JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

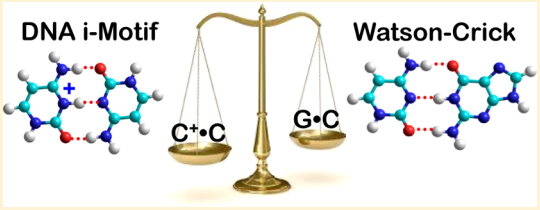
Base-Pairing Energies of Proton-Bound Heterodimers of Cytosine and Modified Cytosines: Implications for the Stability of DNA *i*-Motif Conformations

Bo Yang and M. T. Rodgers*

Department of Chemistry, Wayne State University, Detroit, Michigan 48202, United States

Supporting Information

ABSTRACT: The DNA *i*-motif conformation was discovered in $(CCG)_n \bullet (CGG)_n$ trinucleotide repeats, which are associated with fragile X syndrome, the most widespread inherited cause of mental retardation in humans. The DNA *i*-motif is a four-stranded structure whose strands are held together by proton-bound dimers of cytosine ($C^+ \bullet C$). The stronger base-pairing interactions in $C^+ \bullet C$ proton-bound dimers as compared to Watson−Crick G•C base pairs are the major forces responsible for stabilization of *i*-motif conformations. Methylation of cytosine results in silencing of the FMR1 gene and causes fragile X syndrome. However, the influence of methylation or other modifications such as halogenation of cytosine on the base-pairing energies (BPEs) in the *i*-motif remains elusive. To address this, proton-bound heterodimers of cytosine and 5-methylcytosine, 5-fluorocytosine, 5-bromocytosine, and 5-iodocytosine are probed in detail. Experimentally, the BPEs of proton-bound heterodimers of cytosine and modified cytosines are determined using threshold collision-induced dissociation (TCID) techniques. All modifications at the 5-position of cytosine are found to lower the BPE and therefore would tend to destabilize DNA *i*-motif conformations. However, the BPEs in these proton-bound heterodimers still significantly exceed those of the Watson−Crick G•C and neutral C•C base pairs, suggesting that $C^+ \bullet C$ mismatches are still energetically favored such that *i*-motif conformations are preserved. Excellent agreement between TCID measured BPEs and B3LYP calculated values is found with the def2-TZVPPD and 6-311+G(2d,2p) basis sets, suggesting that calculations at these levels of theory can be employed to provide reliable energetic predictions for related systems.



INTRODUCTION

At least one-third of the human genome consists of repetitive sequences of various types, including low-copy-number repeats (LCRs), long and short interspersed transposon-derived elements (LINEs and SINEs), and tandem repeats.[1] Trinucleotide repeats, known as one of the short tandem repeats, reside in exons and therefore may be beneficial, neutral, or deleterious. Expansions of repeated DNA trinucleotides, a type of mutation where trinucleotide repeats in a certain gene exceed the normal, stable threshold, were found to cause a new class of molecular disease called trinucleotide repeats disorders.[2,3] To date, over 20 neurological disorders are recognized as being associated with trinucleotide repeats expansions, and this number is expected to grow. Fragile X syndrome was the first triplet disease identified and is the most widespread inherited cause of mental retardation in humans. It has previously been shown that the unstable and abnormal expansion of the $(CCG)_n \bullet (CGG)_n$ trinucleotide repeat affects the *fragile X mental retardation 1* (*FMR1*) gene on the X chromosome, resulting in a failure to express the fragile X mental retardation protein (FMRP), which is required for normal neural development.[4] Over the past several years, researchers have concentrated on the challenging task of identifying the mechanism through which the expanded trinucleotide repeat leads to abnormal cellular function. The cause of fragile X syndrome is often related to various noncanonical DNA structures that the single strands of $(CCG)_n$ as well as $(CGG)_n$ may adopt.[3,5,6] The noncanonical DNA *i*-motif conformation was first discovered in 1993 by Gehring et al.[6] and later in $(CCG)_n \bullet (CGG)_n$ trinucleotide repeats.[7] The secondary structure of the DNA *i*-motif is a four-stranded structure in which two parallel-stranded DNA duplexes are zipped together in an antiparallel orientation by intercalated proton-bound dimers of cytosine ($C^+ \bullet C$).[6] Since the discovery of the *i*-motif, the biological roles of *i*-motif structures as well as their potential in pharmaceutical applications have drawn great attention. Recent studies have shown that the structure of the *i*-motif is preserved in the gas phase when electrospray ionization (ESI) is used as the ionization technique,[8] indicating that gas-phase studies can indeed provide insight into solution-phase structures and functions.

The structure of the proton-bound dimer of cytosine is well established and involves binding of the canonical form of neutral cytosine to N3 protonated cytosine via three hydrogen bonds.[6,9,10] However, the 5-position of cytosine is often methylated in vivo, but is not involved in the base-pairing interactions. Therefore, the DNA maintains the same sequence when cytosine is 5-methylated, leading to inheritable changes in the chromatin structure. Indeed, expansion of

Received: September 19, 2013
Published: December 7, 2013

ACS Publications © 2013 American Chemical Society

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

$(CCG)_n$•$(CGG)_n$ repeats beyond 230 trinucleotides have been found to lead to hypermethylation of both the CCG repeats and the *FMR1* promoter region, resulting in transcriptional silencing of the *FMR1* gene and a deficiency of its protein product, FMRP.[11,12] Halogenation at the 5-position could also exert profound biological consequences and influence DNA−protein interactions based on its size and electron-withdrawing properties.[13−16]

