Crystallographic evidence of Watson–Crick connectivity in the base pair of anionic adenine with thymine

Manish Kumar Mishra1,2, Steven P. Kelley1, Volodymyr Smetana2, David A. Dixon2, Ashley S. McNeill2, Anja-Verena Mudring2,3, and Robin D. Rogers1,2,3

1Department of Chemistry and Biochemistry, The University of Alabama, Tuscaloosa, AL 35487; and 2Department of Materials and Environmental Chemistry, Stockholm University, Stockholm 106 91, Sweden

Edited by Richard Eisenberg, University of Rochester, Rochester, New York, and approved June 17, 2020 (received for review April 29, 2020)

Utilizing an ionic liquid strategy, we report crystal structures of salts of free anionic nucleobases and base pairs previously studied only computationally and in the gas phase. Reaction of tetrabutylammonium ([N4444]+) or tetrabutylphosphonium ([P4444]+) hydroxide with adenine (HAd) and thymine (HThy) led to hydrated salts of deprotonated adenine, [N4444][Ad]-2H2O, and thymine, [P4444][Thy]-2H2O, as well as the double salt cocrystal, [P4444][Ad][Thy]-3H2O-2HThy. The cocrystal includes the anionic [Ad(HThy)] base pair which is a stable formation in the solid state that has previously not even been suggested. It exhibits Watson–Crick connectivity as found in DNA but which is unusual for the free neutral base pairs. The stability of the observed anionic bases and their supramolecular formations and hydrates has also been examined by electronic structure calculations, contributing to more insight into how base pairs can bind when a proton is removed and highlighting mechanisms of stabilization or chemical transformation in the DNA chains.

Results and Discussion

[N4444][Ad]-2H2O. In addition to anionic forms of the base pairs, radical anion pairs have been intensively studied computationally and in the gas phase (19–22). Since water is the natural component/medium of biological systems, investigation of the interactions of deprotonated nucleobases with it is of both fundamental and practical interest. Moreover, such interactions

Significance

All genetic information on Earth is encrypted in DNA and RNA with the nucleobases and their pairs being the main information units. There are strict established rules how nucleobases can interact between each other in the DNA. These rules can though be affected by external factors such as radiation causing formation of the deprotonated charged species. Although such species are extremely unstable and low in abundance, they may affect local connectivity and introduce wrong units in the DNA chain, so proper characterization of their interactions is of enhanced importance. Here we could obtain anionic nucleobases in stable form in the solid state, opening the possibility to study them crystallographically and develop theoretical models for real biological systems.
have also been studied in the gas phase for isolated deprotonated nucleobases (23). As discussed below and elsewhere (23, 24), the most acidic gas-phase site in the adenine molecule is the N(9)-H of the imidazole ring (Fig. 1). A synthetic attempt to produce free anionic adenine was undertaken; however, its complete X-ray structure characterization was hindered due to a significant degree of positional disorder in the crystal structure (25).

In this work, we have experimentally obtained and structurally characterized the $[\text{Ad}]^{-}$ anion in the solid state. Our structure is consistent with the one corresponding to the removal of the proton from the site determined to be the most acidic in the gas phase. $[\text{Ad}]^{-}$ anions are bound via water molecules in monodentate and bidentate modes preventing any direct $[\text{Ad}]^{-}$–$[\text{Ad}]^{-}$ connectivity (Fig. 2A and B). It should be noted, though, that the estimated strengths of these three contacts are different varying from medium to the upper edge of the moderate range (26) [$d_{O\cdots N} = 2.819(2)$ to 3.192(2) Å]. The weaker intermolecular interactions of the bidentate H2O are reinforced by additional water bridging [$d_{O\cdots O} = 2.762(2)$ Å]. From the structural viewpoint, the entire anionic layer can be represented on the basis of dimeric units containing two $[\text{Ad}]^{-}$ anions and four water molecules. Each such unit binds to four identical units via direct OH–N hydrogen bonds. Each $[\text{Ad}]^{-}$ anion bridges to three other $[\text{Ad}]^{-}$ by means of at least one water molecule via NH–O, OH–N, and OH–N hydrogen bonds.

