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

Manish Kumar Mishra[a,1], Steven P. Kelley[a,2], Volodymyr Smetana[b], David A. Dixon[a], Ashley S. McNeill[a], Anja-Verena Mudring[b], and Robin D. Rogers[a,b,3]

[a]Department of Chemistry and Biochemistry, The University of Alabama, Tuscaloosa, AL 35487; and [b]Department 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 ($[N_{4444}]^+$) or tetrabutylphosphonium ($[P_{4444}]^+$) hydroxide with adenine (HAd) and thymine (HThy) led to hydrated salts of deprotonated adenine, $[N_{4444}][Ad]\cdot 2H_2O$, and thymine, $[P_{4444}][Thy]\cdot 2H_2O$, as well as the double salt cocrystal, $[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$. 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.

DNA | nucleobase | anionic | hydrate | crystal structure

Since the breakthrough discovery of the double helix structure of DNA by Watson and Crick (1) in the early 1950s, research interest in nucleobases and complementary base pairing chemistry has been a cornerstone of structural biochemistry, pharmaceutical development, protein crystallography, crystal engineering, modern biology, and many other areas (2–4). Current advances in the areas of solid-state chemistry and supramolecular chemistry have already achieved an ability to anticipate the formation of exotic supramolecular assemblies with tailor-made properties. Despite the research interest in the supramolecular chemistry of nucleobases, it is rather surprising that anionic forms of free nucleobases are relatively unexplored and crystal structures of salts of free anionic nucleobases and base pairs are unknown.

Crystal structures would be particularly useful for detailed, quantitative examinations of the interactions between anionic nucleobases, either each other or with other molecules. These are important in a number of biological contexts. Anionic states of nucleic acid bases are involved in various biological processes including DNA damage, electron transfer in DNA, various mutations, Watson–Crick (WC) like anionic base pairs or mispairs, and chemical damage by oxidative species (5–8). In this regard, the chemical properties of the ionic state of the nucleobases, their derivatives, and anionic deprotonated DNA base pairs have been studied in the gas-phase using mass spectrometry, negative ion photoelectron spectroscopy, and also modeled using density functional theory (DFT) electronic structure calculations to gain insight into the mechanism of radiation-induced DNA damage and other biological processes (9–14). Characterization of anionic nucleobases and unstable tautomers in DNA and RNA is a huge challenge because they exist in low abundance, often for only short periods of time, so that they are difficult to observe at the atomic level (7, 15, 16). Thus, detailed, experimentally validated structures of anionic nucleobases may aid their detection in real biological systems.

We recently synthesized low melting salts (i.e., ionic liquids) of anionic acyclovir, an antiviral medicine which is an analog of guanosine, with large bulky tetraalkyl-phosphonium or tetraalkyl-ammonium cations (17, 18). Neutral acyclovir (HAcy) and deprotonated $[Acy]^-$ have appropriate geometries to participate in complementary Watson–Crick-type base pairing. We hypothesized that the same approach would lead to anionic forms of nucleobases. To test this hypothesis, neutral adenine (HAd) and thymine (HThy) were reacted with aqueous tetrabutylammonium hydroxide ($[N_{4444}][OH]$) and tetrabutylphosphonium hydroxide ($[P_{4444}][OH]$) (Scheme 1). In a competitive reaction, we also reacted HAd and HThy with $[P_{4444}][OH]_{(aq)}$ in a 1:1:1 ratio. In all cases, colorless single crystals suitable for single crystal X-ray diffraction (SCXRD) studies formed within a day after cooling to room temperature.

Results and Discussion

$[N_{4444}][Ad]\cdot 2H_2O$. 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.

Author contributions: M.K.M. designed research; M.K.M. and A.S.M. performed research; M.K.M., S.P.K., V.S., D.A.D., A.S.M., A.-V.M., and R.D.R. analyzed data; and M.K.M., S.P.K., V.S., D.A.D., A.-V.M., and R.D.R. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Data deposition: Crystallographic information has been deposited with the Cambridge Structural Database and can be downloaded free of charge from https://www.ccdc.cam.ac.uk/ (accession nos. 1976655, 1976656, and 1976657).