Geometry optimizations and single-point energy calculations performed at the B3LYP/6-311+G(2d,2p)//B3LYP/6-31G* level of theory found that the base-pairing energy (BPE) for the proton-bound dimer of cytosine (C^+•C) is 169.7 kJ/mol, whereas the BPEs of the canonical Watson−Crick G•C and neutral C•C base pairs are only 96.6 and 68.0 kJ/mol, respectively, indicating that the stronger base-pairing interactions in the C^+•C homodimer are likely the major factor that stabilizes the noncanonical DNA *i*-motif conformation. Previously, the structures of the proton-bound homo- and heterodimers of cytosine and several 5-substituted cytosines, (5*x*C)H^+(5*y*C), where *x*, *y* = H, Me, F, and Br, were studied using infrared multiple photon dissociation (IRMPD) action spectroscopy techniques.[10] Comparison of the measured IRMPD action spectra with the linear IR spectra computed for the stable low-energy tautomeric conformations of the proton-bound dimers confirmed that the structure of the proton-bound dimer of cytosine determined in condensed-phase NMR studies[6] is preserved in the gas phase. This work also found that methylation and halogenation at the 5-position of cytosine does not alter the preferred base-pairing interactions in these proton-bound dimers. However, these modifications likely influence the strength of the base-pairing interactions.[10] Therefore, a comprehensive study is required to determine the influence of methylation or halogenation of cytosine on the BPEs. The strength of hydrogen-bonding interactions in similar proton-bound heterodimers[17−19] has been accurately determined using threshold collision-induced dissociation (TCID) techniques, indicating that this technique is capable of providing accurate quantitative determinations of the strength of binding in multiply hydrogen-bonded proton-bound dimers. Thus, we employed the TCID technique to quantitatively determine the BPEs of the ground-state proton-bound homodimers of cytosine and several modified cytosines, (5*x*C)H^+(5*x*C), where *x* = H, Me, F, Br, and I, and excellent agreement was achieved between the B3LYP/def2-TZVPPD calculated and TCID measured BPEs.[20] However, except in cases of hypermethylation, the simultaneous modification of cytosine residues from both strands is likely to be less probable than modification of a single strand. Therefore, we expand our previous TCID studies to the analogous proton-bound heterodimers to determine the BPEs of these species and thus elucidate the effects of modifications on the base-pairing interactions. The proton affinity (PA) of 18C6 was accurately determined from the competitive dissociation of the proton-bound heterodimers of 18C6 and peptidomimetic bases and amino acids in the TCID studies.[21] The determination of the PAs and the preferred protonation sites of the nucleobases contributes to the understanding of the chemical processes that DNA molecules undergo in the condensed phase. However, the PAs of the 5-substituted cytosines have not been reported. Our TCID studies of the proton-bound heterodimers of cytosine and 5-substituted cytosines provide the advantageous sidelight that the relative N3 PAs of the 5-substituted cytosines are also extracted from the experimental data using cytosine as a reference base. The measured values are compared with theoretical results calculated at the B3LYP and MP2(full) levels of theory to evaluate the ability of each level of theory for predicting accurate energetics.

EXPERIMENTAL AND COMPUTATION SECTION

General Procedures. The TCID behavior of four proton-bound heterodimers, (C)H^+(5*x*C), where *x* = Me, F, Br, and I, was studied using a guided ion beam tandem mass spectrometer that has been described in detail previously.[22] The (C)H^+(5*x*C) heterodimers were generated by ESI from solutions containing 0.5−1 mM cytosine, 0.5−1 mM of the 5-substituted cytosine, and 1% (v/v) acetic acid in an approximately 50:50 MeOH/H_2O mixture. The ions are desolvated, focused, and thermalized in a radio frequency (rf) ion funnel and hexapole ion guide collision cell interface. The thermalized ions emanating from the hexapole ion guide are extracted, accelerated, and focused into a magnetic sector momentum analyzer for mass analysis. Mass-selected ions are decelerated to a desired kinetic energy and focused into an rf octopole ion beam guide that acts as an efficient radial trap[23−25] for ions such that scattered reactant and product ions are not lost as they drift toward the end of the octopole. The octopole passes through a static gas cell where the (C)H^+(5*x*C) heterodimers undergo collision-induced dissociation (CID) with Xe[26−28] under nominally single collision conditions, ∼0.05−0.10 mTorr. Product ions and undissociated (C)H^+(5*x*C) heterodimers drift to the end of the octopole where they are focused into a quadrupole mass filter for mass analysis. The ions are detected using a secondary electron scintillation (Daly) detector and standard pulse counting techniques. Cytosine (C), 5-fluorocytosine (5FC), and 5-methylcytosine (5MeC) were purchased from Alfa Aesar (Ward Hill, MA, USA); 5-bromocytosine (5BrC) was purchased from Acros Organics (Fair Lawn, NJ, USA); 5-iodocytosine (5IC) was purchased from MP Biomedicals (Solon, OH, USA).

Theoretical Calculations. The stable low-energy tautomeric conformations of 5*x*C, H^+(5*x*C), and (C)H^+(5*x*C) proton-bound dimers, where *x* = H, Me, F, and Br, have been examined previously by Yang et al. at the B3LYP/6-31G* level as described in detail elsewhere.[10] In the present study, geometry optimizations and frequency analyses of the low-energy tautomeric conformations of 5*x*C, H^+(5*x*C), and the (C)H^+(5*x*C) heterodimers, where *x* = Me, F, Br, and I, were performed using Gaussian 09[29] at the B3LYP/6-31G*, B3LYP/def2-TZVPPD, and MP2(full)/6-31G* levels of theory. The def2-TZVPPD basis set[30] is a balanced basis set on all atoms at the triple-ζ level including polarization and diffuse functions and was obtained from the EMSL basis set exchange library.[31,32] Calculations for 5IC, H^+(5IC), and (C)H^+(5IC) were performed at the same levels of theory. However, the I atom was described using the effective core potentials (ECPs) and valence basis sets developed by Hay and Wadt[33] when the 6-31G* basis set was used, whereas the valence basis set and ECPs developed by Peterson et al.[34] were used with the def2-TZVPPD basis set. The polarizabilities of the neutral nucleobases required for threshold analyses were calculated at the PBE1PBE/6-311+G(2d,2p) level of theory, which has been shown to provide polarizabilities that exhibit better agreement with experimental values than the B3LYP functional employed here for structures and energetics.[35] Relaxed potential energy surface scans along the N3 hydrogen−O2 coordinate of the protonated nucleobase were performed at the B3LYP/6-31G* level of theory to provide candidate structures for transition states (TSs) for dissociation of the ground-state proton-bound heterodimers to produce ground-state neutral and protonated products. The actual TSs were obtained using the quasi-synchronous transit (QST3) method,[36] at the B3LYP/6-31G*, B3LYP/def2-TZVPPD, and MP2(full)/6-31G* levels of theory, using the input from the relevant minima (reactant and products) and an estimate of the TS obtained from the relaxed PES scans. Single-point energy calculations for 5*x*C, H^+(5*x*C), TSs, and the (C)H^+(5*x*C) heterodimers were performed at the B3LYP/6-311+G(2d,2p), B3LYP/def2-TZVPPD, and MP2(full)/6-311+G(2d,2p) levels of theory using geometries optimized at the B3LYP/6-31G*,

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

B3LYP/def2-TZVPPD, and MP2(full)/6-31G* levels, respectively. Frequency analyses at the MP2(full)/def2-TZVPPD require computational resources beyond those available to us; therefore, single-point energy calculations performed at the MP2(full)/def2-TZVPPD made use of the B3LYP/def2-TZVPPD optimized structures. Zero-point energy (ZPE) corrections were determined using vibrational frequencies calculated at the B3LYP and MP2(full) levels of theory and scaled by factors of 0.9804 and 0.9646, respectively.[37] To obtain accurate energetics, basis set superposition errors corrections (BSSEs) are also included in the calculated BPEs using the counterpoise approach.[38,39]

Thermochemical Analysis. The threshold regions of the measured CID cross sections are modeled to extract 0 and 298 K BPEs and relative N3 PAs using procedures developed elsewhere[40−47] that have been found to reproduce CID cross sections well.[18,20,48−50] Details regarding data handling and analysis procedures are described in the Supporting Information.

