

# Different Conformations of 2'-Deoxycytidine in the Gas and Solid Phases: Competition between Intra- and Intermolecular Hydrogen Bonds

*Sanliang Ling<sup>a, b</sup> and Maciej Gutowski<sup>a</sup>*

<sup>a</sup>) Institute of Chemical Sciences, School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom

<sup>b</sup>) Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, United Kingdom

\* E-mail: [s.ling@ucl.ac.uk](mailto:s.ling@ucl.ac.uk); [m.gutowski@hw.ac.uk](mailto:m.gutowski@hw.ac.uk)

## ABSTRACT

Computational results have been reported for 2'-deoxycytidine (dC), its gas phase isomers, tautomers, and their conformers, as well as for the crystalline phase. In addition to the neutral gas phase molecules we have also considered associated radical anions and cations. The structural calculations were performed at the density functional and MP2 levels of theory. Vertical electron ionization energies and excess electron binding energies were determined using electron propagator theory. The  $\alpha$ -anomer proved to be more stable by a fraction of kcal/mol than the biologically relevant canonical  $\beta$ -anomer. The conformational space of canonical dC has been systematically probed. dC in the crystalline phase or DNA structures favours canonical *anti* conformations. These structures were used in past computational studies to model gas phase characteristics of dC. Our findings indicate, however, that the gas phase dC favours *syn* conformations. It has repercussions for earlier interpretations of gas phase experimental results based on these computational results. The thermodynamic dominance of *syn* conformations results from the formation of an intramolecular O5'-H13 $\cdots$ O2 hydrogen bond. The IR spectra of the most stable *syn* and *anti* canonical conformers differ markedly in the region of frequencies corresponding to NH/OH stretching modes. The MP2 value of deprotonation enthalpy of dC of 1411.7 kJ/mol is in very good agreement with the experimental value of 1409 $\pm$ 2.5 kJ/mol. The most stable valence anions are characterized by electron vertical detachment energies (VDE) in the 0.8-1.0 eV range, in good agreement with the experimental VDE of 0.87 eV. The barrier for the glycosidic bond cleavage is significant in the neutral canonical dC, 40.0 kcal/mol, and it is reduced to 22 and 16 kcal/mol for the anionic and cationic radicals of dC, respectively. The cleavage reaction is exothermic by 4 kcal/mol for dC $^-$  and endothermic by 7 and 9 kcal/mol for dC $^+$  and dC, respectively. We decomposed the crystal cohesive energy into repulsive one-body terms associated with the *syn-anti* conformational changes, and the attractive intermolecular interaction term. We exposed that the *syn-anti* conformational changes are very favourable for intermolecular interactions; in particular they make the imino-amino side of the cytosine residue accessible to intermolecular interactions.

## 1. Introduction

Recent years brought significant advances in transferring of thermally labile and nonvolatile biomolecules into the gas phase<sup>1-7</sup> and precise spectroscopic interrogations of these important molecular systems became possible.<sup>8-11</sup> In order to interpret the gas phase spectroscopic data, experimentalists use computational results available in the literature, perform calculations by themselves, or engage computational colleagues in fruitful collaborations. At first glance the computational task is straightforward because density functional theory (DFT) or second-order Møller-Plesset (MP2) calculations are nowadays routinely executed for molecules containing several tens of atoms. There is, however, an intriguing caveat: which structure of the molecule should be used in these calculations? Biomolecules typically support several plausible conformers in addition to chemically better distinguished isomers or tautomers. Computational chemists frequently assume an initial structure extracted from X-ray crystal experiments and apply standard optimization methods to an isolated molecule.<sup>12-14</sup> However, the most common optimization algorithms, which are driven by calculated forces, converge to the closest local minimum rather than to the most stable structure. More sophisticated algorithms, such as basin-hopping,<sup>15</sup> simulated annealing,<sup>16</sup> genetic algorithms,<sup>17</sup> or systematic searches<sup>18</sup> are *not* routinely used. As a result, computational results might be obtained for chemical structures that are not populated in gas phase experimental conditions and this problem is illustrated in the current report.

Here we discuss the case of 2'-deoxycytidine (dC), for which the X-ray crystal structure is known,<sup>19</sup> and experimentally determined gas phase acidity<sup>20</sup> and excess electron binding energy<sup>1</sup> were interpreted in terms of computational results. The molecular structures used in these computational studies were biased by X-ray structural prediction<sup>21-22</sup> or by the structure of DNA.<sup>12-14, 23-24</sup> Our current results unravel that the most stable structure of the neutral gas phase dC differs qualitatively. The reason is that its solid state structure is dictated by a competition between *intra-* and *intermolecular* hydrogen bonds, while the most stable gas phase structure is dictated by an *intramolecular* hydrogen bond. This illustrates that computational results for gas phase molecules based on structures extracted from condense phases might

be misleading and agreement with experiment, or lack thereof, might be fortuitous. The current study was facilitated by our combinatorial/computational tool for systematic screening of conformers (SSC).<sup>25</sup>

For an isolated dC molecule, the structural flexibility could be reflected by various isomers, tautomers, and rotational degrees of freedom, see Figs. 1 and 2. Some rotational degrees of freedom are labelled in Fig. 1 with lower case Greek letters and describe rotations around the C5'-O5', C1'-N1, C3'-O3', and C4'-C5' bonds. Fig. 2 illustrates a terminology which is widely accepted in biochemical community. The orientation of the O5'-H13 group, which is characterized by  $\beta$ , determines whether dC is prearranged to form intra- or intermolecular hydrogen bonds. There are three major  $\beta$  orientations observed in DNA, and they are *gauche*<sup>+</sup>, *gauche*<sup>-</sup> and *trans*, see Fig. 2a. The orientation of the base with respect to the sugar is characterized by  $\chi$ , with two major conformations being *syn* (Fig. 2b) and *anti* (Fig. 2c). The non-planarity of the sugar ring typically takes two forms: C2'-*endo* (Fig. 2d) and C3'-*endo* (Fig. 2e).

In addition to conformational degrees of freedom we consider  $\alpha$ - and  $\beta$ -anomers (see Fig. 2a). The  $\beta$ -anomer of dC, in which the base and the C4'-C5' bond are on the same side of the sugar ring, is common in DNA and for this reason is routinely used in computational studies. Much less is known about the  $\alpha$ -anomer of dC, in which the base and the C4'-C5' bond are on opposite sides of the sugar ring. It was reported that the  $\alpha$ -anomer of dC could also be obtained during direct synthesis of the  $\beta$ -anomer of dC,<sup>26</sup> and recently a method was developed to separate these two isomers.<sup>27</sup> Finally, we considered dC molecules in which cytosine takes the non-canonical 3-imino tautomeric form (see the bottom of Fig. 3). Xia et al. found in their computational density functional theory study that 3-imino dC is less stable than the canonical dC by 2.3 kcal/mol, and they found a transition barrier as high as 39.8 kcal/mol.<sup>28</sup>

Foloppe et al. performed a series of theoretical studies on conformational properties of nucleosides at the MP2 and MP3/6-31G\* levels.<sup>12-14</sup> For the *gauche*<sup>+</sup> conformations of dC, the *anti* conformer was found more stable than the *syn* conformer by more than 7 kcal/mol. The C2'-*endo* or C3'-*endo* sugar puckers differed in stability by less than 1 kcal/mol<sup>12</sup> and a barrier from C3'-*endo/anti* to C2'-*endo/anti* was 3.3 kcal/mol.<sup>14</sup> For the C2'-*endo/gauche*<sup>+</sup> conformations, these authors found a barrier from *anti* to *syn* of 8.4

kcal/mol.<sup>12</sup> A characteristic feature of these computational studies is that the  $\beta$  dihedral angle has been fixed and the O5'H13 group was oriented away from the base.<sup>12</sup> This constraint was dictated by the structure of DNA, in which the O5' atom is engaged in the sugar-phosphate bond and the H13 atom is missing. Foloppe et al. explored variations of energy with respect to only these geometric degrees of freedom that are available when nucleosides are embedded in DNA. Their valuable results might, however, be insufficient to predict structural properties of nucleosides in the gas phase. Indeed, it was found in two other theoretical studies of free canonical dC,<sup>29-30</sup> in which different combinations of *syn/anti*, C2'-*endo*/C3'-*endo*, and *gauche*<sup>+</sup>/*gauche*<sup>-</sup>/*trans* conformations were considered without further geometrical constraints, that the *syn* rather than *anti* conformation is the most stable.

Applying an extended kinetic method, Kumari et al. measured gas phase acidities of four deoxyribonucleosides and the value reported for dC was 1409 kJ/mol.<sup>20</sup> The gas phase acidity of dC is defined as an enthalpy change for the reaction



These authors also performed calculations at the B3LYP/6-311+G\*\*//B3LYP/6-311G\*\* level based on a C3'-*endo/anti* conformation of dC. They found that the H2 site of cytosine is the most susceptible to deprotonation and they reported a gas phase acidity of 1409 kJ/mol, in perfect agreement with experiment. However, the agreement might be fortuitous in view of the limited probing of the conformational space of dC, inaccuracy of the B3LYP functional, and computational inconsistencies (see Section 3.2).

The experimental reports<sup>31-33</sup> from the group of Sanche on single and double strand breaks in DNA induced by low-energy electrons triggered numerous experimental<sup>34</sup> and computational<sup>35</sup> studies on negatively charged biomolecules. The group of Bowen performed pioneering photoelectron spectroscopy (PES) experiments on the gas phase anions of nucleic acid bases,<sup>36-37</sup> nucleosides,<sup>1</sup> and nucleotides.<sup>38</sup> The measured electron vertical detachment energy (VDE) for dC<sup>-</sup> of 0.87 eV<sup>1</sup> was interpreted in terms of preceding computational results obtained in the group of Schaefer.<sup>23</sup> This group performed extensive DFT calculations on anionic nucleosides<sup>23</sup> and the role of low-energy electrons in the glycosidic bond

cleavage.<sup>39</sup> Similarly to Foloppe et al., they also selected conformations of isolated nucleosides as they appeared in X-ray crystal structures of DNA fragments. They characterized a valence bound anion of dC and determined its VDE to be 0.72 eV at the B3LYP/DZP++ level.<sup>23</sup> In view of the fact that B3LYP tends to overestimate the VDEs of nuclei acid bases,<sup>37, 40-41</sup> the structures probed in the PES experiment might be different from the structures used in the calculations.

The glycosidic bond cleavage in anions of dC was studied computationally by several groups.<sup>24, 39</sup> The group of Schaefer determined an activation barrier of 21.6 kcal/mol<sup>39</sup> and a similar result of 22.7 kcal/mol was obtained by Li et al. at the B3LYP/6-31+G\* level.<sup>24</sup> Here we extend these studies by dealing with more stable conformers of dC<sup>-</sup>. We also study the bond cleavage in dC<sup>+</sup>, which could also be generated upon the interaction of DNA with high energy radiation.

The structure of the paper is the following. Our computational approach is discussed in Section 2. The neutral and radical ions of molecular dC, their properties, and energetics of the glycosidic bond cleavage are presented in Section 3.1-3.4. The discussion of crystal structure of dC and the crystal cohesive energy are presented in Section 3.5, followed by a Summary.

## 2. Methods

A desirable approach would be to explore the conformational/isomeric/tautomeric space of a nucleoside without any constraint, though the problem immediately becomes computationally very intensive. It has recently been suggested that combinatorial/computational approaches, which hinge on efficient automation of repetitive computational steps, might be useful when solving structural problems.<sup>42-45</sup> In the current study we rely on the Systematic Screening of Conformers (SSC) tool,<sup>25</sup> which we use to create an initial library of conformers of canonical ( $\beta$ -anomer) dC and we perform a limited exploration of the  $\alpha$ -anomer and the 3-imino tautomer of  $\beta$ -anomer.

When building the library of conformers, we assumed a planar structure of sugar and we performed initial optimizations for 324 initial structures with the B3LYP hybrid functional<sup>46-47</sup> and 6-31G\* basis set.<sup>48</sup> More details on how we used our SSC tool to search for the global minimum of dC can be found in

Ref. 25. These optimizations led to 65 non-planar minima which were further optimized at the B3LYP/6-31++G\*\* level. Among the 28 most stable structures (with a relative energy cutoff of 4 kcal/mol), 16 are C2'-endo and 12 are C3'-endo. For each of them we created a counterpart with the opposite sugar ring puckering and we optimized these new 28 structures at the same level of theory. This approach led to 12 new conformers. Finally, the 39 most stable conformers are further refined at the MP2 level with aug-cc-pVDZ (AVDZ)<sup>49</sup> basis sets to account for the intramolecular dispersion interaction. Valence-bound anionic states of the canonical dC, the  $\alpha$ -anomer (IS), and the 3-imino tautomer (TAU) (see Fig. 3) were studied at the B3LYP/6-31++G\*\* level of theory. The most thorough search was performed for the canonical  $\beta$ -anomer, for which the 28 most stable neutral structures were accepted as the initial anionic structures.