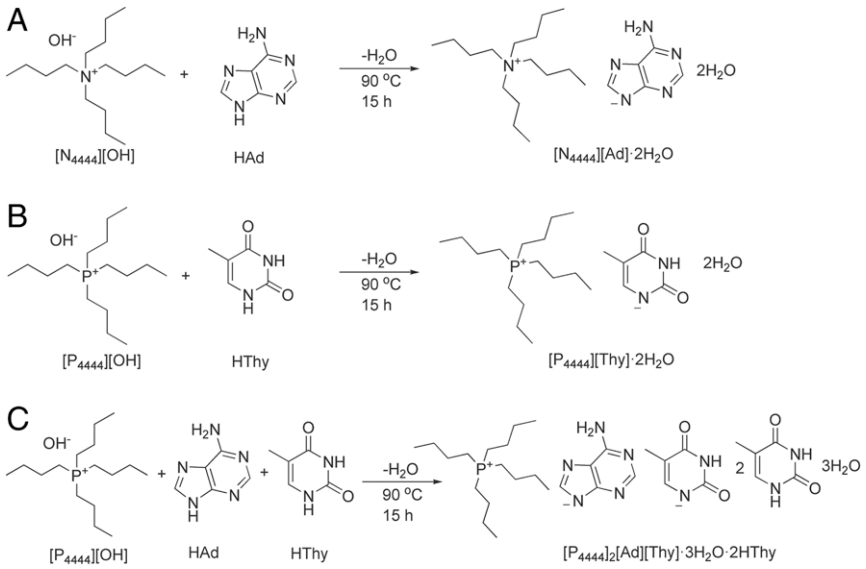
[1]Present address: Department of Pharmaceutics, College of Pharmacy, University of Minnesota, Minneapolis, MN 55455.

[2]Present address: Department of Chemistry, University of Missouri, Columbia, MO 65211.

[3]To whom correspondence may be addressed. Email: rdrogers@ua.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2008379117/-/DCSupplemental.

First published July 17, 2020.



Scheme 1. Reactions leading to crystallization of (*A*) $[N_{4444}][Ad]\cdot 2H_2O$, (*B*) $[P_{4444}][Thy]\cdot 2H_2O$, and (*C*) $[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$. All reagents were mixed in a 1:1 molar ratio (*C* was mixed 1:1:1) at room temperature and then heated to 90 °C for 15 h (see *SI Appendix* for full synthetic details).

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 $[Ad]^-$ 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. $[N_{4444}][Ad]\cdot 2H_2O$ crystallizes in the monoclinic space group $P2_1/c$ with one $[Ad]^-$ anion, one $[N_{4444}]^+$ cation, and two water molecules in the asymmetric unit (*SI Appendix*, Fig. S1). The crystal structure is best described in terms of alternating layers of cations and anions. The latter are polymeric pseudo–two-dimensional (2D) formations of $[Ad]^-$ anions bound via water molecules in monodentate and bidentate modes preventing any direct $[Ad]^-$–$[Ad]^-$ connectivity (Fig. 2 *A* 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 H_2O 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 $[Ad]^-$ anions and four water molecules. Each such unit binds to four identical units via direct OH···N hydrogen bonds. Each $[Ad]^-$ anion bridges to three other $[Ad]^-$ by means of at least one water molecule via NH···O, OH···N, and OH···N^- hydrogen bonds.

The $[N_{4444}]^+$ cations form separate layers with mainly van der Waals intralayer bonding (*SI Appendix*, Fig. S4). Each $[Ad]^-$ is connected with two $[N_{4444}]^+$ cations via weak nonclassical CH···π intermolecular interactions. On the other hand, the $[N_{4444}]^+$ cations are also connected with two $[Ad]^-$ anions via the same CH···π intermolecular interactions and with two water molecules via weak CH···O hydrogen bonds.

$[P_{4444}][Thy]\cdot 2H_2O$. 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

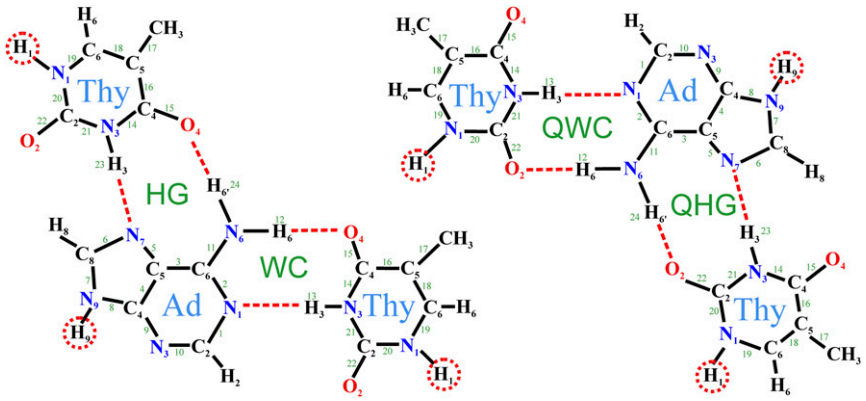


Fig. 1. Visualization of four different types of HAd-HThy nucleobase pairing including alternate nomenclature in the parentheses: WC, Watson–Crick (Cis Watson–Crick/Watson–Crick); HG, Hoogsteen (Cis Watson–Crick/Hoogsteen); QWC, Quasi-Watson–Crick (Trans Watson–Crick/Watson–Crick); and QH, Quasi-Hoogsteen (Trans Watson–Crick/Hoogsteen). In the case of the anionic base pair, N(9) of adenine and/or N(1) of thymine are deprotonated.