RESULTS AND DISCUSSION

Cross Sections for CID. Experimental cross sections were obtained for the interaction of Xe with four (C)H$^+$(5*x*C) heterodimers, where *x* = Me, F, Br, and I. Figure 1a shows representative data for the (C)H$^+$(5FC) complex. The energy-dependent CID cross sections of the other three (C)H$^+$(5*x*C) complexes exhibit similar behavior and are included in the Supporting Information as Figure S1. Loss of intact neutral C or 5*x*C via CID reactions 1 and 2 is observed for all four (C)H$^+$(5*x*C) complexes:

$$(C)H^+(5xC) + Xe \rightarrow H^+(5xC) + C + Xe \quad (1)$$

$$(C)H^+(5xC) + Xe \rightarrow H^+(C) + 5xC + Xe \quad (2)$$

The loss of intact C and production of H$^+$(5MeC) corresponds to the lowest-energy CID pathway for the (C)H$^+$(5MeC) complex. In contrast, the loss of intact 5*x*C and production of H$^+$(C) corresponds to the lowest-energy CID pathway for the (C)H$^+$(5FC), (C)H$^+$(5BrC), and (C)H$^+$(5IC) complexes. This behavior is consistent with fragmentation via IRMPD.[10]

Theoretical Results. As discussed above, theoretical structures for neutral 5*x*C and protonated H$^+$(5*x*C) as well as the proton-bound (C)H$^+$(5*x*C) heterodimers of these species were calculated at the B3LYP and MP2(full) levels of theory using the 6-31G* and def2-TZVPPD basis sets. The three most stable tautomeric conformations of 5FC and H$^+$(5FC) are shown in Figure 2. To differentiate the various stable low-energy tautomeric conformations of these species, lowercase Roman numerals are used to describe the tautomeric conformations of the neutral nucleobase, whereas uppercase Roman numerals with a "+" sign are used to describe the tautomeric conformations of the protonated nucleobase, and both are ordered based on the relative Gibbs free energies at 298 K of the low-energy tautomeric conformations of C and H$^+$(C). Because of the large electronegativity of fluorine as compared to the other halogens, the relative stabilities of the three most stable tautomers of neutral 5FC differ from those of the other nucleobases. All of the other neutral and protonated nucleobases examined here exhibit structures and relative stabilities similar to those of C and H$^+$(C) and are shown in Figure S2 (Supporting Information) for C and H$^+$(C). The B3LYP/def2-TZVPPD optimized ground-state structures of all four proton-bound heterodimers are shown in Figure 3. In all cases, the excess proton binds to the nucleobase with the higher PA in the ground-state structure. The ground-state conformer is designated as **II$^+$●●●i_3a** to indicate that the excited **II$^+$** tautomeric conformation of the protonated nucleobase, H$^+$(5*x*C) or H$^+$(C), binds to the ground-state **i** tautomeric conformation of the neutral base, C or 5*x*C. The underscore **3a** designation indicates that the binding occurs via three hydrogen-bonding interactions and the protonated and neutral bases are bound in an antiparallel configuration, which is the most commonly observed conformation in multistranded DNAs. Previous studies of the analogous proton-bound homodimers have shown that tautomerization to the ground-state products does not occur upon dissociation at threshold energies.[20] However, it is unclear whether or not the same behavior will be observed for the proton-bound heterodimers. Therefore, relaxed potential energy scans and TS calculations were performed to determine the height of the tautomerization barriers. The reaction coordinate diagram for adiabatic and diabatic dissociation of the (C)H$^+$(5FC) complex is shown in

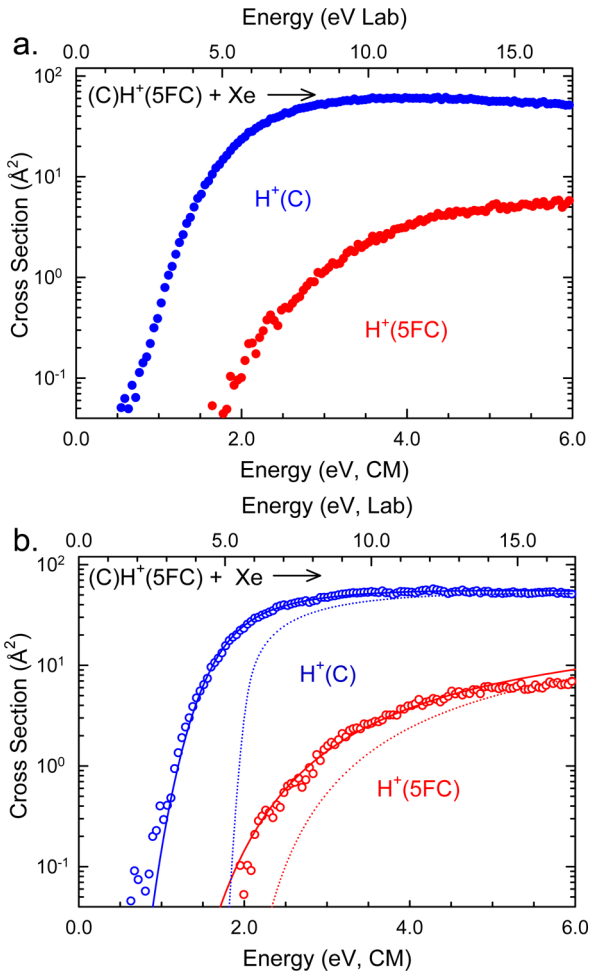


Figure 1. (a) Cross sections for collision-induced dissociation of the (C)H$^+$(5FC) complex with Xe as a function of collision energy in the center-of-mass frame (lower *x*-axis) and laboratory frame (upper *x*-axis). Data are shown for the Xe pressure of ~0.1 mTorr. (b) Zero-pressure-extrapolated cross section for collision-induced dissociation of the (C)H$^+$(5FC) complex with Xe in the threshold region as a function of kinetic energy in the center-of-mass frame (lower *x*-axis) and the laboratory frame (upper *x*-axis). The solid line shows the best fit to the data using the model of eq 3 (Supporting Information) convoluted over the neutral and ion kinetic and internal energy distributions. The dotted line shows the model cross section in the absence of experimental kinetic energy broadening for the (C)H$^+$(5FC) complex with an internal temperature of 0 K.

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

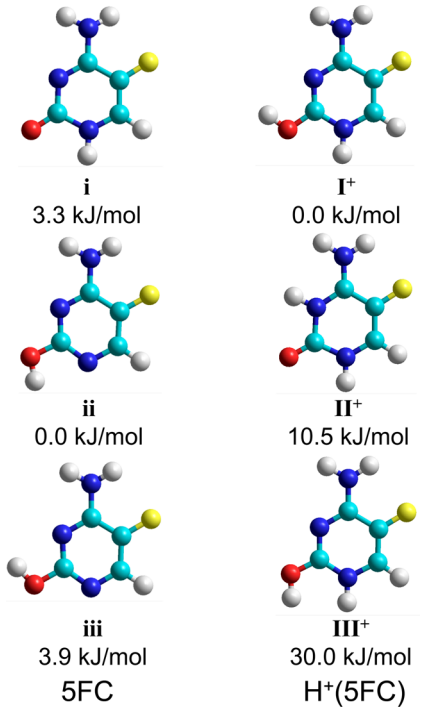


Figure 2. B3LYP/def2-TZVPPD optimized geometries of the three most stable tautomeric conformations of 5-fluorocytosine, 5FC, and protonated 5-fluorocytosine, H^+(5FC). Relative Gibbs free energies at 298 K are calculated at the B3LYP/def2-TZVPPD level of theory with ZPE corrections included.

