

# Base pair analogs in the gas phase

Joseph R. Roscioli and David W. Pratt\*

Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260

Edited by F. Fleming Crim, University of Wisconsin, Madison, WI, and approved September 25, 2003 (received for review June 20, 2003)

**A rotationally resolved electronic spectrum of the gas-phase dimer 2-aminopyridine-2-pyridone, an analog of the adenine-thymine base pair, has been observed and assigned, leading to precise measurements of its moments of inertia and preliminary determinations of its structure. A Watson-Crick configuration results, with N $\cdots$ H—N and N—H $\cdots$ O hydrogen bond lengths of 2.898 and 2.810 Å, respectively. The two bases are found not to be coplanar; a dihedral angle of 6.1° between the base planes is also estimated from the measured moments of inertia. Possible chemical and biological implications of these results are discussed.**

Fifty years have elapsed since x-ray diffraction photographs of fibers pulled from concentrated solutions of DNA revealed that its structure is a double-stranded helix held together by hydrogen bonds between the complementary base pairs adenine (A) and thymine (T), and guanine (G) and cytosine (C) (1–3), providing a molecular-level explanation of heredity (4). Despite this passage of time, the structure of an *isolated* A·T or G·C base pair in the absence of solvent and/or a surrounding medium has yet to be determined. Only the average structure of the base pairs in the condensed phase can be determined from fiber-diffraction patterns of DNA. To be sure, we now know that many different kinds of DNA exist and that, although they share many common features, they can differ significantly in detail. These differences include the number of residues per turn, the translation per residue, the angle between the base planes and the helix axis, and the dihedral angle between base planes (5). We also know that these features may be functionally important, both in gene expression and in the construction of the cell (6), but we do not know the extent to which the differences in DNA structures are a consequence of the condensed-phase environment in which they are found. Gas-phase studies are needed not only to provide benchmark structures but also to quantify the relative importance of intra- and intermolecular interactions in molecular biology.

Toward this end, we report here a preliminary determination of the structure of the base-pair *mimic* 2-aminopyridine-2-pyridone (2AP·2PY) by using the technique of rotationally resolved electronic spectroscopy in the gas phase. 2AP·2PY is found to be held together by two N $\cdots$ H—N and N—H $\cdots$ O hydrogen bonds, identical with those found in the A·T base pair itself. As will be seen, understanding the nature of these interactions provides insight into the structure, stability, and electronic properties of the bonds that hold DNA together.

## Methods

Our experiments were performed in the collision-free environment of a molecular beam (7). Approximately 1 g of 2AP was placed in a quartz sample container at nearly 200°C and entrained in 41 kPa (6 psig) of He. After equilibrium was established,  $\approx$ 2 g of 2PY was added to the sample container, and the spectrum of the 2AP·2PY dimer appeared shortly thereafter. During this time, the mixture was expanded through a 280- $\mu$ m quartz nozzle held at 150°C into a vacuum chamber, skimmed about 2 cm downstream of the nozzle before entering a second differentially pumped chamber, and probed 10 cm downstream of the nozzle with a high-resolution laser (full width at half-maximum  $\approx$  1 MHz in the UV region). The excitation source was an argon-ion-pumped continuous-wave single-frequency, tun-

able, ring dye laser operating with 4-dicyanomethylene-2-methyl-*b*-(*p*-dimethylaminostyryl)-4H-pyran (DCM) and frequency doubled by an intracavity LiIO<sub>3</sub> crystal. Typical powers used were 400  $\mu$ W in the UV region. The spectra were acquired at an acquisition rate of 50 Hz over a 2,000-s scan (step size,  $\approx$ 0.5 MHz). Four signals were collected. The photomultiplier tube-detected fluorescence signal was collected with spatially selective optics by using photon counting and stored on a data-acquisition computer. A signal from a near-confocal interferometer having a mode-matched free spectral range of 599.5040  $\pm$  0.005 MHz in the UV region was collected to perform relative-frequency calibration. The absorption spectrum of I<sub>2</sub> was collected to determine the absolute transition frequencies, which are accurate to  $\pm$ 18 MHz. Finally, the power signal was collected to normalize the photomultiplier-tube signal.