The enthalpies and Gibbs free energies were determined in the ideal gas model based on the most accurate electronic energies, zero-point vibrational corrections and thermal contributions determined in the rigid rotor/harmonic oscillator approximation at T = 298 K and p = 1 atm.

The vertical ionization energies (VIE's) and excess electron VDE's were calculated in two ways: (i) "indirectly" by subtracting the energies of the radical ion and the neutral, and (ii) "directly" using electron propagator theory (EPT) in the P3 approximation,<sup>50</sup> as implemented in Gaussian 03.<sup>51</sup> When calculating VIE's we used the 6-31++G\*\* basis set; for VDE's we used both 6-31++G\*\* and 6-311++G\*\* basis sets recognizing that anionic charge distributions might require a more flexible basis set. The VDE's of valence anions at the EPT/P3 level were calculated as ionisation potentials of anions.

The adiabatic electron affinity (EA) values were determined at the B3LYP/6-31++G\*\* level and include zero-point vibrational corrections. For all conformers of canonical dC<sup>-</sup> we assume that the neutral formed in the process of electron photodetachment can relax to the most stable canonical structure. In other words, the barriers separating various conformers of the neutral canonical dC are assumed to be low. For this reason we referenced the reported EA values with respect to the neutral conformer N2 (see Fig. 3), which is the most stable at the B3LYP/6-31++G\*\* level. For the IS and TAU anions, the EAs are

determined with respect to N2, but also with respect to the “daughter” neutrals as the barriers for interconversions are typically larger for isomers than for conformers.

It is well established that neutral molecules with dipole moments larger than 2.5 D can support dipole-bound anions.<sup>52</sup> The five most stable conformers of neutral canonical dC have dipole moments larger than 6 D. In addition, the dipole moment of IS is as large as 8.5 D. Thus we studied dipole-bound anionic states of dC for two selected structures. One is based on the conformer N2<sup>53</sup> with the B3LYP/6-31++G\*\* dipole moment of 7.1 D, the other is IS. Based on previous experience,<sup>54</sup> we supplemented the original AVDZ basis set with additional basis functions with small exponents to accommodate the diffuse charge distribution of the excess electron. The orientation of the dipole roughly coincides with the O2-C2 bond of cytosine. Thus the five-term *s* and five-term *p* sets of diffuse functions, with a geometric progression ratio of 5.0, were centred at the C5 atom. The B3LYP functional overestimates the VDE values for dipole-bound states.<sup>55</sup> Thus, the geometry optimizations were performed at the MP2 level, which takes into account dispersion-type interaction between the excess electron and electrons of the neutral molecule.<sup>56</sup> In addition, the VDE values were determined at the EPT/P3 level and were calculated as electron affinities of the neutrals. The reported MP2 values of EA for the dipole-bound states do not include zero-point vibrational corrections.

The singly occupied molecular orbitals (SOMOs) of valence and dipole-bound anions are drawn with GaussView. For valence anions, we have used an isovalue of  $0.02 a_0^{-3/2}$ . For dipole-bound anions, we have used the isovalues that reproduce 80% of the excess electron total charge, and the OpenCubMan tool has been used to determine the isovalues.<sup>57</sup>

The glycosidic bond cleavage has been studied for the neutral, anionic and cationic dC. An estimation of the transition state structure was obtained by producing an energy profile for the C1'-N1 bond elongation with a step of 0.1 Å ( $1.47 \text{ \AA} < R_{C1'-N1} < 3.47 \text{ \AA}$ ) through the B3LYP/6-31++G\*\* partial geometry optimizations. The geometries of the neutral and ionic N2 conformers were used as starting points of the energy profiles. We selected the highest energy point on the potential energy profile as an initial guess of the transition state structure, and we performed full transition state geometry optimizations



using the Berny algorithm.<sup>58</sup> We have verified that the resulting stationary points are characterized by only one imaginary vibrational frequency and intrinsic reaction coordinate (IRC)<sup>59</sup> calculations have been performed to confirm the nature of reactants and products.

We have also performed calculations for the molecular crystal of dC to clarify the competition between the intra- and intermolecular hydrogen bonds. The experimental unit cell contains two dC molecules, both in the *anti* conformation.<sup>19</sup> Periodic DFT calculations have been performed using the CP2K code,<sup>60-61</sup> which uses a mixed Gaussian/plane-wave basis set. We employed the double- $\zeta$  polarization quality Gaussian basis sets and a plane-wave cutoff of 600 Ry for the auxiliary grid, in conjunction with the Goedecker-Teter-Hutter pseudopotentials.<sup>62-63</sup> The Perdew-Burke-Ernzerhof (PBE) functional,<sup>64</sup> together with Grimme's D3 van der Waals correction,<sup>65</sup> including the Axilrod-Teller-Muto three-body terms, were used for all total energy calculations and structural optimizations, including lattice parameters and atomic positions. The same settings were used for calculations of dC monomers, which were placed in a  $30 \times 30 \times 30 \text{ \AA}^3$  cubic box, in order to eliminate inter-molecular interactions. The cohesive energy of the molecular crystal of dC was calculated using the N2 conformer (most stable conformer at DFT level) as a reference. Calculation for the crystal was performed with a  $2 \times 2 \times 1$  expansion of the unit cell and sampling of the Brillouin zone limited to the Gamma point. We have used the experimentally observed crystal structure<sup>19</sup> as an initial guess. All calculations were performed at 0 K, i.e. excluding any zero-point or thermal corrections to the energetics and structures.

NWChem 5.0<sup>66</sup> and Gaussian 03<sup>51</sup> were used for molecular and CP2K<sup>60-61</sup> for crystal calculations. Molden<sup>67</sup> and GaussView<sup>68</sup> were used molecular visualizations.

### 3. Results

**3.1. Structures of Neutral dC.** The geometries and relative energies of selected structures of neutral dC are shown in Fig. 3 and Table 1, respectively. The 7 most stable conformers of canonical  $\beta$ -anomer are named according to their relative stability at the MP2 level, e.g., the conformer N2 is the second most stable. For the sake of future discussion we also include: (i) the N12 conformer, which is related to N2

through a *syn-to-anti* transition, (ii) N25, which was extracted from the DNA structure and which was used in previous theoretical studies,<sup>23, 39</sup> (iii) an  $\alpha$ -anomer (IS), (iv) and a 3-imino tautomer (TAU). The main trends in relative stabilities are similar between the MP2 and B3LYP energies, though the exact orderings of conformers differ. The zero-point vibrational and thermal corrections to relative stabilities should definitely be taken into account when considering populations of various conformers in the gas phase. A higher than MP2 level of theory might be required for ultimate predictions.

There are, however, qualitative predictions that are not sensitive to differences between the MP2 and B3LYP methods. First we focus attention on the canonical dC, N1-N7, N12 and N25. The most stable conformers are *syn* rather than *anti*, see Fig. 3 and Table 1. Indeed, the N1-N5 structures are *gauche*<sup>+</sup>/*syn* and they are stabilized by intramolecular hydrogen bonds involving the O5'-H13 proton donor and the O2 proton acceptor. The intramolecular hydrogen bond is weaker in N6 due to the *trans/syn* conformation. The most stable *anti* canonical conformer proves to be N7, which is less stable than N1 by 2-3 kcal/mol. It implies that *anti* conformers might be irrelevant in gas phase conditions. Two other studies<sup>29-30</sup> also predicted that the *syn* conformations might be more stable than the *anti* conformations, but their difference in stability was reduced from 3.4 to 1.0 kcal/mol upon the methodological improvement from B3LYP to MP2, both with the 6-31G\* basis sets.<sup>30</sup>

Another important observation, which applies to the neutral canonical dC and is not sensitive to methodological differences, is that the conformers with the C2'-*endo* sugar puckering should dominate in the gas phase. Indeed, among the seven most stable canonical conformers reported here, four of them (N1, N2, N4 and N6) are C2'-*endo* and the most stable C3'-*endo*, N3, is less stable than N1 by 0.8-2.0 kcal/mol. If one considers the *anti* conformers only then an opposite conclusion emerges: N7 and N25 are C3'-*endo*, in good agreement with the past computational results.<sup>14, 29</sup>

We have demonstrated so far that the C2'-*endo/syn* conformers are dominant in the gas phase and we discussed their stabilities with respect to the C3'-*endo/anti* conformers. Here we discuss the height of barriers for the *syn-to-anti* and C2'-*endo-to-C3'-endo* transitions. We identified two transition state structures illustrated in Fig. 4. The first is denoted as TS2-5 and it connects the N2 and N5 conformers

through a *C2'-endo-to-C3'-endo* transition. The second is denoted as TS2-12 and it connects the N2 and N12 conformers through a *syn-to-anti* transition. The MP2 and B3LYP values of barrier heights are also reported in Fig. 4.

The MP2 barrier for the *syn-to-anti* transition is 8.2 kcal/mol. The barrier heights reported by Foloppe et al.<sup>12</sup> are quite different, which results from geometrical constraints imposed on the orientation of the O5'-H13 group. Three possible pathways for the *C2'-endo-to-C3'-endo* transition were proposed with the one through the O4'-endo conformation having the lowest barrier.<sup>14</sup> Indeed, our TS2-5 is O4'-endo, see Fig. 4, and the MP2 barrier is 3.4 kcal/mol. We conclude that the barriers for the *C2'-endo-to-C3'-endo* transitions are much smaller than for the *syn-to-anti* transitions. The puckering of the sugar ring should be viewed as a dynamical process which favours the *C2'-endo* orientations in the gas phase at standard conditions. The small barrier is reflected by a small imaginary frequency at TS2-5, which is only 49i cm<sup>-1</sup>. Note that the B3LYP barriers are systematically lower than the MP2 barriers, see Fig. 4.

The  $\alpha$ -anomer (IS) is more stable than the canonical forms by a fraction of kcal/mol, see Table 1. Unlike the most stable canonical conformers, which are either *C2'-endo* or *C3'-endo*, IS is O4'-endo. Its stability can be justified only partially by a hydrogen bond between the O3'-H8 proton donor and the O2 proton acceptor. The actual stability of  $\alpha$ -anomers could be even higher, as our conformational search was more restricted than for the  $\beta$ -anomer. Depending on the source conditions,  $\alpha$ -anomers could be formed upon transferring canonical dC to the gas phase.

The 3-imino tautomer (TAU in Table 1 and Fig. 3), is *syn* and it is stabilized by an intramolecular O5'-H13...O2 hydrogen bond. However, it is less stable than N1 by 2-4 kcal/mol. This tautomer was also found less stable by Xia et al.,<sup>28</sup> though these authors considered *anti* conformations only. The actual stability of 3-imino tautomer could be higher, as our conformational search was more restricted than for the canonical amino. A significant barrier of 39.8 kcal/mol for the unimolecular amino-imino tautomerization of dC was found by Xia et al.<sup>28</sup> This barrier could be significantly lowered in binary events involving an intermolecular proton transfer.

**3.2. Properties of Neutral dC.** Our results indicate that *syn* conformers of canonical dC dominate in the gas phase rather than *anti* conformers identified in X-ray structural studies. We would welcome an experimental confirmation of our predictions. Such a confirmation could be based on differences in the infrared (IR) spectra of the most stable *syn* and *anti* conformers, i.e., N1 and N7, respectively. N1 is characterized by a strong intramolecular hydrogen bond between O5'-H13 and O2, see Fig. 3. Inasmuch as both conformers provide similar spectra in the low frequency (0~2000  $\text{cm}^{-1}$ ) region, see Fig. 5a, they differ markedly in the high frequency (2800~4000  $\text{cm}^{-1}$ ) region, see Fig. 5b. In the latter region the IR intensities of different peaks are quite similar for N7, whereas there is a very strong peak for N1 with the calculated IR intensity is 563  $\text{km/mol}$ . This peak is five times stronger than other peaks in this region. The strong peak results from a superposition of two transitions with similar vibrational frequencies. The first transition is associated with the symmetric stretching vibration of the amino group, and the other is the stretching vibration of O5'-H13. The latter vibration is red-shifted in N1 by about 250  $\text{cm}^{-1}$  in comparison with N7, in which there is no intramolecular hydrogen bond between O5'-H13 and O2. We suggest that a strong IR peak in the 3600  $\text{cm}^{-1}$  region can be regarded as a fingerprint of the *syn* conformer.