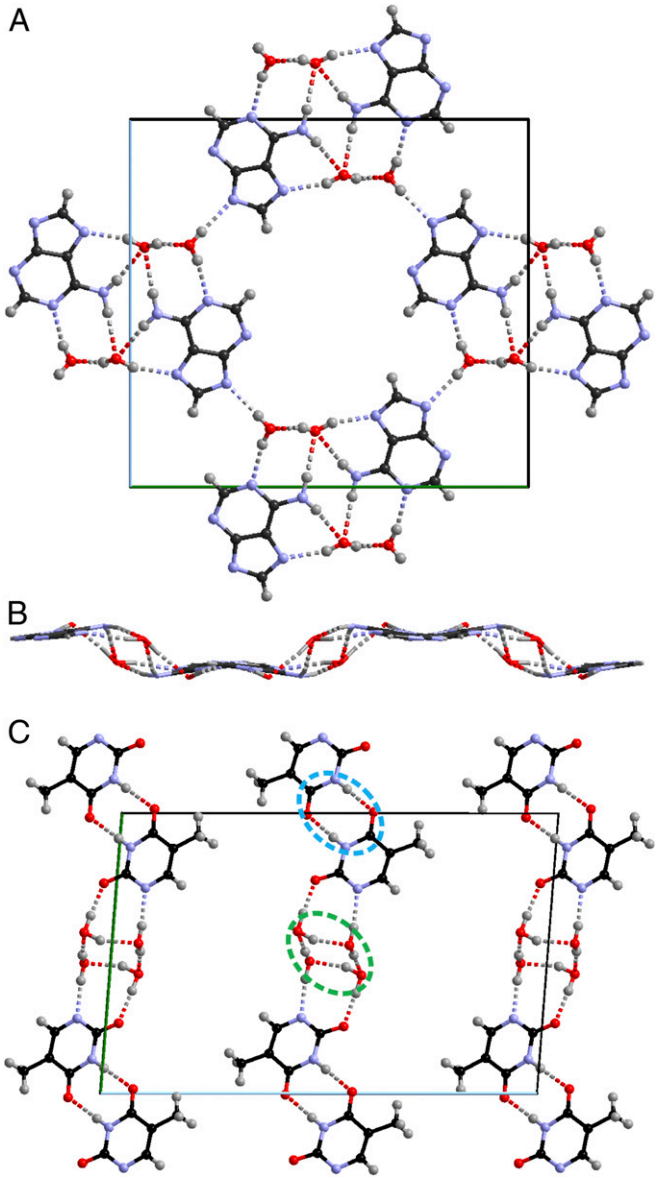


Fig. 2. The 2D sheet of water and $[Ad]^-$ in $[N_{4444}][Ad]\cdot 2H_2O$, projection on (*A*) the *bc* and (*B*) the *ab* plane. (*C*) Polymeric chains of $[Thy_2]^{2-}$ dimers (blue circle) and tetranuclear water bridges (green circle) in $[P_{4444}][Thy]\cdot 2H_2O$. Crystallographic axes are color coded: *b*, green; *c*, blue. $[N_{4444}]^+/[P_{4444}]^+$ cations are omitted for clarity.

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 as the $N(1)^-$ anion is adjacent to only one oxyanion, the N(3) conjugate base is to a higher extent stabilized by the dielectric field associated with water (30).

$[P_{4444}][Thy]\cdot 2H_2O$ crystallizes in the triclinic space group $P\bar{1}$ with two independent but practically identical deprotonated thymine $[Thy]^-$ anions, two $[P_{4444}]^+$ cations, and four water molecules in the asymmetric unit (*SI Appendix*, Fig. S2). The $[Thy]^-$ anion is formed by removal of a proton from the N(1) site (Fig. 1) in the crystalline $[P_{4444}][Thy]\cdot 2H_2O$, and this is consistent with the anions as discussed below and observed in the gas phase (31). The crystal structure is characterized by infinite 1D zigzag chains formed by $[Thy_2]^{2-}$ dimers and tetranuclear water bridges (Fig. 2*C*). The dimers are connected via N(3)H···O hydrogen bonds, while water bridges are interconnected via OH···O hydrogen bonds.

