

# Photochemical selectivity in guanine–cytosine base-pair structures

Ali Abo-Riziq<sup>†</sup>, Louis Grace<sup>†</sup>, Eyal Nir<sup>†</sup>, Martin Kabelac<sup>‡</sup>, Pavel Hobza<sup>‡</sup>, and Mattanjah S. de Vries<sup>†§</sup>

<sup>†</sup>Department of Chemistry and Biochemistry, University of California, Santa Barbara, CA 93106; and <sup>‡</sup>Institute of Organic Chemistry and Biochemistry, Academy of Sciences of the Czech Republic, 166 10 Prague 6, Czech Republic

Communicated by Donald H. Levy, University of Chicago, Chicago, IL, November 17, 2004 (received for review August 19, 2004)

**Prebiotic chemistry presumably took place before formation of an oxygen-rich atmosphere and thus under conditions of intense short wavelength UV irradiation. Therefore, the UV photochemical stability of the molecular building blocks of life may have been an important selective factor in determining the eventual chemical makeup of critical biomolecules. To investigate the role of UV irradiation in base-pairing we have studied guanine (G) and cytosine (C) base pairs in the absence of the RNA backbone. We distinguished base-pair structures by IR–UV hole-burning spectroscopy as well as by high-level correlated *ab initio* calculations. The Watson–Crick structure exhibits broad UV absorption, in stark contrast to other GC structures and other base-pair structures. This broad absorption may be explained by a rapid internal conversion that makes this specific base pair arrangement uniquely photochemically stable.**

*ab initio* computation | DNA base pairs | IR–UV spectroscopy | jet cooling | photochemistry

The DNA bases involved in reproduction have short  $S_1$  excited state lifetimes, of the order of 1 ps or less (1–7). It has been argued that this phenomenon serves to protect these bases against photochemical damage, because after excitation they do not cross to the reactive triplet state; instead, they rapidly internally convert to the electronic ground state (8). This mechanism may have been particularly significant under the conditions of the early earth, when purines and pyrimidines presumably were assembled into the first macromolecular structures, producing RNA. At that time, the earth was exposed to shorter wavelength UV radiation than it is today. For an analysis of possible prebiotic chemistry, it is necessary not only to consider the individual bases but also to study them as they interact and to do so without the RNA backbone (9). We achieve such an analysis by studying clusters of guanine (G) and cytosine (C) in the gas phase by using double resonant laser spectroscopy techniques. The experimental studies are accompanied by state-of-the-art quantum chemical and molecular dynamics calculations of the pairing of G and C. These two bases can form many different hydrogen-bonded structures, of which at least 20 are within 12 kcal/mol of the global minimum. Thus, the familiar Watson–Crick (WC) structure may not be unique, based on energies alone. Here, we report a remarkable difference in excited state properties among the WC structure and other structures. The former exhibits broad UV absorption features; this is in contrast with the sharp UV spectra exhibited by non-WC structures. The broad absorption can be explained by recent theoretical results, predicting a pathway for rapid internal conversion (10). If the difference is solely due to lifetime broadening, these results correspond to a lifetime for the WC structures that is at least 2 orders of magnitude shorter than those of observed non-WC structures. Thus, it appears possible that the WC recognition mechanism involves a structure that was significantly more stable than other base-pair structures because of its photochemistry under the conditions at the time of the origin of life.

## Methods

We placed DNA bases in the gas phase by pulsed laser desorption followed by entrainment in a supersonic expansion of argon (11). This jet cooling served two purposes: (i) The resulting internal temperature of 10–20 K allowed us to perform high-resolution spectroscopy both in the UV and in the IR wavelength ranges. (ii) Formation of clusters in the jet allowed us to study interactions between individual bases. We performed electronic spectroscopy in the UV range of the spectrum by resonance enhanced multiphoton ionization (REMPI). REMPI provided information about the first excited electronic state (the  $S_1$  state) and about the photochemistry. We distinguished between clusters of different bases by detecting the ions in a time-of-flight mass spectrometer.