The $[\text{N4444}]^{+}$ cations form separate layers with mainly van der Waals intralayer bonding (SI Appendix, Fig. S4). Each $[\text{Ad}]^{-}$ is connected with two $[\text{N4444}]^{+}$ cations via weak nonclassical CH–π intermolecular interactions. On the other hand, the $[\text{N4444}]^{+}$ cations are also connected with two $[\text{Ad}]^{-}$ anions via the same CH–π intermolecular interactions and with two water molecules via weak CH–O hydrogen bonds.

$[\text{P4444}][\text{Thy}] \cdot 2\text{H}_2\text{O}$. Whereas the gas-phase photoelectron spectra suggested that Thy$^{-}$ is formed by removal of a proton from the N(1) atom (23, 27) that generally connects to the sugar in the nucleotide, both N(3)$^{-}$ and N(1)$^{-}$ deprotonated monoanions are indistinguishable in aqueous solution (28). Both sites show
distinctive acidities in the gas phase (by 11.0 kcal/mol) as discussed below and in the literature (29), yet they are equivalent in water. Whereas the N(1) conjugate base is more stable in the gas phase (by 11.0 kcal/mol) as discussed below and observed in the gas and in the literature (29), yet they are equivalent in water. Whereas the N(1) conjugate base is more stable in the gas phase (by 11.0 kcal/mol) as discussed below and observed in the gas phase (31). The crystal structure is characterized by infinite 1D hydrogen bond networks, as well as in the context of DNA damage (35-37). Although involving significant structural changes, an identical WC-type mode has also been predicted for the adenine-thymine base pair radical anions (38, 39). Self-organization of neutral adenine and thymine in the solid state indicates primarily the HG mode of base pairing (40-43). The deprotonated guanine- cytosine base pair has been studied computationally (11), but no studies have been, to the best of our knowledge, reported for adenine-thymine or any other combinations.

Cocrystal Base Pairing. [P4444][Ad][Thy]·2H2O crystallizes in the monoclinic acentric space group Cc with one [Ad]− and [Thy]− anions, two [P4444]+ cations, two neutral thymine, and three water molecules in the asymmetric unit (SI Appendix, Fig. S3). The [Ad]− anion is connected to two neutral HThy through NH···O and NH···N hydrogen bonds which leads to the formation of [Ad]− and HThy WC type base pairs and, additionally, quasi-HG base pairs (Figs. 1 and 34, red and blue circles, respectively). Furthermore, it is connected to a third neutral thymine via a direct NH···N bond complemented by a CH···O···N bridge. Although WC-type pairing is usual in DNA, it has never been observed in cocrystals of the neutral HAd–HThy base pair.

There are no direct intermolecular interactions between anionic [Ad]− and [Thy]− in this cocrystal. Anionic [Thy]− forms base pairs with neutral HThy through NH···O dimers, as well as a water bridge (CO···H2O···OC). From a structural motif viewpoint, these can be considered as the four nucleobase unit [HThyAdHThyThy]− (Fig. 34, orange circle). Similar to [P4444][Thy]2H2O, the water molecule is connected to the deprotonated nitrogen of the [Thy]− via OH···N hydrogen bonds. In contrast to the former, in the cocrystal only trimeric water bridges connect the nucleobase tetramers.

Taking into consideration the extended hydrogen bond network and water bridges, we observe face-to-face stacking (7.103 Å) of the helical motifs along the a axis (Fig. 34, green circle, and Fig. 3B). Each helix consists of two [Ad]−, two [Thy]−, three HThy, and four water bridges, while each but one nucleobase is shared between two such motifs. This stacking distance has been observed in many proteins where the aromatic groups of peptides stack in this manner to stabilize the protein structure (44). Despite having the stacking in the helical motif, B-form helices are not found in this structure.

[P4444]+ cations do not play any independent role in the crystal structure of [P4444][Ad][Thy]·2H2O but rather support the helix-like motif of the nucleobase-water formation with the P centers located along the dashed green lines (Fig. 3 and SI Appendix, Fig. S6) and carbon chains filling the space inside of the helices. The mutual cation/anionic environment here follows the general structural complexity. The [Ad]− anion is connected to two [P4444]+ cations via the same weak nonclassical CH···π bonds. The [Thy]− anion is connected to three [P4444]+ cations,
while both neutral HThy molecules are connected to four. Both [P4444]⁺ cations show slightly different connectivity: [P14444]⁺ interacts with 1 [Ad]⁻, 7 HThy, and 4 water molecules and [P2444]⁺ to 4 [Ad]⁺, 4 [Thy]⁻, 2 HThy, and 3 water molecules. The potential contacts to the water molecules are extremely weak.