Kumari et al.<sup>20</sup> reported an experimental deprotonation enthalpy of dC of  $1409 \pm 2.5$   $\text{kJ/mol}$ . Their parallel DFT study, focused on the N7 conformer, unravelled that the H2 site of cytosine is the most susceptible to deprotonation, and led to a remarkably accurate computational value of deprotonation enthalpy of 1409  $\text{kJ/mol}$ . In view of the fact that N7 is unstable with respect to N1 by 1.8-3.2  $\text{kcal/mol}$  (see Table 1) we readdressed the problem and our deprotonation enthalpies obtained for the most stable *syn* and *anti* conformers, i.e., N1 and N7, respectively, are summarized in Table 2. The H2 site of the N1 conformer proves to be the most acidic among the four sites considered and the MP2/AVDZ value of deprotonation enthalpy of 1411.7  $\text{kJ/mol}$  is in excellent agreement with the experimental value, while the B3LYP results are systematically larger by 10-20  $\text{kJ/mol}$ . Our results also indicate that the earlier computational results were flawed<sup>69</sup> and the corrected B3LYP deprotonation enthalpy of the H2 site of N7 is 1445.8 and 1443.7  $\text{kJ/mol}$  with the 6-311G\*\* and 6-31++G\*\* basis sets, respectively. Not only N1

is more stable than N7 but also the [dC-H]<sup>-</sup> product is more stable for N1 by 21 kJ/mol at the MP2/AVDZ level.

The values of VIE, calculated directly in the P3/6-31++G\*\* approximation, are reported in Table 1 for the selected structures of dC. The VIE values for the most stable *syn* conformers cluster in a narrow range of 8.75-8.92 eV and are not sensitive to the puckering of the sugar ring. The VIE value for the most stable *anti* conformer (N7) is, however, significantly smaller and amounts to 8.24 eV. The VIE has not been experimentally measured yet, but our results suggest that its value can be used to discriminate the *anti* and *syn* conformers in the gas phase. The experimental VIE for deoxyribose and cytosine are 10.51 eV<sup>70</sup> and 8.89 eV,<sup>71</sup> respectively. Note that due to the existence of various tautomers in the gas phase,<sup>71</sup> the measured VIE values might not be related exclusively to the canonical tautomer. The computed VIE value for dC of 8.79 eV (conformer N1) is closer to the experimental VIE of cytosine, which implies that the HOMO orbital of dC is localized on cytosine. Indeed, the plots of the Hartree-Fock HOMOs of N1 and N7, see Fig. 6, demonstrate that these  $\pi$  orbitals are localized primarily on the base. We suggest that the larger value of VIE for the *syn* conformers results from a hydrogen bond between O5'-H13 and O2. The protic H13 stabilizes the HOMO orbital of *syn* conformers, which results in their larger VIE values in comparison with the most stable *anti* conformer, N7. This orbital picture is confirmed by a difference in the VIE values determined at the Koopmans' theorem level of 0.6 eV, which is only slightly modified by orbital relaxation and electron correlation effects. Finally, the  $\alpha$ -anomer (IS) is characterized by the largest value of VIE, 8.92 eV.

**3.3. Anions.** The characteristics of valence anions of dC are summarized in Table 3. In addition to the five most stable anions of canonical dC, we consider valence anions of: (i) N25, an *anti* conformer considered in a previous computational study,<sup>23</sup> (ii) the  $\alpha$ -anomer, and (iii) the 3-imino tautomer. The names of valence anions are preceded by the letter "A" and are followed by the name of the neutral the valence anion is related to, e.g., AN4 indicates a valence anion related to the fourth most stable canonical dC, N4. The reported B3LYP/6-31++G\*\* values of EA were calculated with respect to N2, the global minimum of canonical dC at the same level. The valence anions are ordered in Table 3 according to the

descending values of EA. The reported VDE values were calculated indirectly at the B3LYP/6-31++G\*\* level, and directly in the P3/6-311++G\*\* approximation. The plots of SOMOs for AN4 and AIS are reported in Fig. 7 and demonstrate that these are indeed valence anions with the excess electron occupying a  $\pi^*$  orbital localized on the base.

Among the canonical valence anions, there are four conformers with the EA values in the 0.30-0.35 eV range and the VDE values spanning from 0.74 to 0.94 eV. The calculated VDE's are in good agreement with the experimental VDE of 0.87 eV.<sup>1</sup> The reported positive EA values are significant, and larger than the methodological uncertainty of B3LYP/6-31++G\*\*. Thus our results confirm that valence anions of dC are adiabatically bound.<sup>23</sup> The PES spectrum of the anion of dC<sup>1</sup> was interpreted in terms of computational results for the valence anion of N25.<sup>23</sup> These computational results preceded the PES experiment and their objective was to characterize the valence anion of dC in an environment that is characteristic for DNA rather than to predict the gas phase anion PES spectrum. The AN25 *anti* conformer is not a good model for the gas phase valence anion of dC, see Table 3. Its EA is only 0.16 eV, which implies that the equilibrium population of AN25 should be negligible at standard conditions. The P3 value of VDE for AN25 of 0.76 eV is low in comparison with the experimental value.

The dominant stability of the less studied  $\alpha$ -anomer (AIS) continues, this time in terms of the largest values of both EA and VDE. The excess electron attachment did not improve the relative stability of the 3-imino tautomer (ATAU). It is characterized by small values of EA and VDE in comparison with the 3-amino tautomers. We expect that intramolecular hydrogen bonds contribute to the stability of selected valence anions. For example, the protic hydrogens H8 and H13 in AIS and AN4, respectively, stabilize the excess electron localized on the base, see Fig. 7.

The dipole-bound anions supported by N2 and IS are labelled DAN2 (Fig. 8a) and DAIS (Fig. 8b), respectively. The N2 and IS neutrals were selected based on their significant dipole moments and overall thermodynamic stability, see Table 1. The molecular framework distorts upon the excess electron attachment and the dipole moment of the neutral increases by 0.7 and 0.5 D for N2 and IS, respectively, see the bottom of Fig. 8. The MP2(P3) values of VDE are 113(121) and 74(81) meV for DAIS and DAN2,

respectively. Thus these anions are amenable to PES characterization, providing the dipole-bound states could be formed in the anionic beam. The respective MP2 values of EA are 108 and 65 meV, thus the dipole-bound anions are globally less stable than the valence anions characterized in Table 3, but they might be the only bound anionic states in the neighbourhood of the minimum energy structures of the corresponding neutrals. The plots of SOMOs for DAN2 and DAIS were obtained with the isovalues selected to reproduce 80% of the excess charge. The SOMO of DAIS is more compact, which is consistent with the larger value of VDE.

**3.4. Effect of Excess Charge on Thermodynamics and Kinetics of Glycosidic Bond Cleavage.** The glycosidic bond cleavage is an important pathway of DNA damage induced by high energy radiation. Here we present the cleavage for the neutral, anionic, and cationic dC, all studied at the B3LYP/6-31++G\*\* level. The energetics and geometries of the reactants, transition states and products are shown in Fig. 9.

For the neutral dC, we find a significant activation barrier of 40.0 kcal/mol and the cleavage is endothermic by 9.4 kcal/mol (Fig. 9a). Initially we anticipated formation of two radical moieties. However, this cleavage is accompanied by proton transfer from C2'H to the O2 atom of cytosine. In consequence, both products of decomposition are closed-shell moieties with meaningful Lewis structures, and the sugar ring remains practically planar. Alternatively, one could interpret the reaction as a heterolytic bond cleavage followed by intermolecular proton transfer, which neutralizes the decomposition products. The IRC calculation from the reported transition state converges to the neutral conformer N5. This conformer is separated from N2 by TS2-5 with a barrier of 3.4 kcal/mol, see Fig. 4. For this reason we use the N2 conformer as a reactant for the neutral dC.

The barrier for cleavage of glycosidic bond in dC<sup>-</sup> is 22.3 kcal/mol, hence lower than in the neutral dC by 17.7 kcal/mol (Fig. 9b). This is a very significant difference brought in by an excess electron. The barrier height is similar to those reported earlier, 21.6 kcal/mol in Ref. 39 and 22.7 kcal/mol in Ref. 24, both of which were obtained at the B3LYP level for *anti* conformers of dC<sup>-</sup>. The glycosidic bond cleavage

is exothermic for dC<sup>-</sup> by 3.6 kcal/mol, which is another important difference brought in by an excess electron, see also Ref. 39.

Much less is known about the glycosidic bond cleavage in dC<sup>+</sup> (Fig. 9c). The barrier of 16.2 kcal/mol is even smaller than for dC<sup>-</sup>, but the reaction is endothermic by 6.8 kcal/mol. The IRC calculation from the reported transition state converges to the N1 conformer of dC<sup>+</sup>.

**3.5. Effect of Crystalline Environment on the Structure of dC.** The molecular crystal of 2'-deoxycytidine is characterized by a triclinic unit cell with two dC molecules<sup>19</sup> (see Fig. 10): molecule I with a C2'-*endo/anti* conformation, and molecule II with a C3'-*endo/anti* conformation. This periodic structure was optimized with the PBE+D3 functional with lattice vectors and positions of atoms in the unit cell available for relaxation. As expected, only minor relaxation took place, as the initial structure was experimental. The optimized lattice parameters,  $a = 7.28 \text{ \AA}$ ,  $b = 6.88 \text{ \AA}$ ,  $c = 11.13 \text{ \AA}$ ,  $\alpha = 105.2^\circ$ ,  $\beta = 85.2^\circ$ , and  $\gamma = 71.9^\circ$ , are in good agreement with the experimental values,  $a = 7.285 \text{ \AA}$ ,  $b = 6.866 \text{ \AA}$ ,  $c = 11.074 \text{ \AA}$ ,  $\alpha = 104.19^\circ$ ,  $\beta = 84.53^\circ$ , and  $\gamma = 72.26^\circ$ .

In order to rationalise why dC molecules favour *anti* conformations in the crystal lattice we analyse the cohesion energy  $E_{coh}$ :

$$E_{coh} = 2 \times E_{N2} - E_{cryst}, \quad (2)$$

where  $E_{N2}$  is the energy of the fully relaxed, isolated N2 and  $E_{cryst}$  is the crystal energy per unit cell.  $E_{coh}$  can be decomposed into destabilizing one-body terms  $\Delta E_I$  and  $\Delta E_{II}$ , and a stabilizing term  $E_{int}$  resulting from intermolecular interactions in the lattice:

$$E_{coh} = \Delta E_I + \Delta E_{II} + E_{int}, \quad (3)$$

where  $\Delta E_I$  is an energy increase for isolated molecule I upon transformation from the relaxed N2 to the structure of I in the lattice:

$$\Delta E_I = E_{N2} - E_I; \quad (4)$$

and an analogous definition holds for  $\Delta E_{II}$ . The interaction energy in the lattice is:



$$E_{\text{int}} = E_I + E_{II} - E_{\text{cryst}}. \quad (5)$$

Our findings are that the cohesion energy of 95.9 kcal/mol per unit cell decomposes into relatively small destabilizing one-body terms,  $\Delta E_I = -4.5$  kcal/mol and  $\Delta E_{II} = -3.3$  kcal/mol, and a strongly stabilizing  $E_{\text{int}} = 103.7$  kcal/mol. The order of magnitude for  $\Delta E_I$  and  $\Delta E_{II}$  is the same as the energies of *anti* conformers with respect to N2, see Table 1. Therefore, it is the  $E_{\text{int}}$  term which deserves further analysis.

In Fig. 11 we identified intermolecular hydrogen bonds that develop in the crystal structure. In Table 4 we summarize proton acceptors and donors that are involved in the hydrogen bonds identified in Figs. 10 and 11. To make new intermolecular hydrogen bonds more visible we do not show in Fig. 11 the hydrogen bond from Fig. 10. The unit cell was doubled along the A and B vectors in Figs. 11a and 11b, respectively. In Fig. 11c we show an interface between two cells along the C vector.