Supplementing the experimental work, *ab initio* calculations at ground-state B3LYP/6-31G+(d,p) and excited-state CIS/6-31G levels were performed by using the GAUSSIAN 98 (8) suite of electronic structure programs on personal computers.

## Results

Previous studies of this type on isolated nucleic acid bases, albeit at significantly lower resolution, have faced serious obstacles. The pyrimidine bases C and T exhibit broad absorption bands and a near-total absence of fluorescence, even at low temperature or in supersonic jets (9–11), apparently because of rapid nonradiative decay of the excited S<sub>1</sub> state (12). But vibrationally resolved UV spectra have recently been measured for A, G, and the G·G and G·C dimers in the gas phase, by using resonant two-photon ionization (R2PI) detection techniques in supersonic jets (13–15). And Müller *et al.* (16) have recently used both R2PI and fluorescence-based methods to detect S<sub>1</sub>–S<sub>0</sub> vibronic spectra of the 2AP·2PY base-pair analog described here.

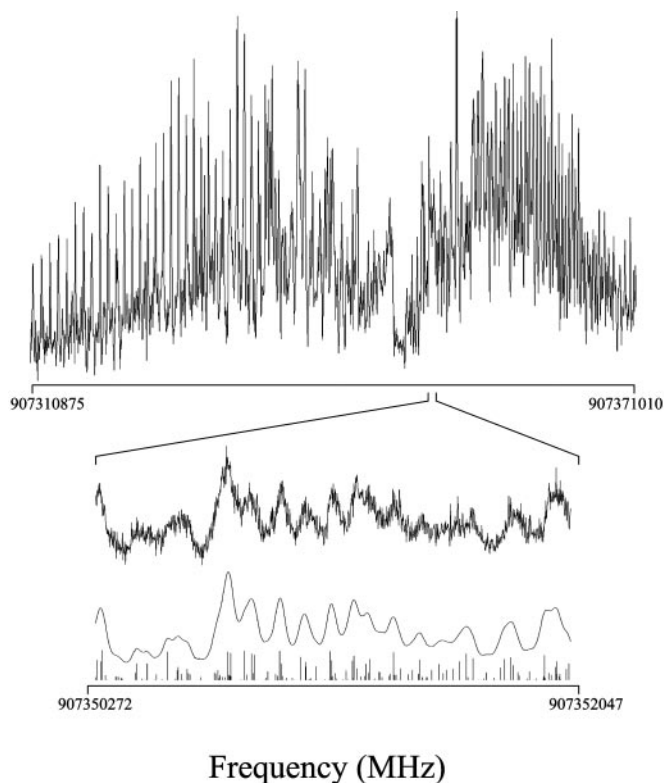
The high-resolution fluorescence excitation spectrum of the presumed S<sub>1</sub>–S<sub>0</sub> electronic origin of the 2AP·2PY dimer is shown in Fig. 1. This spectrum was both frequency- and intensity-fit, with an observed-minus-calculated standard deviation of 3.5 MHz. It was obtained at a relatively high temperature ( $\approx$ 7 K) and therefore contains >15,000 transitions. Because of this congestion, one peak corresponding to a single rovibronic transition could not be identified within the spectrum. However, the intensity fit reveals a Lorentzian linewidth of 38  $\pm$  3 MHz, corresponding to a fluorescence lifetime of 4.2  $\pm$  0.3 ns. The frequency fit of the spectrum yields rotational constants that are conservatively estimated to be accurate to one-tenth of this width or 0.4 MHz. These constants are listed in Table 1, together with the calculated [B3LYP/6-31G+(d,p)] rotational constants of the 2AP·2PY dimer. The theoretical values for 2AP·2PY are all within 3% of the experimental ones, confirming its structure. 2AP·2PY has the Watson-Crick configuration, with two N $\cdots$ H—N and N—H $\cdots$ O hydrogen bonds joining the monomer units together, exactly as in the A·T base pair found in naturally occurring DNA.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TM, transition moment; 2AP, 2-aminopyridine; 2PY, 2-pyridone.