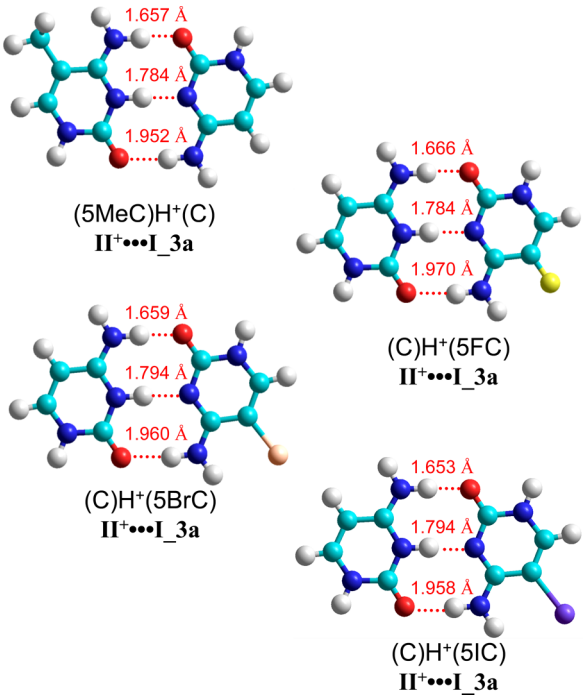


Figure 3. B3LYP/def2-TZVPPD optimized geometries of the ground-state **II**$^+$●●●**i_3a** conformations of four proton-bound heterodimers, (C)H^+(5*x*C), where *x* = Me, F, Br, and I.

Figure 4. Parallel results are obtained for the other three proton-bound heterodimers and are included in the Supporting Information as Figure S3. The relative energies along the PESs for the adiabatic and diabatic dissociation pathways determined

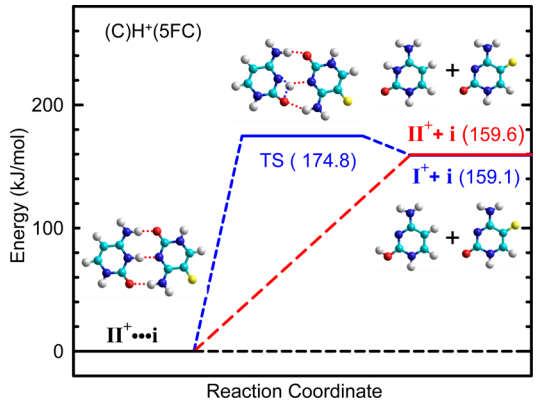


Figure 4. B3LYP/def2-TZVPPD potential energy surfaces for adiabatic and diabatic dissociation of the ground-state **II**$^+$●●●**I_3a** conformation of the (C)H^+(5FC) dimer to produce ground-state neutral, 5FC_**i**, and protonated, H^+(C)_**I**$^+$, products and ground-state neutral, 5FC_**i**, and excited protonated, H^+(C)_**II**$^+$, products, respectively.

at all four levels of theory for all four proton-bound heterodimers are summarized in Table S3 of the Supporting Information. In the TSs of all four proton-bound heterodimers, the excess proton is chelating with the O2 and N3 atoms of the protonated base. As can be seen in the PES of Figure 4, the barrier to tautomerization (174.8 kJ/mol) exceeds the diabatic dissociation energy (159.6 kJ/mol) indicating that, at threshold, tautomerization will not occur. The tautomerization barriers calculated at the B3LYP/def2-TZVPPD level of theory for the other (C)H^+(5*x*C) dimers lie in the range 174.3−176.6 kJ/mol and exceed the diabatic dissociation energies by 7.2−15.3 kJ/mol. The tautomerization barriers are also determined at the B3LYP/6-311+G(2d,2p), MP2(full)/6-311+G(2d,2p), and MP2(full)/def2-TZVPPD levels to ensure that the barriers computed are not highly sensitive to the basis set and level of theory employed. As can be seen in Table S3 (Supporting Information), the computed tautomerization barriers exceed the diabatic dissociation energies for all four proton-bound heterodimers regardless of the level of theory employed, confirming that tautomerization will not occur upon dissociation at threshold energies.

In the most stable conformation of the (5MeC)H^+(C) complex, **II**$^+$●●●**i_3a**, the bridging proton is bound to 5MeC (see Figure 3), suggesting that the PA of 5MeC is greater than that of C, as expected. On the basis of the ground-state structures of the four proton-bound (C)H^+(5*x*C) heterodimers, the relative order of PAs of these four nucleobases follow the order 5MeC > C > 5IC, 5BrC, 5FC. Thus, methyl-substitution of cytosine at the C5 position increases the N3 PA, whereas 5-halo-substitution decreases the N3 PA, consistent with the inductive effects of these substituents. Diabatic BPEs calculated at the B3LYP and MP2(full) levels of theory using the 6-311+G(2d,2p) and def2-TZVPPD basis sets are summarized in Table 1. ZPE and BSSE corrections are also included in the calculated BPEs.

Threshold Analysis. The thresholds for reactions 1 and 2 for four (C)H^+(5*x*C) proton-bound heterodimers are analyzed competitively using the model of eq S3 (Supporting Information). As discussed above, theory suggests that tautomerization will not occur upon CID at threshold energies such that the tautomeric conformations of the neutral and protonated nucleobase products are the same as in the proton-

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

Table 1. Base-Pairing Energies of the (C)H$^+$(5*x*C) Complexes at 0 K in kJ/mol[a]

		B3LYP[b]		MP2(full)[c]	
x	TCID	D_0	$D_{0,\ BSSE}$[d]	D_0	$D_{0,\ BSSE}$[d]
H[e]	169.9 (4.6)	171.7	168.9	155.2	136.7
		170.1	***169.2***	***149.3***	***136.0***
Me	163.6 (5.1)	169.7	166.8	153.9	141.0
		168.0	***167.1***	***145.5***	***132.3***
F	157.2(5.0)	161.2	158.2	145.5	134.0
		160.6	***159.6***	***139.2***	***126.4***
Br	166.1 (3.8)	160.3	157.3	143.7	127.6
		160.5	***159.5***	***136.9***	***122.3***
I	168.1 (3.8)	162.2	159.2	144.5	127.0
		162.6	***161.6***	***138.4***	***123.9***
MAD[f]	4.5 (0.6)[g]	4.7 (1.8)	4.6 (4.0)	16.4 (6.3)	35.3 (6.6)
		3.8 (2.2)	***3.9 (2.6)***	***23.1 (5.9)***	***36.8 (6.7)***

