Influence of 5-Methylation and the 2′- and 3′-Hydroxy Substituents on the Base Pairing Energies of Protonated Cytidine Nucleoside Analogue Base Pairs: Implications for the Stabilities of *i*-Motif Structures

Yakubu S. Seidu, H.A. Roy, and M. T. Rodgers*

Department of Chemistry, Wayne State University, Detroit, MI, 48202, USA

ABSTRACT:

Repetitive nucleic acid sequences, which occur in abundance throughout the mammalian genome, are of enormous research interest due to their potential to adopt fascinating and unusual molecular structures such as the *i*-motif. In remarkable contrast to the DNA double helix, *i*-motif conformations are stabilized by protonated cytosine base pairs, (Cyt)H^+(Cyt), that are centrally located in the core of the *i*-motif and intercalated vertically in an antiparallel fashion. An in-depth understanding of how modifications influence the stability of *i*-motif conformations is a prerequisite to understanding their biological functions and the development of effective means of tuning their stability for specific medical and technological applications. Here, the influence of the 2′- and 3′-hydroxy substituents of the sugar moieties and 5-methylation of the cytosine nucleobases on the base-pairing interactions of protonated cytidine nucleoside analogue base pairs, (*x*Cyd)H^+(*x*Cyd), are examined by complimentary threshold collision-induced dissociation techniques and computational methods. The *x*Cyd nucleosides examined include the canonical DNA and RNA cytidine nucleosides, 2'-deoxycytidine (dCyd) and cytidine (Cyd), as well as several modified cytidine nucleoside analogues, 2',3'-dideoxycytidine (ddCyd), 5-methyl-2'-deoxycytidine (m^5dCyd), and 5-methylcytidine (m^5Cyd). Comparisons among these model base pairs indicate that the 2'- and 3'-hydroxy substituents of the sugar moieties have very little influence on the strength of the base-pairing interactions, whereas 5-methylation of the cytosine nucleobases is found to enhance the strength of the base-pairing interactions. The increase in stability resulting from 5-methylation is only modest, but is more than twice as large for the DNA than RNA protonated cytidine base pair. Overall, present results suggest that canonical DNA *i*-motif conformations should be more stable than analogous RNA *i*-motif conformations, and that 5-methylation of cytosine residues, a significant epigenetic marker, provides greater stabilization to DNA than RNA *i*-motif conformations.

INTRODUCTION

Several decades of research on the broader functionality of the human genome have established that the structure of the genetic material is mostly polymorphic, and among other components, comprises repetitive nucleic acid sequences that occur in abundance throughout the genome.[1,2] Under physiological conditions the B-form double helical structure of DNA characterized by Watson-Crick base pairing is the most thermodynamically stable and commonly known.[1,3] The Watson-Crick duplex can fold into a variety of alternative noncanonical DNA polymorphs whose biochemistry is dominated by noncovalent interactions.[1,2] The *i*-motif is one special class of noncanonical DNA secondary structures that has garnered extensive research interest among laboratories around the world. Nucleic acid sequences that are cytosine (Cyt) rich have the potential to fold into an *i*-motif structure. The *i*-motif conformation is a noncovalently bound four-stranded structure held together by intercalated protonated (Cyt)H^+(Cyt) base pairs that are oriented in an antiparallel fashion with each base pair located in the core of the *i*-motif uniquely positioned face-to-face with its neighbor.[4] In sharp contrast to double-helical DNA structures, *i*-motifs may fold in an intermolecular fashion from the association of two, three or four separate strands, or form an intramolecular structure due to the spatial arrangement of four different Cyt tracts within a single strand. A necessary requirement for the formation of the *i*-motif conformation is N3-protonation of Cyt residues in the oligonucleotide strands, which has been observed under acidic conditions.[5] Recent studies have demonstrated that the DNA *i*-motif conformations can be form at physiological pH in the nuclei of human cells[6] as well as in the nuclei and chromosomes of invertebrates,[7] corroborating the existence of *i*-motif structure in vivo. The prevalence of oligonucleotides with repetitive Cyt-rich sequences in genomic regions such as telomeres,[8] centromeres[9,10] and promoter areas of oncogenes[11-13] have facilitated both in vitro and in vivo laboratory characterization of *i*-motif structures in these regions of the human genome,[6] and stimulated enormous interest in elucidating the functional roles of the *i*-motif in biological systems[14] and its potential for use in nanotechnological applications.[13,15]

Among the myriad of key questions concerning the biochemistry of nucleic acid *i*-motif conformations are the type and optimum number of protonated (Cyt)H^+(Cyt) base pairs required to form a thermodynamically stable *i*-motif architecture and how this number is influenced by posttranscriptional modifications. These pertinent questions can be answered by first focusing on gaining an in-depth knowledge of the factors that govern the structures, stabilities, and dynamics of the noncovalent interactions that stabilize the *i*-motif. At the cellular level, naturally occurring modifications of the nucleobase residues, sugar moieties and phosphate backbone are shown to confer enhanced biostability to nucleic acid structures.[16-19] Furthermore, chemical modifications have also been engineered as a way of increasing the pharmacokinetic and pharmacodynamic properties of small molecules developed as therapeutics for treating genetic diseases[20-22] In living cells, Cyt repeats localized around transcriptional start sites are reported to be susceptible to epigenetic modifications, which are characterized by site-specific methylation of DNA nucleobases.[18,23-25] CpG islands, which are characterized by a pattern of uniquely clustered cytosine and guanine sequences separated by a single phosphate group and commonly found in the promoter regions of the genome are prone to epigenetic modifications.[18] The most studied epigenetic modification is 5-methylation.[24] 5-methylation of Cyt residues at CpG dinucleotide sites has been found to stabilize the noncanonical *i*-motif conformation, therefore increasing the probability of altering gene expression through transcriptional silencing of the corresponding gene.[23] Recent solution-phase studies have also shown that incorporation of four consecutive 5-methylcytosine (m^5Cyt) repeats in a Cyt-rich human telomeric sequence stabilizes the DNA *i*-motif conformations to a greater extent than the corresponding unmodified strand.[19] Given the important roles that 5-methylation plays in stabilizing the *i*-motif and signaling, a comprehensive investigation to further understand the influence of this naturally-occurring modification on the base-pairing interactions of the *i*-motif is needed.

The impact of 5-methylation of Cyt residues on the structure and function of nucleic acids has been widely studied.[18,24] However, information on how such modifications influence hydrogen-bonding interactions among noncanonical DNA polymorphs is still rather limited.[27-30]

Given the purported biological role that genomic *i*-motif conformations play in several disease states,[30] a comprehensive study is therefore needed to determine how 5-methylation of Cyt residues impacts the structure and stabilities of DNA as well as RNA *i*-motif architectures. Our initial efforts toward achieving this broad objective began by reducing the complex structure of an *i*-motif conformation to the simplest model for the core stabilizing interactions, the protonated cytosine base pair, (Cyt)H^+(Cyt). Using complementary threshold collision-induced dissociation (TCID) techniques and computational methods, the base-pairing energy (BPE) of the (Cyt)H^+(Cyt) base pair was measured as 169.9 ± 4.6 kJ/mol in excellent agreement with the B3LYP/6-311+G(2d,2p)//B3LYP/6-31G* predicted value of 168.9 kJ/mol.[31] Significantly, the BPE of the (Cyt)H^+(Cyt) base pair greatly exceeds those of the canonical Watson-Crick guanine-cytosine (Gua)·(Cyt) and noncanonical neutral cytosine-cytosine (Cyt)·(Cyt) base pairs, with B3LYP predicted BPEs of 96.6 kJ/mol and 68.0 kJ/mol, respectively. The much stronger base pairing interactions in the (Cyt)H^+(Cyt) base pair were thus interpreted as important contributors to the stability of genomic *i*-motif conformations. This initial study as well as a series of parallel follow-on investigations introduced increasing complexity to the (Cyt)H^+(Cyt) base pair model for the *i*-motif by examining the effects of a series of modifications at the 1- and 5-positions of the nucleobase.[31-34] Consistent with previous findings that 5-methylation stabilizes noncanonical *i*-motif conformations,[26] the BPE of the (m^5Cyt)H^+(m^5Cyt) base pair was measured as 177.4 ± 5.3 kJ/mol, an increase of 7.5 ± 7.0 kJ/mol in the strength of base pairing. A somewhat smaller increase in the BPE to 173.4 kJ/mol was predicted by theory. The effect of 1-methylation (examined as a simple model for the sugar moiety) was found to be much less significant. The BPEs of the protonated 1-methylcytosine and 1,5-dimethylcytosine base pair analogues, (m^1Cyt)H^+(m^1Cyt) and ($m_2^{1,5}$Cyt)H^+($m_2^{1,5}$Cyt), were measured as 170.7 ± 5.3 kJ/mol and 172.3 ± 5.8 kJ/mol, respectively. Indicating that 1-methylation very slightly enhances the BPEs (by 1.0 ± 7.0 and 2.4 ± 7.4 kJ/mol), but also suggesting that the enhancement in the BPE upon 5-methylation is reduced in the 1-methyl analogues. Theoretical results were not entirely consistent, but overall suggest a modest enhancement in the BPE upon 5-methylation, and a

marginal weakening in the BPE upon 1-methylation. The model systems investigated were extended to the nucleoside level by examining the protonated nucleoside base pairs of 2′-deoxycytidine (dCyd) and 5-methyl-2′-deoxycytidine (m^5dCyd), thus in principle, enabling the effects of both 5-methylation and the sugar moiety on the BPE to be elucidated. The BPEs of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs were measured (and computed) as 159.8 ± 5.2 (163.7) kJ/mol and 162.0 ± 5.7 (166.4) kJ/mol, respectively.[35] These results suggest that 5-methylation slightly enhances the BPE by 2.2 ± 7.7 (2.7) kJ/mol, and that the sugar moiety weakens the BPEs by 10.1 ± 6.9 (5.2) kJ/mol and 15.4 ± 7.8 (6.9) kJ/mol, respectively. As found for 1-methylation, theory predicts a lesser effect of the sugar moiety. As this work was extended to a larger variety of modified cytidine nucleoside analogues and especially results based on the competitive dissociation of mixed base pairs, it became obvious that the BPEs originally reported for the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs were systematically low. This conclusion was further supported when additional calculations found more stable conformers of these base pairs than originally reported and thus larger predicted BPEs. Consequently, the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs are re-examined in the present study.

Given the known potential of 5-methylation of Cyt residues to stabilize *i*-motif conformations, the aim of this study is to use complementary tandem mass spectrometry and computational approaches to systematically characterize symmetric protonated nucleoside base pairs of canonical and modified cytidine (Cyd) nucleoside analogues and use the synergism between TCID measurements and theoretical calculations to provide intrinsic and validated thermochemical data that further enhances our understanding of the role of modifications on the strength of noncanonical binding in genomic *i*-motif model systems. In order to comprehensively explore the effects of naturally-occurring and synthetic modifications on the thermodynamics of base-pairing in model systems for the *i*-motif, here we examine five protonated nucleoside base pairs, (*x*Cyd)H^+(*x*Cyd), where *x*Cyd is Cyd, dCyd, 2',3'-dideoxycytidine (ddCyd), 5-methylcytidine (m^5Cyd), and m^5dCyd. The BPEs of the (*x*Cyd)H^+(*x*Cyd) base pairs are measured by TCID techniques in a custom-built guided ion beam mass spectrometer such that

the influence of 5-methylation and the 2'- and 3'-hydroxy substituents on the structures and BPEs is directly determined. Insights gained from this work are expected to be relevant to and amplified in nucleic acids where the effects of modifications (e.g., 5-hypermethylation) are additive.