\*To whom correspondence should be addressed. E-mail: pratt@pitt.edu.

© 2003 by The National Academy of Sciences of the USA



**Fig. 1.** The central  $\approx 2\text{-cm}^{-1}$  portion of the rotationally resolved fluorescence excitation spectrum of the origin band in the  $S_1$ - $S_0$  electronic spectrum of the 2AP-2PY dimer at 330.3947 nm, recorded in the collision-free environment of a molecular beam by using a high-resolution laser. Three scale expansions below illustrate a simulated spectrum at full resolution (1 MHz) and fits of the experimental spectrum at the experimental resolution of 38 MHz.

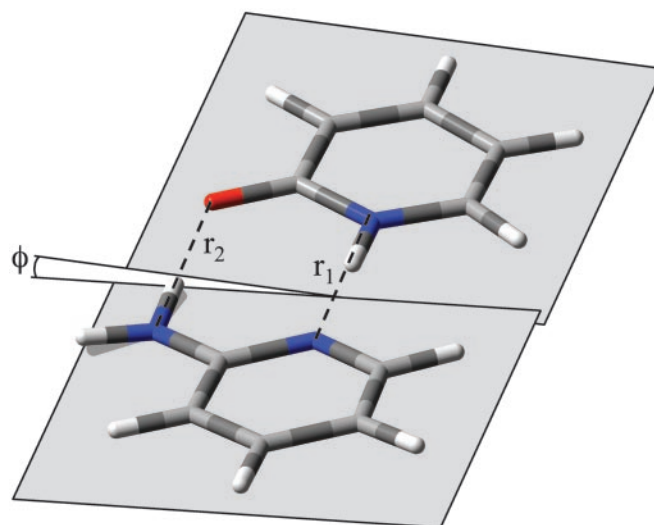
The intensity fit of the spectrum in Fig. 1 also shows that the  $S_1$ - $S_0$  electronic transition moment (TM) makes an angle of  $\pm 62 \pm 3^\circ$  with the  $a$  axis of the dimer. Because the ratio of  $a$ -type intensity and  $b$ -type intensity depends on the square of the TM, only its relative orientation can be determined experimentally. However, if the TM of the 2PY monomer (17) is placed within the inertial axis frame of the dimer, one of the two possible TM orientations is found to coincide to within  $8^\circ$  of that of the monomer. Thus, we conclude that TM orientation in the dimer is rotated by  $62^\circ$  in a counterclockwise direction, away from its  $a$  axis, and that its  $S_1$ - $S_0$  transition is highly localized on the 2PY "side" of the molecule. In agreement with this, the electronic

**Table 1. Inertial parameters of the zero-point vibrational levels of the ground ( $S_0$ ) and first excited ( $S_1$ ) singlet states of the A-T base-pair mimic 2AP-2PY, based on a fit of its fluorescence excitation spectrum at  $\approx 330$  nm (the band origin is at  $30,266.83 \pm 0.02\text{ cm}^{-1}$ )**

Parameter	Experimental		Theoretical*
	$S_0$ state	$S_1$ state	$S_0$ state
A, MHz	1,602.8 (4)	1,596.9 (4)	1,602.1
B, MHz	341.0 (4)	336.9 (4)	332.6
C, MHz	281.5 (4)	278.5 (4)	275.8
$\Delta I$ , $\text{amu}\cdot\text{\AA}^2$	-2.11	-2.19	
$\kappa$	-0.910	-0.911	

A, B, C,  $\Delta I$ , and  $\kappa$  are the three rotational constants, inertial defect, and asymmetry parameter, respectively. amu, Atomic mass units.

\*B3LYP/6-31G+(d,p) values.



