ORGANIC LETTERS

2012
Vol. 14, No. 10
2560–2563

Cytosine Derivatives Form Hemiprotonated Dimers in Solution and the Gas Phase

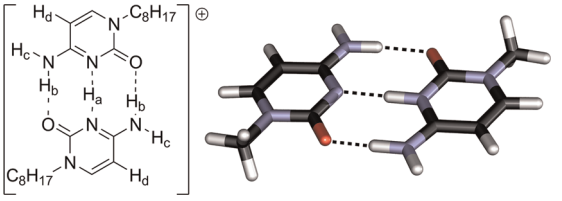
Aaron R. Moehlig, Katherine E. Djernes, V. Mahesh Krishnan, and Richard J. Hooley*

Department of Chemistry, University of California, Riverside, California 92521, United States

richard.hooley@ucr.edu

Received April 4, 2012

ABSTRACT



Hemiprotonated dimers of cytosine derivatives, implicated in the formation of the *i*-motif of DNA, have been created in solution and the gas phase. The mechanism of dimerization has been analyzed by mass spectrometry and multidimensional NMR spectroscopy.

Base pairing in DNA and RNA includes a variety of noncanonical dimers. Under suitable conditions, G-rich repeats within a DNA single strand may form helical domains containing multiple G-quadruplexes consisting of four guanine residues acting as hydrogen bond donors on the Watson−Crick face and hydrogen bond acceptors on the Hoogsteen face.[1] In double-stranded regions of DNA, separation of the G-rich tract in forming a G-quadruplex disconnects hydrogen bonding between the two antiparallel strands.[2,3] Opposite the G-rich strand in such a region, the complementary C-rich strand can also self-associate into hemiprotonated dimers and is thought to do so in promoter regions of oncogenes.[4] NMR,[5,6] X-ray diffraction analysis,[7] and circular dichroism studies[8] have shown that C-rich single strands of DNA form a four stranded complex, in which two parallel-stranded duplexes have their base pairs completely intercalated as hemiprotonated dimers, a structure called the *i*-motif.

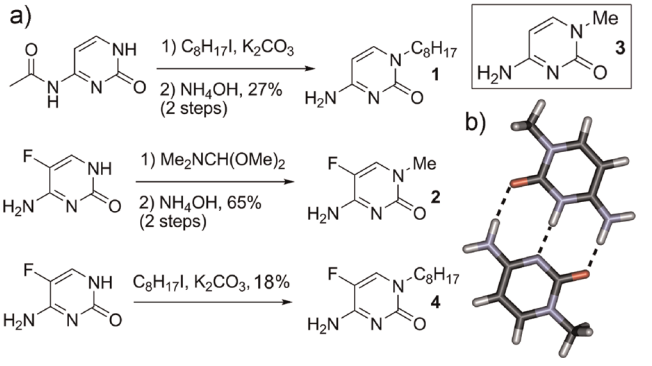


Figure 1. Synthesis of soluble cytosine variants **1**, **2**, and **4** and a minimized structure of $[\mathbf{3}\cdot\mathbf{H}^+\cdot\mathbf{3}]$ (Gaussian; 6-311++G(d,p) forcefield, counterions removed for clarity).

Observation of protonated cytosine dimer pairs *outside* self-assembling nucleotides is challenging, however. Certain charged species are capable of forming hemiprotonated dimers in aprotic solutions. Extremely strong

(1) Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263–282.
(2) Simonsson, T.; Pecinka, P.; Kubista, M. *Nucleic Acids Res.* **1998**, *26*, 1167–1172.
(3) Siddiqul-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 11593–11598.
(4) Sun, D.; Hurley, L. H. *J. Med. Chem.* **2009**, *52*, 2863–2874.
(5) (a) Gehrig, K.; Leroy, J. L.; Gueron, M. *Nature* **1993**, *363*, 561–563. (b) Lieblein, A. L.; Krämer, M.; Dreuw, A.; Fürtig, B.; Schwalbe, H. *Angew. Chem., Int. Ed.* **2012**, *51*, 4067–4070.
(6) Leroy, J. L.; Gueron, M. *Structure* **1995**, *3*, 101–120.
(7) Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. *Acta Crystallogr.* **1999**, *D55*, 422–429.
(8) Holm, A. I. S.; Nielson, L. M.; Kohler, B.; Hoffman, S. V.; Nielson, S. B. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3420–3426.