The mutual interaction environments around the $[Thy]^-$ anions and $[P_{4444}]^+$ cations are slightly different from each other. In contrast to the previous structure with $[Ad]^-$, $[P_{4444}]^+$ cations form not only separate layers but separate $[[Thy]^-/2H_2O]_n$ anionic chains along both *a* and *c* directions (*SI Appendix*, Fig. S5). $[Thy]^-$ anions are connected with three or four $[P_{4444}]^+$ via weak nonclassical CH···π intermolecular interactions and weak CH···O hydrogen bonds. Both $[P_{4444}]^+$ cations are connected to four $[Thy]^-$ anions and a water molecule each.

Cocrystal Base Pairing, $[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$. It has been shown that the DNA double helix can accommodate huge structural variability maintaining Watson–Crick (WC) base pairing (32, 33). Hoogsteen (HG) base pairs in DNA duplexes are frequently characterized as transient with a low population (34). Deviation from the WC base pairing in the double helix is usually connected to external factors, e.g., interactions with additional molecules, 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.

$[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$ crystallizes in the monoclinic acentric space group *Cc* with one $[Ad]^-$ and $[Thy]^-$ anions, two $[P_{4444}]^+$ 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 3*A*, red and blue circles, respectively). Furthermore, it is connected to a third neutral thymine via a direct NH···N bond complemented by a CH···H_2O···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···H_2O···OC). From a structural motif viewpoint, these can be considered as the four nucleobase unit $[HThyAdHThyThy]^{2-}$ (Fig. 3*A*, orange circle). Similar to $[P_{4444}][Thy]\cdot 2H_2O$, 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 (7.703 Å) of the helical motifs along the *a* axis (Fig. 3*A*, green circle, and Fig. 3*B*). 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.

$[P_{4444}]^+$ cations do not play any independent role in the crystal structure of $[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$ 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 cationic/anionic environment here follows the general structural complexity. The $[Ad]^-$ anion is connected to two $[P_{4444}]^+$ cations via the same weak nonclassical CH···π bonds. The $[Thy]^-$ anion is connected to three $[P_{4444}]^+$ cations,

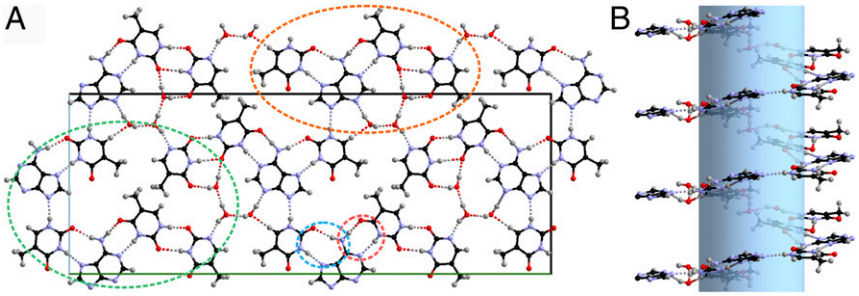


Fig. 3. (*A*) Molecular arrangements and hydrogen bond patterns in the crystal structure of $[P_{4444}]_2$[Ad][Thy]·$3H_2O$·2HThy and (*B*) helical motif along the *a* axis. Crystallographic axes are color coded: *b*, green; *c*, blue. $[P_{4444}]^+$ cations are omitted for clarity.

while both neutral HThy molecules are connected to four. Both $[P_{4444}]^+$ cations show slightly different connectivity: $[P1_{4444}]^+$ interacts with 1 $[Ad]^-$, 7 HThy, and 4 water molecules and $[P2_{4444}]^+$ 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 $-NH_2$ (**B**; *SI Appendix*, Fig. S7) is 7.6 kcal/mol lower in energy than the cis form, whereas the $[Thy]^-$ anion with the deprotonated N located between the two C=O groups (**A**) is 10.5 kcal/mol less favorable. If the effects of solvation are included at the B3LYP/aug-cc-PVDZ/COSMO self-consistent reaction field level, the energy difference for the two different N sites in Thy^- decreases to 2.0 kcal/mol, consistent with experiments (30). The acidity for HThy is essentially the same as that for HAd. These results are fully consistent with the literature and previous experiments (24, 45, 46), and more details are given in *SI Appendix*.