EXPERIMENTAL AND COMPUTATIONAL METHODS

Instrumentation and Experimental Procedures. The energy-dependent CID cross sections of five protonated nucleoside base pairs, (*x*Cyd)H^+(*x*Cyd), of the canonical and several noncanonical cytidine nucleoside analogues are measured using a custom-built guided ion beam tandem mass spectrometer, which has been described in detail previously.[36,37] The protonated (*x*Cyd)H^+(*x*Cyd) base pairs are generated by electrospray ionization (ESI) from a 0.5–1.0 mM solution of the nucleoside and 1% acetic acid in an approximately 50%:50% MeOH:H_2O mixture. The solution is pumped at a flow rate of 0.1–0.3 μL/min through a 35 gauge 304 stainless steel ESI needle operating at a voltage of 1.5–2.0 kV. The droplets emanating from the spray enter the vacuum region of the mass spectrometer through an inlet capillary, biased at 20 to 35 V, and heated to ~100 ºC to promote desolvation of large droplets. The ions emerging from the heated capillary are sampled and focused into a radiofrequency (rf) ion funnel (IF), which is superimposed with a linear dc gradient. Ions are injected from the rf IF into an rf only hexapole ion guide that traps the ions in the radial direction with an rf amplitude of 250 V_{pp}. The ions undergo $>10^4$ collisions with the ambient gases as they drift through the hexapole ion guide. Under these conditions coupled with proper tuning of the source region, the ions emanating from the hexapole ion guide are thermalized to ambient temperature.[38]

The ions generated and thermalized in the source region are effusively collected, and gently extracted from the hexapole ion guide, accelerated, and focused by a series of electrostatic lenses into a magnetic sector momentum analyzer where the (*x*Cyd)H^+(*x*Cyd) base pair is selected in the initial stage of mass analysis. The (*x*Cyd)H^+(*x*Cyd) base pairs are subsequently decelerated to a known kinetic energy via an exponential retarder, focused and injected by a

second series of electrostatic lenses into an rf only octopole ion guide where the ions are radially trapped.[39-41] The octopole ion guide passes through a static collision gas cell in which Xe is present at low pressure, ~0.05, 0.10, and 0.20 mTorr. At these pressures, nominally single-collision conditions prevail as the (*x*Cyd)H^+(*x*Cyd) base pairs undergo CID with Xe as they drift through the collision cell.[42,43] Xe is chosen as the neutral reactant to induce dissociation because it is heavy, monoatomic, highly polarizable, and chemically unreactive, ensuring efficient kinetic-to-internal energy transfer in the (*x*Cyd)H^+(*x*Cyd) base pair upon collision with Xe.[44,45] The application of a dc voltage to all eight rods of the rf octopole controls the axial translational energy of the ion beam. After collision, the CID products and any undissociated (*x*Cyd)H^+(*x*Cyd) base pairs drift to the end of the octopole where they are extracted, focused into a quadrupole mass filter for the final stage of mass analysis, and detected with a Daly detector using standard pulse-counting techniques.

Theoretical Calculations. Our experimental studies are augmented with chemical computations from which we obtain structural information and molecular parameters such as harmonic vibrational frequencies, rotational constants, and isotopic molecular polarizabilities used in thermochemical modeling of our experimental CID cross sections. This synergism between experiment and theory enables more accurate and precise threshold determinations and simultaneous validation of both experimental and theoretical results.

Simulated Annealing Protocol. In the first stage of our theoretical workflow, a molecular mechanics simulated annealing process using the Amber 3 force field as implemented in HyperChem software[46] was used to generate potential low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs and the neutral *x*Cyd and protonated H^+(*x*Cyd) nucleosides. A three-stage simulated annealing process was employed, with each cycle beginning and ending at 0 K, lasting 0.8 ps, and achieving a simulation temperature of 1000 K. At least 300 cycles of heating and cooling were performed for each initial structure so as to provide sufficient sampling of conformational space. All unique conformers identified by the simulated annealing procedures

and within 30 kJ/mol of the most stable conformer found were subjected to further study using high-level quantum mechanical approaches.

Geometry Optimization, Frequency Analysis and Single-Point Energy Calculations. In the second stage of our theoretical workflow, the unique conformers identified by the simulated annealing and energy filtering outlined above were subjected to high-level quantum mechanical calculations using the Gaussian 16 suite of programs.[47] Geometry optimizations and frequency analyses were performed at the B3LYP/6-311+G(d,p) level of theory for the (*x*Cyd)H^+(*x*Cyd) base pairs and the neutral *x*Cyd and protonated H^+(*x*Cyd) nucleosides. Single point energy calculations using the B3LYP/6-311+G(d,p) optimized geometries were performed at the B3LYP/6-311+G(2d,2p), B3LYP/def2-TVZPPD, B3P86/6-311+G(2d,2p), and M06-2X/6-311+G(2d,2p) levels of theory. To obtain reliable energetics and accurate thermodynamic information, zero-point energy corrections were applied to the single point energies to yield 0 K thermodynamic values. Basis set superposition error corrections were also included in the computed BPEs using the counterpoise method.[48,49]

A series of dihedral angles extracted from the B3LYP/6-311+G(d,p) optimized geometries are used to determine three important structural parameters as previously described.[50] The pseudorotation phase angle (*P*, which describes the sugar puckering) is characterized by the five dihedral angles of the sugar ring: ∠C4′O4′C1′C2′, ∠O4′C1′C2′C3′, ∠C1′C2′C3′C4′, ∠C2′C3′C4′O4′, and ∠C3′C4′O4′C1′). The classical sugar puckering described in terms of envelope designations, Cn'-endo and Cn'-exo, where n denotes the number of the atom that deviates most from the plane of the sugar ring and endo and exo indicate the direction of deviation, toward or away from the 5′-hydroxy substituent, respectively. The sugar puckering is alternatively defined as envelope (E) and twist (T) designations, the latter describes structures in which two atoms within the sugar ring pucker in opposite directions from the plane defined by the remaining three atoms. Superscripts are used to designate endo configurations, whereas subscripts denote exo puckers. Values that precede the E and T designations indicate the atom numbers of major deviator, whereas values after the T designations indicate the atom number of

the minor deviator, respectively. The glycosidic bond angle which is characterized by ∠O4′C1′N1C2, determines the orientation (either *anti* for values between 90º and 270º or *syn* for values in the range from −90º to 90º) of the nucleobase residues. The orientation of the 5-hydroxy substituent is characterized by the ∠O4′C4′C5′O5′ dihedral angle. The 5-hydroxy orientation is defined as gauche$^+$ when the O5′ atom lies between O4′ and C3′ atoms (corresponding to values between 120º and 240º), gauche$^-$ when the O5′ atom points away from the O4′ atom (values between 240º and 360º), or trans when the O5′ atom points away from C3 atom (values between 0º and 120º).

Isotropic Molecular Polarizabilities. Among the parameters needed for thermochemical analysis of TCID data for noncovalently bound complexes such as the (*x*Cyd)H$^+$(*x*Cyd) base pairs of interest here are the isotropic molecular polarizabilities (α) of the neutral CID products, which are the neutral *x*Cyd nucleosides. The isotropic molecular polarizabilities of the B3LYP/6-311+G(d,p) ground conformers of the neutral *x*Cyd nucleosides were computed at the PBE0/6-311+G(2d,2p) level of theory as polarizabilities of polyatomic molecules and ions calculated using the PBE0 functional have been found to be in good agreement with experimentally measured polarizabilities.[51]

Electrostatic Potential Maps. The electrostatic potential (ESP) of the ground conformers of the (*x*Cyd)H$^+$(*x*Cyd) base pairs were computed at the B3LYP/6-311+G(d,p) level of theory and color-mapped onto an isosurface of 0.06 a.u. of the total SCF electron density using GaussView6.[47] The ESP maps visualize the distribution of positive and negative charges on the hydrogen-bond donor and acceptor sites of the Cyt residues involved in the base-pairing interactions as well as those associated with sugar-sugar and sugar-base stabilizing interactions. Importantly, comparisons among the ESP maps for these systems elucidate the effects of the 5-methyl and 2′- and 3′-hydroxy substituents on the stability of the (*x*Cyd)H$^+$(*x*Cyd) base pairs.

Noncovalent Interaction Maps. Noncovalent interactions (NCI) within the ground conformers of the (*x*Cyd)H$^+$(*x*Cyd) base pairs were mapped using Multi-WFN 3.7.[52] Visualization of the 3D and 2D NCI plots generated was performed using the Visual Molecular

Dynamics (VMD) program[53] and IrfanView-image viewer, respectively. The NCI formalism is based on spikes in the reduced density gradient (RDG) function in the vicinity of low electron density.[54] Generally, the values of the RDG increase significantly (positively) as the electron density exponentially decays to zero. However, in regions dominated by both covalent and noncovalent interactions, the values of the RDG become smaller, and eventually approach zero. The electron density alone provides information about the strength of interactions found in the (*x*Cyd)H^+(*x*Cyd) base pairs. Therefore, to differentiate between the different types of interactions (attractive and repulsive) and simultaneously evaluate the strength of these interactions, the electron density is customarily multiplied by the sign of the second eigenvalue of the density Hessian.[55] In this manner, the 2D NCI plots enable identification and characterization of the strength of the noncovalent interactions in these (*x*Cyd)H^+(*x*Cyd) base pairs as chemically intuitive RDG isosurfaces that recover both attractive (hydrogen bonds, van der Waals) and repulsive (steric clashes) interactions. A conventional scalar coloring scheme is chosen to rank the NCIs, where red is used for repulsive (destabilizing) interactions and blue for attractive (stabilizing) interactions. Green and yellow isosurfaces represent delocalized weak attractive and repulsive noncovalent interactions, respectively.[54,56] Very intense colors are associated with higher local electron density and therefore correspond to stronger interactions.

Quantum Theory of Atoms in Molecules (QTAIM). The nature of the bonding interactions in the (*x*Cyd)H^+(*x*Cyd) base pairs was further investigated using the QTAIM topological analysis performed with the Multi-WFN 3.7 program[52] Within the QTAIM theory,[57] topological parameters are evaluated by partitioning the total electron density into an atomic ensemble and using the density gradient as a criterion for the existence of bonding paths. The Laplacian of the density at the bond critical points is used as a descriptor of local charge concentration, which gives information for the two interacting atoms in real space. The caveat in characterizing noncovalent interactions with the QTAIM approach is that their representation of bond paths is not very intuitive. This limitation is tackled by combining the QTAIM approach with NCI analysis, which together provide enhanced visualization and more informative results.

Data Handling Procedures. The Beer–Lambert law, written in terms of parameters appropriate for our TCID experiments is used to convert measured precursor and product ion intensities to absolute cross sections as described previously.[58] Under most operating conditions, where appreciable momentum is transferred to the product ions, the sum of the uncertainties in the absolute cross sections measured are ~ 20%, which represents roughly equal contributions from both random errors in pressure measurement and the collision cell length. Conversion from the laboratory frame ion kinetic energies, E_{lab}, to the relative collision energy in the center-of-mass frame, E_{cm}, is accomplished using the stationary target assumption given by:

$$E_{cm} = E_{lab}\, m\, /(m + M) \quad (1)$$

where m is the mass of the neutral collision gas, Xe, and M is the mass of the (xCyd)H^+(xCyd) base pair. The octopole ion guide allows the use of retarding potential analysis to measure the precursor ion beam absolute zero of energy and the kinetic energy distribution as previously described.[58] The first derivative of the normalized ion beam intensity is fitted to a Gaussian distribution to extract its center and full width at half-maximum (fwhm). The nearly Gaussian fit gives the origin for the laboratory frame energy, $E_{\mathrm{lab}} = 0$ eV, and the width of the kinetic energy distribution of the ion beam, which typically lies in the range from 0.2 to 0.4 eV (lab), respectively.

The shapes of measured CID cross sections are sensitive to pressure effects caused by multiple ion-neutral collisions. This pressure dependence is straightforwardly tackled by routinely performing TCID experiments under nominally single-collision conditions and at multiple Xe pressures of ~ 0.05, 0.1 and 0.2 mTorr. These pressure regimes are used to ensure that the cross sections subjected to thermochemical analysis correspond to rigorously single-collision conditions by linearly extrapolating to zero pressure of the Xe reactant.[59,60]

TCID Data Analysis. The procurement of available quantitative data and accurate thermochemical information from modeling of TCID data requires implementation of certain important procedures capable of mitigating both experimental and analytical factors that may convolute the raw experimental data acquired. The required conditions needed to obtain reliable

CID thresholds include properly accounting for the kinetic and internal energy distributions of the (*x*Cyd)H$^+$(*x*Cyd) and Xe reactants,[60,61] the effects of multiple ion-neutral collisions,[59] and lifetime effects of the activated (*x*Cyd)H$^+$(*x*Cyd) base pair,[62,63] which may exceed the time-of-flight of the experiment.