IR–UV hole-burning is a highly sensitive structural probe because it allows us to measure OH and NH stretch frequencies that shift significantly upon hydrogen bonding. In the hole-burning technique, we employed an IR “burn” laser pulse followed after 100 ns by a UV “probe” laser pulse. The latter was tuned to a resonance in the  $S_0 \leftarrow S_1$  transition and led to ion formation after absorption of a second UV photon, either from the same laser (one-color REMPI) or from another laser (two-color REMPI). When the IR laser was tuned to a resonance, it depleted the ground state, which caused a decrease of the ion signal. In this way we could record ground-state IR spectra of mass-selected and optically selected clusters. Furthermore, even when the UV absorption was broad, it was possible to obtain sharp IR spectra with this technique.

Calculations were performed at the approximate resolution of identity by means of the MP2 (RI-MP2) method with the TZVPP ([5s3p2d1f/3s2p1d]) basis set [TURBOMOLE 5.6 (Quantum Chemistry Group, Karlsruhe, Germany)]. Final stabilization energies were corrected for the basis set superposition error and the deformation energies of the monomers. The presented stabilization energies were close to the true stabilization energies calculated at the CCSD(T) complete basis set limit (12). Harmonic vibration frequencies were calculated at the RI-MP2/TZVPP level and were scaled by the factor 0.956, determined from the comparison of theoretical and experimental spectra of isolated guanine and cytosine.

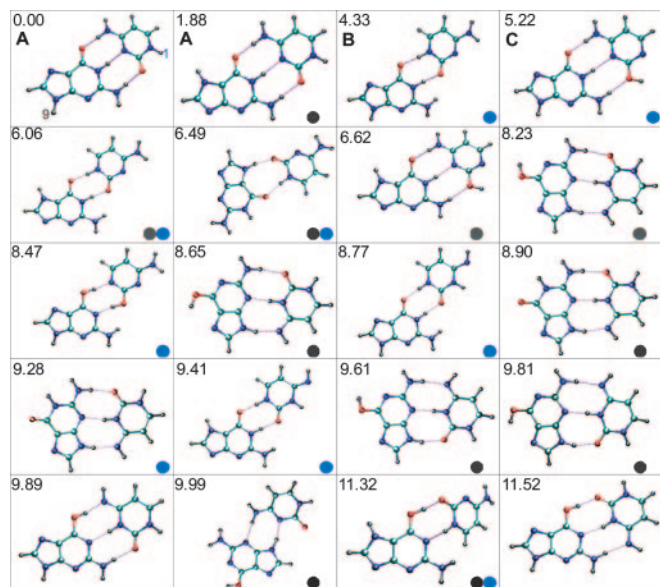
## Results

Fig. 1 shows the 20 lowest-energy structures as calculated for the GC base pair. For the adenine–thymine base pair (AT), these calculations do not predict the WC structure to be the lowest in energy. For AT, the global minimum is more stable by  $\approx 3$  kcal/mol and contains the H-bond pattern between the N3 atom of adenine and the N1 atom of thymine and between the N9 atom of adenine and the O2 atom of thymine (13). The difference between the global minimum and the WC structure becomes

Abbreviations: REMPI, resonance enhanced multiphoton ionization; WC, Watson–Crick.

<sup>§</sup>To whom correspondence should be addressed. E-mail: devries@chem.ucsb.edu.

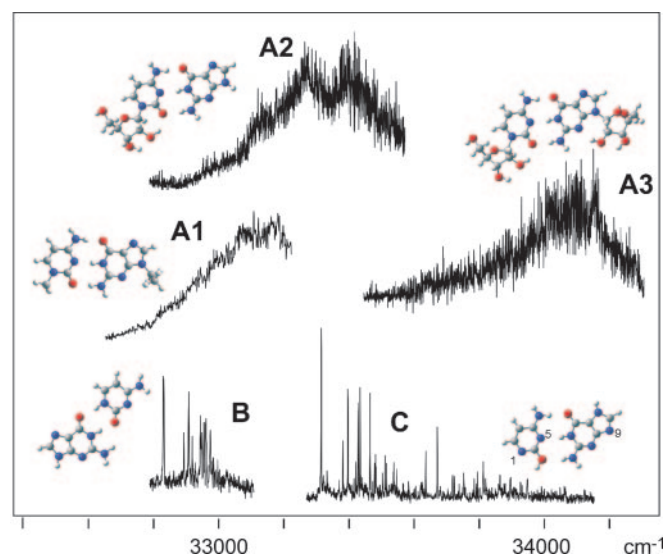
© 2004 by The National Academy of Sciences of the USA



