td>2.71</td>
<td>-28.35</td>
<td>-0.37</td>
<td>-12.68</td>
<td>-0.19</td>
</tr>
<tr>
<td>7-Br-DAG:C</td>
<td>cWW</td>
<td>-31.09</td>
<td>2.71</td>
<td>-28.38</td>
<td>-0.40</td>
<td>-12.70</td>
<td>-0.21</td>
</tr>
<tr>
<td>7-NH2-DAG:C</td>
<td>cWW</td>
<td>-32.15</td>
<td>3.01</td>
<td>-29.14</td>
<td>-1.14</td>
<td>-12.86</td>
<td>-0.37</td>
</tr>
<tr>
<td>7-NH2-DAG:C</td>
<td>cWW</td>
<td>-29.97</td>
<td>4.34</td>
<td>-25.63</td>
<td>2.35</td>
<td>-12.47</td>
<td>0.02</td>
</tr>
<tr>
<td>G:iC</td>
<td>tWW</td>
<td>-33.95</td>
<td>4.02</td>
<td>-29.93 (-30.35)</td>
<td>0.00</td>
<td>-13.24</td>
<td>0.00</td>
</tr>
<tr>
<td>DAG:iC</td>
<td>tWW</td>
<td>-33.37</td>
<td>3.88</td>
<td>-29.49 (-30.78)</td>
<td>0.46 (-0.43)</td>
<td>-13.17</td>
<td>0.07</td>
</tr>
<tr>
<td>7-E-DAG:iC</td>
<td>tWW</td>
<td>-33.90</td>
<td>3.49</td>
<td>-30.41</td>
<td>-0.48</td>
<td>-13.36</td>
<td>-0.12</td>
</tr>
<tr>
<td>7-T-DAG:iC</td>
<td>tWW</td>
<td>-32.40</td>
<td>3.26</td>
<td>-29.13</td>
<td>0.80</td>
<td>-13.25</td>
<td>-0.01</td>
</tr>
<tr>
<td>7-F-DAG:iC</td>
<td>tWW</td>
<td>-34.15</td>
<td>3.20</td>
<td>-30.94</td>
<td>-1.01</td>
<td>-13.34</td>
<td>-0.10</td>
</tr>
<tr>
<td>7-Cl-DAG:iC</td>
<td>tWW</td>
<td>-34.19</td>
<td>3.39</td>
<td>-30.79</td>
<td>-0.86</td>
<td>-13.38</td>
<td>-0.14</td>
</tr>
<tr>
<td>7-Br-DAG:iC</td>
<td>tWW</td>
<td>-34.25</td>
<td>3.47</td>
<td>-30.78</td>
<td>-0.85</td>
<td>-13.41</td>
<td>-0.17</td>
</tr>
<tr>
<td>7-NH2-DAG:iC</td>
<td>tWW</td>
<td>-35.34</td>
<td>4.17</td>
<td>-31.17</td>
<td>-1.24</td>
<td>-13.55</td>
<td>-0.31</td>
</tr>
<tr>
<td>7-NH2-DAG:iC</td>
<td>tWW</td>
<td>-32.96</td>
<td>4.64</td>
<td>-28.31</td>
<td>1.62</td>
<td>-13.19</td>
<td>0.05</td>
</tr>
<tr>
<td>iG:C</td>
<td>tWW</td>
<td>-34.08</td>
<td>3.18</td>
<td>-30.90 (-31.30)</td>
<td>0.00</td>
<td>-13.77</td>
<td>0.00</td>
</tr>
<tr>
<td>DAiG:C</td>
<td>tWW</td>
<td>-35.60</td>
<td>3.23</td>
<td>-32.36 (-33.02)</td>
<td>-1.46 (-1.72)</td>
<td>-13.97</td>
<td>-0.20</td>
</tr>
<tr>
<td>7-E-DAiG:C</td>
<td>tWW</td>
<td>-35.20</td>
<td>3.22</td>
<td>-31.98</td>
<td>-1.08</td>
<td>-14.21</td>
<td>-0.44</td>
</tr>
<tr>
<td>7-T-DAiG:C</td>
<td>tWW</td>
<td>-35.86</td>
<td>3.49</td>
<td>-32.37</td>
<td>-1.47</td>
<td>-14.28</td>
<td>-0.51</td>
</tr>
<tr>
<td>7-F-DAiG:C</td>
<td>tWW</td>
<td>-35.57</td>
<td>3.16</td>
<td>-32.40</td>
<td>-1.50</td>
<td>-14.18</td>
<td>-0.41</td>
</tr>
<tr>
<td>7-Cl-DAiG:C</td>
<td>tWW</td>
<td>-35.43</td>
<td>3.16</td>
<td>-32.27</td>
<td>-1.37</td>
<td>-14.26</td>
<td>-0.49</td>
</tr>
<tr>
<td>7-Br-DAiG:C</td>
<td>tWW</td>
<td>-35.39</td>
<td>3.20</td>
<td>-32.18</td>
<td>-1.28</td>
<td>-14.30</td>
<td>-0.53</td>
</tr>
<tr>
<td>7-NH2-DAiG:C</td>
<td>tWW</td>
<td>-35.13</td>
<td>3.01</td>
<td>-32.11</td>
<td>-1.21</td>
<td>-14.40</td>
<td>0.63</td>
</tr>
<tr>
<td>7-NH2-DAiG:C</td>
<td>tWW</td>
<td>-34.64</td>
<td>3.47</td>
<td>-31.17</td>
<td>-0.27</td>
<td>-13.83</td>
<td>-0.06</td>
</tr>
</tbody>
</table>