The zero-pressure-extrapolated cross sections for dissociation of the base-pairing interaction within the (*x*Cyd)H$^+$(*x*Cyd) base pairs via loss of an intact nucleoside are subjected to thermochemical analysis using an empirical threshold function of the form:

$$\sigma(E) = \sigma_0 \sum_i g_i (E + E_i - E_0)^n / E \qquad (2)$$

where σ_0 is an energy independent scaling parameter, E is the relative translational energy of the reactants, $E_{0,}$ is the reaction endothermicity or threshold for dissociation corresponding to 0 K, and n is an adjustable fitting parameter that describes the efficiency of translational-to-internal energy transfer upon collision.[64] The summation is over the ro-vibrational states of the (*x*Cyd)H$^+$(*x*Cyd) base pair with excitation energies, E_i, and fractional populations, g_i, where Σg_i = 1. The fractional populations, g_i, of the internal degrees of freedom are calculated based on a Maxwell–Boltzmann distribution at 300 K. The number and densities of ro-vibrational states are directly counted with the Beyer–Swinehart–Stein–Rabinovitch algorithm and used to evaluate the internal energy distribution for the reactants.[65-67]

Large energized complexes having many degrees of freedom may possess sufficient internal energy to undergo dissociation, however there is an increasing probability that the average time needed for the dissociation to occur may exceed the time scale of the experiment, which is ~100 μs on our custom-built GIBMS instrument. This leads to a shift in the observed threshold for dissociation to higher translational energies, causing a phenomenon termed "kinetic" shift. We explicitly account for the lifetime of the energized complexes in the analysis of the CID cross sections by incorporating Rice–Ramsperger–Kassel–Marcus (RRKM) statistical theory for unimolecular dissociation into eq 2, as described by eq 3.[62,63] The rationale for such a treatment is sound because the measured apparent thresholds for large energized complexes

without RRKM formalism can be appreciably high and exhibit curvature that cannot be reproduced without incorporation of such lifetime modeling in the analysis.[63]

$$\sigma_j(E) = \frac{n\sigma_0}{E}\sum_i g_i \int_{E_0-E_i}^{E} (E-\varepsilon)^{n-1} P_D dE \quad (3)$$

In this equation ε is the energy deposited into internal degrees of freedom of the (*x*Cyd)H^+(*x*Cyd) base pair upon collision with the neutral Xe reactant, such that it has an internal energy of $E^* = \varepsilon + E_i$. The probability of dissociation, P_D, can be expressed in terms of the RRKM unimolecular dissociation rate coefficient, $k(E^*)$, as $P_D = (1 - \exp\{-k(E^*)T\})$. The RRKM unimolecular dissociation rate coefficient is calculated using eq 4.

$$k(E^*) = sN^{\dagger}(E^* - E_0) / h\rho(E^*) \quad (4)$$

where s is the reaction degeneracy, $N^{\dagger}(E^* - E_0)$ is the sum of the ro-vibrational states of the transition state (TS), *h* is Planck's constant, and $\rho(E^*)$ is the density of ro-vibrational states for the energized (*x*Cyd)H^+(*x*Cyd) base pair at the energy available E^*. Whenever the rate of decomposition is much faster than the average experimental time window, eq 3 reduces to eq 2.

The set of ro-vibrational frequencies for the energized (*x*Cyd)H^+(*x*Cyd) base pair (EM) and transition state (TS) for dissociation are incorporated in the calculation of the RRKM unimolecular rate coefficients. Theoretical calculations of the precursor ion provide the required molecular parameters for the energized (*x*Cyd)H^+(*x*Cyd) base pair. CID of the (*x*Cyd)H^+(*x*Cyd) base pair results in noncovalent dissociation of the base-pairing interaction to produce the neutral *x*Cyd and protonated H^+(*x*Cyd) nucleosides. The threshold energy determined from modeling of the measured CID cross section thus corresponds to the BPE of the (*x*Cyd)H^+(*x*Cyd) base pair. Such noncovalent bond cleavage occurs via a loose phase space limit (PSL)[63] TS, which postulates that the TS is located at the centrifugal barrier for dissociation. Consequently, there is no reverse activation barrier in excess of the reaction endothermicity, such that the molecular parameters for the TS are simply those of the H^+(*x*Cyd) and *x*Cyd products. Thus, by combining the PSL formalism with the RRKM statistical treatment of the lifetime of the dissociating ions, accurate threshold energies are determined and are denoted as E_0(PSL) to differentiate these

values from the kinetically shifted threshold energies, $E_{0,}$ extracted from analyses that exclude the RRKM lifetime analysis.

Threshold energies are determined by fitting the threshold regions of the CID cross sections to the model function of eq 3. The CID cross section model of eq 3 takes into account effects of the internal energy of the (*x*Cyd)H^+(*x*Cyd) base pair, Doppler broadening by Xe, and the relative kinetic energy distribution of the (*x*Cyd)H^+(*x*Cyd) base pair and Xe reactants by convoluting the model cross sections with these energy distributions. The convoluted cross section is then compared with the experimental cross section over a broad range of the collision energies examined, while simultaneously optimizing the fitting parameters, σ_0, E_0 or E_0(PSL), and *n*, to give the best least-squares fit for each model. The statistical uncertainties associated with the σ_0, E_0 or E_0(PSL), and n fitting parameters are conservatively estimated from variations in these parameters determined across different data sets, and by modeling the zero-pressure cross sections using vibrational frequencies for the TS and EM that have been scaled by ±10% and by scaling the assumed experimental time-window for dissociation up and down by a factor of 2. Because all sources of energy are accounted for in these analyses, the measured thresholds, E_0(PSL), correspond to the BPE at 0 K for the (*x*Cyd)H^+(*x*Cyd) base pair.

RESULTS

Cross Sections for Collision-Induced Dissociation. Kinetic energy-dependent CID cross sections for the interaction of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs with Xe over the range of collision energies from ~0 to 6 eV are compared in Figure 1. A similar comparison is shown for all five (*x*Cyd)H^+(*x*Cyd) base pairs in Figure S1 of the Supporting Information. The major CID pathway observed involves loss of an intact neutral *x*Cyd nucleoside as shown in the CID reactions 5.

$$(x\text{Cyd})H^+(x\text{Cyd}) + \text{Xe} \rightarrow H^+(x\text{Cyd}) + x\text{Cyd} + \text{Xe} \quad (5)$$

The endothermic process of reaction 5 involves heterolytic cleavage of the intermolecular hydrogen bonds responsible for the binding in these (*x*Cyd)H^+(*x*Cyd) base pairs. At elevated

collision energies, the primary H^+(*x*Cyd) product is formed with sufficient internal energy that it undergoes sequential dissociation. Two sequential dissociation pathways, both of which involve glycosidic bond cleavage occur in competition. In all cases the excess proton is preferentially retained by the departing base, reaction 6, whereas retention of the excess proton by the sugar moiety is much less favorable (by more than an order of magnitude), reaction 7.

$$H^+(x\text{Cyd}) \rightarrow H^+(x'\text{Cyt}) + (x\text{Cyd}–x'\text{Cyt}) \quad (6)$$

$$H^+(x\text{Cyd}) \rightarrow H^+(x\text{Cyd}–x'\text{Cyt}) + x'\text{Cyt} \quad (7)$$

The fragmentation patterns observed for the (*x*Cyd)H^+(*x*Cyd) base pairs parallel those observed in previous studies of protonated base pairs of cytosine nucleobase and cytidine nucleoside analogues as well as that of iosloated protonated cytidine nucleoside analogues using TCID[31-35,68] and IRMPD[69-74] techniques.

Threshold Analysis. The threshold regions of the CID cross sections for reaction 5 were analyzed by modeling with the empirical threshold law of eq 3 that includes RRKM and eq 2 without RRKM, respectively. Modeling of reaction 5 is based on a loose PSL TS.[63] The PSL TS model has been shown to provide the most accurate correction of kinetic shifts observed for CID reactions of noncovalently bound complexes.[75-83] The molecular parameters (vibrational frequencies and rotational constants) that describe the energized (*x*Cyd)H^+(*x*Cyd) base pair and corresponding PSL TSs used for the threshold analyses are listed in Tables S1 and Table S2. The optimized modeling parameters obtained from these analyses are summarized in Table 1. Fits to the CID cross sections of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs are illustrated in Figure 2. A similar comparison for all five (*x*Cyd)H^+(*x*Cyd) base pairs is shown in Figure S2. Figure 2 and Figure S2 show that the experimental cross sections of reaction 5 are reproduced with high fidelity throughout the threshold regions using the loose PSL TS model. These analyses yield E_0(PSL) threshold values of 1.73 ± 0.05, 1.74 ± 0.05, 1.75 ± 0.07, 1.77 ± 0.05, and 1.83 ± 0.06 eV for the (Cyd)H^+(Cyd), (dCyd)H^+(dCyd), (ddCyd)H^+(ddCyd), (m^5Cyd)H^+(m^5Cyd), and (m^5dCyd)H^+(m^5dCyd) base pairs, respectively. Also listed in Table 1 are entropies of activation at 1000 K, $\Delta S^{\ddagger}$(PSL), which are calculated from the molecular

constants used in the RRKM lifetime modeling. These values give a measure of the size and complexity of the system as well as the looseness of the PSL TS employed. The entropies of activation at 1000 K are large and positive and vary from 84 to 94 $J \cdot mol^{-1}K^{-1}$, consistent with expectations for dissociation via a loose PSL TS.[75-83] Modeling of the data without accounting for lifetime effects gives an estimate of the extent of kinetic shifting associated with the finite experimental time window. Without inclusion of the RRKM formalism in the modeling, the threshold values obtained are 1.57, 1.41, 1.38, 1.89 and 1.93 eV higher for the (Cyd)H^+(Cyd), (dCyd)H^+(dCyd), (ddCyd)H^+(ddCyd), (m^5Cyd)H^+(m^5Cyd), and (m^5dCyd)H^+(m^5dCyd) base pairs, respectively. These significant and relatively large kinetic shifts are consistent with the size of these (*x*Cyd)H^+(*x*Cyd) base pairs. The kinetic shift increases with the size (or number of vibrational modes) of the base pair. Noticeably, the kinetic shift decreases as the 2'- and 3'-hydroxy substituents are removed from (Cyd)H^+(Cyd) to the (dCyd)H^+(dCyd) and (ddCyd)H^+(ddCyd) base pairs, respectively. The kinetic shift is substantially increased in the 5-methylated analogues, (m^5Cyd)H^+(m^5Cyd) and (m^5dCyd)H^+(m^5dCyd), relative to the canonical base pairs. Although, the (m^5Cyd)H^+(m^5Cyd) base pair is larger in size (2 oxygen atoms or six additional vibrational modes) than the (m^5dCyd)H^+(m^5dCyd) base pair, the kinetic shift associated with the latter is slightly larger (by 0.04 eV), clearly indicating that it is not only the size of the (*x*Cyd)H^+(*x*Cyd) base pair that influences the kinetic shift, but an interplay between the size and strength of the noncovalent interactions cleaved in the dissociation.

B3LYP Optimized Geometries of Ground and Stable Structures. Ground and low-energy structures of the protonated nucleoside base pairs, (*x*Cyd)H^+(*x*Cyd), protonated nucleosides, H^+(*x*Cyd), and neutral nucleosides, *x*Cyd, were calculated as described in the Theoretical Calculations section. The ground conformations of all five (*x*Cyd)H^+(*x*Cyd) base pairs are displayed in Figure 3, whereas Table S3 compares the geometric parameters of these conformations. All low-energy conformers (within 10 kJ./mol Gibbs energy) of the ground conformers are compared in Figure S3. As can be seen in Figure 3 and Figure S3, the ground and low-energy conformers of all five (*x*Cyd)H^+(*x*Cyd) base pairs are bound by three hydrogen-

bonding interactions. The central hydrogen-bond, N3H$^+$···N3, is ionic due to protonation at the most basic site, N3 of the Cyt residue, whereas the two terminal hydrogen bonds, N4H···O2, are conventional hydrogen bonds.

The presence of both the 2'-, and 3'-hydroxy substituents in the (Cyd)H$^+$(Cyd) base pair, and its 5-methylated analogue, (m^{5}Cyd)H$^+$(m^{5}Cyd), enables the formation of an O2′H···O3′ sugar-sugar intramolecular hydrogen bond in each nucleoside of the base pair. Removal of the 2'- and 3'-hydroxy substituents from (Cyd)H$^+$(Cyd) to the (dCyd)H$^+$(dCyd) and (ddCyd)H$^+$(ddCyd) base pairs, respectively, results in a slight lengthening of the two terminal hydrogen-bonding interactions and a very slight decrease in the length of the central ionic hydrogen bond. 5-Methylation of the (Cyd)H$^+$(Cyd) base pair leads to a slight shortening of the upper terminal hydrogen bond, a marginal elongation of the central ionic hydrogen bond, and with no effect on the lower terminal hydrogen bond. The effects of 5-methylation of the (dCyd)H$^+$(dCyd) base pair are similar with a slight shortening of both terminal hydrogen bonds, and a marginal elongation of the central ionic hydrogen bond. Overall, the presence of the 2'-hydroxy and/or 5-methyl substituent(s) strengthens the terminal hydrogen bonds.