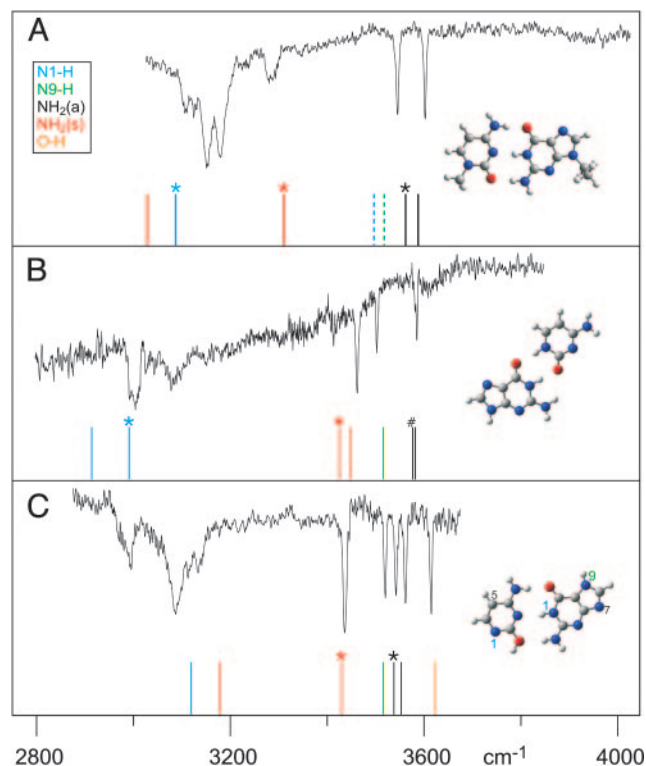
**Fig. 1.** Twenty lowest-energy hydrogen-bonded GC structures. Stabilization energy for each is indicated in kcal/mol relative to that of the lowest-energy configuration, the WC structure (11). Blue circles indicate structures that are not possible with 1-substituted cytosine, and gray circles indicate structures that are not possible with 9-substituted guanine.

even higher at the free-energy surface, and the WC structure is populated insignificantly at room temperature.

Fig. 2 shows the UV-REMPI spectra of five different base-pair structures. Two of these spectra exhibit sharp spectra with many peaks (B and C), whereas the other three (A1–3) are very broad. We show below that the broad spectra are due to WC base pairs, whereas the sharp spectra are not. Spectra A1–3 were obtained by desorbing different mixtures of substituted bases and were recorded at their respective cluster masses as follows: A1, 9-ethylguanine with 1-methylcytosine; A2, cytosine with gua-



**Fig. 2.** REMPI spectra recorded at the mass of guanine-cytosine (B and C), 9-ethylguanine-1-methylcytosine (A1), guanine-cytidine (A2), and guanosine-cytidine (A3). All spectra are one-color, two-photon spectra, except B, which is a two-color spectrum with 266 nm for the ionization laser. All spectra were recorded at the parent mass of the respective cluster. The cluster structures in the figure were optimized at the DFT (B3LYP/6-311G\*\*) level.



**Fig. 3.** IR-UV hole-burning spectra of structures A1, B, and C from Fig. 2 (A–C). The vertical bars are frequencies obtained from *ab initio* calculations (15). The colors indicate the respective modes. When there are two possibilities, asterisks indicate modes that reside on the guanine. “#” indicates two  $\text{NH}_2$  antisymmetric stretches that are within  $7 \text{ cm}^{-1}$  from each other. (a), antisymmetric; (s), symmetric.

nine; A3, cytosine with guanosine. Spectrum B was recorded with two-color REMPI (the ionizing laser was at 266 nm), whereas all other spectra were recorded with one-color REMPI. We also scanned the wavelength range from  $31,000\text{--}35,000 \text{ cm}^{-1}$  with two-color REMPI to ascertain that we did not miss any structures because of an ionization potential of more than twice the one-color photon energy. Spectra B and C both were obtained by desorbing a mixture of guanine and cytosine and were recorded at the mass of the guanine-cytosine cluster. We know that these are spectra of two different cluster isomers, based on UV-UV hole-burning measurements.