Computational Results. To provide better insights into the connectivity preferences and supramolecular formation, a range of structures have been optimized in the gas phase. The optimizations were done using the standard approaches in Gaussian16. The initial structures for the monomers and dimers were constructed using a graphical user interface. For the dimers, different hydrogen bonding orientations were chosen. The starting geometries for the trimers, tetramers, and pentamers were taken from the crystal structure with adjusted hydrogen positions. Canonical neutral HAd with the acidic proton trans to the NH₂ axis. Crystallographic axes are color coded: b, green; c, blue. [P4444]⁺ cations are omitted for clarity.

Fig. 3. (A) Molecular arrangements and hydrogen bond patterns in the crystal structure of [P4444][Ad][Thy]·3H₂O·2HThy and (B) helical motif along the a axis. Crystallographic axes are color coded: b, green; c, blue. [P4444]⁺ cations are omitted for clarity.

Table 1. Complexation energies for dimers at the different computational levels at 298 K in kcal/mol

| Reaction | B3LYP/aD | G3(MP2) | G3MP2(B3) |
|----------|----------|---------|-----------|
| [Ad]⁻ + HAd → [HAdAd]⁻ | ΔHgas | ΔGgas | ΔHgas | ΔGgas | ΔHgas | ΔGgas |
| [Ad]⁻ + HAd → [HAdAd]⁻ | −27.4 | −15.7 | −28.8 | −19.2 | −29.2 | −18.8 |
| [Ad]⁻ + HAd → [HAdAd]⁻ | −13.8 | −2.6 | −14.9 | −5.9 | −15.4 | −5.4 |
| [Thy]⁻ + HThy → [HThyThy]⁻ | −22.0 | −11.6 | * | * | −23.8 | −14.0 |
| [Thy]⁻ + HThy → [HThyThy]⁻ | −19.7 | −9.6 | −20.3 | −12.5 | −21.3 | −11.6 |
| [Thy]⁻ + HThy → [HThyThy]⁻ | −11.4 | −1.0 | * | * | −14.3 | −1.9 |
| [Thy]⁻ + HThy → [HThyThy]⁻ | −10.4 | −0.1 | −12.1 | −3.6 | −12.6 | −3.2 |
| [Ad]⁻ + HThy → [HAdThy]⁻ | −22.9 | −12.0 | −25.0 | −13.3 | −25.2 | −12.9 |
| [Ad]⁻ + HThy → [HAdThy]⁻ | −13.3 | −2.3 | −14.6 | −4.7 | −15.4 | −4.6 |
| HAdB + HThy → [HAdHThy]⁻ (HG) | −12.3 | −0.2 | −14.2 | −4.6 | −14.4 | −3.9 |
| HAdB + HThy → [HAdHThy]⁻ (WC) | −11.6 | 0.4 | −13.4 | −3.9 | −13.7 | −3.3 |

*Did not converge.
Finally, we optimized the pentamer [HThy][Ad]·3H2O formed by adding an additional HThy. In this case, there is no proton transfer, and the optimized structure of the isolated pentamer is the same as that in the crystal showing that the additional HThy governs the proton transfer between [HThy] and [Ad]−. The close similarity between the structures in the gas phase and in the solid state for a large enough cluster suggests that this is an inherent binding motif for these types of complexes and that the observed structures are not being strongly directed by crystal lattice forces.

The structures of hydrated anions have also been optimized in the gas phase to better understand the base pair connectivity modes. Up to four waters were used to hydrate the [Ad] + and [Thy]− anions (Table 3). Although our results predict more binding than the values reported by Wincel (23), they are consistent with the fact that the hydration energies are essentially the same for the two anions.