[a]Present results, uncertainties are listed in parentheses. [b]Calculated at the B3LYP level of theory including ZPE corrections. Values obtained using the 6-311+G(2d,2p) basis set are shown in standard font, whereas those computed using the def2-TZVPPD basis set are in bold italics. [c]Calculated at the MP2(full) level of theory using the B3LYP/def2-TZVPPD optimized geometries and including ZPE corrections. Values obtained using the 6-311+G(2d,2p) basis set are shown in standard font, whereas those computed using the def2-TZVPPD basis set are in bold italics. [d]Also includes BSSE corrections. [e]Values taken from reference 20. [f]Mean absolute deviation (MAD) between TCID and calculated values. [g]Average experimental uncertainty (AEU).

bound heterodimers, **II**$^+$ and **i**. In this case, a loose PSL TS model[46] is applied. The results of these analyses are summarized in Table 1, and a representative fit for the (C)H$^+$(5FC) complex is shown in Figure 1b. The other three (C)H$^+$(5*x*C) complexes exhibit similar behavior and are included in the Supporting Information as Figure S1. For the (C)H$^+$(5*x*C) complexes, where *x* = F, Br, and I, the experimental cross sections for reactions 1 and 2 are accurately reproduced using the loose PSL TS model[46] for the (C)H$^+$(5*x*C)_**II**$^+$●●●**i_3a** → H$^+$(C)_**II**$^+$ + 5*x*C_**i** and (C)H$^+$(5*x*C)_**II**$^+$●●●**i_3a** → H$^+$(5*x*C)_**II**$^+$ + C_**i** CID pathways, respectively, confirming our assumption that tautomerization does not occur upon dissociation at or near threshold energies and indicating that the ground-state (C)H$^+$(5*x*C)_**II**$^+$●●●**i_3a** structures are accessed in the experiments. The experimental cross sections for reactions 1 and 2 of the (5MeC)H$^+$(C) complex are best reproduced using the loose PSL TS model[46] for the (5MeC)H$^+$(C)_**II**$^+$●●●**i_3a** → H$^+$(5MeC)_**II**$^+$ + C_**i** and (5MeC)H$^+$(C)_**II**$^+$●●●**i_3a** → H$^+$(C)_**II**$^+$ + 5MeC_**i** CID pathways, respectively, indicating that the ground-state structure (5MeC)H$^+$(C)_**II**$^+$●●●**i_3a** is accessed in the experiments. Table S4 (Supporting Information) includes threshold values, E_{01} and E_{02}, for dissociation pathways 1 and 2, respectively. The threshold for the first dissociation channel, E_{01}, represents the BPE of the complex, whereas the difference between E_{01} and E_{02} represents the difference in the N3 PA of C and 5*x*C. The N3 PA of 5MeC is 16.1 ± 2.0 kJ/mol greater than the N3 PA of C, whereas the relative N3 PAs of 5FC, 5BrC, and 5IC are 22.9 ± 1.9, 18.3 ± 2.2, and 10.4 ± 1.4 kJ/mol lower than that of C, respectively.

The entropy of activation, $\Delta S^{\dagger}$, is a measure of the looseness of the TS and also a reflection of the complexity of the system. $\Delta S^{\dagger}$ is largely determined from the molecular constants used to model the energized molecule and the TS, but also depends on the threshold energy, E_0. The $\Delta S^{\dagger}$(PSL) values at 1000 K are listed in Table S4 (Supporting Information) and vary between 95 and 110 J K^{-1} mol^{-1} across these systems. These values are consistent with the noncovalent nature of the binding and the loose PSL TSs used to describe these systems.

DISCUSSION

Comparison of Experiment and Theory. The BPEs of the four proton-bound (C)H$^+$(5*x*C) heterodimers at 0 K measured here by TCID techniques are summarized in Table 1. Also listed in Table 1 are the BPEs of the proton-bound heterodimers calculated at the B3LYP and MP2(full) levels of theory using the 6-311+G(2d,2p) and def2-TZVPPD basis sets, and including ZPE and BSSE corrections. The measured and calculated BPE of the (C)H$^+$(C) complex is also included for comparison.[20] The agreement between the measured and B3LYP/def2-TZVPPD calculated BPEs is illustrated in Figure 5, whereas results for all four levels of theory are illustrated in Figure S4 of the Supporting Information. Overall, the B3LYP results exhibit better agreement with the measured BPEs, whereas the MP2(full) values are systematically low. The mean absolute deviations (MADs) between theory and experiment for the B3LYP/def2-TZVPPD and B3LYP/6-311+G(2d,2p)

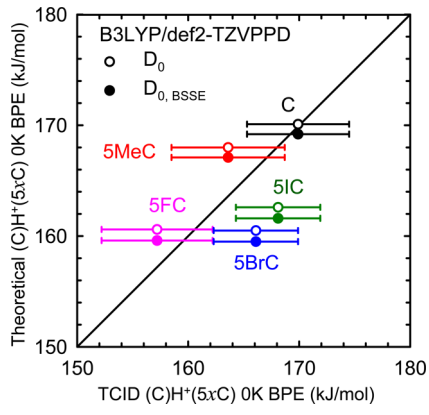


Figure 5. TCID measured (C)H$^+$(5*x*C) BPEs at 0 K (in kJ/mol), where *x* = H, Me, F, Br, and I, plotted versus B3LYP/def2-TZVPPD calculated values. The solid circles (●) represent theoretical values that include BSSE corrections, and the open circles (○) represent values without BSSE corrections. The black solid diagonal line indicates the values for which calculated and measured dissociation energies are equal.