**Fig. 2.** Model structure of the 2AP-2PY dimer illustrating the Watson-Crick configuration found experimentally and geometrical parameters derived from the fit of the spectrum in Fig. 1.  $r_1$  and  $r_2$  are the hydrogen bond distances  $\text{N}\cdots\text{H}-\text{N}$  and  $\text{N}-\text{H}\cdots\text{O}$ , and  $\phi$  is the (nonzero) dihedral angle.

origin of 2AP-2PY is shifted by only  $\approx 400\text{ cm}^{-1}$  with respect to that of 2PY itself (17).

Still, the fact that the TMs of 2PY and 2AP-2PY are not exactly aligned suggests some rearrangement of electrons on dimer formation. *Ab initio* calculations support this view. The B3LYP/6-31G+(d,p) and CIS/6-31G calculations show that only the highest occupied and lowest unoccupied molecular orbitals of 2PY-2AP play a significant role in its  $S_1$ - $S_0$  transition. These two molecular orbitals are largely localized on 2PY, but both are partially delocalized onto the 2AP side of the molecule. Covalent interactions between the two monomer units is one possible explanation of this behavior (18). Another is state mixing; vibronic coupling of the  $S_1$  state (localized on 2PY) and the  $S_2$  state (localized on 2AP) could lead to the observed rotation of the TM.

## Discussion

Three rotational constants are clearly inadequate to determine the structure of a large molecule like 2AP-2PY in either electronic state. However, these constants are primarily sensitive to the separation of the two monomer units and to the dihedral angle between the two base planes; they are only secondarily sensitive to the detailed geometry of the two rings. The rotational constants of the two monomer units have been determined by both microwave (19, 20) and high-resolution UV methods (21), and these are well reproduced by *ab initio* calculations. Therefore, we base our preliminary determination of the ground-state geometry of 2AP-2PY on the [B3LYP/6-31G+(d,p)] theoretical structures of 2AP and 2PY and on the further assumption that these are unchanged on complex formation. A coordinate system was then defined that represents the approach of the two monomer units along the  $\text{N}\cdots\text{H}-\text{N}$  direction, as shown in Fig. 2.  $r_1$  is the (heavy-atom) distance between the two nitrogen atoms,  $r_2$  is the analogous distance between the amino nitrogen and oxygen atoms, and  $\phi$  is the torsional (or C-N-N-C dihedral) angle between the two planes. Unique values of these three parameters were then obtained by varying them until the calculated values of A, B, and C matched the observed ones, within experimental error. This analysis yields the hydrogen bond lengths  $r_1$  ( $\text{N}\cdots\text{H}-\text{N}$ ) =  $2.898 \pm 0.002\text{ \AA}$  and  $r_2$  ( $\text{N}-\text{H}\cdots\text{O}$ ) =  $2.810 \pm 0.004\text{ \AA}$  and the dihedral angle  $\phi = 6.1 \pm 0.2^\circ$ .

**Table 2. Hydrogen bond lengths, in Å, in 2AP·2PY and related systems**

Bond	2AP·2PY		2AP·2AP Cryst (22)	2PY·2PY Cryst (23)	SAU Cryst (24)
	Gas phase	Calc (16)			
N···H—N	2.90	2.94	3.07	—	2.82
N—H···O	2.81	2.88	—	2.76	2.95

Calc, calculated; Cryst, crystal; SAU, sodium adenyl-3',5'-uridine.

Table 2 compares the experimental values of the hydrogen bond lengths in ground-state 2AP·2PY with theoretical estimates (16) and with experimental values for the 2AP dimer (22), the 2PY dimer (23), and the Watson–Crick A·T base pair found in sodium adenyl-3',5'-uridine (SAU) by Seeman *et al.* (24). The latter experimental values are based on x-ray diffraction studies of crystals. Theory gives a good account of the properties of the hydrogen bonds in 2AP·2PY. At the B3LYP/6-311++G (2d,2p) level (16), the two (heavy-atom) distances are quite different, as observed in the gas phase. The theoretical values of  $r_1$  and  $r_2$  are 0.04 and 0.07 Å longer, respectively. [Symmetry (or the lack thereof) seems to play an important role in those interactions, because the “asymmetric” dimer 2-hydroxypyridone·2PY also has N···H—N and O—H···O bonds of significantly different lengths (2.86 and 2.45 Å, respectively) (ref. 25; unpublished data).] Experimentally, the N···H—N bond in 2AP·2PY is found to be 0.17 Å shorter than that in 2AP·2AP, but 0.08 Å longer than that in SAU. The N—H···O bond in 2AP·2PY is 0.05 Å longer than that in 2PY·2PY but 0.14 Å shorter than that in SAU. These data suggest that crystal-packing effects on hydrogen bond lengths are significant.