**Impact of the translocation of the N7 and C8 atoms in the heterocyclic ring of DAG and DAiG.**

The geometry of the optimized DAG:C cWW pair is almost identical to the corresponding unmodified G:C cWW pair, with differences in H-bond lengths within 0.01 Å. The interaction of the
modified DAG:C cWW pair is comparable to the G:C cWW pair, with $E_{\text{Mod}}$ values of 0.32 kcal/mol and 0.03 kcal/mol in the gas phase and in water, respectively.

As for the tWW optimized geometries of the DAG:iC and DAiG:C pairs, they are also geometrically very similar to the reference base pairs (G:iC and iG:C tWW, respectively), both in gas and in water, with differences in H-bond lengths within 0.02 Å. However, while the DAG:iC tWW pair is slightly less stable than the corresponding unmodified pair both in gas phase and in water, with $E_{\text{Mod}}$ of 0.45 and 0.07 kcal/mol, the DAiG:C tWW base pair is moderately more stable with $E_{\text{Mod}}$ of -1.44 and -0.20 kcal/mol in gas phase and in water, respectively.

Figure 2. Stick representation of the base pairs including a modified 7-deaza-8-azaguanine H-bonded to cytosine in the cWW geometry. Values in parentheses correspond to the optimized distances in water and values without parentheses correspond to optimized distances in the gas phase. All distances are in Å.
Figure 3. Stick representation of the base pairs including a modified 7-deaza-8-azaguanine H-bonded to isocytosine in the tWW geometry. The values in parentheses correspond to the optimized distances in water and values without parentheses correspond to optimized distances in the gas phase. All distances are in Å.

**Canonical Watson-Crick base pairs: Impact of different substituents on the C7 atom of DAG.**

We have modeled 7 different functional groups (other than hydrogen), with varying electronic properties, at the C7 atom of the DAG base with the modifications starting from ethynyl, resulting in 7-ethynyl-7-deaza-8-azaguanine (7-E-DAG) and triazole, resulting in 1,2,3 trizole-7-deaza-8-azaaguanine (7-T-DAG) bases. The optimized geometries of 7-E-DAG:C cWW and 7-T-DAG:C cWW pairs are virtually unaffected by the modification (compared to the unmodified G:C cWW pair). As for the interaction energies, for 7-E-DAG:C, it is comparable to that of the unmodified G:C pair (E_{Mod} of -0.19/-0.17 kcal/mol in gas phase/water). For the 7-T-DAG:C cWW pair, instead, the E_{Mod} of +0.87 kcal/mol in gas phase indicates a moderate destabilization as compared to the unmodified G:C cWW pair. This mild destabilization can be explained with a small reduction in the electron density around the O6 atom (indicated by blue curves), making it a poorer H-bonding acceptor (see Figure 5).
Figure 4. Stick representation of the base pairs including a modified 7-deaza-8-azaisoguanine (DAiG) H-bonded to cytosine in the tWW geometry. The values in parentheses correspond to the optimized distances in water and values without parentheses correspond to optimized distances in the gas phase. All distances are in Å.

Intrigued by the above results, we decided to stress more the basic DAG:C skeleton by considering the effect of a strong electron withdrawing group, –NO$_2$, and of a strong electron donating group, –NH$_2$, on the C7 atom of DAG. Finally, to have a more comprehensive picture, we functionalized the C7 atom of the DAG base with different halogen substituents, –F, –Cl and –Br.