The protonated and neutral *x*Cyd nucleosides in the (*x*Cyd)H$^+$(*x*Cyd) base pairs adopt an antiparallel configuration, as generally found in double-stranded nucleic acid structures. The sugar puckering of the ground conformers of the (*x*Cyd)H$^+$(*x*Cyd) base pairs where *x*Cyd = Cyd, dCyd, ddCyd, m^{5}Cyd, and m^{5}dCyd are **C2'-endo**·C2'-endo, **C3'endo**·C3'-endo, **C3'endo**·C3'-endo, **C2'endo**·C3'-endo, and **C3'endo**·C2'-endo (see Figure 3), respectively. Here, the sugar puckering of the protonated H$^+$(*x*Cyd) nucleoside is indicated first in boldface font, followed by that of the neutral *x*Cyd nucleoside in standard font. 5-Methylation is found to alter the preferred sugar puckering of the neutral nucleoside, while retaining the puckering of the protonated nucleoside in the canonical nucleoside base pairs. Although, the removal of the 2'- and 3'-hydroxy substituents makes minor contributions to the energetics, their absence changes the sugar puckering from **C2'-endo**·C2'-endo in the (Cyd)H$^+$(Cyd) base pair to **C3'endo**·C3'-endo in the (dCyd)H$^+$(dCyd) and (ddCyd)H$^+$(ddCyd) base pairs.

Parallel quantum chemical calculations were also performed for the neutral *x*Cyd and protonated H^+(*x*Cyd) nucleosides. The stable low-energy conformers of the neutral *x*Cyd and protonated H^+(*x*Cyd) nucleosides along with their relative Gibbs energies at 298 K are shown in Figures S4 and S5, respectively. The geometric parameters of the ground conformers of the neutral nucleosides are listed in Table S4, whereas the geometric parameters of the stable ground conformers of the N3- and O2-protonated nucleosides are listed in Tables S5 and S6, respectively. Our computed results for the low-energy conformers of the N3-and O2-protonated nucleosides are consistent with previous results in which the B3LYP functional was found to favor N3 over O2 protonation.[70]

Key Structural Parameters. In addition to the geometric parameters of the ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs summarized in Table S3, three additional structural parameters including the pseudorotation phase angle, glycosidic bond angle and 5-hydroxy orientations are of particular interest. These key structural parameters of the ground and very low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs are visually compared in the polar plots of Figure 4; a similar comparison of these parameters for all stable conformers computed within 10 kJ/mol of the ground conformers is shown in Figure S6. Among the ground and very low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs only C2'-endo and C3'-endo conformations are observed. The 2'- and 3'-hydroxy groups do not significantly influence the preferred *P*, whereas 5-methylation is found to slightly increase *P* in the C3'-endo conformers and marginally decrease *P* in the C2'-endo conformers. Not surprisingly, the glycosidic bond angles of the (*x*Cyd)H^+(*x*Cyd) base pairs exclusively adopt anti conformations as this is required for base pairing. The C2′-endo conformers exhibit larger glycosidic bond angles (~230º) than C3′-endo conformers (~205º). 5-Methylation leads to a slight increase (~1−3º) in the glycosidic bond angles. In all cases, the 5'-hydroxy orientations are gauche$^+$, as this orientation provides stabilization via formation of noncanonical C6H···O5′ hydrogen-bonding interactions in both nucleosides of the base pair. Only modest variation in the pseudorotation phase angles and glycosidic bond angles are found among the low-energy conformers compared in Figure S6.

Notably, a few conformers exhibiting C1′-exo or C4′-exo sugar puckers are found. While all of the most stable conformers exhibit gauche$^+$ 5-hydroxy orientations, conformers with gauche$^-$ and trans orientations are also found among the low-energy conformers compared in Figure S6. These conformers lack the noncanonical C6H···O5′ hydrogen-bonding interactions; their relative stabilities suggest that these interactions provide ~4–6 kJ/mol to the overall stabilization of the base pair.

These three key structural parameters of the neutral and protonated nucleosides are similarly summarized in the polar plots of Figure S7. As found for the protonated (*x*Cyd)H$^+$(*x*Cyd) base pairs, the polar plots for the neutral *x*Cyd and protonated H$^+$(*x*Cyd) nucleosides show that among the stable low-energy conformers computed there is a strong and approximately equal preference for either C2'-endo or C3'-endo sugar puckering. The ground N3-protonated conformers of H$^+$(dCyd), H$^+$(ddCyd), H$^+$(m^{5}Cyd) and H$^+$(m^{5}dCyd) species exhibit C3'-endo sugar puckering, whereas H$^+$(Cyd) exhibits C2'-endo sugar puckering. Parallel results are found for the O2-protonated conformers of these nucleosides. In addition to these common sugar puckers, a good number of conformers exhibit C3'-exo sugar puckering, and a few with C1'-exo, C4'-endo or C4'-exo sugar puckerings are also found. These less common sugar puckers are also less stable and primarily of importance in the interconversion between C2'-endo and C3'-endo conformations. Most of the conformers exhibit a preference for the *anti* nucleobase orientation over *syn*. The ground N3- and O2-protonated conformers exclusively exhibit an *anti* nucleobase orientation. Similar to the (*x*Cyd)H$^+$(*x*Cyd) base pairs, which exclusively adopt the gauche$^+$ orientation, a majority (~60%) of the low-energy conformers of the protonated and neutral *x*Cyd nucleosides exhibit a preference for a gauche$^+$ 5'-hydroxy orientation, which enables the C6H···O5′ hydrogen-bonding interaction. Almost equal populations (~20% each) of the low-energy conformers of the protonated and neutral nucleosides adopt gauche$^-$ or trans orientations (Figure S3).

ESP Maps of Optimized Geometries of (*x*Cyd)H$^+$(*x*Cyd) Ground Structures: ESP maps showing the 3D spatial arrangement of the basic and acidic sites and the charge distribution

of these sites provide explanation for the observed trend in hydrogen-bond distances observed in the ground conformers of Figure 3. Shown in Figure S8 is a representation of ESP maps of the ground conformers of the (*x*Cyd)H$^+$(*x*Cyd) base pairs and atomic charges obtained from Mülliken population analysis at the B3LYP/6-311+G(d,p) level of theory. The ESP maps show regions in the molecular plane of the ground structures with negative charges encompassing the N3 and O2 atoms, very consistent with both theoretical and experimental observation that show the N3 and O2 positions to be the most greatly favored sites for protonation.[70] It is essential to keep in mind that, these systems are nearly symmetric nucleoside base pairs in which switching of the proton between the N3 atoms of the Cyt residues will lead to an inversion of the two terminal N4H···O2 hydrogen bonds. Generally, among the hydrogen-bond acceptors, the ESP maps reveal that the O2 atom of the neutral nucleoside is more basic, followed by the N3 atom of the neutral nucleoside, and least for the O2 atom of the protonated nucleoside. The relatively high concentration of negative charge on the O2 atom of the neutral nucleoside makes it a better hydrogen-bond acceptor, consistent with the formation of shorter hydrogen bonds (upper terminal N4H···O2 hydrogen bonds of Figure 3). Conversely, the low partial negative charge on the O2 atom of the protonated nucleoside makes it the least basic site in each system leading to longer N4H···O2 hydrogen bonds (bottom terminal hydrogen bonds of Figure 3). The negative charge concentration of the basic sites are found to be systematically enhanced in the 5-methylated systems, (m^{5}Cyd)H$^+$(m^{5}Cyd), and (m^{5}dCyd)H$^+$(m^{5}dCyd). Similarly, the removal of 2'-, and 3'-hydroxy substituents from (Cyd)H$^+$(Cyd) to the (dCyd)H$^+$(dCyd) and (ddCyd)H$^+$(ddCyd) base pairs also enriches the partial charges of the basic sites.

NCI and QTAIM Analyses of the (*x*Cyd)H$^+$(*x*Cyd) Base Pairs. For further characterization of the dynamic and electronic structures of the bonding interactions in the ground conformers of the (*x*Cyd)H$^+$(*x*Cyd) base pairs (Figure 3), we performed NCI and QTAIM analyses, which are sufficiently accurate to draw qualitative conclusions. The 3D plots generated from the NCI analysis are shown together with the topology maps obtained from the QTAIM analysis in Figure 5 for the (dCyd)H$^+$(dCyd) and (m^{5}dCyd)H$^+$(m^{5}dCyd) base pairs. A similar

comparison is shown for all five (*x*Cyd)H^+(*x*Cyd) base pairs in Figure S9 of the Supporting Information. The combined plots reveal the presence of the intermolecular hydrogen-bonding interactions between the Cyt residues, along with additional intramolecular interactions between the sugar moieties and Cyt residues as well as within the sugar moiety. The hydrogen-bonding interactions associated with the base-pairing interactions are seen as two small blue and a third green disk-shaped regions. The density values indicated by the color of the isosurfaces, indicate that the upper terminal, shorter N4H···O2 and central ionic N3H^+···N3 hydrogen bonds are stronger than the bottom terminal, longer N4H···O2 hydrogen bond. That is, blue isosurfaces in the NCI plots indicate regions of relatively high electron density between the interacting atoms and stronger stabilizing interactions, whereas the green isosurface indicates lower electron density and a weaker stabilizing interaction. The NCI plots also reveal the presence of weak intramolecular van der Waal attractions between the Cyt and sugar moieties of each nucleoside. These also appear as green isosurfaces, and account for the formation of the noncanonical C6H···O5′ hydrogen-bonding interactions. NCI analyses further highlight weak O2′H···O3′ hydrogen-bonding interactions in the sugar moiety for the (Cyd)H^+(Cyd) and (m^5Cyd)H^+(m^5Cyd) base pairs, and additional unique, but weaker N4H···CH_3 intramolecular interactions, which are formed above the plane of the Cyt residues of the (m^5Cyd)H^+(m^5Cyd) and (m^5dCyd)H^+(m^5dCyd) base pairs. Finally, red isosurfaces are found in the center of the ring structure of the Cyt residues, accounting for the presence of steric clashes among the ring atoms. The structural effects of the 2′ and 3′-hydroxy and 5-methyl substituents on the inter- and intramolecular interactions were further elucidated by examining the intensities of the blue isosurfaces between the Cyt residues as shown in Figure 5 (and Figure S9). In the (ddCyd)H^+(ddCyd), (m^5Cyd)H^+(m^5Cyd), and (m^5dCyd)H^+(m^5dCyd) systems, a more intense isosurface is evident, and this is indicated by an opening in the center of the upper blue isosurfaces (that can seen in the views twisted out of the plane), qualitatively suggesting an increase in electron density. However, this ‘hole’ is very faint in the canonical (dCyd)H^+(dCyd) base pair, and conspicuously absent in the canonical (Cyd)H^+(Cyd) base pair, indicative of

progressive electron density depletion. In the 5-methylated analogues, the intensity of the terminal (upper) blue isosurface is further increased with concomitant widening of the 'hole', qualitatively implying a dramatic increase in electron density between the interacting atoms. Qualitatively, the bond path produced from topological analysis of Bader's theory of atom in molecules[57] is used as a signature for mapping the biologically relevant interactions found in the ground structures of the (*x*Cyd)H^+(*x*Cyd) base pairs. The bond paths are shown as solid yellow lines in Figure 5 (and Figure S9). A bond path is found to interconnect the following interactions: N3H^+···N3, N4H···O2, C6H···O5′, and O2′H···O3′ and passes through the RDG isosurfaces of each interaction. Although NCI analysis indicates a weak green isosurface for the C2′H···O2 and N4H···CH_3 interactions, no bond path was found interconnecting the interacting atoms. The corresponding 2D NCI scatter maps for the ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs as shown in Figure S10 are plotted as the RDG versus the electron density. Three unique spikes are distinguishable in the 2D scatter maps, which are specified by the color-filled isosurface. The black horizontal line indicates the RDG plotted in Figure 5 (and Figure S9), such that the segments just crossing the peaks represent the RDG isosurface. The blue filled regions correspond to the stronger N3H^+···N3 and N4H···O2 hydrogen-bonding interactions. Weak C6H···O5′ van der Waals interactions are shown as green regions, whereas the red regions refer to interactions within the Cyt and sugar rings where there is strong steric repulsion.