10.1021/ol300861r
Published on Web 05/09/2012

association is observed for $[AAA\text{-}DDD]^+$ pairs,[9] but lower association constants for cytosine pairs can be expected due to unfavorable secondary interactions.[10] Protonation of species displaying a DAA hydrogen bonding motif allows conversion to the corresponding DDA^+ species, which can form heterodimers with molecules displaying a suitable AAD motif. Cytosine itself can be crystallized as a hemiprotonated dimer,[11] but the solution-phase properties of this dimer are unknown. The bridging proton between two molecules of 1-methylcytosine has been reported to be shared between the two imino nitrogens in the gas-phase hemiprotonated dimer.[12] In this work, we describe the synthesis of protonated dimeric pairs of cytosine derivatives (Figure 1) and analyze their properties in the solution and gas phases.

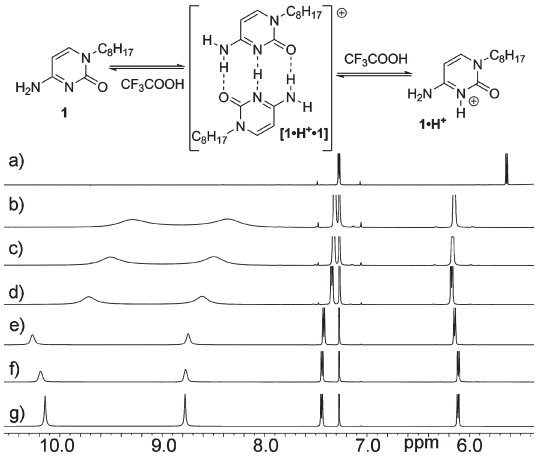


Figure 2. Solution-phase dimerization of **1** (500 MHz, $CDCl_3$, 32 mM, 298 K). NMR data of $\mathbf{[1\cdot H^+\cdot 1]}$ with differing concentrations of CF_3CO_2H: (a) **1** alone; (b) 0.40 equiv; (c) 0.50 equiv; (d) 0.60 equiv; (e) 0.70 equiv; (f) 1.50 equiv; (g) 2.00 equiv.

To confer solubility on the dimer in aprotic solvents, *N*-octylcytosine **1** was synthesized by a variant of a published procedure.[13] Octylation of commercially available N^4-acetylcytosine gives a combination of mono- and bis-alkylation products, which can be easily separated after saponification of the amide, giving clean **1** in 27% yield over two steps. The electron-withdrawn counterpart, 5-fluoro-1-octylcytosine **4**, could be accessed directly from 5-fluorocytosine, albeit in low yield. 5-Fluoro-1-methylcytosine **2** was synthesized according to literature procedures,[14] and 1-methylcytosine **3** is commercially available. DFT geometry optimizations (B3LYP/6-311++G**, Figure 1b) show that the bridging protons of $\mathbf{[2\cdot H^+\cdot 2]}$, $\mathbf{[3\cdot H^+\cdot 3]}$, and $\mathbf{[2\cdot H^+\cdot 3]}$ are not shared equally between the two imino nitrogens. The barriers for the proton to traverse from one side of the dimer to the other are 4.1, 4.4, and 5.3 kcal/mol for $\mathbf{[2\cdot H^+\cdot 2]}$, $\mathbf{[3\cdot H^+\cdot 3]}$, and $\mathbf{[2\cdot H^+\cdot 3]}$, respectively. Gas-phase IR spectra of various heterodimeric ions support the notion that the proton is situated on one side of the dimer in the gas phase.[15]

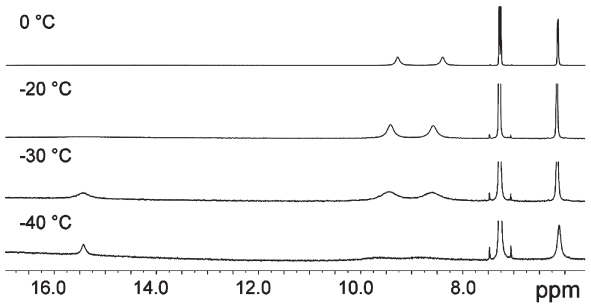


Figure 3. Downfield region of the ^1H NMR spectrum of $\mathbf{[1\cdot H^+\cdot 1]}$ at low temperatures (500 MHz, $CDCl_3$).