Fig. 3 shows the resulting IR spectra, recorded with the UV laser tuned to resonant wavelengths in spectra A1, B, and C of Fig. 2 and with the corresponding masses gated in the mass spectrometer. The figure compares the spectra with calculated vibrational frequencies for specific cluster structures. Spectrum A in Fig. 3 exhibits an excellent fit with the calculated IR frequencies for the WC structure. The mismatch for N1-H and symmetric  $\text{NH}_2$  frequencies can be ascribed to hydrogen-bonding effects, consistent with the width of the IR peaks. However, the qualitative feature of the red-shift of several hundred wave numbers due to hydrogen bonding of the specific hydrogen is well reproduced. The dotted lines indicate the frequencies that would have been expected for the N1-H stretch in cytosine and the N9-H stretch in guanine, had those positions not been blocked. We also calculated the frequencies for the base pair including the substituents at a lower level [DFT (B3LYP/6-311G\*\*)] and found them to be virtually the same as in the higher-level calculation for the unsubstituted canonical pair. Furthermore, no OH stretch frequency is observed in this spectrum. We can therefore conclude that both bases must be in

the keto form, rather than in the enol tautomeric form. This conclusion also was confirmed by a preliminary measurement on guanine–cytosine (A2) in the C–O stretch frequency region at the FELIX free electron laser facility of the FOM Institute in Rijnhuizen, The Netherlands. The substituents further restrict possible structures even among keto–keto combinations, blocking, for example, structures of the form of those we assign to spectrum B, as shown below. Of the remaining possible assignments, all non-WC ones are much higher in energy than the WC structure (compare Fig. 1) and can clearly be excluded based on the IR frequencies. We do not compare intensities, because, in our hole-burning experiments, the IR transitions are saturated.

Based on the IR data, we can assign spectrum B to a doubly hydrogen-bonded structure, as shown in Fig. 3. We assign spectrum C to a triply hydrogen-bonded structure very similar to WC but with cytosine in the enol form, rather than in the keto form. For structures B and C, methylation of cytosine in the 5 position should have no effect, and indeed the REMPI spectra of clusters of G with 5-methyl-C (not shown) closely resemble those for B and C. However, substitution in the 1 position blocks both these structures. We note that spectra A1–3 in Fig. 2 all involve cytosine with a substituent in the 1 position.

## Discussion

The broad UV spectra corresponding to the WC structures are several hundred wave numbers wide, whereas the sharp UV spectra of the non-WC clusters (spectra B and C) consist of peaks with widths  $<1$  wave number (the laser band width is  $\approx 0.3$   $\text{cm}^{-1}$ ). The difference, if solely due to lifetime broadening, corresponds to lifetimes of the order of  $10^{-14}$  s for the WC structures and  $>10^{-12}$  s for the non-WC structures. The intensities are also significantly lower for the broader spectra. This finding is consistent with a shorter intermediate state lifetime, which reduces two-photon absorption efficiency. In fact, the signal for A2 and A3 was too weak to obtain IR–UV hole-burning spectra with sufficient signal to noise. It is possible that this effect forms the reason that, as yet, we have not observed the WC structure with the unsubstituted bases. However, a quantitative comparison of intensities is difficult, because we do not know relative efficiencies of cluster formation. Furthermore, a good correlation between intensity and lifetime may not exist, because a short lifetime (due to internal conversion) rapidly transfers the system to the ground state with considerable vibrational energy. Therefore, it still may be possible to ionize the cluster after internal conversion, depending on the Franck–Condon overlap and on the time scale of subsequent intramolecular vibrational redistribution.

We cannot exclude possible alternative explanations for differences in line widths. Broadening can be due to an increased density of states in the excited state. Such a situation can arise from excitation high into the vibrational manifold. However, this circumstance would require a drastic change in Franck–Condon overlap, suggesting a large geometry change, which would be unusual for a molecule consisting of a polycyclic structure. The possibility should be considered that 9-substitution of guanine could affect the electronic states (14). All 9-substituted guanines that we have studied as monomers, including four different guanosines, exhibit sharp UV spectra. These all involve the enol tautomer (15–17). Thus, if the 9-substitution plays a role, it does so only for the keto tautomer, which is required to form the WC structure. So, even if 9-substitution is involved, the effect is still selective for the WC structure. In other words, whether as a result of substitution, tautomer form, hydrogen bonding, or a combination of these, the WC structure exhibits different electronic spectroscopy from all other base-pair structures we observed so far. Furthermore, we note that spectrum A2, with unsubstituted guanine, shows the same broadening as the other A structures. We have also studied GG base pairs with 9-meth-

ylguanine, for which we found the triply hydrogen-bonded keto–keto structure to exhibit a sharp UV spectrum (18).