DISCUSSION

Comparison of Experimental and Theoretical (*x*Cyd)H^+(*x*Cyd) BPEs. Proper interpretation and validation of the BPEs of the studied *i*-motif model systems drawn from cytidine and cytidine analogues and measured by the TCID technique is accomplished by comparison with BPEs predicted from computational work involving the B3LYP, B3P86, and M06-2X functionals. The 0 K measured BPEs are compared with the corresponding theoretical values in Table 2. Also included in Table 2 are results previously reported for the

(dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs.[35] The BPEs predicted using B3LYP with the 6-311+G(2d,2p) and def2-TZVDDP basis sets are compared to the TCID measured BPEs in Figure 6. Similar comparisons of the BPEs predicted by B3P86 and M06-2X with the 6-311+G(2d,2p) basis set are shown in Figure S11. Excellent agreement between the B3LYP calculated and TCID measured BPEs is found as depicted in Figure 6. The superior performance of the B3LYP functional is quantitatively seen in the comparison of the respective mean absolute deviations (MADs) between the BPEs predicted by each level of theory versus those experimentally determined and the average experimental uncertainty (AEU) as provided in Table 2. The MADs between the B3LYP/6-311+G(2d,2p) and B3LYP/def2-TZVPPD calculated and TCID measured BPEs are 0.7 ± 0.3 and 1.4 ± 1.7 kJ/mol, respectively. The MAD in each case is much smaller than the AEU of 5.2 ± 0.6 kJ/mol, indicating that B3LYP provides an excellent description of the noncovalent interactions responsible for the binding of these (*x*Cyd)H^+(*x*Cyd) base pairs with the 6-311+G(2d,2p) basis set performing marginally better than the def2-TVZPPD basis set. The MADs between experiment and the B3P86 and M06-2X levels of theory are much larger 10.6 ± 1.9 and 12.3 ± 2.0 kJ/mol, more than two times larger than the AEU in the BPEs. As seen in Figure S11, both the B3P86 and M06-2X functionals systematically overestimate the BPEs by an average of 10.6 and 12.3 kJ/mol, respectively. These discrepancies are a proper reflection of the poorer ability of the B3P86 and M06-2X functionals for accurately describing the noncovalent interactions that contribute to the stability of these base pairs. Although there is no gold standard in the choice of computational method for evaluating the structures and energetics of noncovalently-bound complexes, the B3LYP/6-311+G(2d,2p) and B3LYP/def2-TVZPPD have been shown to be reliable in describing similar hydrogen-bonding interactions in parallel protonated nucleobase pairs previously investigated.[31-34]

The BPEs measured here for the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs exceed those previously reported[35] by 7.9 and 14.6 kJ/mol (Table 2), respectively. The excellent agreement between theory and experiment found here clearly suggest that the present results are more reliable. Further, present results are internally consistent with results from competitive

dissociation of mixed (*x*Cyd)H^+(*y*Cyd) base pairs being examined in parallel studies, whereas the values reported earlier[35] are not. The most likely explanation for the differences in the BPEs measured here for the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs, and those previously reported is that the source conditions employed in the former work had not completely thermalized the precursor ion population. As a result, the BPEs determined were underestimated in the previous work.

Conversion from 0 to 298 K Thermodynamics. The TCID experiments performed here were performed at room temperature as is typical of most laboratory measurements. However, the thresholds extracted from our thermochemical analysis remove the experimental broadening and thus provide 0 K bond enthalpies. To enable facile comparison to room temperature observations and values typically reported in the literature, the 0 K BPEs determined here are converted to enthalpies and Gibbs energies at 298 K using standard statistical thermodynamics formulas in conjunction with computed parameters (vibrational frequencies and rotational constants) determined for the B3LYP/6-311+G(d,p) optimized geometries. The vibrational frequencies and average vibrational energies are listed in Table S1, whereas Table S2 lists the associated rotational constants. Listed in Table 3 are 0 and 298 K enthalpy, Gibbs energy, and enthalpic and entropic corrections for all systems investigated. The uncertainties in the enthalpies and Gibbs energies reported in Table 3 are based on variation in the computed values arising from scaling the harmonic vibrational frequencies up and down by 10%.

Influence of the 2'-, and 2',3'-Hydroxy Substituents on the Base Pairing Interactions. The TCID measured BPEs of the (Cyd)H^+(Cyd), (dCyd)H^+(dCyd), and (ddCyd)H^+(ddCyd) base pairs are 166.9 ± 4.5 kJ/mol, 167.7 ± 5.3 kJ/mol, and 168.9 ± 5.3 kJ/mol, respectively (Table 2). Thus, the 2'- and 3'-hydroxy substituents weaken the base pairing interaction by 1.2 ± 7.5 kJ/mol and 0.8 ± 7.0 kJ/mol, respectively. B3LYP/6-311+G(2d,2p) theory also predicts a 2.0–2.3 kJ/mol weakening in the base-pairing interactions arising from the 2'- and 3'-hydroxy substituents, whereas the other levels of theory suggest that 3'-hydroxy substituent increases the BPE by 0.8–1.2 kJ/mol while the 2'-hydroxy substituent weakens the

BPE 0.1–0.6 kJ/mol. The observed trend can be understood by examining the ESP maps (see Figure S8). Absence of both the 2'-, and 3'-hydroxy substituents in the (dCyd)H^+(dCyd) and (ddCyd)H^+(ddCyd) base pairs is found to increase the negative charge concentration of the N3 atoms in the neutral nucleosides of these base pairs, leading to a shortening of the central ionic hydrogen bond in (dCyd)H^+(dCyd) and (ddCyd)H^+(ddCyd) as compared with (Cyd)H^+(Cyd) as shown in Figure 3. While a systematic ~1 kJ/mol decrease per hydroxy substituent is measured, the uncertainties in these values are ~5 times larger. Overall, the trends in the measured and computed BPEs of the (*x*Cyd)H^+(*x*Cyd) base pairs indicate that the 2'- and 3'-hydroxy substituents have very little impact on the base-pairing interactions.

Influence of 5-Methylation on the Base-Pairing Interactions. The TCID measured BPEs of the (Cyd)H^+(Cyd), (dCyd)H^+(dCyd), (m^5Cyd)H^+(m^5Cyd), and (m^5dCyd)H^+(m^5dCyd) base pairs are 166.9 ± 4.5 kJ/mol, 167.7 ± 5.3 kJ/mol, 170.8 ± 4.7 kJ/mol, and 176.6 ± 6.0 kJ/mol, respectively (Table 2). Thus, 5-methylation of the Cyt residues enhances the base-pairing interaction by 3.9 ± 6.5 kJ/mol of the RNA analogue, and by 8.9 ± 8.0 kJ/mol of the DNA analogue. Theory also predicts an increase in the base-pairing interaction of 3.4–6.7 kJ/mol for the RNA analogue, and an increase in the base-pairing interaction of 7.5–12.0 kJ/mol for the DNA analogue. As summarized, the BPEs of the 5-methylated analogues are higher than the corresponding canonical counterparts, suggesting that 5-methylation strengthens the base pairing interactions. As shown in Table 2, this observation corroborates the results of previously studied protonated nucleobase pairs of cytosine, (Cyt)H^+(Cyt), and 5-methylcytosine, (m^5Cyt)H^+(m^5Cyt).[31] in which 5-methylation was found to increase the BPE by 7.5 ± 7.0 kJ/mol. As an electron donors, the methyl substituents increase the electron density within the Cyt residues with concomitant stabilization of the positive charge on the exocyclic N3-atom of the protonated nucleoside. This proposition is corroborated by complementary ESP maps that reveal an increase in the charge density of the Cyt N3 and O2 atoms of the neutral and protonated nucleosides. Although, 5-methylation enhances the electron density of both of the N3 atoms, the enhancement is more substantial for the N3 atom of the protonated nucleoside leading to shorter

N3–H^+ bond distances (Table S3) with concomitant lengthening of the central ionic hydrogen bonds (Figure 3). The observed shorter N3–H^+ bond at one of the Cyt residues in the protonated nucleoside base pairs is consistent with the concept of a double-well potential. Apparently, there is no variation in the N3–H^+ and O2–H^+ bond lengths in the ground conformers of the N3- and O2-protonated nucleosides as shown in Tables S5 and S6. Furthermore, a noticeable increase in electron density of the O2 atoms leads to a shortening of the two terminal hydrogens bonds in the (m^5Cyd)H^+(m^5Cyd) and (m^5dCyd)H^+(m^5dCyd) base pairs as compared with the canonical base pairs (Figure 3), and this should compensate for the slight weakening in the central ionic hydrogen bonds and explain the observed enhancement in the BPEs.

Influence of Polarizability on Measured BPEs. In order to gain a detailed understanding of the factors that may contribute to the base-pairing interactions, the measured BPEs of the (*x*Cyd)H^+(*x*Cyd) base pairs are compared to the isotropic molecular polarizabilities of the neutral *x*Cyd nucleosides in Figure 7. As can clearly be seen in the figure a roughly linear correlation between the measured BPEs and neutral polarizability volume is associated with 5-methylation, whereas an anti correlation is associated with the 2′- and 3′-hydroxy substituents, indicating that polarizability is playing a role in the stabilization of the (*x*Cyd)H^+(*x*Cyd) base pairs. 5-Methylation is found to enhance both the polarizability and BPE, consistent with the increase in size of the neutral methylated *x*Cyd nucleosides and thus leading to an enhancement in the ion-dipole interactions in the 5-methylated (*x*Cyd)H^+(*x*Cyd) base pairs. Although, each hydroxy substituent enhances the polarizability, the BPE is found to slightly decrease, indicating that the 2′- and 3′-hydroxy substituents weaken the BPEs due to steric hindrance.

Implications of Modification on the Stability of Genetic *i*-Motif Conformations. In order to comprehensively uncover the biological significance of the nucleic acid *i*-motif, an understanding of the energetics and structural changes associated with specific modifications of the functional subunits of the *i*-motif architecture is important. The potential enhancement in stability of DNA and RNA *i*-motif conformations via chemical modifications of its building

blocks has been methodically and accurately investigated previously by our research group at the nucleobase level.[31-34]

Both the TCID measured and theoretically predicted BPEs indicate that 5-methylation of the Cyt residues of the (Cyd)H^+(Cyd) base pair results in strengthening of the base-pairing interactions, the BPE increases by 3.9 ± 6.5 kJ/mol (TCID) and 3.4–6.7 kJ/mol (theory). These results suggest that 5-methylation increases the stability of RNA *i*-motif conformations, and it is accordingly expected that 5-hypermethylated *i*-motif structures would exhibit greater thermal stability and thus be less susceptible to unfolding to single-stranded RNA as compared with *i*-motif conformations formed from the unmodified RNA. Again, the TCID measured and theoretically predicted BPEs indicate that 5-methylation of the Cyt residues of the (dCyd)H^+(dCyd) base pair increases the strength of binding by 8.9 ± 8.0 kJ/mol (TCID) and 7.5–12.0 kJ/mol (theory). These results suggest that 5-methylation also increases the stability of DNA *i*-motif conformations, and it is accordingly expected that 5-hypermethylated *i*-motif structures would exceedingly be more resistant to thermal unfolding into a DNA duplex as compared with the *i*-motif conformations formed from the unmodified canonical DNA.

Analysis of our results for model systems comprising the unmodified RNA, (Cyd)H^+(Cyd), which possesses both 2'- and 3'-hydroxy substituents, and its DNA counterpart, (dCyd)H^+(dCyd), in which the 2'-hydroxy substituent have been removed as well as the (ddCyd)H^+(ddCyd) analogue that lacks both the 2'- and 3'-hydroxy substituents, leads to the observation that the presence of the hydroxy substituents lead to very minor destabilization of the *i*-motif conformations. For instances, the absence of 2'-hydroxy substituents in (dCyd)H^+(dCyd) model system marginally increases the BPE by about 0.8 ± 7.0 kJ/mol, whereas the increase in BPE is doubled, 2.0 ± 7.0 kJ/mol, for the (ddCyd)H^+(ddCyd) analogue. Although the magnitude of the increase in BPE is marginal, the result is somewhat indicative of the considerable stability that the DNA *i*-motif conformation may have over its RNA analogue where pairing interactions contribute to the overall stability. The lower stability of the RNA *i*-motif could be attributed to crowding caused by the steric hindrance experienced by the 2'-hydroxy substituents in the

narrow grooves of rather compact *i*-motif conformations. Our results corroborate with previous solution-phase reports that show that DNA forms more stable *i*-motif conformations compared to their RNA counterparts.[84,85] Thus, in a previous study aimed at examining the potential of *i*-motif formation by an 18-mer DNA with its corresponding 18-mer RNA sequence, melting temperatures of 54 ºC and 25 ºC respectively, were reported indicating a more stable DNA than RNA *i*-motif structure.[84] Likewise, at a strand concentration of 9.8 mM, melting temperatures of 12 ºC and 69 ºC were reported for the RNA $[r(UC_5)]_4$ and DNA $[r(TC_5)]_4$ *i*-motif conformations respectively.[84]

Consistent with previous studies, present findings indicate that a DNA *i*-motif comprised of 5-hypermethylated symmetric Cyt is 5.8 ± 7.6 kJ/mol per base pair more stable than an RNA *i*-motif of the analogous sequence. The higher stabilization of the DNA *i*-motif upon 5-methylation is primarily due to the electronic effect of the methyl substituents, which promote charge stabilization and also compensate for stronger base-pairing interactions with concomitant enhancement in the thermodynamic stability of the DNA *i*-motif conformations. These results further demonstrate that for DNA and RNA oligonucleotides containing the same number of protonated Cyt base pairs, 5-methylation would tend to stabilize the DNA *i*-motif conformations more than the RNA counterparts. As 5-hypermethylation of DNA can be severe in diseased states associated with $d(CCG)_n \cdot d(CGG)_n$ trinucleotide sequences, which are implicated as the cause of Fragile X syndrome among human populations, these finding suggest that the DNA *i*-motif may play a role.[85] This extrapolation of our results is based on parallel studies conducted in solution that probed a series of unmodified oligonucleotides containing Cyd and their 5-methylated analogues and reported a change in the melting temperature, T_m, from 67.7 ± 0.5 ºC to 75.9 ± 0.3 ºC respectively, unambiguously suggesting that duplexes containing 5-methylated Cyt residues are thermally more stable.[86] Most recently, solution-phase studies have shown that 5-methylation of both Cyt residues in symmetric base pairs within the stem region of the *i*-motif also leads to an increase in stability.[87] Stabilization of *i*-motif conformations upon 5-permethylation was also seen in the enhancement in pKa of *i*-motif conformations from 6.1 to

6.3,[88] indicating that 5-methylation not only stabilizes *i*-motif conformations, but that the associated base pairing alters the pH dependent conditions that purportedly governs the formation of the *i*-motif architecture.