Focusing on the optimized geometry of the base pairs, we again found very small deviations: H-bond lengths within 0.03 Å as compared to the G:C base pair, with all the introduced functional groups (–NO$_2$, –NH$_2$, –F, –Cl, –Br). Moving to the gas phase energies, a small stabilizing effect with E$_{Mod}$ within -0.40 kcal/mol was observed for the base pairs presenting a halogen at the C7 position of the DAG base. In contrast, a high destabilization, with E$_{Mod}$ of +2.35 kcal/mol, is calculated for –NH$_2$, which can be correlated with the decreased electronic density around the O6 atom, making it a poorer H-bond acceptor, along with the decreased electronic density around the N1 atom of DAG, making it a weaker H-bond donor. Both the effects contribute to the lower stability of the 7-NH$_2$-
DAG:C cWW pair compared to the unmodified G:C pair. A substantial stabilization, with an \( E_{\text{Mod}} \) of -1.14 in gas phase, is instead predicted for the NO\(_2\) substituent. It can also be related to changes in the base electron density, specifically to the increased density around the O6 and N1/N2 atoms of 7-NO\(_2\)-DAG making them better H-bond acceptor and donors, respectively, as compared to the unmodified G base (Figure 5).

**Figure 5.** Electron density difference, in the base plane, between DAG bases presenting different substituents on the C7 atom, and the G base. Density difference curves are plotted between -0.02 and 0.02 a.u., with a spacing of 0.001 a.u. Blue (red) lines refer to negative (positive) density difference curves, i.e., to areas where the C7-substituted DAG presents reduced (increased) electron density as compared to the natural G base.

The above results showed that a strong electron withdrawing group on C7, −NO\(_2\), stabilizes the DAG:C pair, whereas a strong electron donating group, −NH\(_2\), destabilizes it. On these grounds, we decided to plot the interaction energies of all the substituents under analysis *versus* their Hammett constant (\( \sigma_p \)), reflecting their electron donor/withdrawing capability, in order to determine whether an insightful correlation can be obtained. As the Hammett constant is not available for triazole, this substituent was not included in our analysis.[23] Interestingly, a remarkable linear correlation, with \( r^2=0.91 \), over the Hammett scale was found (see Figure 7), clearly indicating that the impact of the substituents at the C7 position of the DAG moiety is totally related to their electronic properties, as transmitted through the \( \sigma \)-bonds skeleton.
This analysis clearly indicates that the G:C pair is remarkably robust to sustain the skeleton 7-deaza-8-aza modification. However, depending on their electronic properties, substituents at the C7 position of the DAG heterocycle can significantly impact it.

**Figure 6.** Electron density difference, in the base plane, between DAiG bases presenting different substituents on the C7 atom, and the iG base. Density difference curves are plotted between \(-0.02\) and 0.02 a.u., with a spacing of 0.001 a.u. Blue (red) lines refer to negative (positive) density difference curves, i.e., to areas where the C7-substituted DAiG presents reduced (increased) electron density as compared to the canonical iG base.

**Reverse Watson-Crick base pairs: Impact of different substituents on the C7 atom of DAG.** In this section, we will discuss the geometric and energetic stability of the DAG:iC tWW base pair and of the 7 tWW base pairs deriving from the C7-functionalization of DAG. The G:iC tWW pair is our reference system here. Geometries of all the DAG modifications are extremely similar to that of the reference unmodified system (the G:iC tWW pair, see Figure 3), with deviations in the H-bond lengths within 0.03 Å. The effect of the 7-substituted DAG bases on the stability of the corresponding tWW pairs with iC is analogous to that observed previously for their cWW pairs with C, both in gas phase and in water. In fact, the DAG:iC tWW pair gets slightly destabilized, with an
EMod of +0.46 kcal/mol, while a remarkable destabilization is observed for 7-NH₂-DAG:iC tWW and a significant stabilization is found for 7-NO₂-DAG:iC tWW, as compared to the unmodified G:iC base pair, with an EMod of +1.62 kcal/mol, and -1.24 kcal/mol, respectively. Again a linear correlation, with r²=0.77, was found between the interaction energies and the Hammett constant (see Figure 7), suggesting that also in this case the impact of the substituent on the C7 position of the DAG moiety is related to its electronic property, as transmitted through the σ-bonds skeleton.

Figure 7. Correlation between the interaction energy of the base pairs presenting different substituents at position 7 of the heterocycle of DAG, and the Hammett constant (σP) of these substituents.

Reverse Watson-Crick base pairs: Impact of different substituents on the C7 atom of DAiG.