A model that affects the lifetimes of individual bases has been proposed by several authors in which excitation to a  $\pi\pi^*$  state is followed by a curve crossing to a “dark” or “doorway” state, which, in turn, couples with the ground state, providing a pathway for rapid internal conversion (4, 8, 19, 20). This process presumably provides photochemical stability by preventing population of the reactive triplet state. A similar process can play a role in base pairs, albeit possibly involving different excited states. Recently, Sobolewski and Domcke (27) have reported detailed computations for the GC base pair in the WC configuration, identifying two doorway states that can mediate the internal conversion. One of those is a  $^1\pi\pi^*$  state that can be thought of as a proton transfer state (10). According to Sobolewski and Domcke’s calculations, as the N2 hydrogen moves to the other base together with an out-of-plane deformation of the cytosine, the ground state potential increases, creating a conical intersection with the doorway state at an increased internuclear distance. Two conical intersections are involved: the first coupling  $S_1$  and the doorway state and the second coupling the doorway state to the ground state. Critical to this process is the barrier height for reaching the first conical intersection (between  $S_1$  and the doorway state). If this intersection is above the excitation energy, we should expect a discrete spectrum and a long lifetime; in the reverse situation there would be a broad spectrum associated with rapid transfer to the doorway state. Thus, in this model, a small change in the relative potential energy, of the order of a few kcal/mol, which changes the barrier height, can make a difference of orders of magnitude in lifetime.

A precedent for the lifetime depending on the barrier height can be found in adenine, where REMPI spectroscopy shows a very narrow structured spectrum followed by a continuum to the blue. This observation suggests that the barrier to the first conical intersection is very small but nonzero. The doorway state in the case of adenine is of either  $n\pi^*$  or  $\pi\sigma^*$  character. The latter is associated with motion along an NH coordinate (10, 20, 21). In contrast to adenine (which is 6-aminopurine), 2-aminopurine has a lower  $S_1$  state (as witnessed by a red-shifted and extensively structured spectrum) and has a very long fluorescence lifetime, corresponding to a high barrier. We note that Kang *et al.* (22) observed a lifetime shortening for adenine upon clustering with water molecules (but not for adenine dimers), whereas He *et al.* (23) observed a similar effect for thymine. It seems possible that solvation of a single base by water has similar effects on the electronic states as solvation by a complementary base.

The relevance of gas phase studies of base-pair interactions is twofold. First, these are, at the moment, the only available experimental data for testing the highest level theoretical modeling and calculations. Secondly, gas phase studies are the only way to scrutinize the details of the interactions between individual base pairs, because in solution one cannot obtain data on individual structures. In fact, in aqueous solution, G and C tend to predominantly form a stacked structure, with a small fraction in a WC hydrogen-bonded structure (24). Nevertheless, the question is what chemistry is possible when hydrogen-bonded structures do form, which must happen at some point in order to lead to a replication mechanism. In a recent study, Woutersen and Cristalli (25) forced A–U pairs into a single WC structure by using substituted bases and a different solvent. They found that excitation of the NH bond that is involved in the hydrogen bonding leads to a strongly enhanced vibrational relaxation. This observation would be consistent with a model in which that particular coordinate would be excited by the internal conversion. To further study the role of the solvent, we have started experiments on clusters of bases and base pairs with water. In these experiments, we will in principle be able to study the role of water as a function of the number of added solvent molecules. Preliminary results suggest a lifetime shortening of guanine mono-

mers upon addition of two or more water molecules and little effect on GG dimers, even after addition of six or more water molecules. We also have started experiments with single photon ionization, to compare ionization efficiencies when bypassing the intermediate state. Other experiments that would be helpful in further testing possible models include femtosecond pump-probe experiments to attempt to measure the excited state lifetime directly and calculations of excited state lifetimes of different structures.