To further deepen our understanding and unravel the influence of modifications on the stability of genomic *i*-motif conformations, additional studies involving a greater variety of naturally-occurring and synthetic cytidine nucleoside analogues that substantially expand the scope of this work are being pursued.

CONCLUSIONS

Methylated nucleosides play important roles in the functions of nucleic acids, and 5-methylation of Cyt residues is a particularly important inheritable, dynamic and common epigenetic signature that occurs in several regions of the human genome. The stability of DNA and RNA genomic *i*-motif conformations can be intrinsically altered through modifications. In order to elucidate the effects of 5-methylation of Cyt residues and investigate the effects of the 2'- and 3'-hydroxy substituents on the base-pairing interactions, the kinetic energy-dependent cross sections for CID of five model protonated nucleoside base pairs of the canonical and several modified cytidine analogues, $(Cyd)H^+(Cyd)$, $(dCyd)H^+(dCyd)$, $(ddCyd)H^+(ddCyd)$, $(m^5Cyd)H^+(m^5Cyd)$ and $(m^5dCyd)H^+(m^5dCyd)$, are determined using a custom-built guided ion beam mass spectrometer. The dominant dissociation channel for five $(xCyd)H^+(xCyd)$ base pairs studied involves cleavage of the three hydrogen bonds responsible for the base pairing in these protonated nucleoside base pairs. The BPEs of these model systems relevant to genomic *i*-motif structures were determined by threshold analysis of the CID cross sections, while properly accounting for the effects of the kinetic and internal energy distributions of the reactants, multiple ion-neutral collisions, and the lifetime of the activated base pairs using a loose PSL TS model. Molecular parameters needed to aid threshold analysis of our experimental data, as well as structures and theoretical estimates for the BPEs, were determined from theoretical calculations performed at B3LYP/6-311+G(2d,2p)//B3LYP/6-311+G(d,p), B3LYP/def2-

TZVPPD//B3LYP/6-311+G(d,p), B3P86/6-311+G(2d,2p)//B3P86/6-311+G(d,p), and M06-2X/6-311+G(2d,2p)//M06-2X/6-311+G(d,p) levels of theory including zero-point energy and counterpoise corrections. Excellent agreement was found between the TCID measured and BPEs calculated at the B3LYP level of theory, suggesting that the B3LYP functional describes the intrinsic properties of the noncovalent interactions in these (*x*Cyd)H^+(*x*Cyd) base pairs that dominate the stabilization of i-motif conformations. In contrast, the B3P86 and M062X levels of theory systematically overestimate the strength of the base pairing.

Present results indicate that the thermodynamics of *i*-motif forming capability of the RNA and DNA model systems investigated systemically change in accordance with the presence or absences of the 2'- and 3'-hydroxy substituents of the sugar moieties and 5-methylation of the nucleobase residue. Overall, both our experimental results and theoretical findings confirm that the DNA models exhibit a greater enhancement in the base-pairing interactions compared with the RNA model systems, and that the impact of 5-methylation on the base-pairing interaction is greater than the influence of the 2'-hydroxy and 3'-hydroxy substituents. Further, the trends in measured BPEs indicate that polarizability is a contributing factors that influences the BPEs of the (*x*Cyd)H^+(*x*Cyd) base pairs with 5-methylation providing a larger influence on the binding than the 2'- and 3'-hydroxy substituents.

Theoretical analysis using combined NCI and AIM approaches reveals the existence of a systemic network of canonical and noncanonical hydrogen bonds in these model base pairs that include N3H^+···N3, N4H···O2, C6H···O5′, and O2′H···O3′ hydrogen-bonding interactions. Particularly, the C6H···O5′ interactions between the sugar moieties and the Cyt residues are important noncanonical hydrogen bonds in biological macromolecules. The role that these interactions play in determining the stabilization of the three-dimensional structure of genomic *i*-motif conformations could be significant owing to the large number of these noncanonical hydrogen bonds that may be presence in large Cyd-rich oligonucleotide sequences. These noncanonical nucleobase-sugar hydrogen-bonding interactions are found to contribute to the stabilization of all five (*x*Cyd)H^+(*x*Cyd) base pairs examined.

Collectively, the present results coupled with our previous reports provide compelling evidence, that 5-hypermethylation of the Cyt residues in *i*-motif functional units occurs to further enhance the base-pairing interactions and therefore leads to enhanced stability of *i*-motif architectures. Importantly, our results show that 5-hypermethylation enhances the stability of DNA *i*-motif conformations to a greater extent than RNA sequences. The shift in the competitive equilibrium towards the formation of a stable *i*-motif conformation over the canonical Watson-Crick duplex is primarily due to the high thermodynamic stability of the symmetric canonical (*x*Cyd)H^+(*x*Cyd) base pairs, which is enhanced upon 5-methylation of the Cyt residues. Overall, present results lead to a very fascinating inference that the accumulated effect of 5-hypermethylation could be dramatic in diseased states where hundreds of trinucleotide repeats and protonated cytosine base pairs may be present.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/article #. Tables of average internal energies, vibrational frequencies, rotational constants, and geometric parameters, of the B3LYP/6-311+G(d,p) ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs, and the neutral and N3- and O2-protonated forms of the *x*Cyd nucleosides; figures comparing the CID cross sections for the (*x*Cyd)H^+(*x*Cyd) protonated nucleoside base pairs; zero-pressure-extrapolated CID cross sections for the (*x*Cyd)H^+(*x*Cyd) protonated nucleoside base pairs and fits to the data using eq 3; stable B3LYP/6-311+G(d,p) conformations of the (*x*Cyd)H^+(*x*Cyd) base pairs, and the neutral and N3- and O2-protonated forms of the *x*Cyd nucleosides and relative Gibbs energies at 298 K; ESP and NCI maps superimposed on the ground conformations of the (*x*Cyd)H^+(*x*Cyd) protonated nucleoside base pairs and the corresponding reduced electron density gradient scatter maps; comparison of the B3P86 and M06-2X predicted BPEs versus TCID measured BPEs of the (*x*Cyd)H^+(*x*Cyd) protonated nucleoside base pairs (PDF)

AUTHOR INFORMATION

Corresponding Author

M. T. Rodgers – *Department of Chemistry, Wayne State University, Detroit, Michigan, 48202, United States;* orcid.ord/0000-0002-5614-0948; Phone (313) 577-2431; Email: mrodgers@chem.wayne.edu

Authors

Yakubu S. Seidu – *Department of Chemistry, Wayne State University, Detroit, Michigan, 48202, United States;* orcid.ord/0000-0002-8347-7263

H. A. Roy – *Department of Chemistry, Wayne State University, Detroit, Michigan, 48202, United States;* orcid.ord/0000-0002-8347-7263

ACKNOWLEDGEMENTS

Financial support for this work was provided by the National Science Foundation, grant CHE-1709789. Computational resources were provided by Wayne State University C&IT. The authors acknowledge support from the Wayne State University via a Thomas C. Rumble Graduate Research Fellowship for SYS and a Summer Dissertation Fellowship for HAR.

REFERENCES

(1) Kaushik, M.; Kaushik, S.; Roy, K.; Singh, A.; Mahendru, S.; Kumar, M.; Chaudhary, S.; Ahmed S.; Kukreti, S. A Bouquet of DNA Structures: Emerging Diversity. *Biochem. Biophys. Rep*. **2016**, *5*, 388–395.

(2) Choi, J.; Majima, T. Conformational Changes on non-B DNA. *Chem. Soc. Rev*. **2011**, *40*, 5893–5909.

(3) Mir, B.; Solés, X.; González, C.; Escaja, N. The Effect of the Neutral Cytidine Protonated Analogue Pseudoisocytidine on the Stability of *i*-Motif Structures. *Sci. Rep*. **2017**, *7*, 2772.

(4) Saha, S.; Krishnan, Y. pH Sensitive DNA Devices. In *DNA Congugates and Sensors*; Fox, K.R., Brown, T., Eds; *RSC Biomolecular Sciences*. **2012**, 166–189.

(5) Day, H. A.; Pavlou, P.; Waller, Z. A. E. i-Motif DNA: Structure, Stability and Targeting with Ligands. *Bioorg. Med. Chem*. **2014**, *22*, 4407-4418.

(6) Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D. *i*-Motif DNA Structures are Formed in the Nuclei of Human Cells. *Nat. Chem*. **2018**, *10*, 631–637.

(7) Tang, W.; Niu, K.; Yu, G.; Jin, Y.; Zhang, X.; Peng, Y.; Chen, S.; Deng, H.; Li, S.; Wang, J. In Vivo Visualization of the i-Motif DNA Secondary Structure in the Bombyx Mori Testis. *Epigenetics & Chromatin*. **2020**, *13*, 1-12.

(8) Phan, A. T.; Guéron, M.; Leroy, J.-L. The Solution Structure and Internal Motions of a Fragment of the Cytidine-Rich Strand of the Human Telomere. *J. Mol. Biol*. **2000**, *299*, 123–44.

(9) Garavís, M.; Escaja, N.; Gabelica, V.; Villasante, A.; González, C. Centromeric Alpha-Satellite DNA Adopts Dimeric *i*-Motif Structures Capped by AT Hoogsteen Base Pairs. *Chem. Eur. J*. **2015**, *21*, 9816–9824.

(10) Nonin-Lecomte, S.; Leroy. J. L. Structure of a C-Rich Strand Fragment of the Human Centromeric Satellite III: A pH-Dependent Intercalation Topology. *J. Mol. Biol*. **2001**, *309*, 491–506.

(11) Brooks, T. A.; Kendrick, S.; Hurley, L. Making Sense of G-Quadruplex and *i*-Motif Functions in Oncogene Promoters. *FEBS J*. **2010**, *277*, 3459–69.

(12) Chaudhary, S.; Kaushik, M.; Ahmed, S.; Kukreti, S. Exploring Potential of i-Motif DNA Formed in the Promoter Region of GRIN1 Gene for Nanotechnological Applications. *Results in Chem*. **2020**, *2*, 100086.

(13) Thorne, B. N.; Ellenbroek, B. A.; Day, D. J. Evaluation of i-Motif Formation in the Serotonin Transporter-Linked Polymorphic Region. *Chembiochem*. **2021**, *22*, 349-353.