The $[HAdAd]^-$, $[HThyThy]^-$, $[HAdThy]^-$, and $[AdHThy]^-$ dimers were optimized at the DFT B3LYP/aD level and at the composite correlated molecular orbital theory G3(MP2) and G3MP2(B3) levels; all three methods give qualitatively comparable results (Table 1). The formation of $[HAdAd]^-$ from HAd and $[Ad]^-$ is exothermic by ∼29 kcal/mol for the most stable structure with the NH on the five-member ring interacting with a nitrogen on the anionic five-member ring and an NH_2 group interacting with a six-member ring (*SI Appendix*, Fig. S8). The most stable $[HThyThy]^-$ dimeric structure has both CH_3 groups being maximally distant from the interacting C=O groups (*SI Appendix*, Fig. S9). In contrast to $[HAdAd]^-$, there is a second low-energy configuration that is just 2.4 kcal/mol higher in energy. Note that the structures of the isolated gas-phase dimer monoanions differ from the structure observed for the $[ThyThy]^{2-}$ hydrated dianion dimers in the crystal. The most stable structure of the mixed anionic $[HAdThy]^-$ dimer adopts the QHG mode (Fig. 1 and *SI Appendix*, Fig. S10) but contains deprotonated $[Thy]^-$, whereas the structure with deprotonated $[Ad]^-$ adopts the WC mode and is 10 kcal/mol higher. That is surprisingly large as both compounds have the same gas-phase acidity. For the neutral HAdHThy dimer both WC and HG modes are roughly comparable in the gas phase with the latter being slightly lower. The binding energy for the neutral dimer is about half that of the anionic one.

The $[HThyAdHThy]^-$ moiety observed in the crystal structure optimized to $[ThyHAdHThy]^-$, with a proton transfer to the $[Ad]^-$ from a HThy (*SI Appendix*, Fig. S11). This result of proton transfer is consistent with the lowest-energy structure for the mixed dimer (Table 2), and the resulting connectivity modes (coexisting QHG and WC) perfectly match the experimental observation despite different localization of the proton. The addition of HThy to the mixed dimer is exothermic by −10 kcal/mol for the most stable mixed dimer structure. An alternative structure that does not involve proton transfer from the neutral thymine to an adenine anion is $[AdHThyAd]^{2-}$. For the isolated gas-phase trimer structure, the connectivity modes between the base pair are HG and QWC, which is opposite to the crystal structure with counterions and water molecules present. We also optimized the structure of the tetramer $[HThy][Ad]^-[HThy][Thy]^-·3H_2O$ starting from the crystal structure which led to $[Thy]^-[HAd][HThy][Thy]^-·3H_2O$ with proton transfer from the [HThy] to the adjacent $[Ad]^-$.

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

	B3LYP/aD		G3(MP2)		G3MP2(B3)	
Reaction	ΔH_{gas}	ΔG_{gas}	ΔH_{gas}	ΔG_{gas}	ΔH_{gas}	ΔG_{gas}
$[Ad]^-$ + HAd **B** → $[HAdAd]^-$ (1)	−27.4	−15.7	−28.8	−19.2	−29.2	−18.8
$[Ad]^-$ + HAd **B** → $[HAdAd]^-$ (2)	−13.8	−2.6	−14.9	−5.9	−15.4	−5.4
$[Thy]^-$ **B** + HThy → $[HThyThy]^-$ (1)	−22.0	−11.6	*	*	−23.8	−14.0
$[Thy]^-$ **B** + HThy → $[HThyThy]^-$ (2)	−19.7	−9.6	−20.3	−12.5	−21.3	−11.6
$[Thy]^-$ **B** + HThy → $[HThyThy]^-$ (3)	−11.4	−1.0	*	*	−14.3	−1.9
$[Thy]^-$ **B** + HThy → $[HThyThy]^-$ (4)	−10.4	−0.1	−12.1	−3.6	−12.6	−3.2
$[Ad]^-$ + HThy → $[HAdThy]^-$ (1, QHG)	−22.9	−12.0	−25.0	−13.3	−25.2	−12.9
$[Ad]^-$ + HThy → $[AdHThy]^-$ (2, WC)	−13.3	−2.3	−14.6	−4.7	−15.4	−4.6
HAd **B** + HThy → [HAdHThy] (HG)	−12.3	−0.2	−14.2	−4.6	−14.4	−3.9
HAd **B** + HThy → [HAdHThy] (WC)	−11.6	0.4	−13.4	−3.9	−13.7	−3.3

*Did not converge.