N-Octylcytosine **1** is easily soluble in $CDCl_3$. Upon titration of CF_3CO_2H to a solution of **1** in anhydrous $CDCl_3$, large changes in the chemical shift of the NH_2 protons were observed. The largest shifts ($\Delta\delta$ = 4.7 ppm) were observed after addition of 0.7 molar equiv of CF_3CO_2H (Figure 2g). Increased concentration of CF_3CO_2H leads to a *decreased* shift for the NH_2 protons. At room temperature, no additional proton resonance was observed. At concentrations of CF_3CO_2H above 0.7 molar equiv, $\mathbf{[1\cdot H^+\cdot 1]}$ is converted to $\mathbf{1H^+}$. To confirm the presence of $\mathbf{[1\cdot H^+\cdot 1]}$ in solution and analyze the nature of the central proton, the titration was repeated at low temperature. Upon cooling to 273 K, a broad peak was observed at ~15.5 ppm, corresponding to the central H^+ of the hemiprotonated dimer.[16] The peak sharpened considerably after cooling to 233 K (see Figure 3 and the Supporting Information).

The internal peak is sufficiently sharp at 243 K to allow low-temperature 2D NMR analysis of the dimer (Figure 4). A 2D NOESY spectrum of $\mathbf{[1\cdot H^+\cdot 1]}$ with 500 ms mixing time shows the exchange processes occurring in the

(9) (a) Blight, B. A.; Camara-Campos, A.; Djurdjevic, S.; Kaller, M.; Leigh, D. A.; McMillan, F. M.; McNab, H.; Slawin, A. M. Z. *J. Am. Chem. Soc.* **2009**, *131*, 14116–14122. (b) Djurdjevic, S.; Leigh, D. A.; McNab, H.; Parsons, S.; Teobaldi, G.; Zerbetto, F. *J. Am. Chem. Soc.* **2007**, *129*, 476–477. (c) Bell, D. A.; Anslyn, E. A. *Tetrahedron* **1995**, *51*, 7161–7172.

(10) Jorgensen, W. L.; Pranata, J. *J. Am. Chem. Soc.* **1990**, *112*, 2008–2010.

(11) (a) Kistenmacher, T. J.; Rossi, M.; Marzilli, L. G. *Biopolymers* **1979**, *17*, 2581–2585. (b) Kruger, T.; Bruhn, C.; Steinborn, D. *Org. Biomol. Chem.* **2004**, *2*, 2513–2516. (c) Bosnjakovic, N.; Spasojevic de Brie, A. *J. Phys. Chem. A* **2010**, *114*, 10664–10675. (d) Murata, T.; Enomoto, Y.; Saito, G. *Solid State Sci.* **2008**, *10*, 1364–1368.

(12) Oomens, J.; Moehlig, A. R.; Morton, T. H. *J. Phys. Chem. Lett.* **2010**, *1*, 2891–2897.

(13) Lafitte, V. G. H.; Aliev, A. E.; Horton, P. N.; Hursthouse, M. B.; Bala, K.; Golding, P.; Hailes, H. C. *J. Am. Chem. Soc.* **2006**, *128*, 6544–6545.

(14) Helfer, D. L.; Hosmane, R. S.; Leonard, N. J. *J. Org. Chem.* **1982**, *46*, 4803–4804.

(15) Moehlig, A. R. Ph.D. Thesis, University of California, Riverside, 2011.

(16) This is consistent with the chemical shift of the central proton in proton-bound dimers of linear diamines: Coles, M. P.; Aragon-Saez, P. J.; Oakley, S. H.; Hitchcock, P. B.; Davidson, M. G.; Maksic, Z. B.; Vianello, R.; Leito, I.; Kaljuarnd, I.; Apperley, D. C. *J. Am. Chem. Soc.* **2009**, *131*, 16858–16868.

mixture. Rapid exchange occurs between the terminal NH_2 protons H_b and H_c. As this exchange cannot occur rapidly in the **[1·H$^+$·1]** dimer, the exchange must occur in a monomeric state, corroborating the rapid on/off rate of dimer formation observed in the ^{1}H NMR spectra. The central proton (H_a, Figure 4) shows a strong exchange crosspeak not with the terminal NH_2 (H_b/H_c), but with the residual water in the system (from added CF_3CO_2H). Small NOE crosspeaks can be observed between the central proton H_a and H_b, allowing identification of the individual NH_2 resonances. NOE spectra taken with a shorter mixing time (10 ms, see the Supporting Information) show exchange peaks between the central proton and water, but no crosspeaks between H_b and H_c, indicating proton transfer from the cytosine NH_2 is far slower than protonation and dimer formation.