(14) Brown, S. L.; Kendrick, S. The *i*-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharmaceuticals*. **2021**, *14*, 96.
(15) Shi, L.; Peng, P.; Zheng, J.; Wang, Q.; Tian, Z.; Wang, H.; Li, T. I-Motif/Miniduplex Hybrid Structures Bind Benzothiazole Dyes with Unprecedented Efficiencies: A Generic Light-up System for Label-Free DNA Nanoassemblies and Bioimaging. *Nucleic Acids Res*. **2020**, *48*, 1681-1690.
(16) Evich, M.; Spring-Connell, A. M.; Germann, M. W. Impact of Modified Ribose Sugars on Nucleic Acid Conformation and Function. *Heterocycl. Commun*. **2017**, *23*, 155-165.
(17) Chen, K.; Zhao, B. S.; He, C. Nucleic Acid Modifications in Regulation of Gene Expression. *Cell Chem. Biol*. **2016**, *23*, 74–85.
(18) Pérez, A.; Castellazzi, C. L.; Battistini, F.; Collinet, K.; Flores, O.; Deniz, O.; Ruiz, M. L.; Torrents, D.; Eritja, R.; Soler-López, M. Impact of Methylation on the Physical Properties of DNA. *Biophys. J*. **2012**, *102*, 2140–2148.
(19) McKenzie, L. K.; El-Khoury, R.; Thorpe, J. D.; Damha, M. J.; Hollenstein, M. Recent Progress in Non-Native Nucleic Acid Modifications. *Chem. Soc. Rev*. **2021**, *50*, 5126-5164.
(20) Jordheim L. P.; Durantel, D.; Zoulim, F.; Dumontet. C. Advances in the Development of Nucleoside and Nucleotide Analogues for Cancer and Viral Diseases. *Nat. Rev. Drug. Discov.* **2013**, *12*, 447–464.
(21) Corey, D. R. Chemical Modification: The Key to Clinical Application of RNA Interference? *J. Clin. Invest*. **2007**, *117*, 3615–22.
(22) Gallo, M.; Montserrat, J.; Iribarren, A. M. Design and Applications of Modified Oligonucleotides. *Bra. J. Med. Biol. Res*. **2003**, *36*, 143–151.
(23) Morgan, R. K.; Molnar, M. M.; Batra, H.; Summerford, B.; Wadkins, R. M.; Brooks, T. A. Effects of 5-Hydroxymethylcytosine Epigenetic Modification on the Stability and Molecular Recognition of VEGF *i*-Motif and G-Quadruplex Structures. *J. Nucleic Acids*. **2018**, 1–14.
(24) Breiling, A.; Lyko, F. Epigenetic Regulatory Functions of DNA Modifications: 5-Methylcytosine and Beyond. *Epigenetics Chromatin*. **2015**, *8*, 1–9.
(25) Xu, B.; Devi, G.; Shao, F. Regulation of Telomeric *i*-Motif Stability by 5-Methylcytosine and 5-Hydroxymethylcytosine Modification. *Org. Biomol. Chem*. **2015**, *13*, 5646–5651.
(26) Abou Assi, H.; Lin, Y. C.; Serrano, I.; González, C.; Damha, M. J. Probing Synergistic Effects of DNA Methylation and 2'-β-Fluorination on *i*-Motif Stability. *Chem. Eur. J*. **2018**, *24*, 471–477.
(27) Karabıyık, H.; Sevinçek, R.; Karabıyık, H. π-Cooperativity Effect on the Base Stacking Interactions in DNA: Is there a Novel Stabilization Factor Coupled with Base Pairing H-Bonds? *Phy. Chem. Chem. Phys*. **2014**, *16*, 15527–15538.
(28) Lieblein, A. L.; Krämer, M.; Dreuw, A.; Fürtig, B.; Schwalbe, H. The Nature of Hydrogen Bonds in Cytidine···H^+···Cytidine DNA Base Pairs. *Angew. Chem. Int. Ed.* **2012**, *51*, 4067–4070.
(29) Berger, I.; Egli, M.; Rich, A. Inter-strand CH···O Hydrogen Bonds Stabilizing Four-Stranded Intercalated Molecules: Stereoelectronic Effects of O4' in Cytosine-Rich DNA. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 12116–12121.
(30) Mitas, M. Trinucleotide Repeats Associated with Human Disease. *Nucleic Acids Res.* **1997**, *25*, 2245–2253.
(31) Yang, B.; Wu, R. R.; Rodgers, M. T. Base-Pairing Energies of Proton-Bound Homodimers Determined by Guided Ion Beam Tandem Mass Spectrometry: Application to Cytosine and 5-Substituted Cytosines. *Anal. Chem.* **2013**, *85*, 11000–11006.

(32) Yang, B.; Moehlig, A. R.; Frieler, C. E.; Rodgers, M. T. Base-Pairing Energies of Proton-Bound Dimers and Proton Affinities of 1-Methylated Cytosines: Model Systems of the Effects of the Sugar Moiety on the Stability of DNA *i*-Motif Conformations. *J. Phys. Chem. B.* **2015**, *119*, 1857–1868.
(33) Yang, B.; Wu, R. R.; Rodgers, M. T. Base-Pairing Energies of Proton-Bound Dimers and Proton Affinities of 1-Methyl-5-Halocytosines: Implications for the Effects of Halogenation on the Stability of the DNA *i*-Motif. *J. Am. Mass Spectrom*. **2015**, *26*, 1469–1482.
(34) Yang, B.; Rodgers, M.T. Base-Pairing Energies of Proton-Bound Heterodimers of Cytosine and Modified Cytosines: Implications for the Stability of DNA *i*-Motif Conformations. *J. Am. Chem. Soc.* **2014**, *136*, 282–290.
(35) Yang, B.; Rodgers, M. T. Base-Pairing Energies of Protonated Nucleoside Base Pairs of dCyd and m^5dCyd: Implications for the Stability of DNA *i*-Motif Conformations. *J. Am. Mass Spectrom*. **2015**, *26*, 1394–1403.
(36) Rodgers, M. T. Substituent Effects in the Binding of Alkali Metal Ions to Pyridines Studied by Threshold Collision-Induced Dissociation and Ab Initio Theory: The Methylpyridines. *J. Phys. Chem. A* **2001**, *105*, 2374–2383.
(37) Muntean, F.; Armentrout, P. B. Guided Ion Beam Study of Collision-Induced Dissociation Dynamics: Integral and Differential Cross Sections. *J. Chem. Phys*. **2001**, *115*, 1213–1228.
(38) Coates, R. A.; Armentrout, P. B. Thermochemical Investigations of Hydrated Nickel Dication Complexes by Threshold Collision-Induced Dissociation and Theory. *J. Phys. Chem*. *A* **2017**, *121*, 3629–3646.
(39) Teloy, E.; Gerlich, D. Integral Cross Sections for Ion-Molecule Reactions I. The Guided Ion Beam Technique. *Chem. Phys.* **1974**, *4*, 417–427.
(40) 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**.
(41) Gerlich, D. State-Selected and State-to-State Ion-Molecule Reaction Dynamics, Part I, Experiment, Ng, C.-Y., Baer, M. Eds., *Adv. Chem. Phys.* **1992**, *82*, 1–176.
(42) Dalleska, N. F.; Honma, K.; Armentrout, P. B. Stepwise Solvation Enthalpies of Protonated Water Clusters: Collision-Induced Dissociation as an Alternative to Equilibrium Studies. *J. Am. Chem. Soc.* **1993**, *115*, 12125–12131.
(43) Hales, D. A.; Armentrout, P. B. Effect of Internal Excitation on the Collision-Induced Dissociation and Reactivity of Co_2^+. *J. Cluster Sci.* **1990**, *1*, 127–142.
(44) Aristov, N.; Armentrout, P. B. Collision-Induced Dissociation of Vanadium Monoxide Ion. *J. Phys. Chem*. **1986**, *90*, 5135–5140.
(45) Dalleska, N. F.; Honma, K.; Sunderlin, L. S.; Armentrout, P. B. Solvation of Transition Metal Ions by Water. Sequential Binding Energies of $M^+(H_2O)_x$ (x = 1–4) for M = Ti to Cu Determined by Collision-Induced Dissociation. *J. Am. Chem. Soc*. **1994**, *116*, 3519–3528.
(46) Wolinski, K.; Hinton, J.; Wishart, D.; Sykes, B.; Richards, F.; Pastone, A.; Saudek, V.; Ellis, P.; Maciel, G.; McIver, J. HyperChem Computational Chemistry Software Package, Version 5.0, Hypercube *Inc*: *Gainsville*, FL, **1997**.
(47) 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**.
(48) Boys, S. F.; Bernardi, F. The Calculation of Small Molecular Interactions by the Differences of Separate Total Energies. Some Procedures with Reduced Errors. *Mol. Phys.* **1970**, *19*, 553–566.

(49) van Duijneveldt, F. B.; van Duijneveldt-van de Rijdt, J. G. C. M.; van Lenthe, J. H. State of the Art in Counterpoise Theory. *Chem. Rev.* **1994**, *94*, 1873–1885.
(50) Altona, C.; Sundaralingam, M. Conformational Analysis of the Sugar Ring in Nucleosides and Nucleotides. New Description Using the Concept of Pseudorotation. *J. Am. Chem. Soc*. **1972**, *94*, 8205-8212.
(51) Smith, S. M.; Markevitch, A. N.; Romanov, D. A.; Li, X.; Levis, R. J.; Schlegel, H. B. Static and Dynamic Polarizabilities of Conjugated Molecules and Their Cations. *J. Phys. Chem. A* **2004**, *108*, 11063–11072.
(52) Lu, T.; Chen, F. Multiwfn: A Multifunctional Wavefunction Analyzer. *J. Comput. Chem*. **2012**, *33*, 580–592.
(53) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics*. **1996**, *14*, 33–38.
(54) Contreras-García, J.; Johnson, E. R.; Keinan, S.; Chaudret, R.; Piquemal, J.-P.; Beratan, D. N.; Yang, W. NCIPLOT: A Program for Plotting Noncovalent Interaction Regions. *J. Chem. Theory Comput*. **2011**, *7*, 625–632.
(55) Contreras-García, J.; Yang, W.; Johnson E. R. Analysis of Hydrogen-Bond Interaction Potentials from the Electron Density: Integration of Noncovalent Interaction Regions, *J. Phys. Chem.* **2011**, *115*, 12983–12990.
(56) Boto, R. A.; Peccati, F.; Laplaza, R.; Quan, C.; Carbone, A.; Piquemal, J.-P.; Maday, Y.; Contreras-García, J. NCIPLOT4: Fast, Robust, and Quantitative Analysis of Noncovalent Interactions. *J. Chem. Theory Comput.* **2020**, *16*, 4150–4158.
(57) Bader, R.F.W. Atoms in Molecules: A Quantum Theory. Clarendon Press, Oxford Reed, A. E.; Curtiss L.A.; Weinhold, F. *Chem. Rev.* **1990**, *88*, 899.
(58) Ervin, K. M.; Armentrout, P. Translational Energy Dependence of $Ar^+ + XY \rightarrow ArX^+ + Y$ (XY= H_2, D_2, HD) from Thermal to 30 eV cm. *J. Chem. Phys.* **1985**, *83*, 166−189.
(59) Hales, D. A.; Lian, L.; Armentrout, P. B. Collision-Induced Dissociation of Nb_n^+ (n = 2 - 11): Bond Energies and Dissociation Pathways. *Int. J. Mass Spectrom. Ion Process*. **1990**, *102*, 269.
(60) Schultz, R. H.; Crellin, K. C.; Armentrout, P.B. Sequential Bond Energies of Iron Carbonyl $Fe(CO)_x^+$ (x = 1−5): Systematic Effects on Collision-Induced Dissociation Measurements. *J. Am. Chem. Soc.* **1991**, *113*, 8590−8601.
(61) Chantry, P. Doppler Broadening in Beam Experiments. *J. Chem. Phys.* **1971**, *55*, 2746−2759.
(62) Khan, F. A.; Clemmer, D. E.; Schultz, R. H.; Armentrout, P. B. Sequential Bond Energies of $Cr(CO)_x^+$, x = 1–6. *J. Phys. Chem.* **1993**, *97*, 7978–7987.
(63) Rodgers, M. T.; Ervin, K. M.; Armentrout, P. B. Statistical Modeling of Collision-Induced Dissociation Thresholds. *J. Chem. Phys.* **1997**, *106*, 4499–4508.
(64) Muntean, F.; Armentrout, P. B. Guided Ion Beam Study of Collision-Induced Dissociation Dynamics: Integral and Differential Cross Sections. *J. Chem. Phys.* **2001**, *115*, 1213–1228.
(65) Stein, S. E.; Rabinovitch, B. S. On the Use of Exact State Counting Methods in RRKM Rate Calculations. *Chem. Phys. Lett.* **1977**, *49*, 183–188.
(66) Beyer, T.; Swinehart, D. Algorithm 448: Number of Multiply-Restricted Partitions. *Commun. ACM*, **1973**, *16*, 379.
(67) Stein, S. E.; Rabinovitch, B. S. Accurate Evaluation of Internal Energy Level Sums and Densities Including Anharmonic Oscillators and Hindered Rotors. *J. Chem. Phys.* **1973**, *58*, 2438–2445.