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

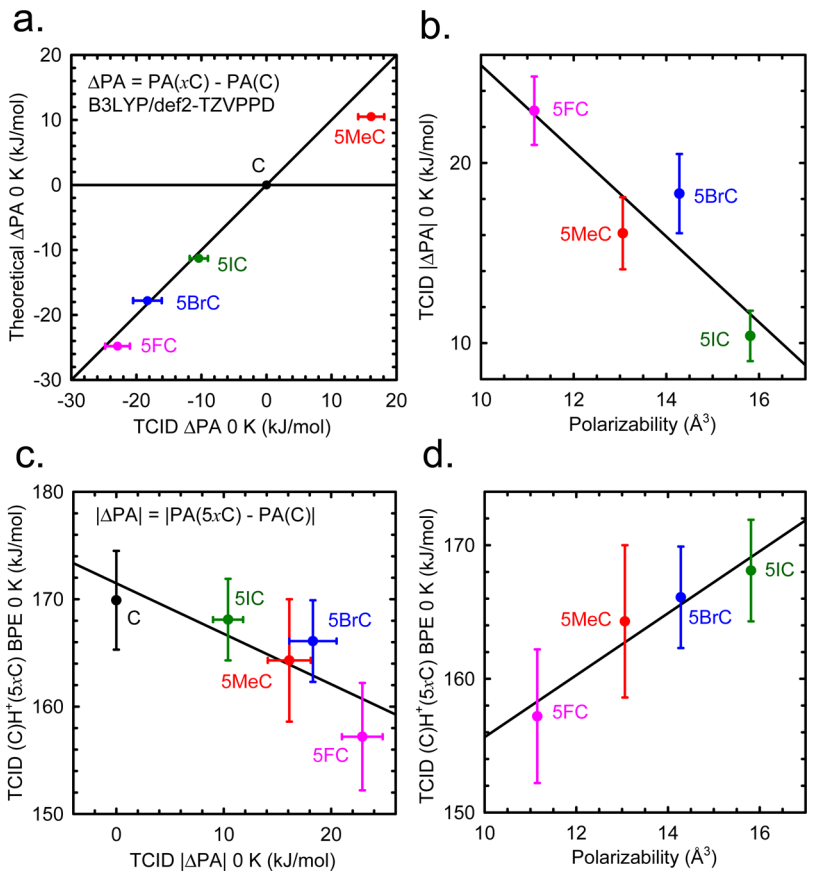


Figure 6. (a) B3LYP/def2-TZVPPD calculated versus TCID measured relative N3 PAs of 5*x*C at 0 K (in kJ/mol), where *x* = H, Me, F, Br, and I. (b) TCID measured relative N3 PAs of 5*x*C (in kJ/mol) plotted versus the calculated polarizability volumes of 5*x*C, where *x* = Me, F, Br, and I. (c) TCID measured (C)H^+(5*x*C) BPEs at 0 K (in kJ/mol) versus measured relative N3 PAs of 5*x*C and C, where *x* = H, Me, F, Br, and I. (d) TCID measured (C)H^+(5*x*C) BPEs at 0 K (in kJ/mol) versus calculated polarizability volumes of 5*x*C, where *x* = Me, F, Br, and I. The lines in parts b–d are linear regression fits to the data.

levels of theory are 3.9 ± 2.6 and 4.6 ± 4.0 kJ/mol, respectively. These MADs are very similar to the average experimental uncertainty (AEU) in these values, 4.5 ± 0.6 kJ/mol, suggesting that the B3LYP level of theory accurately describes the hydrogen-bonding interactions in these proton-bound (C)H^+(5*x*C) heterodimers, with the def2-TZVPPD results being slightly more accurate. The MP2(full) level of theory produces parallel results, but the absolute BPEs computed are systematically low. The MADs between the MP2(full)/def2-TZVPPD and MP2(full)/6-311+G(2d,2p) results and the measured values are 36.8 ± 6.7 and 35.3 ± 6.6 kJ/mol, respectively, significantly greater than the MADs for the B3LYP values and the AEU. The agreement between the MP2(full) calculated and TCID measured values improves to 23.1 ± 5.9 and 16.4 ± 6.3 kJ/mol when BSSE corrections are not included. Parallel behavior was observed in previous GIBMS studies of the analogous proton-bound homodimers of C and 5*x*C.[20] This behavior is also consistent with previous theoretical studies of hydrogen-bonded complexes,[51−58] which have shown that at least triple-ζ-quality basis sets are required to accurately describe systems where there can be significant intramolecular noncovalent interactions, and the BSSE corrections can get rather large for MP2 calculations when flexible but still unsaturated basis sets are used.

The agreement between the measured and B3LYP/def2-TZVPPD calculated relative N3 PAs is shown in Figure 6a. As can be seen in the figure, the B3LYP/def2-TZVPPD level of theory provides accurate estimates for the relative N3 PAs of C and 5*x*C. The MAD between theory and experiment is 2.2 ± 2.3 kJ/mol, just slightly larger than the AEU in these values, 1.9 ± 0.3 kJ/mol. On the basis of the comparisons between theory and experiment, it is clear that B3LYP theory accurately describes the energetics of the base-pairing interactions in these proton-bound heterodimers, whereas the MP2(full) values are systematically low for all four heterodimers.

Influence of Modifications on the N3 PA and BPEs. As can be seen in Figure 6a, 5-methylation leads to an increase in the N3 PA of C, whereas 5-halogenation leads to a decrease in the N3 PA. This is the expected behavior and is easily understood based on the electronic properties of these modifications. The methyl substituent is an electron-donating moiety and therefore increases the electron density within the aromatic ring, leading to stabilization of the positive charge associated with the excess proton. The halogens are electron withdrawing and therefore decrease the electron density within the aromatic rings, resulting in destabilization of the positive charge associated with the excess proton. The correlation between the polarizabilities of the nucleobases and the absolute difference in N3 PAs of 5*x*C and C is shown in Figure 6b. A linear regression fit through the data is also shown. It is clear that the absolute difference in the N3 PAs of C and 5*x*C decreases as the polarizability of 5*x*C increases.

The measured and calculated BPEs of the four proton-bound (C)H^+(5*x*C) heterodimers at 0 K are listed in Table 1. In all

dx.doi.org/10.1021/ja409515v | J. Am. Chem. Soc. 2014, 136, 282−290

cases, 5-substitution results in a decrease in the BPE as compared to that of the $(C)H^+(C)$ homodimer,[20] indicating that all modifications at the C5 position of a single nucleobase of the proton-bound dimer weaken the base-pairing interactions and should result in destabilization of the DNA *i*-motif. The correlation between the BPEs and relative N3 PAs of C and 5*x*C is illustrated in Figure 6c. As can be seen in the figure, the BPEs of the proton-bound heterodimers decrease as the absolute difference in the relative N3 PAs of C and 5*x*C increase as a result of the unequal sharing of the excess proton in these proton-bound heterodimers. Thus, the influence of modifications on the BPEs should directly correlate with its influence on the N3 PA. The correlation between the polarizabilities of 5*x*C and the TCID measured BPEs of the proton-bound $(C)H^+(5xC)$ heterodimers is shown in Figure 6d. A linear regression fit through all data is also shown. Clearly, the BPEs of the proton-bound heterodimers increase as the polarizabilities of 5*x*C increase.

Implications for the Stability of DNA *i*-Motif Conformations. The base-pairing interactions in the proton-bound dimer of cytosine are the major forces responsible for stabilization of DNA *i*-motif conformations. Previous TCID studies of proton-bound homodimers of cytosine and modified cytosines found that methylation of cytosine on both strands increases the BPEs of the proton-bound dimer and therefore tends to stabilize DNA *i*-motif conformations.[20] This also indicates that hypermethylation of CCG repeats, which is the cause of fragile X syndrome, occurs to further stabilize *i*-motif conformations. In contrast, present results indicate that methylation of a single cytosine destabilizes the proton-bound dimer and thus would tend to destabilize DNA *i*-motif conformations. In both proton-bound homo- and heterodimers, halogenation of cytosine weakens the base-pairing interactions in the proton-bound dimer. However, the BPEs of all four $(C)H^+(5xC)$ heterodimers are still much greater than those of canonical Watson−Crick G•C and neutral C•C base pairs, suggesting that DNA *i*-motif conformations are still favored over conventional base pairing. Thus, although modifications at the C5 position tend to weaken the base-pairing interactions in the proton-bound dimers of cytosine, the decreases in the BPEs are small such that *i*-motif conformations should be stable to modifications. Only in the case of hypermethylation are the base-pairing interactions enhanced, and this leads to the diseased state associated with the fragile X syndrome.