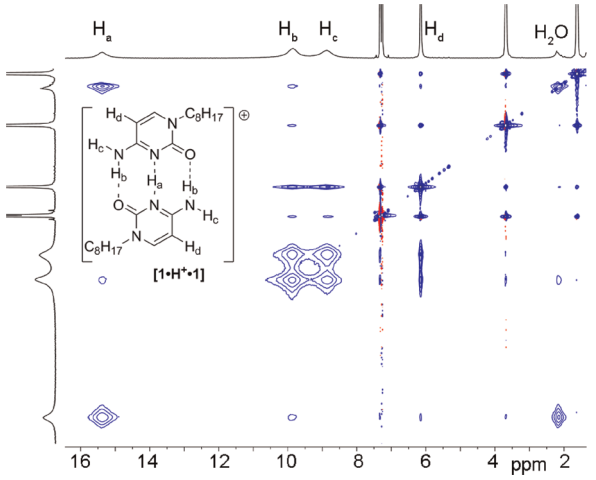


Figure 4. 2D NOESY Spectrum of **[1·H$^+$·1]** at low temperature (600 MHz, $CDCl_3$, 233 K).

The presence of multiple species in rapid exchange with each other from the complex equilibrium between the acid, **1**, **[1·H$^+$·1]**, **1H$^+$** and adventitious water in the sample makes accurate determination of the association constant of **[1·H$^+$·1]** in solution impractical. As the uncertainty in solution stems from the broad NH peak, gas-phase analysis of the association is more useful. Further information on the properties of the hemiprotonated dimers can be obtained by gas-phase MS analysis of the system and most notably the propensity for *hetero*dimer formation between different cytosine monomers. Hemiprotonated dimers were formed in the gas phase by mixing solutions containing the individual cytosine monomers and injecting the solution into an electrospray ionization source connected to a mass spectrometer. *N*-Methyl variants **2** and **3** were used in the MS analysis to minimize the effect of different ionization energies in the quantitative analysis.[17]

Because only a small percentage of molecules are ionized using electrospray ionization, the injection ratio can be directly compared to the ratio of the resultant ionic species.[18] The difference in electronic energies between the sum of isolated protonated and neutral monomers and the hemiprotonated dimer can be used to predict the relative abundance of the dimers in the gas phase. Density functional theory calculations were performed on the various hemiprotonated dimers. To account for the difference in basicity between **2** and **3**, the dimers that include the less basic monomer, **2**, have been adjusted by the difference in the calculated proton affinity between **2** and **3** (6.4 kcal/mol). Gas-phase IR experiments suggest that only a single tautomer of the **[2·H$^+$·3]** heterodimer is present in the gas phase following electrospray ionization.[15] The relative binding enthalpies (tabulated in the Supporting Information) show favored formation of the **[3·H$^+$·3]** homodimer over the **[2·H$^+$·3]** heterodimer, which is in turn more favorable than the **[2·H$^+$·2]** homodimer.

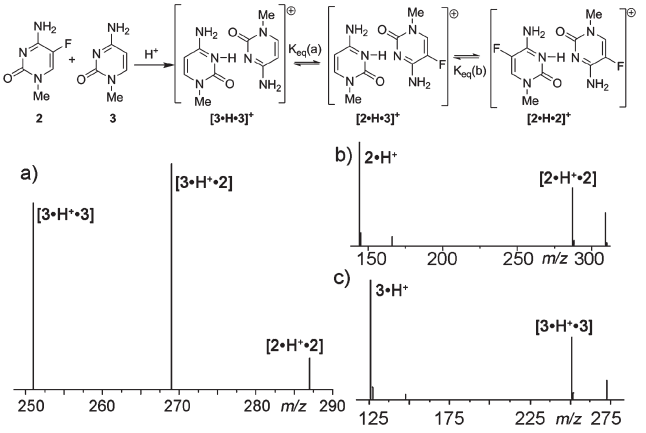


Figure 5. Gas phase analysis of **2·H$^+$·2**, **2·H$^+$·3** and **3·H$^+$·3** complexes: (a) mixture of **2** and **3**; (b) **2** alone; (c) **3** alone. Minor peaks correspond to ^{13}C isotopes and MNa^+ ions.

Stock solutions (0.05 M) of **2** and **3** were prepared in 50:50 CH_3OH/water to determine the equilibrium constants between the various hemiprotonated dimers containing **2** and **3**. Five monomer ratios (70/30, 60/40, 50/50, 40/60, and 30/70) were prepared and analyzed by ESI-MS (see Figure 5 and Table 1). Five ions are observed in the mass spectrum: two correspond to the protonated monomers **2·H$^+$** and **3·H$^+$**, two correspond to the hemiprotonated homodimers **[2·H$^+$·2]** and **[3·H$^+$·3]**, and one corresponds to the hemiprotonated heterodimer **[2·H$^+$·3]**. The peak intensity ratios of the dimers can be used to determine the relative equilibrium constants (K_{eq}) via the scheme shown in Figure 5. The average ratio of the hemiprotonated dimer intensities observed in the mass spectrum relative to that of the **[2·H$^+$·2]** homodimer is shown in Table 1 for each of the five ratios. As expected,

(17) *N*-Octylcytosine **1** shows similar MS data to **3**. See the Supporting Information for spectra.