Table 2. Complexation energies for oligomers at the B3LYP/aD level at 298 K in kcal/mol

Reaction	ΔH_{gas}	ΔG_{gas}
$[Ad]^-$ + 2HThy → $[HThyHAdThy]^-$	−33.0	−10.7
$[Thy]^-$ **B** + HAd **B** + HThy → $[HThyHAdThy]^-$	−42.0	−18.7
$[HAdThy]^-$ (2) + HThy → $[HThyHAdThy]^-$	−19.8	−8.3
$[HThyAd]^-$ (1) + HThy → $[HThyHAdThy]^-$	−10.1	1.3
HAd **A** + HThy + $[Thy]^-$ **A** + $[Thy]^-$ **B** + $3H_2O$ → $[Thy]^-[HAd][HThy]$ $[Thy]^-\cdot 3H_2O$	−69.3	−8.4
$[Ad]^-$ + 3HThy + $[Thy]^-$ **B** + $3H_2O$ → $[HThy]_2[Ad]^-[HThy][Thy]^-\cdot 3H_2O$	−71.7	0.8

Finally, we optimized the pentamer $[HThy]_2[Ad]^-[HThy][Thy]^-\cdot 3H_2O$ 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 NH_2 group on the six-member ring. For the monohydrated $[Thy]^-$, the H_2O is a proton donor to a C=O and to the N^- center on the ring. The addition of a second H_2O results in a drop of 1 kcal/mol as compared to adding the first H_2O for $[Ad]^-$ and 3.5 kcal/mol for $[Thy]^-$. Addition of two further H_2O molecules to $[Ad]^-$ averages to −9 kcal/mol per H_2O and for $[Thy]^-$ −10 kcal/mol.

Comparing the addition of four H_2O 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 H_2O 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 H_2O 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 H_2O in $[Ad]^-$ **A** as discussed above.

The structure of the dianionic $[Thy_2]^{2-}$ dimer solvated by two terminal H_2O 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 H_2O that can be used for further hydrogen bonding to another species. The resulting dimeric complex (*SI Appendix*, Fig. S15) shows the ring of four H_2O molecules between the two $[Thy_2]^{2-}$ dimers that holds the entire complex together.

Table 3. Energies of hydration for $[Ad]^-$ and $[Thy]^-$ at the G3(MP2) level at 298 K in kcal/mol

Reactants	Product	ΔH_{gas}	$\Delta\Delta H_{gas}$	ΔG_{gas}
$[Ad]^-$				
1 H_2O	$[Ad]^-H_2O$ **A**	−13.9	0.0	−5.6
1 H_2O	$[Ad]^-H_2O$ **B**	−13.5	0.4	−4.8
2 H_2O	$[Ad]^-(H_2O)_2$	−26.9	−13.0	−9.9
4 H_2O	$[Ad]^-(H_2O)_4$	−44.8	−17.9 ($2H_2O$)	−11.9
$[Thy]^-$ **A**				
1 H_2O	$[Thy]^-H_2O$	−25.2	0.0	−16.6
2 H_2O	$[Thy]^-(H_2O)_2$	−36.0	−10.8	−18.8
4 H_2O	$[Thy]^-(H_2O)_4$	−55.7	−19.7 ($2H_2O$)	−22.4
$[Thy]^-$ **B**				
1 H_2O	$[Thy]^-H_2O$	−14.3	0.0	−6.1
2 H_2O	$[Thy]^-(H_2O)_2$	−25.1	−10.8	−8.2
4 H_2O	$[Thy]^-(H_2O)_4$	−44.8	−19.7 ($2H_2O$)	−11.9

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 backup 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 $[P_{4444}][OH]$ (40 wt % in water) and tetrabutylammonium hydroxide $[N_{4444}][OH]$ (55 wt % in water) were purchased from Fisher Scientific. All chemicals were used as received unless otherwise stated.

Crystallization Procedures.

$[N_{4444}][Ad]\cdot 2H_2O$. Adenine (1 mmol; 0.135 g) and liquid $[N_{4444}][OH]$ (55 wt % in water) (1 mmol, 0.476 g) were mixed 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 block-shaped crystals of $[N_{4444}][Ad]\cdot 2H_2O$ formed in the reaction vessel, and the vessel was allowed to cool to room temperature.

$[P_{4444}][Thy]\cdot 2H_2O$. Thymine (1 mmol; 0.126 g) and liquid $[P_{4444}][OH]$ (40 wt % in water) (1 mmol, 0.691 g) were mixed 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 block-shaped crystals of $[P_{4444}][Thy]\cdot 2H_2O$ formed in the reaction vessel, and the vessel was allowed to cool to room temperature.