To further probe the influence of modifications on the stability of DNA *i*-motif conformations, other factors that play roles in stabilizing/destabilizing these noncanonical structures such as nucleobase-stacking interactions and steric effects associated with nucleobase orientation and the folding of the nucleic strands must also be considered. Follow-up work to examine how these base-pairing interactions evolve in increasingly larger model systems including proton-bound dimers of the analogous 2′-deoxycytidine nucleosides and nucleotides and extending to $(CCG)_n$ trinucleotide repeats that are associated with the formation of *i*-motif conformations and fragile X syndrome are being pursued. Present results have shown that the B3LYP level of theory provides accurate estimates for the energetics of binding in such proton-bound dimers and therefore may be suitable for investigating larger and more biologically relevant model systems. Information provided by this work including structures, the energy-dependent dissociation behavior, and relative stabilities of these proton-bound dimers should also facilitate experiments and data interpretation for studies of these improved model systems.

■ CONCLUSIONS

The stability of DNA *i*-motif conformations is of interest, and methylation and halogenation are important modifiers of that stability. We have examined the effects of these modifications on the BPEs of four proton-bound heterodimers, $(C)H^+(5xC)$, where *x* = Me, F, Br, and I, in a guided ion beam tandem mass spectrometer using TCID techniques. The primary CID pathway observed for all four heterodimers corresponds to the cleavage of the three hydrogen bonds responsible for the binding of these species, resulting in loss of the intact neutral nucleobase with the lower N3 PA. A second CID pathway occurring at higher threshold energy corresponds to loss of the intact neutral nucleobase with the higher N3 PA. Thresholds for these competitive CID reactions are determined after careful consideration of the effects of the kinetic and internal energy distributions of the $(C)H^+(5xC)$ and Xe reactants, multiple collisions with Xe, and the lifetime of the activated $(C)H^+(5xC)$ heterodimers using a loose PSL TS model. The ground-state structures and theoretical estimates for the BPEs of the $(C)H^+(5xC)$ heterodimers and N3 PAs of C and 5*x*C are determined from theoretical calculations performed at the B3LYP and MP2(full) levels of theory using the 6-311+G(2d,2p) and def2-TZVPPD basis sets. Very good agreement between experimental and theoretical values is found for the B3LYP levels of theory, especially for the B3LYP/def2-TZVPPD level of theory, whereas MP2(full) theory produces values that are systematically low, suggesting that the B3LYP functional can provide reliable energetic predictions for related systems that bind via hydrogen bonding such that theoretical studies of larger improved models for the DNA *i*-motif can be profitably studied using these techniques.

Modifications at the C5 position of cytosine clearly affect the N3 PA. The N3 PAs follow the order 5MeC > C > 5IC > 5BrC > 5FC, indicating that the electron-donating methyl substituent stabilizes the positive charge resulting from the excess proton and increases the N3 PA, whereas electron-withdrawing halogens destabilize the positive charge associated with the excess proton and lower the N3 PA. The influence of modifications on the strength of the base-pairing interactions correlates well with the proton sharing properties of these proton-bound heterodimers. The BPEs of all four $(C)H^+(5xC)$ heterodimers are lower than the BPE of the $(C)H^+(C)$ homodimer, indicating that any modifications at the 5-position generally weaken the base-pairing interactions in the proton-bound heterodimers and would therefore tend to destabilize DNA *i*-motif conformations. However, the effects are sufficiently small suggesting that *i*-motif conformations should be stable to modifications. The linear correlation between the BPEs and the relative N3 PAs of C and 5*x*C suggests that the effects of other modifications on the BPE can be estimated based on their effects on the N3 PA.

■ ASSOCIATED CONTENT

Supporting Information

Description of data handling, thermochemical analysis, and thermal correction procedures; complete citation for ref 29; tables of vibrational frequencies, average vibrational energies at 298 K, and rotational constants of the neutral and protonated nucleobases as well as the proton-bound heterodimers; tables of relative energies determined at the B3LYP/def-TZVPPD level

of theory along the PESs for adiabatic and diabatic dissociation of three (C)H$^+$(5xC) proton-bound heterodimers, where x = Me, Br, and I; tables of fitting parameters of eq 3 and enthalpies and free energies of the base-pairing interactions at 298 K; figures showing cross sections for CID, thermochemical analyses of zero-pressure extrapolated CID cross sections, and PESs determined at the B3LYP/def-TZVPPD level of theory for adiabatic and diabatic dissociation of the ground-state proton-bound heterodimers (C)H$^+$(5xC) complexes, where x = Me, Br, and I; figures showing the three most stable tautomeric conformations of C and H$^+$(C) and their relative energies calculated at the B3LYP/6-311+G(2d,2p)//B3LYP/6-31G* level of theory with ZPE corrections included; figures showing comparisons of TCID measured and theoretical BPEs calculated at the B3LYP and MP2(full) levels of theory using def2-TZVPPD and 6-311+G(2d,2p) basis sets. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

mrodgers@chem.wayne.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

Financial support for this work was provided by the National Science Foundation, grant CHE-0911191. We would also like to thank Wayne State University C&IT for computer time. B.Y. also gratefully acknowledges support from a Thomas C. Rumble Graduate Fellowship at Wayne State University.