The water for the monohydrated structure [Ad]−·A (SI Appendix, Fig. S12) binds to the two accessible N atoms on the five- and six-member ring and is the proton donor. The water in structure [Ad]−·B is a proton donor to an N on the five-member ring and a proton acceptor from the NH2 group on the six-member ring. For the monohydrated [Thy]−, the H2O is a proton donor to a C=O and to the N− center on the ring. An addition of a second H2O results in a drop of 1 kcal/mol as compared to adding the first H2O for [Ad]− and 3.5 kcal/mol for [Thy]−. Addition of two further H2O molecules to [Ad]− averages to −9 kcal/mol per H2O and for [Thy]− −10 kcal/mol.

Comparing the addition of four H2O molecules to [Ad]− with the results from the crystal structure (SI Appendix, Fig. S13), one may notice that the hydrogen bonding of the top three molecules is essentially the same in the gas-phase structure as in the crystal. The fourth H2O molecule on the bottom of the image has one hydrogen bond to the anion, while the remaining hydrogen points away to interact with additional molecules in the crystal lattice. In the gas phase, this H2O molecule rotates to form two hydrogen bonds to N groups on the five- and six-member rings of [Ad]− due to no alternative. The interaction of the fourth water molecule is essentially the same as binding one H2O in [Ad]−·A as discussed above.

The structure of the dianionic [Thy]2−·2 dimer solvated by two terminal H2O molecules is essentially the same in the crystal and in the free gas-phase molecular complex (SI Appendix, Fig. S14). Note that in both cases, there is one dangling H on a terminal H2O that can be used for further hydrogen bonding to another species. The resulting dimeric complex (SI Appendix, Fig. S15) shows the ring of four H2O molecules between the two [Thy]2− dimers that holds the entire complex together.

Conclusions

The explicit molecular mechanism of how radiation or low-energy electrons damage DNA is still under intense study, and the results of the current study provide insights into how base pairs can bind when a proton is removed, highlighting mechanisms of stabilization or chemical transformation in the DNA chains. The anionic [AdHThy]− base pair exhibits Watson–Crick connectivity as found in DNA but is unusual for the free neutral base pairs. This deprotonated anionic [AdHThy]− base pair structure is a stable formation in the solid state which has not even been previously suggested. The combined experimental and computational study shows that this type of binding in anionic base pairs is inherent to the base pair and not driven by external forces. The results also show the role that waters of solvation can play in controlling the base pair binding, which is very important due to the major role of water in biological systems. The conventional Watson–Crick connectivity between adenine and thymine certainly possesses backbone stabilization mechanisms in the charged state that becomes evident in both protonated and deprotonated forms.

Materials and Methods

Chemicals. Adenine (99%) and thymine (99%) were bought from Sigma-Aldrich, Co. LLC. Tetrabutylphosphonium hydroxide [P4444][OH] (40 wt % in water) and tetrabutylammonium hydroxide [N4444][OH] (55 wt % in water) were purchased from Fisher Scientific. All chemicals were used as received unless otherwise stated.

Crystallization Procedures.

Finally, we optimized the pentamer [HThy][Ad]·3H2O formed by adding an additional HThy. In this case, there is no proton transfer, and the optimized structure of the isolated pentamer is the same as that in the crystal showing that the additional HThy governs the proton transfer between [HThy] and [Ad]−. The close similarity between the structures in the gas phase and in the solid state for a large enough cluster suggests that this is an inherent binding motif for these types of complexes and that the observed structures are not being strongly directed by crystal lattice forces.

Table 3. Energies of hydration for [Ad] + [Ad] + 2H2O [Ad]−·3H2O at the G3(MP2) level at 298 K in kcal/mol

| Reactants | Product | ΔHgas | ΔGgas |
|-----------|---------|-------|-------|
| [Ad]− | [HThyHAdThy]− | −33.0 | −10.7 |
| [Thy]− | [HThyHAdThy]− | −42.0 | −18.7 |
| [HThyHAdThy]− | [HThyHAdThy]− | −19.8 | −6.3 |
| [HThyHAdThy]− | [HThyHAdThy]− | −10.1 | 1.3 |
| [HThyHAdThy]− | [HThyHAdThy]− | −69.3 | −8.4 |
| [HThyHAdThy]− | [HThyHAdThy]− | −71.7 | 0.8 |