The most remarkable is the cyclic hydrogen bond illustrated in Fig. 11c, which involves two N3 proton acceptors and two N4-H1 proton donors of the molecules I and II. A transformation *syn-to-anti* compromises the intramolecular O5'H13...O2 hydrogen bond as quantified by the  $\Delta E_I$  and  $\Delta E_{II}$  terms. The same transformation, however, “unfolds” the cytosine ring and allows the N3-C4-N4H<sub>2</sub> fragment to engage in intermolecular interactions. The strength of the resulting cyclic hydrogen bond between the neighbouring cytosine residues is significant, ca. 19 kcal/mol, when estimated based on the cytosine dimer.<sup>72</sup> Our interpretation is consistent with the finding of Thompson and Day, who concluded that molecules can be distorted by up to 5 kcal/mol by crystal packing forces, providing these distortions allow for stronger intermolecular interactions.<sup>73</sup> In fact we attempted to construct a crystal lattice of dC with two *syn* molecules per unit cell. Our search was by no means complete, but the most stable crystal structure found by us was less stable than the experimental structure (two *anti* molecules per unit cell) by 34 kcal/mol per unit cell. Clearly the energy penalty for the conformation change from *syn* to *anti* is overwritten by very favourable intermolecular interactions.

#### 4. Summary

We have studied the structure of 2'-deoxycytidine (dC): from the gas phase neutral and radical ionic species to the molecular crystal. The molecular structural calculations were completed at the density functional and the second order Møller-Plesset level. The crystal of dC was studied with the PBE+D3 functional. The molecular vertical excess electron binding energies and ionisation potentials were determined using electron propagator theory in the P3 approximation. The conformational space of dC was explored using our tool for systematic screening of conformers.<sup>25</sup>

Our results indicate that the neutral canonical dC in the gas phase favours *syn* rather than *anti* conformations. The stability of *syn* conformations results from the formation of an intramolecular O5'-H13...O2 hydrogen bond. We have also demonstrated that the C2'-*endo* sugar puckering is favourable in the gas phase. We have identified barriers for the *syn*-to-*anti* and C2'-*endo*-to-C3'-*endo* transitions at 8.2 and 3.4 kcal/mol, respectively. The  $\alpha$ -anomer of dC proved to be more stable by a fraction of kcal/mol than the  $\beta$ -anomer. The latter is the dominant building block of DNA. We have demonstrated that the IR spectra of the most stable *syn* and *anti* canonical conformers differ markedly in the region of NH/OH vibrations. The VIEs for the most stable *syn* conformers cluster in a narrow range of 8.75-8.92 eV whereas the VIE of the most stable *anti* conformer is smaller and amounts to 8.24 eV. The difference was interpreted through an orbital stabilizing effect resulting from the intramolecular O5'-H13...O2 hydrogen bond in *syn* conformers. The MP2 value of deprotonation enthalpy of dC of 1411.7 kJ/mol is in very good agreement with the experimental value of  $1409 \pm 2.5$  kJ/mol.<sup>20</sup>

Our results have confirmed that valence anions of dC are adiabatically bound.<sup>23</sup> The most stable valence anions proved to be the  $\alpha$ -anomer and *syn* canonical conformers. The experimental VDE of 0.87 eV<sup>1</sup> is in very good agreement with the calculated VDE's for the canonical *syn* conformers. Two dipole bound anions of dC have been characterized, one based on the  $\alpha$ -anomer and another on a *syn* canonical structure. The respective EPT/P3 values of VDE are 121 and 81 meV. Thus these anions are amenable to PES characterization, providing the dipole-bound states could be formed in the anionic beam.

The glycosidic bond cleavage has been studied for the neutral, anionic, and cationic dC. The glycosidic bond cleavage in neutral dC encounters a significant barrier of 40.0 kcal/mol that is reduced by 18 and 24 kcal/mol for anionic and cationic dC, respectively. The reaction is exothermic for dC<sup>-</sup> by 4 kcal/mol and endothermic by 9 and 7 kcal/mol for dC and dC<sup>+</sup>, respectively. For the neutral dC the cleavage is accompanied by proton transfer from C2'H to the O2 atom of cytosine.

Finally we have analysed the problem why dC molecules favor *anti* conformations in the crystalline phase but *syn* conformations in the gas phase. We have demonstrated that intermolecular hydrogen bonds favor *anti* conformers in the experimental structure of the molecular crystal of dC. We suggested that cytosine's very basic N3 proton acceptor and the N4H1 proton donor would not be able to engage in intermolecular hydrogen bonds if the crystal was built from *syn* conformers. In the case of *anti* conformers these sites can engage in intermolecular hydrogen bonds. Indeed, we have identified a cyclic hydrogen bonded structural motif, which involves two N3 proton acceptors and two N4H1 proton donors.

## **Acknowledgements**

The authors thank Dr. Maciej Haranczyk for helpful discussions. The authors thank financial support from ScotCHEM, which supported the PhD fellowship of S.L at the early stage of this project. This work was conducted within the framework of the COST Action CM1301 (CELINA). The research used resources of the National Energy Research Scientific Computing Centre (NERSC), which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. Computing resources were also available through (i) the HW cluster at Heriot-Watt University, (ii) a Computational Grand Challenge Application grant from the Molecular Sciences Computing Facility in the Environmental Molecular Sciences Laboratory (EMSL). Pacific Northwest National Laboratory is operated by Battelle for the U.S. DOE under Contract DE-AC06-76RLO1830.

**Supporting Information Available:** Complete author list for refs. 51 and 66; Cartesian coordinates of the structures discussed in the current study. This material is available free of charge via the Internet at <http://pubs.acs.org>.

**Table 1.** Relative MP2/AVDZ and B3LYP/6-31++G\*\* energies (in kcal/mol) with zero-point vibrational and thermal corrections ( $\Delta E_{0,\text{vib}}$ ,  $\Delta H_{298}$ ,  $\Delta G_{298}$ ), dipole moments ( $\mu_N^{B3LYP}$  in Debye) and P3/6-31++G\*\* vertical ionization energies (VIE in eV) for selected structures of neutral dC.

structure	$E^{B3LYP}$	$E^{MP2}$	$E^{B3LYP}$ + $\Delta E_{0,\text{vib}}$	$E^{B3LYP}$ + $\Delta H_{298}$	$E^{B3LYP}$ + $\Delta G_{298}$	$\mu_N^{B3LYP}$	VIE
N1	0.000	0.000	0.000	0.000	0.000	6.4	8.79
N2	-0.003	0.109	0.053	0.021	0.115	7.1	8.80
N3	1.244	2.009	1.053	1.223	0.804	7.0	8.84
N4	1.808	2.049	1.689	1.781	1.487	8.3	8.92
N5	1.349	2.097	1.059	1.290	0.661	7.5	8.84
N6	3.308	2.733	3.299	3.362	3.305	6.1	8.75
N7	2.485	3.185	2.149	2.418	1.751	7.0	8.24
N12	2.983	3.716	2.473	2.808	1.290	4.6	8.50
N25	3.590	4.810	3.078	3.520	1.445	6.1	8.58
TAU(3-imino)	3.165	2.262	3.822	3.538	4.037	4.9	8.77
IS( $\alpha$ -anomer)	-0.937	-0.129	-0.661	-0.629	-0.961	8.5	8.92

**Table 2.** Gas phase acidity (in kJ/mol) of conformers N1 and N7.

Conformer	Site of deprotonation	gas phase acidity		
		B3LYP/6-31++G**	Ref. <sup>a</sup>	MP2/AVDZ
N1	H1	1452.3	-	-
	H2	1429.1	1429.4	1411.7
	H8	1510.6	-	-
	H13	1469.7	-	-
N7	H2	1443.7	1445.8	1432.6
		Expt. <sup>b</sup>	1409±2.5	

<sup>a</sup> B3LYP/6-311+G\*\*//B3LYP/6-311G\*\* theoretical method used in Ref. 20.

<sup>b</sup> experimental value from Ref. 20.

**Table 3.** The B3LYP/6-31++G\*\* EA and VDE values (eV) for valence anions of dC. The P3/6-311++G\*\* VDE values (eV) for valence anions of dC. The EAs are calculated with respect to N<sub>2</sub>. For AIS and ATAU, additional values of EA (in parentheses) calculated with respect to the corresponding neutrals.

Structure	EA <sup>B3LYP</sup>	VDE <sup>B3LYP</sup>	VDE <sup>P3</sup>
AIS	0.46 (0.43)	0.90	0.98
AN4	0.35	0.86	0.94
AN2	0.34	0.76	0.84
AN1	0.33	0.68	0.74
AN5	0.30	0.82	0.92
AN9	0.28	0.88	0.97
AN25 <sup>a</sup>	0.16	0.71	0.76
ATAU	0.07 (0.23)	0.28	0.01
Expt. <sup>b</sup>	~0.5	0.87	

<sup>a</sup> AN25 corresponds to the structure used in Ref. 23.

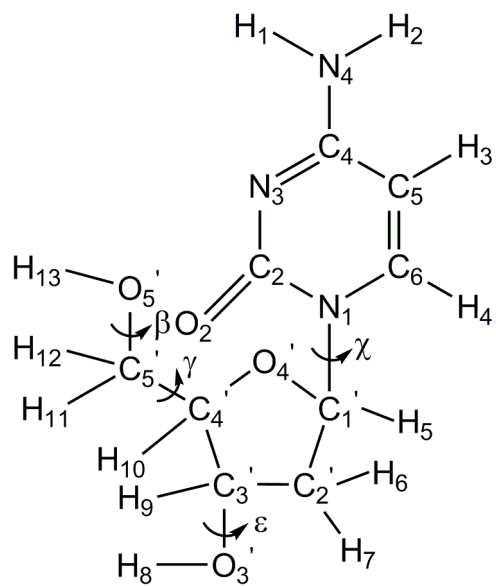
<sup>b</sup> Experimental results from Ref. 1.

**Table 4.** Intermolecular hydrogen bonds in the crystal structure. “L” and “R” indicate unit cells on the left and right hand side, respectively. “A”, “B” and “C” indicate that hydrogen bonds are formed between monomers in neighbouring unit cells along lattice vector A, B and C, respectively.

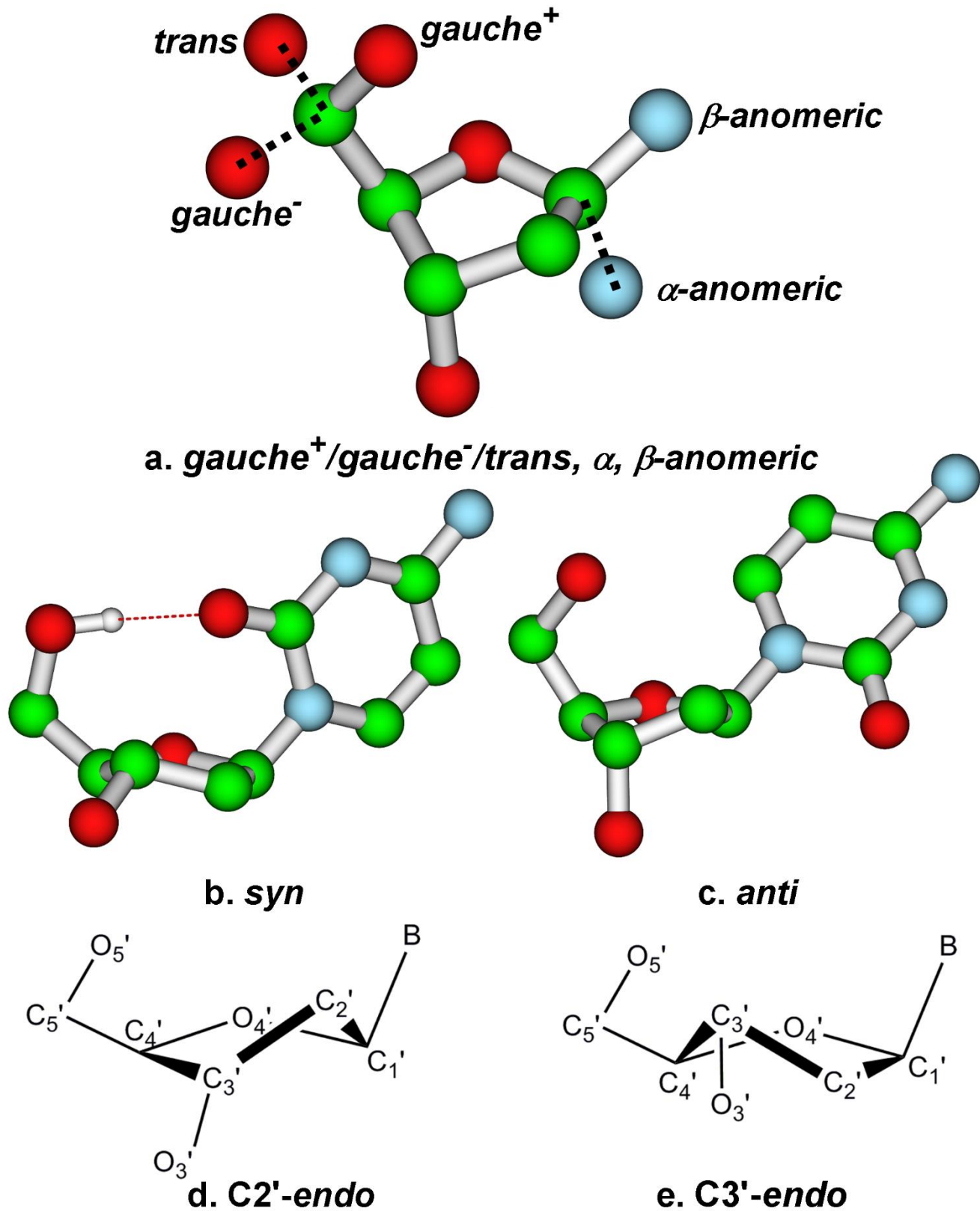
Figure	Hydrogen bond	
	Donor	Acceptor
10 (dCanti)	O5'-H13 (II)	O4' (I)
11a (dCanti_A)	O5'-H13 (I_L)	O2 (I_R)
11b (dCanti_B)	N4-H2 (II_L)	O2 (II_R)
	N4-H2 (I_R)	O2 (I_L)
	O3'-H8 (I_L)	O5' (I_R)
	O3'-H8 (II_R)	O3' (I_L)
11c (dCanti_C)	N4-H1 (I)	N3 (II)
	N4-H1 (II)	N3 (I)