REFERENCES

(1) Jasinska, A.; Krzyzosiak, W. J. *FEBS Lett.* **2004**, *567*, 136−141.
(2) Wells, R. D. *J. Biol. Chem.* **1996**, *271*, 2875−2878.
(3) Mitas, M. *Nucleic Acids Res.* **1997**, *25*, 2245−2253.
(4) Pieretti, M.; Zhang, F.; Fu, Y.-H.; Warren, S. T.; Oostra, B. A.; Caskey, C. T.; Nelson, D. L. *Cell* **1991**, *66*, 817−822.
(5) Darlow, J. M.; Leach, D. R. F. *J. Mol. Biol.* **1998**, *275*, 3−16.
(6) Gehring, K.; Leroy, J.-L.; Guéron, M. A. *Nature* **1993**, *363*, 561−565.
(7) Fotjík, P.; Vorlícková, M. *Nucleic Acids Res.* **2001**, *29*, 4684−4690.
(8) Rosu, F.; Gabelica, V.; Joly, L.; Grégoire, G.; Pauw, E. D. *Phys. Chem. Chem. Phys.* **2010**, *12*, 13448−13454.
(9) Oomens, J.; Moehlig, A. R.; Morton, T. H. *J. Phys. Chem. Lett.* **2010**, *1*, 2891−2897.
(10) Yang, B.; Wu, R. R.; Berden, G.; Oomens, J.; Rodgers, M. T. *J. Phys. Chem. B* **2013**, *117*, 14191−14201.
(11) Sutcliffe, J. S.; Nelson, D. L.; Zhang, F.; Pieretti, M.; Caskey, C. T.; Saxe, D.; Warren, W. T. *Hum. Mol. Genet.* **1992**, *1*, 397−400.
(12) Oberlé, I.; Rousseau, F.; Heitz, D.; Kretz, C.; Devys, D.; Hanauer, A.; Boué, J.; Bertheas, M. F.; Mandel, J. L. *Science* **1991**, *252*, 1097−1102.
(13) Valinluck, V.; Liu, P.; Kang, J. I., Jr.; Burdzy, A.; Sowers, L. C. *Nucleic Acids Res.* **2005**, *33*, 3057−3064.
(14) Vanetich, K. M.; Santi, D. V. *Adv. Exp. Med. Biol.* **1998**, *244*, 113−124.
(15) Smith, S. S.; Kaplan, B. E.; Sowers, L. C.; Newman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 4744−4748.
(16) Klimasauskas, S.; Kumar, S.; Roberts, R. J.; Cheng, X. *Cell* **1994**, *76*, 357−369.
(17) Chen, Y.; Rodgers, M. T. *J. Am. Chem. Soc.* **2012**, *134*, 2313−2324.
(18) Chen, Y.; Rodgers, M. T. *J. Am. Chem. Soc.* **2012**, *134*, 5863−5875.
(19) Chen, Y.; Rodgers, M. T. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 2020−2030.
(20) Yang, B.; Wu, R. R.; Rodgers, M. T. *Anal. Chem.* **2013**, *85*, 11000−11006.
(21) Chen, Y.; Rodgers, M. T. *Anal. Chem.* **2012**, *84*, 7570−7577.
(22) Rodgers, M. T. *J. Phys. Chem. A* **2001**, *105*, 2374−2383.
(23) Teloy, E.; Gerlich, D. *Chem. Phys.* **1974**, *4*, 417−427.
(24) Gerlich, D. Inhomogeneous RF Fields: A Versatile Tool for the Study of Processes with Slow Ions. Diplomarbeit, University of Freiburg, Federal Republic of Germany, 1971.
(25) Gerlich, D. Inhomogeneous RF fields: a versatile tool for the study of processes with slow ions. In *Advances in Chemical Physics: State-Selected and State-to-State Ion-Molecule Reaction Dynamics, Part I. Experiment*; Ng, C.-Y., Baer, M., Eds.; Wiley: Weinheim, Germany, 1992; Vol. *82*, pp 1−176.
(26) Dalleska, N. F.; Honma, K.; Armentrout, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 12125−12131.
(27) Aristov, N.; Armentrout, P. B. *J. Phys. Chem.* **1986**, *90*, 5135−5140.
(28) Hales, D. A.; Armentrout, P. B. *J. Cluster Sci.* **1990**, *1*, 127−142.
(29) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; et al. *Gaussian 09*, Revision A.1; Gaussian, Inc.: Wallingford, CT, 2009.
(30) Weigend, F.; Ahlrichs, R. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297−3305.
(31) Feller, D. *J. Comput. Chem.* **1996**, *17*, 1571−1586.
(32) Schuchardt, K. L.; Didier, B. T.; Elsethagen, T.; Sun, L.; Gurumoorthi, V.; Chase, J.; Li, L.; Windus, T. L. *J. Chem. Inf. Model.* **2007**, *47*, 1045−1052.
(33) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299−310.
(34) Peterson, K. A.; Figgen, D.; Goll, E.; Stoll, H.; Dolg, M. *J. Chem. Phys.* **2003**, *119*, 11113−11123.
(35) Smith, S. M.; Markevitch, A. N.; Romanov, D. A.; Li, X.; Levis, R. J.; Schlegel, H. B. *J. Phys. Chem. A* **2000**, *108*, 11063−11072.
(36) Peng, C.; Schlegel, H. B. *Isr. J. Chem.* **1993**, *33*, 449−454.
(37) Foresman, J. B.; Frisch, Æ. *Exploring Chemistry with Electronic Structures Methods*, 2nd ed.; Gaussian: Pittsburg, PA, 1996; p64.
(38) Boys, S. F.; Bernardi, R. *Mol. Phys.* **1979**, *19*, 553−566.
(39) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. *Chem. Rev.* **1994**, *94*, 1873−1885.
(40) Ervin, K. M.; Armentrout, P. B. *J. Chem. Phys.* **1985**, *83*, 166−189.
(41) Muntean, F.; Armentrout, P. B. *J. Chem. Phys.* **2001**, *115*, 1213−1228.
(42) Beyer, T. S.; Swinehart, D. S. *Commun. ACM* **1973**, *16*, 379.
(43) Stein, S. E.; Rabinovitch, B. S. *J. Chem. Phys.* **1973**, *58*, 2438−2445.
(44) Stein, S. E.; Rabinovitch, B. S. *Chem. Phys. Lett.* **1977**, *49*, 183−188.
(45) Khan, F. A.; Clemmer, D. E.; Schultz, R. H.; Armentrout, P. B. *J. Phys. Chem.* **1993**, *97*, 7978−7987.
(46) Rodgers, M. T.; Ervin, K. M.; Armentrout, P. B. *J. Chem. Phys.* **1997**, *106*, 4499−4508.
(47) Chesnavich, W. J.; Bowers, M. T. *J. Phys. Chem.* **1979**, *83*, 900−905.
(48) Austin, C. A.; Chen, Y.; Rodgers, M. T. *Int. J. Mass Spectrom.* **2012**, *330*, 27−34.
(49) Armentrout, P. B.; Yang, B.; Rodgers, M. T. *J. Phys. Chem. B* **2013**, *117*, 3771−3781.
(50) Nose, H.; Chen, Y.; Rodgers, M. T. *J. Phys. Chem. A* **2013**, *117*, 4316−4330.
(51) Sim, F.; St-Amant, A.; Pápai, I.; Salahub, D. R. *J. Am. Chem. Soc.* **1992**, *114*, 4391−4400.
(52) Laasonen, K.; Parrinello, M.; Car, R.; Lee, C.; Vanderbilt, D. *Chem. Phys. Lett.* **1993**, *207*, 208−213.
(53) Kim, K.; Jordan, K. D. *J. Phys. Chem.* **1994**, *98*, 10089−10094.
(54) Del Bene, J. E.; Person, W. B.; K. Szczepaniak, K. *J. Phys. Chem.* **1995**, *99*, 10705−10707.

(55) Novoa, J. J.; Sosa, C. *J. Phys. Chem.* **1995**, *99*, 15837−15845.
(56) Hobza, P.; Šponer, J.; Reschel, T. *J. Comput. Chem.* **1995**, *16*, 1315−1325.
(57) Suhai, S. *J. Phys. Chem.* **1995**, *99*, 1172−1181.
(58) Valdés, H.; Klusák, V.; Pitoňák, M.; Exner, O.; Starý, I.; Hobza, P.; Rulisek, L. *J. Comput. Chem.* **2008**, *29*, 861−870.