| Reaction | ΔHgas | ΔGgas |
|-----------|-------|-------|
| [Ad]− + 2H2O → [HThyHAdThy]− | −33.0 | −10.7 |
| [Thy]− + [HThyHAdThy]− | −42.0 | −18.7 |
| [HThyHAdThy]− (1) + [HThyHAdThy]− | −19.8 | −6.3 |
| [HThyHAdThy]− (2) + [HThyHAdThy]− | −10.1 | 1.3 |
| [HThyHAdThy]− + 3H2O → [HThyHAdThy]−·3H2O | −69.3 | −8.4 |
| [HThyHAdThy]− + 3H2O → [HThyHAdThy]−·3H2O | −71.7 | 0.8 |

Materials and Methods

Chemicals. Adenine (99%) and thymine (99%) were bought from Sigma-Aldrich, Co. LLC. Tetrabutylphosphonium hydroxide [P4444][OH] (40 wt % in water) and tetrabutylammonium hydroxide [N4444][OH] (55 wt % in water) were purchased from Fisher Scientific. All chemicals were used as received unless otherwise stated.
mixed together into an empty borosilicate glass culture tube (20 mL) at room temperature and slowly homogenized by hand grinding with a glass stirring rod. The obtained mixture was placed in a heated sand bath at 90 °C for 15 h. Colorless plate-shaped crystals of [P4444][Ad][Thy]·3H2O·2HThy formed in the reaction vessel, and the vessel was allowed to cool to room temperature.

**SCXRD.** The single crystals of the salts were isolated directly from each reaction mixture. SCXRD data were collected on a Bruker D8 Advance diffractometer with a Photon 100 CMOS area detector and an iLu microfocus X-ray source using Mo-Kα radiation. Crystals were coated with Paratone oil and cooled to 100 K under a cold stream of nitrogen using an Oxford cryostat (Oxford Cryosystems). Hemispheres of data out to a resolution of at least 0.80 Å were collected by a strategy of ϕ and ω scans. Unit cell determination, data collection, data reduction, correction for absorption, structural solution, and refinement were all conducted using the Apex3 software suite. Hydrogen atoms bonded to nitrogen and oxygen atoms were located from the difference map. Their coordinates were allowed to refine while their thermal parameters were constrained to ride on the carrier atoms. Hydrogen atoms bonded to carbon atoms were placed in calculated positions, and their coordinates and thermal parameters were constrained to ride on the carrier atoms. Related crystallographic information has been deposited with the Cambridge Structural Database and can be downloaded from https://www.ccdc.cam.ac.uk/ (accession nos. CCDC 1976655, 1976656, and 1976657).

**Powder X-Ray Diffraction.** Powder X-ray diffraction (PXRD) data were collected on a Bruker D8 Advance equipped with a Lyneye linear position sensitive detector (Bruker AXS). The bulk semisolid sample of [P4444][Ad][Thy]·3H2O·2HThy was smeared directly onto the silicon wafer of a proprietary low-background sample holder. Data were collected using a continuous coupled θ/2θ scan with Ni-filtered Cu-Kα radiation. Diffraction data were measured across a 2θ range of 5 to 30°. The collected diffractogram was compared with the diffractograms calculated from the SCXRD data of [P4444][Ad][Thy]·3H2O·2HThy, [PAAD][2H2O], [P4444][H2O], [P4444][Thy]·2H2O, adenine, and, thymine. The diffractograms indicate that the bulk solid is a mixture of [P4444][Ad][Thy]·3H2O·2HThy, adenine, and thymine. PXRD data for [N4444][Ad]·2H2O and [P4444][Thy]·2H2O were not collected due to the low viscosity of their bulk samples.

**Computational Methods.** All structures were optimized at the density functional theory (DFT) level using Gaussian 16 (47). These geometries were initially optimized (48–50) with the B3LYP (51, 52) exchange-correlation functional using the D2VP2 (53) basis set, followed by optimization at the B3LYP/aug-cc-pVDZ (ad) (54, 55) level. For adenine and thymine monomers and dimers, improved energetics were obtained at the composite correlated molecular orbital theory G3(MP2) (56) and/or G3(MP2) (57) levels as these methods are shown to perform better (58) in the prediction of bond energies, acidities, and through-space interactions compared to the most commonly used DFT functionals. Gas-phase acidities are defined as the change in free energy at 298 K for the deprotonation reaction (1).