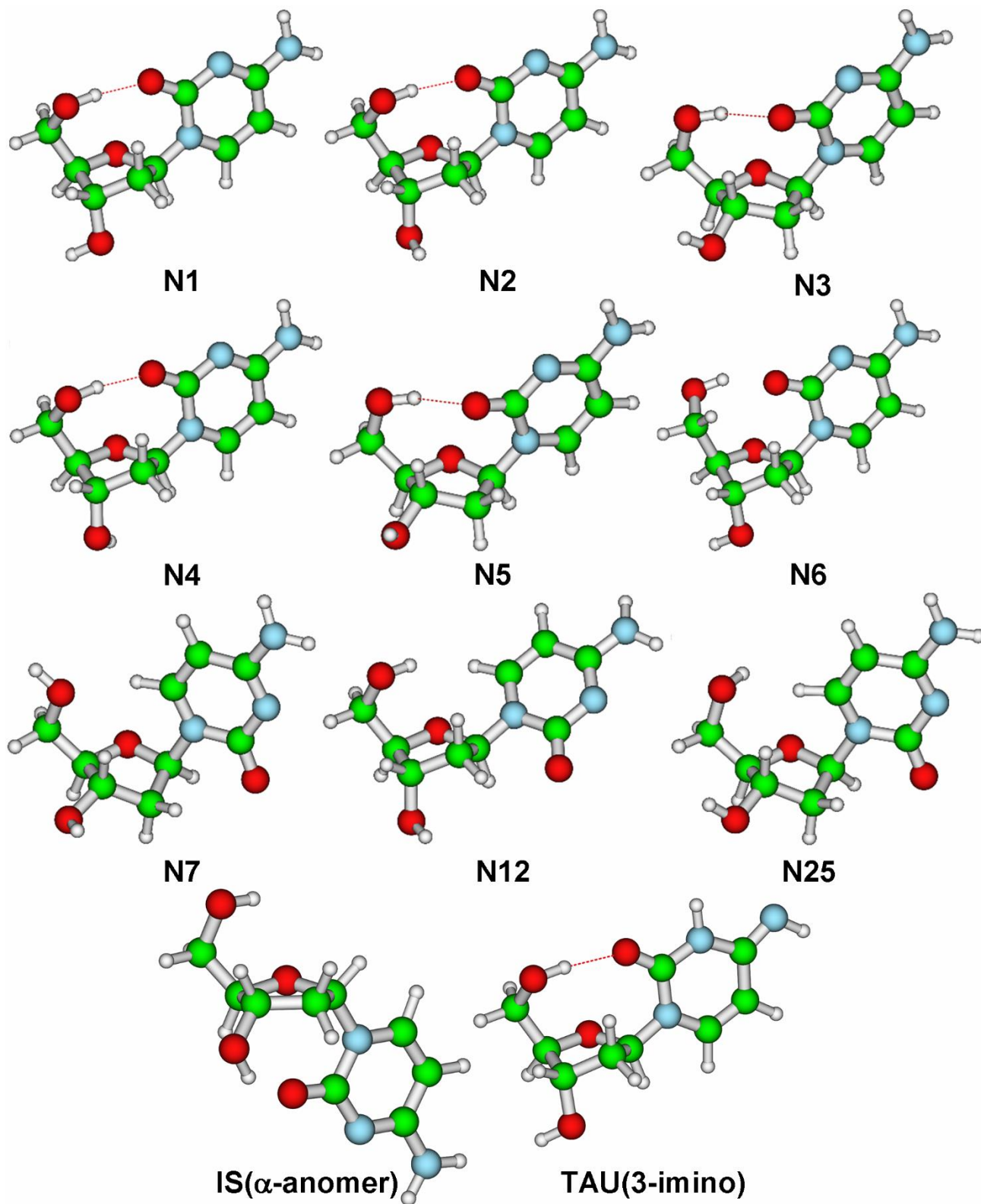
**Figure 1.** Schematic molecular structure of 2'-deoxycytidine.



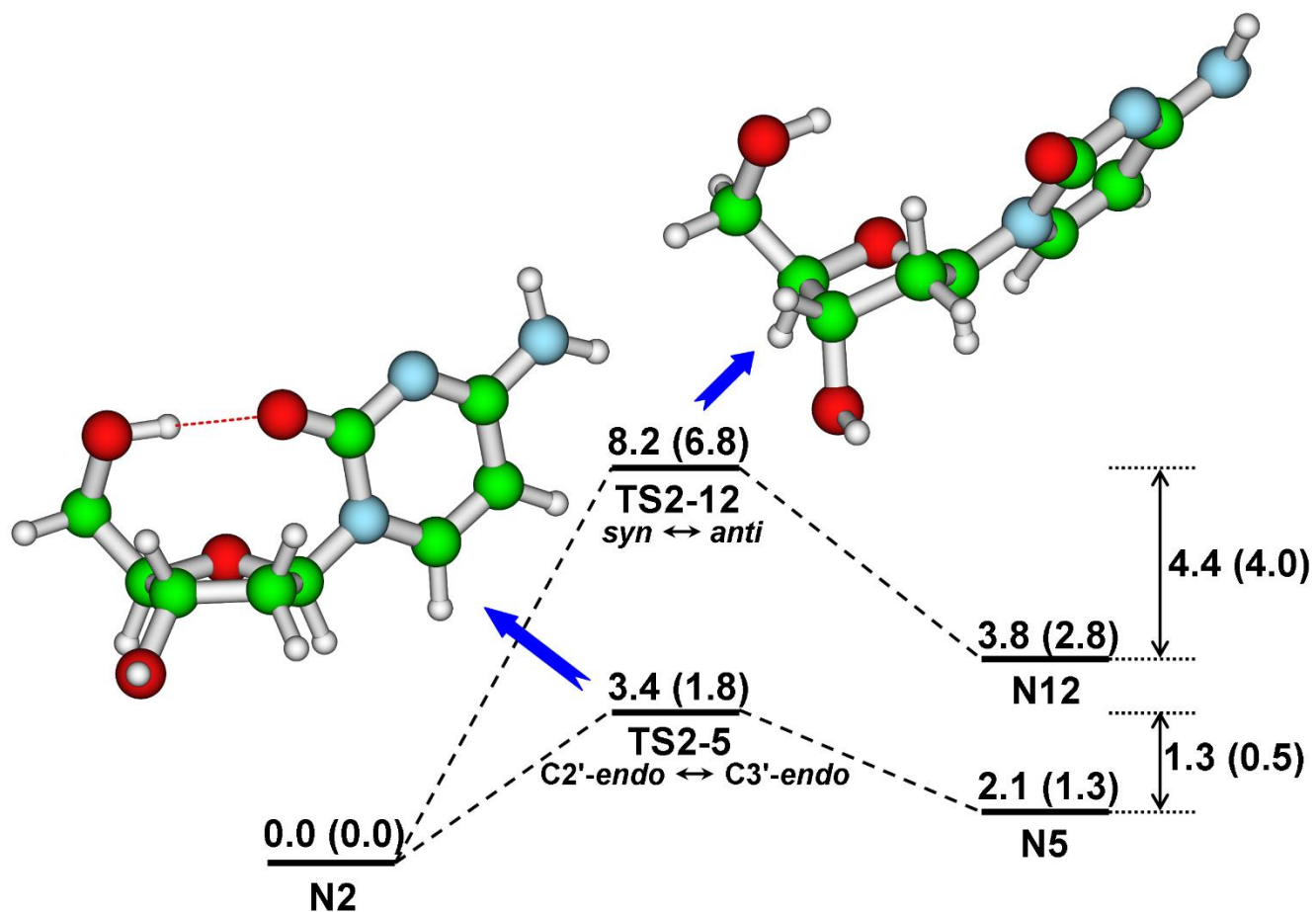
**Figure 2.** A schematic definition of selected degrees of freedom of dC.



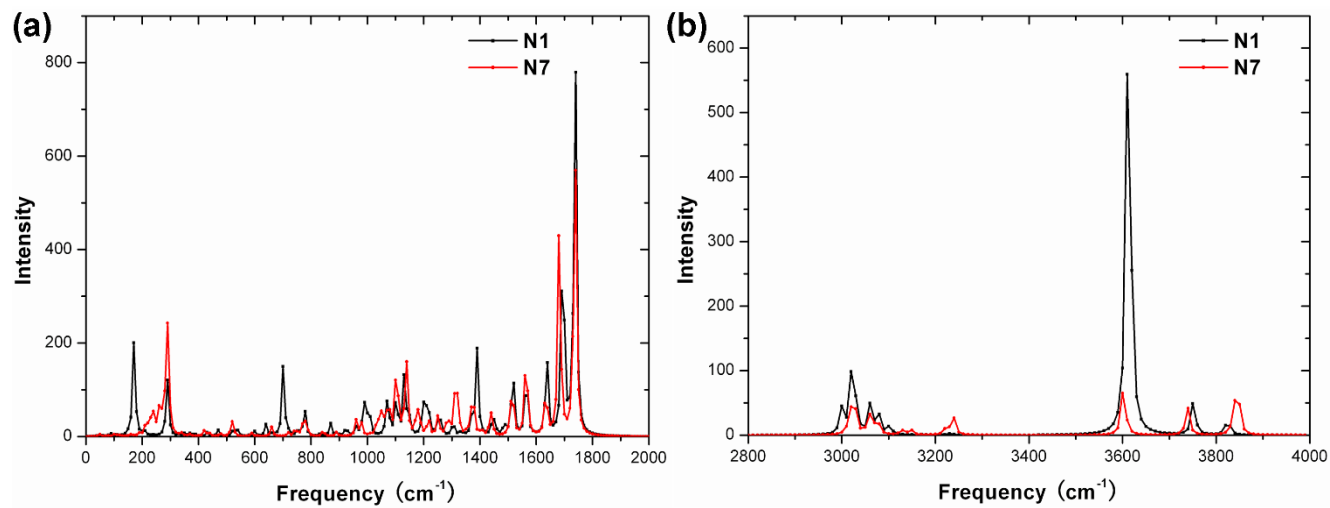
**Figure 3.** The MP2/AVDZ geometries of selected structures of neutral 2'-deoxycytidine. The dotted lines indicate the intramolecular hydrogen bonds.