$[P_{4444}]_2[Ad][Thy]\cdot 3H_2O\cdot 2HThy$. Adenine (1 mmol; 0.135 g), thymine (1 mmol; 0.126 g), and liquid $[P_{4444}][OH]$ (40 wt % in water) (1 mmol, 0.691 g) were

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 $[P_{4444}]_2$[Ad][Thy]·3H_2O·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 IμS 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 free of charge 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 Lynxeye linear position sensitive detector (Bruker AXS). The bulk semisolid sample of $[P_{4444}]_2$[Ad][Thy]·3H_2O·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 $[P_{4444}]_2$[Ad][Thy]·3H_2O·2HThy, $[N_{4444}]$[Ad]·2H_2O, $[P_{4444}]$[Thy]·2H_2O, adenine, and, thymine. The diffractograms indicate that the bulk solid is a mixture of $[P_{4444}]_2$[Ad][Thy]·3H_2O·2HThy, adenine, and thymine. PXRD data for $[N_{4444}]$[Ad]·2H_2O and $[P_{4444}]$[Thy]·2H_2O 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 DZVP2 (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 G3MP2(B3) (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).

$$HA \rightarrow H^+ + A^-. \quad [1]$$

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.

1. J. D. Watson, F. H. C. Crick, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738 (1953).
2. B. E. Tropp, *Molecular Biology: Genes to Proteins*, (Jones & Bartlett Publishers, Sudbury, MA, 2012).
3. G. R. Desiraju, *Crystal Engineering: The Design of Organic Solids*, (Elsevier, Amsterdam, 1989).
4. J. Lehn, *Supramolecular Chemistry: Concepts and Perspectives*, (VCH, Weinheim, 1995).
5. B. D. Michael, P. O'Neill, Molecular biology. A sting in the tail of electron tracks. *Science* **287**, 1603–1604 (2000).
6. S. Steenken, Purine bases, nucleosides, and nucleotides: Aqueous solution redox chemistry and transformation reactions of their radical cations and e- and OH adducts. *Chem. Rev.* **89**, 503–520 (1989).
7. W. Wang, H. W. Hellinga, L. S. Beese, Structural evidence for the rare tautomer hypothesis of spontaneous mutagenesis. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17644–17648 (2011).
8. I. J. Kimsey, K. Petzold, B. Sathyamoorthy, Z. W. Stein, H. M. Al-Hashimi, Visualizing transient Watson-Crick-like mispairs in DNA and RNA duplexes. *Nature* **519**, 315–320 (2015).
9. B. Boudaïffa, P. Cloutier, D. Hunting, M. A. Huels, L. Sanche, Resonant formation of DNA strand breaks by low-energy (3 to 20 eV) electrons. *Science* **287**, 1658–1660 (2000).
10. Y. A. Berlin, A. L. Burin, M. A. Ratner, Charge hopping in DNA. *J. Am. Chem. Soc.* **123**, 260–268 (2001).
11. M. C. Lind, P. P. Bera, N. A. Richardson, S. E. Wheeler, H. F. Schaefer 3rd, The deprotonated guanine-cytosine base pair. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7554–7559 (2006).
12. S. Kim, M. C. Lind, H. F. Schaefer 3rd, Structures and energetics of the deprotonated adenine-uracil base pair, including proton-transferred systems. *J. Phys. Chem. B* **112**, 3545–3551 (2008).
13. P. P. Bera, H. F. Schaefer 3rd, (G-H)*-C and G-(C-H)* radicals derived from the guanine.cytosine base pair cause DNA subunit lesions. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6698–6703 (2005).
14. J. Berdys, I. Anusiewicz, P. Skurski, J. Simons, Damage to model DNA fragments from very low-energy (<1 eV) electrons. *J. Am. Chem. Soc.* **126**, 6441–6447 (2004).
15. S. Xia, W. H. Konigsberg, Mispairs with Watson-Crick base-pair geometry observed in ternary complexes of an RB69 DNA polymerase variant. *Protein Sci.* **23**, 508–513 (2014).
16. M. D. Topal, J. R. Fresco, Complementary base pairing and the origin of substitution mutations. *Nature* **263**, 285–289 (1976).
17. R. D. Rogers, "Nucleoside analog salts with improved solubility and methods of forming same." US patent 20160002240 A1 (2014).
18. J. L. Shamshina *et al*., Acyclovir as an ionic liquid cation or anion can improve aqueous solubility. *ACS Omega* **2**, 3483–3493 (2017).