### Summary

We have recorded so far 27 different base-pair combinations and structures (26). Of these, 24 exhibit sharp UV spectra. Only the

structures assigned as WC have broad UV spectra. Whatever the mechanism, the central observation is the special photochemistry for the WC structure. This finding suggests that, when constructing possible scenarios for prebiotic chemistry, the photochemistry of the different structures may need to be taken into account. This finding also raises the intriguing possibility that, under the conditions of the early earth, there was a special stability for the very structure that became key to the recognition mechanism for transfer of the genetic code.

This work is based on research supported by National Science Foundation Grant CHE-0244341 and Grant 203/05/0009 (to P.H. and M.K.) from the Grant Agency of the Czech Republic.

1. Daniels, M. & Hauswirth, W. (1971) *Science* **171**, 675–677.
2. Callis, P. R. (1983) *Annu. Rev. Phys. Chem.* **34**, 329–357.
3. Peon, J. & Zewail, A. H. (2001) *Chem. Phys. Lett.* **48**, 255–262.
4. Kang, H., Jung, B. & Kim, S. K. (2003) *J. Chem. Phys.* **118**, 6717–6719.
5. Luhrs, D. C., Viallon, J. & Fischer, I. (2001) *Phys. Chem. Chem. Phys.* **3**, 1827–1831.
6. Crespo-Hernandez, C. E., Cohen, B., Hare, P. M. & Kohler, B. (2004) *Chem. Rev.* **104**, 1977–2019.
7. Kang, H. Y., Jung, B. Y. & Kim, S. K. (2003) *J. Chem. Phys.* **118**, 11336–11336.
8. Broo, A. (1998) *J. Phys. Chem. A* **102**, 526–531.
9. Nir, E., Kleiner-mann, K. & de Vries, M. S. (2000) *Nature* **408**, 949–951.
10. Sobolewski, A. L. & Domcke, W. (2004) *Phys. Chem. Chem. Phys.* **6**, 2763–2771.
11. Meijer, G., de Vries, M. S., Hunziker, H. E. & Wendt, H. R. (1990) *Appl. Phys. B* **51**, 395–403.
12. Jurecka, P. & Hobza, P. (2003) *J. Am. Chem. Soc.* **125**, 15608–15613.
13. Kabelac, M. & Hobza, P. (2001) *J. Phys. Chem. B* **105**, 5804–5817.
14. Langer, H. & Doltsinis, N. L. (2003) *Phys. Chem. Chem. Phys.* **5**, 4516–4518.
15. Chin, W., Mons, M., Dimicoli, I., Piuze, F., Tardivel, B. & Elhanine, M. (2002) *Eur. Phys. J. D* **20**, 347–355.
16. Mons, M., Dimicoli, I., Piuze, F., Tardivel, B. & Elhanine, M. (2002) *J. Phys. Chem. A* **106**, 5088–5094.
17. Nir, E., Hunig, I., Kleiner-mann, K. & de Vries, M. S. (2004) *Chemphyschem* **5**, 131–137.
18. Nir, E., Janzen, C., Imhof, P., Kleiner-mann, K. & de Vries, M. S. (2002) *Phys. Chem. Chem. Phys.* **4**, 740–750.
19. Sobolewski, A. L. & Domcke, W. (1999) *Chem. Phys. Lett.* **300**, 533–539.
20. Sobolewski, A. L. & Domcke, W. (2002) *Eur. Phys. J. D* **20**, 369–374.
21. Sobolewski, A. L., Domcke, W., Dedonder-Lardeux, C. & Jouvet, C. (2002) *Phys. Chem. Chem. Phys.* **4**, 1093–1100.
22. Kang, H., Lee, K. T. & Kim, S. K. (2002) *Chem. Phys. Lett.* **359**, 213–219.
23. He, Y. G., Wu, C. Y. & Kong, W. (2004) *J. Phys. Chem. A* **108**, 943–949.
24. Kabelac, M. & Hobza, P. (2001) *Chemistry: A Eur. J.* **7**, 2067–2074.
25. Woutersen, S. & Cristalli, G. (2004) *J. Chem. Phys.* **121**, 5381–5386.
26. Nir, E., Plutzer, C., Kleiner-mann, K. & de Vries, M. (2002) *Eur. Phys. J. D* **20**, 317–329.
27. Sobolewski, A. L. & Domcke, W. (2004) *Phys. Chem. Chem. Phys.* **6**, 2763–2771.