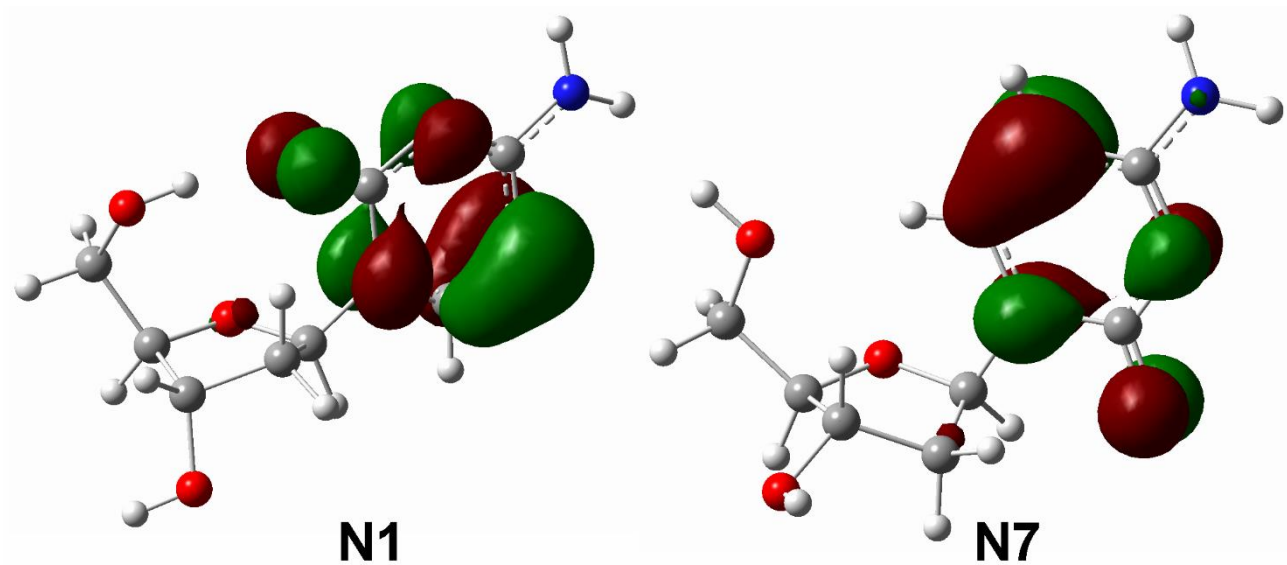
**Figure 4.** Interconversions between the *syn/anti*, and *C2'-endo/C3'-endo* conformers. Relative energies and transition barriers, which are determined at the MP2/AVDZ and B3LYP/6-31++G\*\* (in parentheses) levels, are in kcal/mol. Only transition state structures are visualized.



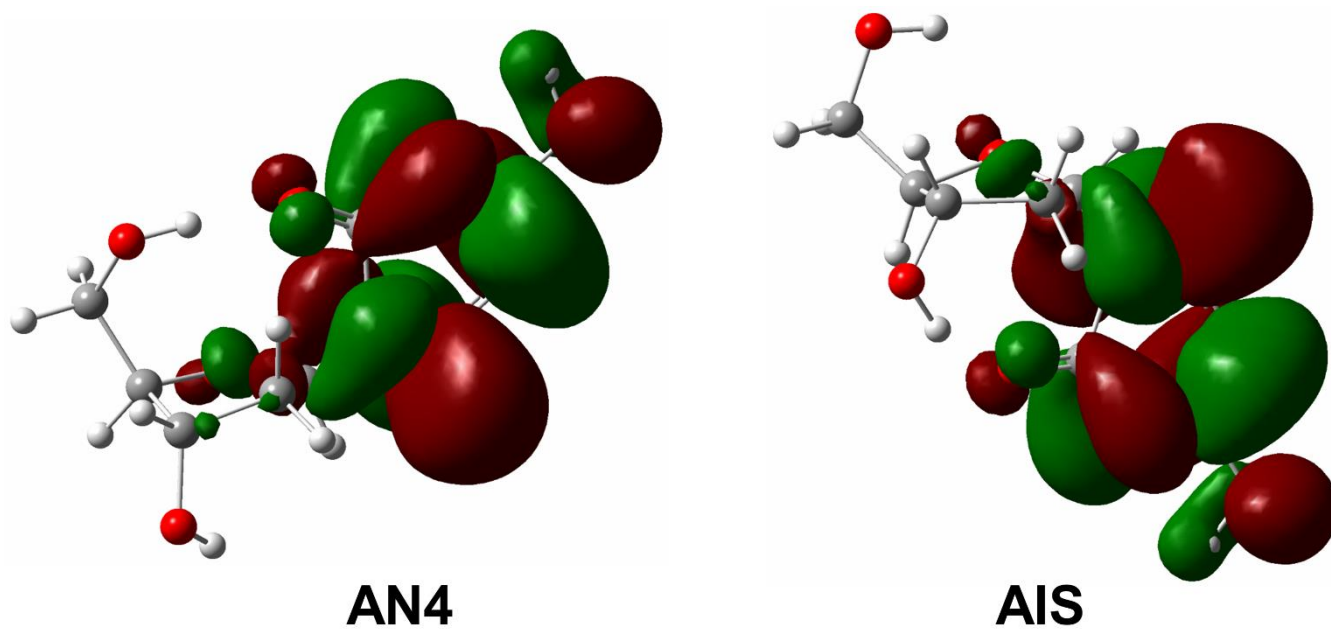
**Figure 5.** IR spectra of conformers N1 (in black) and N7 (in red), determined at the B3LYP/6-31++G\*\* level, in the (a) low frequency, 0~2000  $\text{cm}^{-1}$ , and (b) high frequency, 2800~4000  $\text{cm}^{-1}$  range.



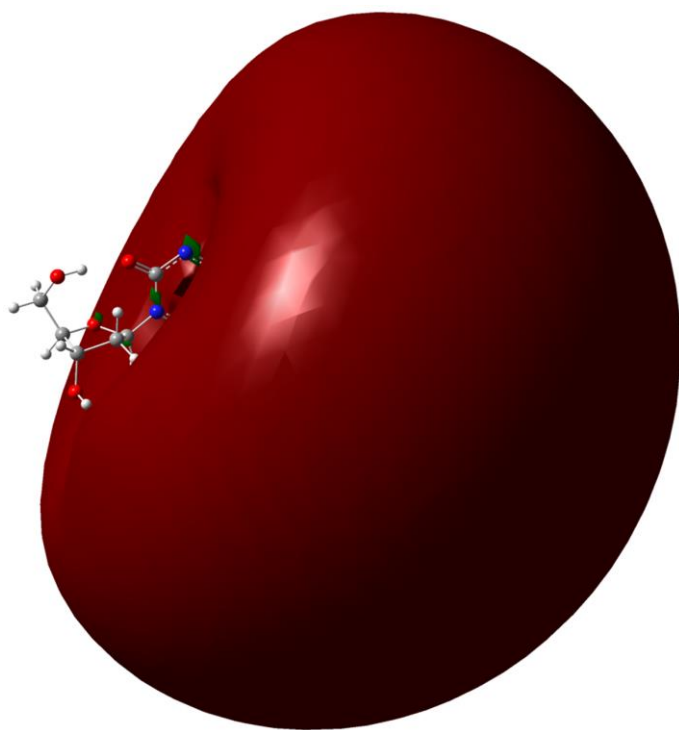
**Figure 6.** The isosurfaces of HOMO for N1 and N7 determined at the HF/6-31++G\*\* level, with an isovalue of  $0.05 a_0^{-3/2}$ .



**Figure 7.** The isosurfaces of SOMO for AN4 and AIS determined at the B3LYP/6-31++G\*\* level, with an isovalue of  $0.02 a_0^{-3/2}$ .



**Figure 8.** The isosurfaces of the HF/AVDZ+5sp SOMOs for dipole-bound anions (a) DAN2 and (b) DAIS. 80% of the total excess charge is reproduced for each anion. The VDE (MP2 and P3) and EA (MP2) values in eV. The MP2 dipole moments (D) of the neutral calculated at the optimal geometries of the neutral and the anion.

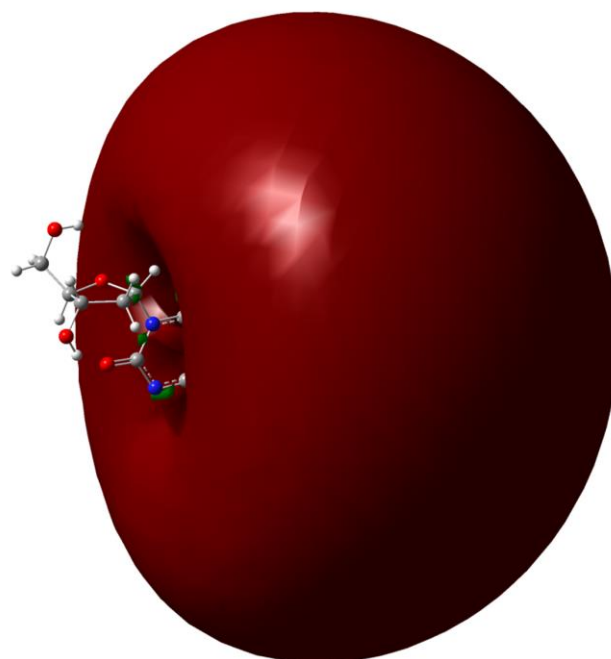


**(a) DAN2**

$$\text{VDE}^{\text{MP2/P3}} = 0.074/0.081 \text{ eV}$$

$$\text{EA} = 0.065 \text{ eV}$$

$$\mu_N^{\text{MP2}}(\text{N2}) = 6.7 \text{ D}, \mu_N^{\text{MP2}}(\text{DAN2}) = 7.4 \text{ D}$$



**(b) DAIS**

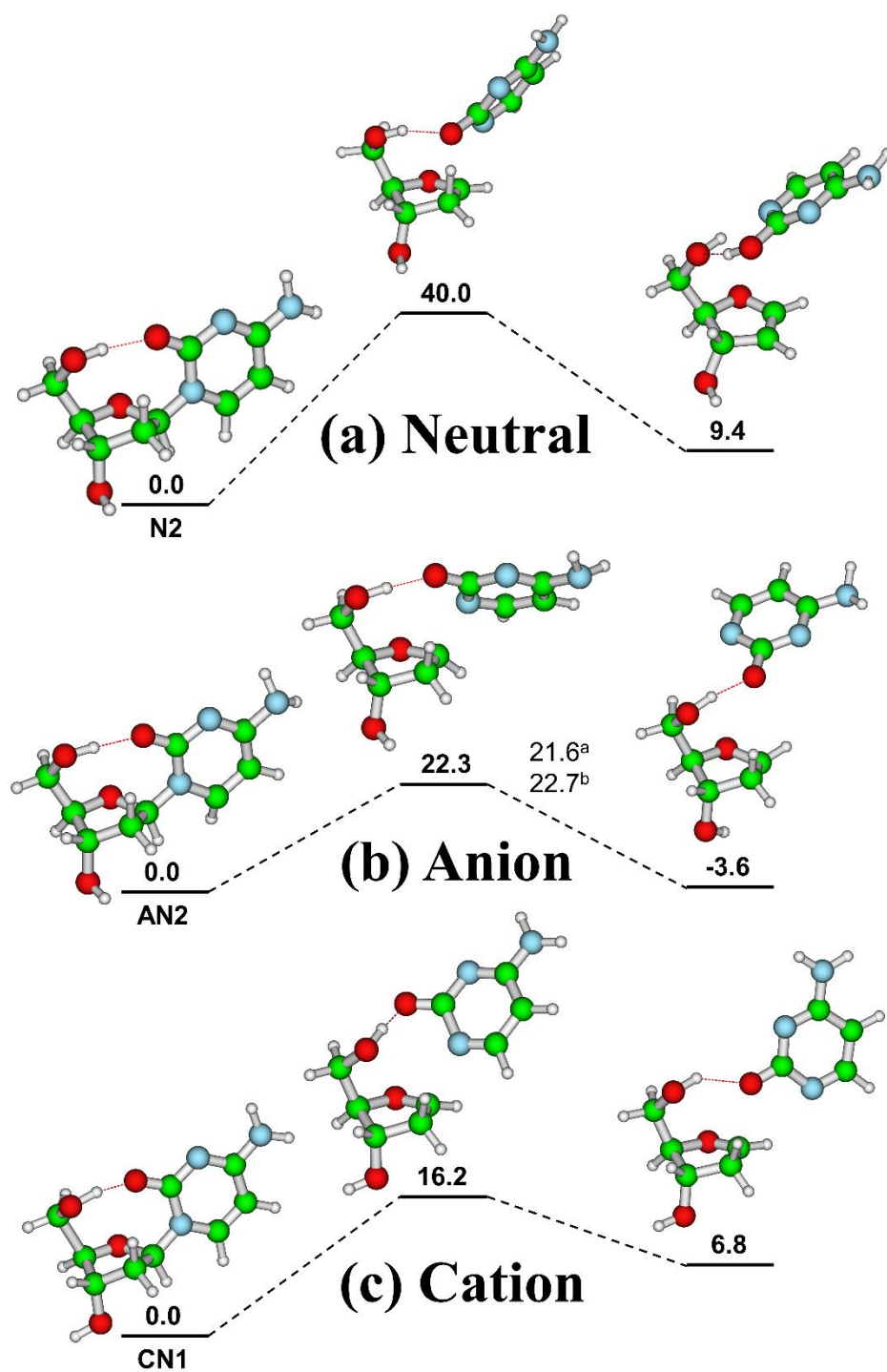
$$\text{VDE}^{\text{MP2/P3}} = 0.113/0.121 \text{ eV}$$

$$\text{EA} = 0.108 \text{ eV}$$

$$\mu_N^{\text{MP2}}(\text{IS}) = 8.0 \text{ D}, \mu_N^{\text{MP2}}(\text{DAIS}) = 8.5 \text{ D}$$