A survey of 88 crystal structures with similar hydrogen-bonding motifs in the Cambridge Structural Database (26) reveals average bond lengths of  $r_1$  (N···H—N) =  $2.92 \pm 0.10$  Å and  $r_2$  (N—H···O) =  $3.00 \pm 0.09$  Å. The gas-phase value of  $r_1$  (2.90 Å) lies well within the distribution of x-ray values, but the gas-phase value of  $r_2$  (2.81 Å) lies well outside this distribution.

More surprising, perhaps, is the experimental finding [supported by theory (16)] that 2AP·2PY is not a planar molecule in either electronic state. Most models of DNA assume that the complementary base pairs occupy a common plane. But, in 2AP·2PY, at least, this is not the case, as is immediately apparent from the inertial defects ( $\Delta I = I_c - I_a - I_b$ ) that are derived from

the fit of its high-resolution spectrum. Both electronic states of 2AP·2PY have  $\Delta I$  values on the order of  $-2 \mu\text{Å}^2$ , indicating a significant deviation from planarity. (Values of  $\Delta I \approx 0$  are expected for a planar molecule.) The calculated ground-state structure of 2AP·2PY has  $\Delta I \approx -2.17 \mu\text{Å}^2$ . Examination of this structure suggests that the source of the nonplanarity is the NH<sub>2</sub> group of 2AP. Ground-state 2AP has  $\Delta I \approx -0.26 \mu\text{Å}^2$  (21), supporting this view. Indeed, aniline itself is a nonplanar molecule (27). Recent measurements of the vibrational TM angles in adenine suggest that it, too, is nonplanar in the gas phase with the NH<sub>2</sub> group tilted, 20°, out-of-plane (28). Thus, the statement that “clear, direct experimental evidence about the nonplanarity of isolated bases is still missing due to the (lack of) resolution of the available experimental techniques” (ref. 29, p. 3256) is now moot.

The biological consequences of this nonplanarity may be significant. Some interstrand NH<sub>2</sub>-group contacts in B-DNA crystal structures appear to be stabilized by NH<sub>2</sub>-group pyramidalization and interstrand bifurcated hydrogen bonds (29). In 2AP itself, Kydd and Mills (20) established that the angle between the plane made by the NH<sub>2</sub> group and the ring is nearly 32° in the gas phase. In the crystal structure (22), this angle is reduced to  $\approx 15^\circ$ . This difference indicates that the geometry of the NH<sub>2</sub> group highly depends on the local environment of the molecule. The flexibility offered by the NH<sub>2</sub> group is advantageous for the dimer, because the NH<sub>2</sub> group will most likely deform in such a way that the N—H···O bond is linear, creating an overall structure that is bound by two linear hydrogen bonds by an induced fit.

More globally, an incipient chirality apparently is created when a planar base connects to a nonplanar one, giving the base pair a right- or left-handed sense, on its own. It is believed that helices form when a series of subunits bind to each other in a regular way to take advantage of a  $\pi$ -stacking of the bases. Is it possible that the *direction* of the helix twist can be traced to the nonplanar character of the purine bases? Could the degrees of this nonplanarity (and twist) be influenced by an attached solvent molecule or other environmental factors? And are the properties of the real A·T base pair similar to those of the mimic described here? Further experiments and calculations will be necessary to answer these intriguing questions.

We thank J. M. Rosenberg, P. E. Siska, and G. A. Stewart for their encouragement. This work was supported by National Science Foundation Grant CHE-9987048.