Considering that the iG:C tWW pair is the reference system here, we first focused on the translocation of the N7 and C8 atoms in the iG residue, resulting in DAiG, which can H-bond to C in the tWW geometry giving the DAiG:C tWW base pair (see Figure 4). Next, the 7 functional groups discussed before have been modeled at the C7 atom of the DAiG base in order to look at their impact on the modified base H-bonding with C. Looking at the optimized geometries (Figure 4), it is evident that they are well maintained with all the introduced functional groups, as compared to the reference
iG:C system, with very small deviations in the H-bond lengths, within 0.04 Å. Moving to energetics, a stabilization with an E\text{Mod} of -1.46 kcal/mol is predicted for the DAiG:C tWW pair, as compared to the reference iG:C tWW geometry calculated at RIMP2/aug-cc-pVTZ level of theory. A similar stabilization effect has been obtained with all the 6 considered bases bearing an electron-withdrawing functional group (ethynyl, triazole, -NO2, -F, -Cl, -Br), with E\text{Mod} values ranging between -1.50 and -1.08 kcal/mol. This results from an increased electron density around the N6 and O2 atoms (Figure 6). On the contrary, the only electro-donating substituent here, -NH2, while having a slight stabilizing effect as compared to the unmodified iG base, clearly destabilizes the tWW geometry as compared to the DAiG modification (E\text{Mod} of -0.27 kcal/mol). The correlation between the interaction energies of the tWW base pairs involving C7-substituted DAiG bases and Hammett constant of the substituents is poor, with r²=0.42 (not shown), most probably as a consequence of the interaction energy values being quite flat for all the electron-withdrawing functional groups. Remarkably, we observed a similar behavior for the DAA base, where the impact of substituents at the C7 position was shown to be unrelated to their electronic properties. We notice that both DAA and DAiG feature a N6 amino group, which can directly interact with the C7-substituents, as indicated by the modification in the electron density around it (Figure 6).[14]

Conclusions

In the present work, we have examined the impact that a series of non-natural modifications have on the geometry and stability of the G:C cWW base pair and of the G:iC and iG:C tWW base pairs. We started from the 8-aza-7-deaza modification of the G and iG skeleton, resulting in DAG and DAiG, respectively, and then modeled 7 different substituents on the C7 atom of the modified DAG and DAiG moieties. Our calculations clearly indicate that the 7-deaza-8-aza modifications and all the considered 7-substituents have a negligible impact on the geometry of the above base pairs, with H-bond distances basically matching those of the unmodified base pairs. As for the base pairs interaction energy, it is instead dependent on the physicochemical features of the substituents. As a general trend, all the electron withdrawing 7-substituents tend to stabilize the above base pairs. On the contrary, the only strong electron donating substituent we simulated, -NH2, has a clear destabilizing effect, both as compared to DAG/DAiG and to G/iG, with the only exception of the iG:C tWW pair, where it has a slight stabilizing effect as compared to iG. These findings are in
general agreement with thermodynamic data experimentally available for some of the considered modifications.\textsuperscript{11e, g}

The observed effect of the above modifications on the corresponding base pairs can be explained with the enhanced/reduced H-bond donor/acceptor capabilities of atoms involved in base-base H-bonds, due to the perturbed electron density distribution, as shown by the electron density maps we report. Remarkably, for the 7-substituted DAG:C cWW and DAG:iC tWW systems, we could also point out a clear linear correlation between the base pair interaction energy and the Hammett constant of the 7-substituent on DAG, reflecting its electron-donating/withdrawing properties. Therefore, while all the explored modifications are expected to be accommodated in both parallel and antiparallel nucleic acid duplexes without perturbing their geometry, the strength of the corresponding base pair (and duplex) can in principle be tuned by incorporating different substituents with varying electronic properties at the C7 position, where the higher the electron-withdrawing character, the higher the stability. This can be of considerable interest, as modulating nucleic acids stability while maintaining their structure and fidelity in biological processes is one of the scopes of chemical labeling\textsuperscript{29}, besides being at the basis of several biotechnological and biomedical processes\textsuperscript{30}, including the optimization of RNA-based drugs\textsuperscript{31}.

**Supporting Information.**  
Details of Quantum chemical calculations, electron density analysis and the electron density difference maps for the base pairs and the Cartesian coordinates of all the structures discussed in this work.

**Acknowledgments.**  
LC thanks the King Abdullah University of Science and Technology (KAUST) for financial support. For computer time, this research used the resources of the Supercomputing Laboratory at KAUST.

**References:**  


Material for the Table of Contents

versus..