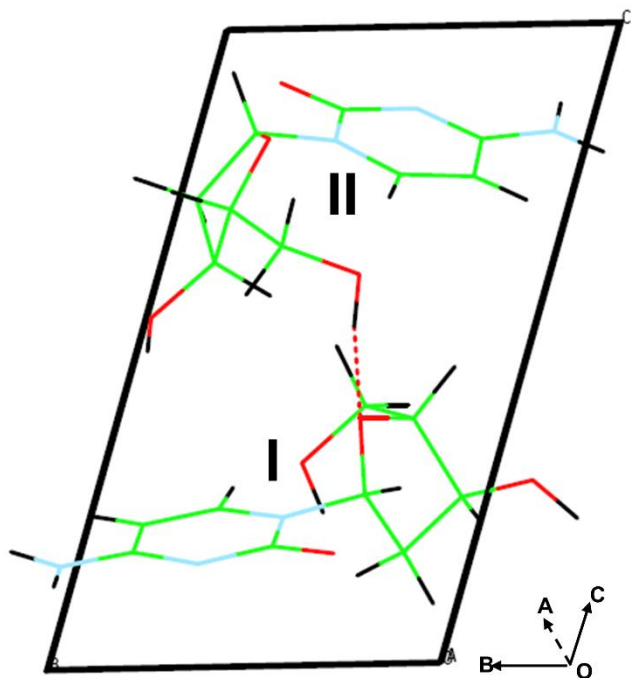
**Figure 9.** The geometries, relative energies, and transition barriers for glycosidic bond cleavage reactions of (a) neutral, (b) anionic, and (c) cationic species of dC. The B3LYP/6-31++G\*\* level of theory, all energies in kcal/mol.



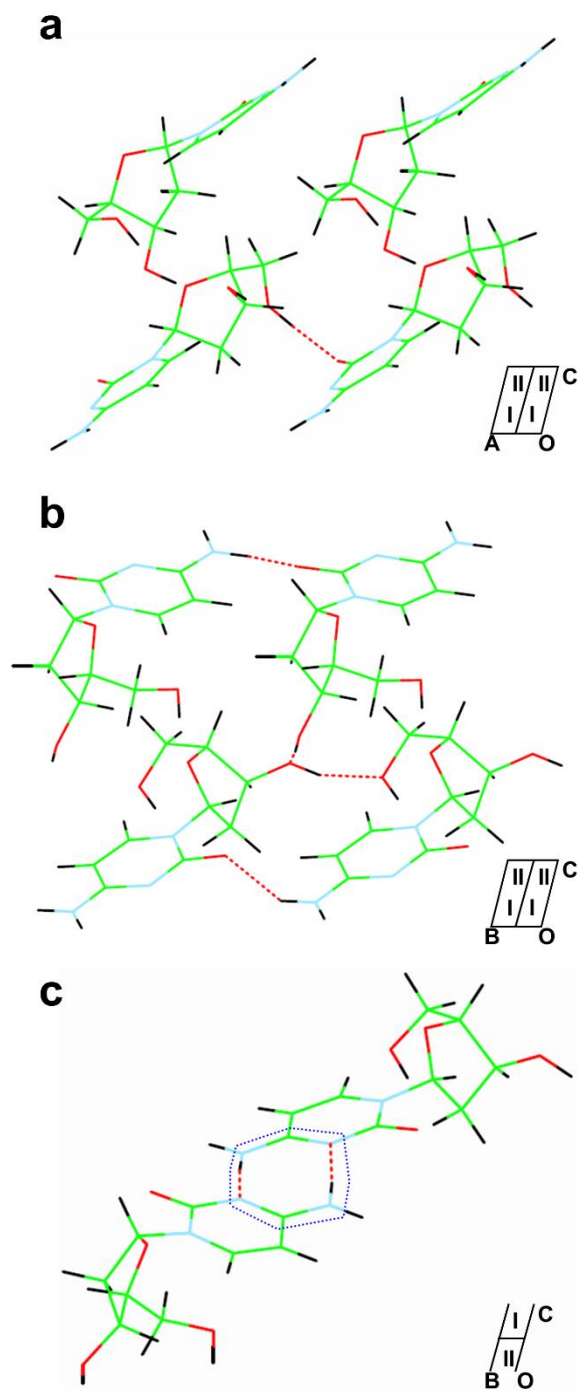
<sup>a</sup> Theoretical value taken from Ref. 39.

<sup>b</sup> Theoretical value taken from Ref. 24.

**Figure 10.** Crystal structure of dC, with schematics in the right bottom showing different lattice vectors.



**Figure 11.** Intermolecular hydrogen bonds in the crystal structure of dC, with schematics in the right bottom showing neighbouring unit cells along different lattice vectors.



## References

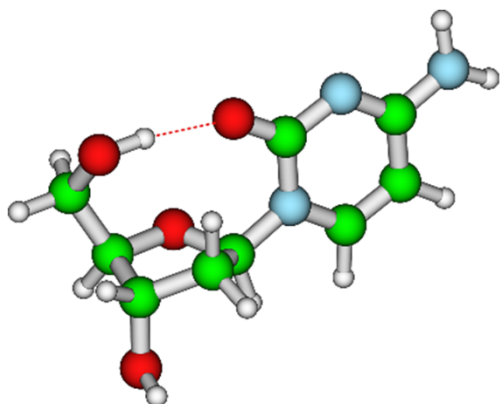
1. Stokes, S. T.; Li, X.; Grubisic, A.; Ko, Y. J.; Bowen, K. H., Intrinsic Electrophilic Properties of Nucleosides: Photoelectron Spectroscopy of Their Parent Anions. *J. Chem. Phys.* **2007**, *127*, 084321(1)-084321(6).
2. Schermann, J.-P., *Spectroscopy and Modeling of Biomolecular Building Blocks*; Elsevier: Amsterdam ; Oxford, 2007, p 479.
3. Rizzo, T. R.; Park, Y. D.; Levy, D. H., A Molecular-Beam of Tryptophan. *J. Am. Chem. Soc.* **1985**, *107*, 277-278.
4. Meijer, G.; de Vries, M. S.; Hunziker, H. E.; Wendt, H. R., Laser Desorption Jet-Cooling Spectroscopy of Para-Amino Benzoic-Acid Monomer, Dimer, and Clusters. *J. Chem. Phys.* **1990**, *92*, 7625-7635.
5. Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M., Electrospray Ionization-Principles and Practice. *Mass Spectrom. Rev.* **1990**, *9*, 37-70.
6. Hillenkamp, F.; Karas, M.; Beavis, R. C.; Chait, B. T., Matrix-Assisted Laser Desorption Ionization Mass-Spectrometry of Biopolymers. *Anal. Chem.* **1991**, *63*, A1193-A1202.
7. Bald, I.; Dabkowska, I.; Illenberger, E., Probing Biomolecules by Laser-Induced Acoustic Desorption: Electrons at near Zero Electron Volts Trigger Sugar-Phosphate Cleavage. *Angew. Chem. Int. Ed.* **2008**, *47*, 8518-8520.
8. Wyttenbach, T.; Witt, M.; Bowers, M. T., On the Question of Salt Bridges of Cationized Amino Acids in the Gas Phase: Glycine and Arginine. *Int. J. Mass spectrom.* **1999**, *183*, 243-252.
9. Price, W. D.; Jockusch, R. A.; Williams, E. R., Is Arginine a Zwitterion in the Gas Phase? *J. Am. Chem. Soc.* **1997**, *119*, 11988-11989.
10. Chappo, C. J.; Paul, J. B.; Provencal, R. A.; Roth, K.; Saykally, R. J., Is Arginine Zwitterionic or Neutral in the Gas Phase? Results from IR Cavity Ringdown Spectroscopy. *J. Am. Chem. Soc.* **1998**, *120*, 12956-12957.
11. Alonso, J. L.; Cocinero, E. J.; Lesarri, A.; Sanz, M. E.; Lopez, J. C., The Glycine-Water Complex. *Angew. Chem. Int. Ed.* **2006**, *45*, 3471-3474.
12. Foloppe, N.; Hartmann, B.; Nilsson, L.; MacKerell Jr, A. D., Intrinsic Conformational Energetics Associated with the Glycosyl Torsion in DNA: A Quantum Mechanical Study. *Biophys. J.* **2002**, *82*, 1554-1569.
13. Foloppe, N.; MacKerell Jr, A. D., Intrinsic Conformational Properties of Deoxyribonucleosides: Implicated Role for Cytosine in the Equilibrium among the A, B, and Z Forms of DNA. *Biophys. J.* **1999**, *76*, 3206-3218.
14. Foloppe, N.; Nilsson, L.; MacKerell Jr, A. D., Ab Initio Conformational Analysis of Nucleic Acid Components: Intrinsic Energetic Contributions to Nucleic Acid Structure and Dynamics. *Biopolymers* **2002**, *61*, 61-76.
15. Wales, D. J.; Doye, J. P. K., Global Optimization by Basin-Hopping and the Lowest Energy Structures of Lennard-Jones Clusters Containing up to 110 Atoms. *J. Phys. Chem. A* **1997**, *101*, 5111-5116.
16. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by Simulated Annealing. *Science* **1983**, *220*, 671-680.
17. Holland, J. H., *Adaptation in Natural and Artificial Systems*; MIT Press: Cambridge, 1992.
18. Ling, S.; Yu, W.; Huang, Z.; Lin, Z.; Haranczyk, M.; Gutowski, M., Gaseous Arginine Conformers and Their Unique Intramolecular Interactions. *J. Phys. Chem. A* **2006**, *110*, 12282-12291.
19. Young, D. W.; Wilson, H. R., Crystal and Molecular Structure of 2'-Deoxycytidine. *Acta Crystallogr. Sect. B* **1975**, *B31*, 961-965.
20. Kumari, S.; Devi, C. L.; Prabhakar, S.; Bhanuprakash, K.; Vairamani, M., Estimation of Gas-Phase Acidities of Deoxyribonucleosides: An Experimental and Theoretical Study. *J. Am. Soc. Mass. Spectrom.* **2010**, *21*, 136-143.