1. Watson, J. D. & Crick, F. H. C. (1953) *Nature* **171**, 737–738.
2. Wilkins, M. H. F., Stokes, A. R. & Wilson, H. R. (1953) *Nature* **171**, 738–740.
3. Franklin, R. E. & Gosling, R. G. (1953) *Nature* **171**, 740–741.
4. Watson, J. D. & Crick, F. H. C. (1953) *Nature* **171**, 964–967.
5. Stryer, L. (1995) *Biochemistry* (Freeman, New York), 4th Ed.
6. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1994) *Molecular Biology of the Cell* (Garland, New York).
7. Majewski, W. A., Pfanstiel, J. F., Plusquellic, D. F. & Pratt, D. W. (1995) in *Laser Techniques in Chemistry*, eds Myers, A. B. & Rizzo, T. R. (Wiley, New York), pp. 101–148.
8. Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E., Burant, J. C., *et al.* (1998) GAUSSIAN 98 (Gaussian, Pittsburgh), Revision A.4.
9. Fujii, M., Tamura, T., Mikami, N. & Ito, M. (1986) *Chem. Phys. Lett.* **126**, 583–587.
10. Tsuchiya, Y., Tamura, T., Fujii, M. & Ito, M. (1988) *J. Phys. Chem.* **92**, 1760–1765.
11. Brady, B. B., Peteanu, L. A. & Levy, D. H. (1988) *Chem. Phys. Lett.* **147**, 538–543.
12. Kang, H., Lee, K. T., Jung, B., Ko, Y. J. & Kim, S. K. (2002) *J. Am. Chem. Soc.* **124**, 12958–12959.
13. Nir, E., Grace, L., Brauer, B. & de Vries, M. S. (1999) *J. Am. Chem. Soc.* **121**, 4896–4897.
14. Nir, E., Janzen, Ch., Imhof, P., Kleinerhmanns, K. & de Vries, M. S. (2001) *J. Chem. Phys.* **115**, 4604–4611.
15. Nir, E., Kleinerhmanns, K. & de Vries, M. S. (2000) *Nature* **408**, 949–951.
16. Müller, A., Talbot, F. & Leutwyler, S. (2002) *J. Am. Chem. Soc.* **124**, 14486–14494.
17. Held, A., Champagne, B. B. & Pratt, D. W. (1991) *J. Chem. Phys.* **95**, 8732–8743.
18. Jeffrey, G. A. (1997) *An Introduction to Hydrogen Bonding* (Oxford Univ. Press, Oxford).
19. Hatherly, L. D., Brown, R. D., Godfrey, P. D., Pierlot, A. P., Caminati, W., Damiani, D., Melandri, S. & Favero, L. B. (1993) *J. Phys. Chem.* **97**, 46–51.
20. Kydd, R. A. & Mills, I. M. (1972) *J. Mol. Spectrosc.* **42**, 320–326.
21. Borst, D. R., Roscioli, J. R. & Pratt, D. W. (2002) *J. Phys. Chem. A* **106**, 4022–4027.
22. Chao, M., Schempp, E. & Rosenstein, R. D. (1975) *Acta Crystallogr. B* **31**, 2922–2924.
23. Yang, H. W. & Craven, B. M. (1998) *Acta Crystallogr. B* **54**, 912–920.
24. Seeman, N. C., Rosenberg, J. M., Suddath, F. L., Kim, J. J. P. & Rich, A. (1976) *J. Mol. Biol.* **104**, 109–144.
25. Borst, D. R., Roscioli, J. R., Pratt, D. W., Florio, G. M., Zwier, T. S., Müller, A. & Leutwyler, S. (2002) *Chem. Phys.* **283**, 341–354.
26. Allen, F. H. (2002) *Acta Crystallogr. B* **58**, 380–388.
27. Sinclair, W. E. & Pratt, D. W. (1996) *J. Chem. Phys.* **105**, 7942–7956.
28. Dong, F. & Miller, R. E. (2002) *Science* **298**, 1227–1230.
29. Hobza, P. & Sponer, J. (1999) *Chem. Rev.* **99**, 3247–3276.