19. M. Harańczyk, M. Gutowski, X. Li, K. H. Bowen, Bound anionic states of adenine. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 4804–4807 (2007).
20. S. Ptasinska, S. Denifl, P. Scheier, E. Illenberger, T. D. Märk, Bond- and site-selective loss of H atoms from nucleobases by very-low-energy electrons (<3 eV). *Angew. Chem. Int. Ed. Engl.* **44**, 6941–6943 (2005).
21. H. Abdoul-Carime, S. Gohlke, E. Illenberger, Site-specific dissociation of DNA bases by slow electrons at early stages of irradiation. *Phys. Rev. Lett.* **92**, 168103 (2004).
22. M. K. Mishra *et al*, Crystallographic information file for [N4444][Ad]·2H2O. Cambridge Structural Database. https://www.ccdc.cam.ac.uk/structures/Search?ccdc=1976655. Deposited 9 January 2020.
23. H. Wincel, Microhydration of deprotonated nucleobases. *J. Am. Soc. Mass Spectrom.* **27**, 1383–1392 (2016).
24. S. Sharma, J. K. Lee, Acidity of adenine and adenine derivatives and biological implications. A computational and experimental gas-phase study. *J. Org. Chem.* **67**, 8360–8365 (2002).
25. T. J. Kistenmacher, The crystal and molecular structure of the tetraphenylarsonium salt of the monoanion of adenine $[C_{24}H_{20}As]^+[C_5N_5H_4]^-$.3H2O. *Acta Crystallogr. B* **29**, 1974–1979 (1973).
26. T. Steiner, The hydrogen bond in the solid state. *Angew. Chem. Int. Ed. Engl.* **41**, 49–76 (2002).
27. M. K. Mishra *et al*, Crystallographic information file for [P4444][Thy]·2H2O. Cambridge Structural Database. https://www.ccdc.cam.ac.uk/structures/Search?ccdc=1976656. Deposited 9 January 2020.
28. R. Shapiro, S. Kang, Uncatalyzed hydrolysis of deoxyuridine, thymidine, and 5-bromodeoxyuridine. *Biochemistry* **8**, 1806–1810 (1969).
29. M. Liu *et al*., Gas-phase thermochemical properties of pyrimidine nucleobases. *J. Org. Chem.* **73**, 9283–9291 (2008).
30. M. A. Kurinovich, J. K. Lee, The acidity of uracil from the gas phase to solution: The coalescence of the N1 and N3 sites and implications for biological glycosylation. *J. Am. Chem. Soc.* **122**, 6258–6262 (2000).
31. B. F. Parsons *et al*., Anion photoelectron imaging of deprotonated thymine and cytosine. *Phys. Chem. Chem. Phys.* **9**, 3291–3297 (2007).
32. M. T. Record Jr. *et al*., Double helical DNA: Conformations, physical properties, and interactions with ligands. *Annu. Rev. Biochem.* **50**, 997–1024 (1981).
33. M. K. Mishra *et al*, Crystallographic information file for [P4444]2[Ad][Thy]·3H2O·2HThy. Cambridge Structural Database. https://www.ccdc.cam.ac.uk/structures/Search?ccdc=1976657. Deposited 9 January 2020.
34. E. N. Nikolova *et al*., Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498–502 (2011).
35. M. Kitayner *et al*., Diversity in DNA recognition by p53 revealed by crystal structures with Hoogsteen base pairs. *Nat. Struct. Mol. Biol.* **17**, 423–429 (2010).
36. H. Yang, Y. Zhan, D. Fenn, L. M. Chi, S. L. Lam, Effect of 1-methyladenine on double-helical DNA structures. *FEBS Lett.* **582**, 1629–1633 (2008).
37. F. C. Seaman, L. Hurley, Interstrand cross-linking by bizelesin produces a Watson-Crick to Hoogsteen base-pairing transition region in d(CGTAATTACG)2. *Biochemistry* **32**, 12577–12585 (1993).
38. N. A. Richardson, S. S. Wesolowski, H. F. Schaefer, The adenine–thymine base pair radical anion: Adding an electron results in a major structural change. *J. Phys. Chem. B* **107**, 848–853 (2003).
39. I. Al-Jihad, J. Smets, L. Adamowicz, Covalent anion of the canonical adenine–thymine base pair. Ab initio study. *J. Phys. Chem. A* **104**, 2994–2998 (2000).
40. M. C. Etter, S. M. Reutzel, C. G. Choo, Self-organization of adenine and thymine in the solid state. *J. Am. Chem. Soc.* **115**, 4411–4412 (1993).

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. Row, 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.* **798**, 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. L. 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).