21. Colson, A. O.; Besler, B.; Sevilla, M. D., Ab-Initio Molecular-Orbital Calculations on DNA Radical Ions. 3. Ionization-Potentials and Ionization Sites in Components of the DNA Sugar-Phosphate Backbone. *J. Phys. Chem.* **1993**, *97*, 8092-8097.
22. Lu, L.-H.; Su, C.-C.; Hsieh, T.-J., Density Functional Calculations and Experimental Studies of the Sidearm Remote Controlling Effect in Eight Lariat Crown Ethers. *J. Mol. Struct.* **2007**, *831*, 151-164.
23. Richardson, N. A.; Gu, J. D.; Wang, S. Y.; Xie, Y. M.; Schaefer, H. F., DNA Nucleosides and Their Radical Anions: Molecular Structures and Electron Affinities. *J. Am. Chem. Soc.* **2004**, *126*, 4404-4411.
24. Li, X.; Sanche, L.; Sevilla, M. D., Base Release in Nucleosides Induced by Low-Energy Electrons: A DFT Study. *Radiat. Res.* **2006**, *165*, 721-729.
25. SSC: A Tool for Constructing Libraries for Systematic Screening of Conformers. <http://sscf.sf.net> (accessed Sept 28, 2016). Ling, S.; Gutowski, M. *J. Comput. Chem.* **2011**, *32*, 2047-2054.
26. Fox, J. J.; Yung, N. C.; Wempen, I.; Hoffer, M., Nucleosides. XII. Direct Synthesis of 2'-Deoxycytidine and Its  $\alpha$ -Anomer. *J. Am. Chem. Soc.* **1961**, *83*, 4066-4070.
27. Garcia, J.; Diaz-Rodriguez, A.; Fernandez, S.; Sanghvi, Y. S.; Ferrero, M.; Gotor, V., New Concept for the Separation of an Anomeric Mixture of  $\alpha/\beta$ -D-Nucleosides through Regioselective Enzymatic Acylation or Hydrolysis Processes. *J. Org. Chem.* **2006**, *71*, 9765-9771.
28. Xia, F.; Xie, H.; Cao, Z., Density Functional Study of Protonation of Deoxynucleosides: Electrophilic Active Sites and Proton Affinities. *Int. J. Quantum Chem.* **2008**, *108*, 57-65.
29. Shishkin, O. V.; Pelmeshnikov, A.; Hovorun, D. M.; Leszczynski, J., Molecular Structure of Free Canonical 2'-Deoxyribonucleosides: A Density Functional Study. *J. Mol. Struct.* **2000**, *526*, 329-341.
30. Kolandaivel, P.; Knapp-Mohammady, M.; Suhai, S., Studies on Structure and Conformational Stability of Free Canonical 2'-Deoxyribonucleosides: Approximate SCC-DFTB and LMP2 Methods. *Int. J. Quantum Chem.* **2004**, *99*, 28-38.
31. Boudaiffa, B.; Cloutier, P.; Hunting, D.; Huels, M. A.; Sanche, L., Resonant Formation of DNA Strand Breaks by Low-Energy (3 to 20 eV) Electrons. *Science* **2000**, *287*, 1658-1660.
32. Huels, M. A.; Boudaiffa, B.; Cloutier, P.; Hunting, D.; Sanche, L., Single, Double, and Multiple Double Strand Breaks Induced in DNA by 3-100 eV Electrons. *J. Am. Chem. Soc.* **2003**, *125*, 4467-4477.
33. Martin, F.; Burrow, P. D.; Cai, Z. L.; Cloutier, P.; Hunting, D.; Sanche, L., DNA Strand Breaks Induced by 0-4 eV Electrons: The Role of Shape Resonances. *Phys. Rev. Lett.* **2004**, *93*, 068101(1)-068101(4).
34. Sanche, L., Low Energy Electron-Driven Damage in Biomolecules. *Eur. Phys. J. D* **2005**, *35*, 367-390.
35. Simons, J., How Do Low-Energy (0.1-2 eV) Electrons Cause DNA-Strand Breaks? *Acc. Chem. Res.* **2006**, *39*, 772-779.
36. Hendricks, J. H.; Lyapustina, S. A.; deClercq, H. L.; Snodgrass, J. T.; Bowen, K. H., Dipole Bound, Nucleic Acid Base Anions Studied Via Negative Ion Photoelectron Spectroscopy. *J. Chem. Phys.* **1996**, *104*, 7788-7791.
37. Li, X.; Bowen, K. H.; Haranczyk, M.; Bachorz, R. A.; Mazurkiewicz, K.; Rak, J.; Gutowski, M., Photoelectron Spectroscopy of Adiabatically Bound Valence Anions of Rare Tautomers of the Nucleic Acid Bases. *J. Chem. Phys.* **2007**, *127*, 174309(1)-174309(6).
38. Stokes, S. T.; Grubisic, A.; Li, X.; Ko, Y. J.; Bowen, K. H., Photoelectron Spectroscopy of the Parent Anions of the Nucleotides, Adenosine-5'-Monophosphate and 2' Deoxyadenosine-5'-Monophosphate. *J. Chem. Phys.* **2008**, *128*, 044314(1)-044314(5).
39. Gu, J.; Xie, Y.; Schaefer, H. F., Glycosidic Bond Cleavage of Pyrimidine Nucleosides by Low-Energy Electrons: A Theoretical Rationale. *J. Am. Chem. Soc.* **2005**, *127*, 1053-1057.
40. Bachorz, R. A.; Rak, J.; Gutowski, M., Stabilization of Very Rare Tautomers of Uracil by an Excess Electron. *Phys. Chem. Chem. Phys.* **2005**, *7*, 2116-2125.
41. Haranczyk, M.; Rak, J.; Gutowski, M., Stabilization of Very Rare Tautomers of 1-Methylcytosine by an Excess Electron. *J. Phys. Chem. A* **2005**, *109*, 11495-11503.

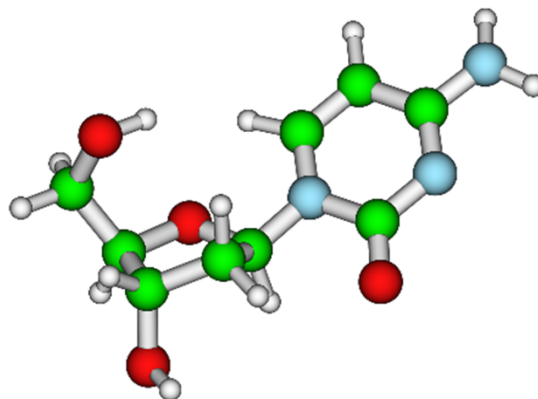
42. Haranczyk, M.; Gutowski, M., Quantum Mechanical Energy-Based Screening of Combinatorially Generated Library of Tautomers. Tautgen: A Tautomer Generator Program. *J. Chem. Inf. Model.* **2006**, *47*, 686-694.
43. Haranczyk, M.; Gutowski, M., Combinatorial–Computational–Chemoinformatics (C3) Approach to Finding and Analyzing Low-Energy Tautomers. *J. Comput. Aided Mol. Des.* **2010**, *24*, 627-638.
44. Haranczyk, M.; Puzyn, T.; Sadowski, P., ConGENER - A Tool for Modeling of the Congeneric Sets of Environmental Pollutants. *QSAR & Comb. Sci.* **2008**, *27*, 826-833.
45. Keolopile, Z. G.; Gutowski, M.; Haranczyk, M., Discovery of Most Stable Structures of Neutral and Anionic Phenylalanine through Automated Scanning of Tautomeric and Conformational Spaces. *J. Chem. Theory Comput.* **2013**, *9*, 4374-4381.
46. Lee, C.; Yang, W.; Parr, R. G., Development of the Colle-Salvetti Correlation-Energy Formula into a Functional of the Electron Density. *Phys. Rev. B* **1988**, *37*, 785-789.
47. Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J., Ab Initio Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
48. Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self-Consistent Molecular Orbital Methods. XII. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257-2261.
49. Kendall, R. A.; Dunning, T. H.; Harrison, R. J., Electron Affinities of the First-Row Atoms Revisited. Systematic Basis Sets and Wave Functions. *J. Chem. Phys.* **1992**, *96*, 6796-6806.
50. Ortiz, J. V., Electron Propagator Theory: An Approach to Prediction and Interpretation in Quantum Chemistry. *WIREs Comput. Mol. Sci.* **2013**, *3*, 123-142.
51. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. Gaussian 03, Revision C.02, Gaussian, Inc., Wallingford CT, 2004.
52. Simons, J., Molecular Anions. *J. Phys. Chem. A* **2008**, *112*, 6401-6511.
53. N2 is less stable than N1 by only 0.1 kcal/mol at MP2/AVDZ level of theory, but it has a dipole moment larger by 0.7 D.
54. Skurski, P.; Gutowski, M.; Simons, J., How to Choose a One-Electron Basis Set to Reliably Describe a Dipole-Bound Anion. *Int. J. Quantum Chem.* **2000**, *80*, 1024-1038.
55. Rak, J.; Skurski, P.; Gutowski, M., An Ab Initio Study of the Betaine Anion-Dipole-Bound Anionic State of a Model Zwitterion System. *J. Chem. Phys.* **2001**, *114*, 10673-10681.
56. Gutowski, M.; Skurski, P., Dispersion Stabilization of Solvated Electrons and Dipole-Bound Anions. *J. Phys. Chem. B* **1997**, *101*, 9143-9146.
57. Open-source Cubefile Manipulator Program (OpenCubMan) is available free of charge at the SourceForge archive: <http://opencubman.sourceforge.net> (accessed Sept 28, 2016); Haranczyk, M.; Gutowski, M. Visualization of Molecular Orbitals and the Related Electron Densities. *J. Chem. Theory Comput.* **2008**, *4*, 689-693.
58. Schlegel, H. B., Optimization of Equilibrium Geometries and Transition Structures. *J. Comput. Chem.* **1982**, *3*, 214-218.
59. Fukui, K., The Path of Chemical Reactions - the IRC Approach. *Acc. Chem. Res.* **1981**, *14*, 363-368.
60. VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J., Quickstep: Fast and Accurate Density Functional Calculations Using a Mixed Gaussian and Plane Waves Approach. *Comput. Phys. Commun.* **2005**, *167*, 103-128.
61. Hutter, J.; Iannuzzi, M.; Schiffmann, F.; VandeVondele, J., CP2K: Atomistic Simulations of Condensed Matter Systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 15-25.
62. Goedecker, S.; Teter, M.; Hutter, J., Separable Dual-Space Gaussian Pseudopotentials. *Phys. Rev. B* **1996**, *54*, 1703-1710.
63. Krack, M., Pseudopotentials for H to Kr Optimized for Gradient-Corrected Exchange-Correlation Functionals. *Theor. Chem. Acc.* **2005**, *114*, 145-152.

64. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865-3868.
65. Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H., A Consistent and Accurate Ab Initio Parametrization of Density Functional Dispersion Correction (DFT-D) for the 94 Elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104(1)-154104(19).
66. Bylaska, E. J.; de Jong, W. A.; Kowalski, K.; Straatsma, T. P.; Valiev, M.; Wang, D.; Apra, E.; Windus, T. L.; Hirata, S.; Hackler, M. T.; et al. "NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.0" (2006), Pacific Northwest National Laboratory, Richland, Washington 99352-0999, USA.
67. Schaftenaar, G.; Noordik, J. H., Molden: A Pre- and Post-Processing Program for Molecular and Electronic Structures. *J. Comput. Aided Mol. Des.* **2000**, *14*, 123-134.
68. GaussView, Version 5, Dennington, Roy; Keith, Todd; Millam, John. Semichem Inc., Shawnee Mission, KS, 2009.
69. Notice that the theoretical value of deprotonation enthalpy of N7-H2 from our calculation is different from the one reported in Ref. 20 at the same level of theory which these authors used. This is because these authors double counted the contributions from zero-point vibrations. In Gaussian 03, the thermal correction to enthalpy includes the contribution from zero-point vibrations. The corrected value is 1445.8 kJ/mol.
70. Ptasinska, S.; Denifl, S.; Scheier, P.; Mark, T. D., Inelastic Electron Interaction (Attachment/Ionization) with Deoxyribose. *J. Chem. Phys.* **2004**, *120*, 8505-8511.
71. Trofimov, A. B.; Schirmer, J.; Kobaychev, V. B.; Potts, A. W.; Holland, D. M. P.; Karlsson, L., Photoelectron Spectra of the Nucleobases Cytosine, Thymine and Adenine. *J. Phys. B: At., Mol. Opt. Phys.* **2006**, *39*, 305-329.
72. Hobza, P.; Šponer, J., Structure, Energetics, and Dynamics of the Nucleic Acid Base Pairs: Nonempirical Ab Initio Calculations. *Chem. Rev.* **1999**, *99*, 3247-3276.
73. Thompson, H. P. G.; Day, G. M., Which Conformations Make Stable Crystal Structures? Mapping Crystalline Molecular Geometries to the Conformational Energy Landscape. *Chem. Sci.* **2014**, *5*, 3173-3182.

## Intra- vs. Intermolecular H-Bonds



Gas phase (*syn*)



Solid (*anti*)