**ACKNOWLEDGMENTS.** This work is supported in part by the US Department of Energy (DOE) Basic Energy Sciences (BES), Heavy Elements program under Award DE-SC0019220 (R.D.R.), and the DOE BES Geosciences program by a subcontract to D.A.D. (computational work) from Pacific Northwest National Laboratory. This research was supported in part by the Swedish Research Council Tage Erlander professorship to R.D.R. (Swedish Research Council [VR] Grant 2018-00233) and Göran Gustafsson prize by the Royal Swedish Academy of Science to A.-V.M.
41. M. D. King, W. Ouellette, T. M. Korter, Noncovalent interactions in paired DNA nucleobases investigated by terahertz spectroscopy and solid-state density functional theory. J. Phys. Chem. A 115, 9467–9478 (2011).

42. K. Hoogsteen, The structure of crystals containing a hydrogen-bonded complex of 1-methylthymine and 9-methyladenine. Acta Crystallogr. 12, 822–823 (1959).

43. S. Chandrasekhar, T. R. R. Naik, S. K. Nayak, T. N. G. Rov, Crystal structure of an intermolecular 2:1 complex between adenine and thymine. Evidence for both Hoogsteen and “quasi-Watson-Crick” interactions. Bioorg. Med. Chem. Lett. 20, 3530–3533 (2010).

44. S. K. Burley, G. A. Petsko, Aromatic-aromatic interaction: A mechanism of protein structure stabilization. Science 229, 23–28 (1985).

45. L. M. Salter, G. M. Chaban, Theoretical study of gas phase tautomerization reactions for the ground and first excited electronic states of adenine. J. Phys. Chem. A 106, 4251–4256 (2002).

46. E. C. M. Chen, C. Herder, E. S. Chen, The experimental and theoretical gas phase acidities of adenine, guanine, cytosine, uracil, thymine and halouracils. J. Mol. Struct. 788, 126–133 (2006).

47. M. J. Frisch et al., Gaussian 16 (Revision A.03, Gaussian, Inc., Wallingford CT, 2016).

48. H. B. Schlegel, Optimization of equilibrium geometries and transition structures. J. Comput. Chem. 3, 214–218 (1982).

49. X. Li, M. J. Frisch, Energy-represented direct inversion in the iterative subspace within a hybrid geometry optimization method. J. Chem. Theory Comput. 2, 835–839 (2006).

50. C. Peng, P. Y. Ayala, H. B. Schlegel, M. J. Frisch, Using redundant internal coordinates to optimize equilibrium geometries and transition states. J. Comput. Chem. 17, 49–56 (1996).

51. A. D. Becke, Density-functional thermochemistry. III. The role of exact exchange. J. Chem. Phys. 98, 5648–5652 (1993).

52. C. Lee, W. Yang, R. G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. Phys. Rev. B Condens. Matter 37, 785–789 (1988).

53. N. Godbout, D. R. Salahub, J. Andzelm, E. Wimmer, Optimization of Gaussian-type basis sets for local spin density functional calculations. Part I. Boron through neon, optimization technique and validation. Can. J. Chem. 70, 560–571 (1992).

54. T. H. Dunning, Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. J. Chem. Phys. 90, 1007–1023 (1989).

55. R. A. Kendall, T. H. Dunning, R. J. Harrison, Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions. J. Chem. Phys. 96, 6796–6806 (1992).

56. L. A. Curtiss, P. C. Redfern, K. Raghavachari, V. Rassolov, J. A. Pople, Gaussian-3 theory using reduced Møller-Plesset order. J. Chem. Phys. 110, 4703–4709 (1999).

57. A. G. Baboul, L. A. Curtiss, P. C. Redfern, K. Raghavachari, Gaussian-3 theory using density functional geometries and zero-point energies. J. Chem. Phys. 110, 7650–7657 (1999).

58. M. I. Stover et al., Fundamental thermochemical properties of amino acids: Gas-phase and aqueous acidities and gas-phase heats of formation. J. Phys. Chem. B 116, 2905–2916 (2012).