(18) Kebarle, P.; Verkerk, U. H. *Mass Spectrom. Rev.* **2009**, *28*, 898–917.

the relative intensity of the hemiprotonated dimer peak containing that monomer also increases. The intensity ratios of the dimers from the mixtures of **2/3** reflect the relative basicities of the monomers. In the case of the 50/50 mixture of **2/3**, the signal for **[3·H$^+$·3]** homodimer is over 6 times more intense than **[2·H$^+$·2]**. This agrees with the analysis of the calculated binding enthalpies of the two homodimers that predicts the **[3·H$^+$·3]** homodimer to be 6.9 kcal/mol more stable.

K_{eq}(a) is measured to be 0.94 (for graphical plots, see the Supporting Information), indicating a slight preference for **[3·H$^+$·3]** under electrospray conditions, but indicating that heterodimer formation is energetically accessible. K_{eq}(b) is much lower, measured to be 0.18, indicating the strong favorability of **[2·H$^+$·3]** over homodimer **[2·H$^+$·2]**. This supports the notion that the basicity of the cytosine ring nitrogen plays a major role in the stability of the dimers. The hemiprotonated dimers containing **3** are more stable (and more abundant in the gas phase) because this increased basicity stabilizes the cationic dimer. The relative intensities of the dimers in the ESI-MS support this notion as the abundance of the **[2·H$^+$·2]** ion is less than that of the other dimers under all injection ratios. At lower concentrations of **3**, heterodimer **[2·H$^+$·3]** has a higher observed abundance due to the fact that the equilibrium between **[2·H$^+$·3]** and **[3·H$^+$·3]** is close to unity and the equilibrium constant from **[2·H$^+$·3]** and **[2·H$^+$·2]** is small.

Table 1. Average Peak Intensity Ratios in the ESI-MS Spectra of the Hemiprotonated Dimers Containing **2** and **3**, Normalized to the Ratio of the **2·2** Homodimer

3/2 ratio	intensity (**[3·H$^+$·3]**)	intensity (**[2·H$^+$·3]**)	intensity (**[2·H$^+$·2]**)
70/30	24.4	15.6	1.0
60/40	12.0	9.8	1.0
50/50	6.1	7.0	1.0
40/60	2.6	4.3	1.0
30/70	1.0	2.5	1.0

To determine whether the relative affinities of the heterodimer mixtures could be analyzed in solution, 5-fluoro-1-octylcytosine **4** was combined with 1-octylcytosine **1** and CF_3CO_2H in $CDCl_3$ (for spectra, see the Supporting Information). Formation of the **[4·H$^+$·4]** homodimer occurs under similar acidic conditions as for the formation of **[1·H$^+$·1]**. The **4**, **[4·H$^+$·4]**, **4H$^+$** equilibrium exchange is rapid as before, and the central proton can be observed as a very broad peak at ~16 ppm at 243 K. This peak is far broader than for **[1·H$^+$·1]** at the same temperature, however, which is consistent with the lowered favorability of dimerization of fluorinated cytosine species in the gas phase. Dimer formation is observed in solution upon combination of **4** with **1** in the presence of 0.5 mol.-equiv. CF_3CO_2H in $CDCl_3$, but the broadness of the NH peaks prevents any observation of individual heterodimers.

In conclusion, organic soluble cytosine derivatives have been synthesized, and their hemiprotonated dimers observed in chloroform solution upon treatment with strong acid and in the gas phase by ESI-MS. Dimer formation is controlled by the basicity of the cytosine derivative, with the 5-fluoro counterparts showing a lower affinity for acidic self-assembly. This is the first observation of cytosine hemiprotonated dimers in solution outside DNA strands, and provides a greater understanding of the formation of uncommon architectures in biological hydrogen bonded systems. Further study of targeted self-assembly of these derivatives is underway in our laboratory.

Acknowledgment. R.J.H. acknowledges the donors of the American Chemical Society Petroleum Research Fund and UC Riverside for partial support of this research. A. M. acknowledges financial support from the NSF (CHE-0848517). We thank Prof. T. H. Morton, UC Riverside, for stimulating discussions.

Supporting Information Available. Experimental details and full characterization of new compounds. This material is available free of charge via the Internet at http://pubs.acs.org

The authors declare no competing financial interest.