(68) Wu, R. R.; Rodgers, M. T. O2 Protonation Controls Threshold Behavior for N-Glycosidic Bond Cleavage of Protonated Cytosine Nucleosides. *J. Phys. Chem. B* **2016**, *120*, 4803–4811.
(69) Yang, B.; Wu, R. R.; Berden, G.; Oomens, J.; Rodgers, M. T. Infrared Multiple Photon Dissociation Action Spectroscopy of Proton-Bound Dimers of Cytosine and Modified Cytosines: Effects of Modifications on Base-Pairing Interactions. *J. Phys. Chem. B* **2013**, *117*, 14191–14201.
(70) Wu, R. R.; Yang, B.; Berden, G.; Oomens, J.; Rodgers, M. T. N3 and O2 Protonated Tautomeric Conformations of 2'-Deoxycytidine and Cytidine Coexist in the Gas Phase. *J. Phys. Chem. B.* **2015**, *119*, 5773–5784.
(71) He, C. C.; Hamlow, L. A.; Devereaux, Z. J.; Zhu, Y.; Nei, Y.-w.; Fan, L.; McNary, C.; Maitre, P.; Steinmetz, V.; Schindler, B. et al. Structural and Energetic Effects of 2′-Ribose Methylation of Protonated Purine Nucleosides", *J. Phys. Chem. B* **2018**, *122*, 9147–9160.
(72) Devereaux, Z. J.; Roy, H. A.; He, C. C.; Zhu, Y.; Cunningham, N. A; Hamlow, L. A.; Berden, G.; Oomens, J.; Rodgers, M. T. Influence of 2′-fluoro modification on glycosidic bond stabilities and gas-phase ion structures of protonated pyrimidine nucleosides. *J. Fluorine Chem.* **2019**, *219*, 10–22.
(73) Hamlow, L. A.; He, C. C.; Devereaux, Z. J.; Roy, H. A.; Cunningham, N.; Soley, E. O.; Berden, G.; Oomens, J.; Rodgers, M. T. Gas-Phase Structures of Protonated Arabino Nucleosides. *Int. J. Mass Spectrom.* **2019**, *438*, 124–134.
(74) Hamlow, L. A; Devereaux, Z. J.; Roy, H. A.; Cunningham, N.; Berden, G.; Oomens, J.; Rodgers, M. T. Impact of the 2′- and 3′-Sugar Hydroxyl Moieties on Gas-Phase Nucleoside Structure. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 832–845.
(75) Rannulu, N. S.; Rodgers, M. T. Noncovalent Interactions of Cu^+ with N-Donor Ligands (Pyridine, 4,4-Dipyridyl, 2,2-Dipyridyl, and 1,10-Phenanthroline): Collision-Induced Dissociation and Theoretical Studies. *J. Phys. Chem. A* **2007**, *111*, 3465–3479.
(76) Hallowita, N.; Carl, D. R.; Armentrout, P. B; Rodgers, M. T. Dipole Effects on Cation-π Interactions: Absolute Bond Dissociation Energies of Complexes of Alkali Metal Cations to N-Methylaniline and N,N-Dimethylaniline. *J. Phys. Chem. A* **2008**, *112*, 7996–8008.
(77) Vitale, G.; Valina, A. B.; Huang, H.; Amunugama, R.; Rodgers, M. T. Solvation of Copper Ions by Acetonitrile. Structures and Sequential Binding Energies of $Cu^+(CH_3CN)_x$, x =1–5 from Collision-Induced Dissociation and Theoretical Studies. *J. Phys. Chem. A* **2001**, *105*, 11351–11364.
(78) Ruan, C.; Yang, Z.; Rodgers, M. T. Influence of the d Orbital Occupation on the Nature and Strength of Copper Cation–π Interactions: Threshold Collision-Induced Dissociation and Theoretical Studies. *Phys. Chem. Chem. Phys.* **2007**, *9*, 5902–5918.
(79) Rannulu, N. S.; Amunugama, R.; Yang, Z.; Rodgers, M. T. Influence of s and d Orbital Occupation on the Binding of Metal Ions to Imidazole. *J. Phys. Chem. A* **2004**, *108*, 6385–6396.
(80) Ruan, C.; Rodgers, M. T. Cation-π Interactions: Structures and Energetics of Complexation of Na^+ and K^+ with the Aromatic Amino Acids, Phenylalanine, Tyrosine and Tryptophan. *J. Am. Chem. Soc.* **2004**, *126*, 14600–14610.
(81) Huang, H.; Rodgers, M. T. Sigma versus Pi Interactions in Alkali Metal Ion Binding Affinities of Azoles: Threshold Collision-Induced Dissociation and Ab Initio Theory Studies. *J. Phys. Chem. A* **2002**, *106*, 4277–4289.

(82) Rodgers, M. T.; Armentrout, P. B. Absolute Alkali Metal Ion Binding Affinities of Several Azoles Determined by Threshold Collision-Induced Dissociation. *Int. J. Mass Spectrom.* **1999**, *187*, 359–380.
(83) Ruan, C.; Yang, Z.; Hallowita, N.; Rodgers, M. T. Cation-π Interactions with a Model for the Side Chain of Tryptophan: Structures and Absolute Binding Energies of Alkali Metal Cation-Indole Complexes. *J. Phys. Chem. A* **2005**, *109*, 11539–11550.
(84) Lacroix, L.; Mergny, J.-L.; Leroy, J.-L.; Hélène, C. Inability of RNA to Form the *i*-Motif: Implications for Triplex Formation. *Biochem*, **1996**, *35*, 8715–8722.
(85) Poeta, L.; Drongitis, D.; Verrillo, L.; Miano, M. G. DNA Hypermethylation and Unstable Repeat Diseases: A Paradigm of Transcriptional Silencing to Decipher the Basis of Pathogenic Mechanisms. *Genes*, **2020**, *11*, 684.
(86) Valinluck, V.; Wu, W.; Liu, P.; Neidigh, J. W.; Sowers, L. C. Impact of Cytosine 5-Halogens on the Interaction of DNA with Restriction Endonucleases and Methyltransferase. *Chem. Res. Toxicol*. **2006**, *19,* 556–62.
(87) Školáková, P.; Badri, Z.; Foldynová-Trantírková, S.; Ryneš, J.; Šponer, J.; Fojtová, M.; Fajkus, J.; Marek, R.; Vorlíčková, M.; Mergny, J.-L. Composite 5-Methylations of Cytosines Modulate *i*-Motif Stability in a Sequence-Specific Manner: Implications for DNA Nanotechnology and Epigenetic Regulation of Plant Telomeric DNA. *Biochim. Biophys. Acta Gen. Subj*. **2020**, *1864*, 129651.
(88) Bhavsar-Jog, Y. P.; Van Dornshuld, E.; Brooks, T. A.; Tschumper, G. S.; Wadkins, R. M. Yogini, P.; Eric, B-J.; Dornshuld, V.; Brooks, T. A.; Tschumper, J. S. et al. Epigenetic Modification, Dehydration, and Molecular Crowding Effects on the Thermodynamics of *i*-Motif Structure Formation from C-Rich DNA. *Biochem*. **2014**, *53*, 1586–1594.

Table 1. Threshold Dissociation Energies at 0 K, Entropies of Activation at 1000 K, Fitting Parameters of Equations 2 and 3, and Kinetic Shifts of Protonated Cytidine Nucleoside Analogue Base Pairs.[a]

(*x*Cyd)H$^+$(*x*Cyd)	E_0(PSL)[b] (eV)	$\Delta S^{\ddagger}$(PSL)[b] (J·mol^{-1}K^{-1})	σ[b]	n[b]	E_0[c] (eV)	Kinetic Shift (eV)
(Cyd)H$^+$(Cyd)	1.73 (0.05)	84 (3)	118.8 (3.2)	1.0 (0.11)	3.30 (0.11)	1.57
(dCyd)H$^+$(dCyd)	1.74 (0.05)	91 (4)	37.4 (2.6)	1.3 (0.06)	3.15 (0.09)	1.41
(ddCyd)H$^+$(ddCyd)	1.75 (0.07)	94 (4)	62.4 (3.8)	1.2 (0.09)	3.13 (0.08)	1.38
(m^{5}Cyd)H$^+$(m^{5}Cyd)	1.77 (0.05)	89 (3)	80.0 (1.7)	0.72 (0.02)	3.66 (0.08)	1.89
(m^{5}dCyd)H$^+$(m^{5}dCyd)	1.83 (0.06)	91 (4)	53.5 (2.1)	0.70 (0.03)	3.76 (0.09)	1.93

[a] Present results, uncertainties are listed in parentheses. [b] Average values obtained based on fits using a loose PSL TS. [c]No RRKM lifetime analysis included.

Table 2. Base-Pairing Energies of Protonated Cytidine Nucleoside Analogue Base Pairs at 0 K in kJ/mol[a].

System	TCID	Level of Theory[b]			
		B3LYP[c]	B3LYP[d]	B3P86[e]	M06-2X[f]
(Cyd)H$^+$(Cyd)	166.9 (4.5)	166.4	166.8	177.3	179.1
(dCyd)H$^+$(dCyd)	167.7 (5.3)	168.7	167.4	177.7	179.2
	159.8 (5.2)[g]	163.7[g]			
(ddCyd)H$^+$(ddCyd)	168.9 (5.3)	168.4	166.2	176.9	178.4
(m^{5}Cyd)H$^+$(m^{5}Cyd)	170.8 (4.7)	169.8	171.6	184.0	184.6
(m^{5}dCyd)H$^+$(m^{5}dCyd)	176.6 (6.0)	176.2	176.5	188.0	191.2
	162.0 (5.7)[g]	166.4[g]			
(Cyt)H$^+$(Cyt)	169.9 (4.6)[h]	168.9[h]			
(m^{5}Cyt)H$^+$(m^{5}Cyt)	177.4 (5.3)[h]	173.4[h]			
AEU/MAD[i]	5.2 (0.6)	0.7 (0.3)	1.4 (1.7)	10.6 (1.9)	12.3 (2.0)

[a]Present results except as noted, uncertainties are listed in parentheses. [b]All theoretical BPEs are computed based on the ground conformations of structures optimized at the B3LYP/6-311+G(d,p) level of theory including ZPE and BSSE corrections. [c]Calculated at the B3LYP/6-311+G(2d,2p) level of theory. [d]Calculated at the B3LYP/def2-TVZPPD level of theory. [e]Calculated at the B3P86/6-311+G(2d,2p) level of theory. [f]Calculated at the M062X/6-311+G(2d,2p) level of theory. [g]Values taken from reference 28. [h]Values taken from reference 24.

[i]Average experimental uncertainty (AEU) in the BPEs measured in the present study. Mean absolute deviation (MAD) between the measured and calculated BPEs determined in the present study.

Table 3. Enthalpies and Gibbs Energies of Base-Pairing of Protonated Cytidine Nucleoside Analogue Base Pairs at 298 K in kJ/mol[a]

System	ΔH_0	ΔH_0[b]	$\Delta H_{298} - \Delta H_0$[b]	ΔH_{298}	ΔH_{298}[b]	$T\Delta S_{298}$[b]	ΔG_{298}	ΔG_{298}[b]
(Cyd)H^+(Cyd)	166.9 (4.5)	166.4	0.3 (0.1)	167.2 (4.5)	166.7	51.5 (1.7)	115.7 (4.8)	115.2
(dCyd)H^+(dCyd)	167.7 (5.3)	168.7	0.2 (0.1)	167.9 (5.3)	168.9	52.7 (1.5)	115.2 (5.5)	116.2
(ddCyd)H^+(ddCyd)	168.9 (5.3)	168.4	0.1 (0.1)	169.0 (5.3)	168.5	53.4 (1.5)	115.6 (5.5)	115.1
(m^5Cyd)H^+(m^5Cyd)	170.8 (4.7)	169.8	0.1 (0.1)	170.9 (4.7)	169.9	52.9 (1.7)	118.0 (5.0)	117.0
(m^5dCyd)H^+(m^5dCyd)	176.6 (6.0)	176.2	0.3 (0.1)	176.9 (6.0)	176.5	52.5 (1.4)	124.4 (6.2)	124.0

[a]Present results, uncertainties are listed in parentheses. [b]Density functional theory calculations at the B3LYP/6-311+G(2d,2p)// B3LYP/6-311+G(d,p) level of theory with frequencies scaled by 0.9804.

FIGURE CAPTIONS

Figure 1. Cross sections for CID of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs with Xe as a function of collision energy in the center-of-mass frame (lower *x*-axis) and laboratory frame (upper *x*-axis), parts a and b. Data are shown for a Xe pressure of ~0.2 mTorr.

Figure 2. Zero-pressure-extrapolated cross sections for CID of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs 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), parts a and b. The solid lines show the best fits to the data using the model of eq 3 convoluted over the neutral and ion kinetic and internal energy distributions. The dotted lines show the model cross sections in the absence of experimental kinetic energy broadening for the base pairs with an internal temperature of 0 K.

Figure 3. B3LYP/6-311+G(d,p) optimized geometries of the ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs, where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The base pairs are oriented such that the protonated nucleoside is shown on the left. The nucleobase orientation, 5-hydroxy orientation, and sugar puckering of the protonated nucleoside is indicated first in boldface font followed by those of the neutral nucleoside in standard font. The hydrogen-bond lengths are also shown

Figure 4. Pseudorotation phase angles (*P*), glycosidic bond angles, and 5-hydroxy orientations of the ground and low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs. Conformers in which the values are the same for the protonated and neutral nucleosides of the base pair are indicated with solid symbols. Conformers in which the values differ for the protonated and neutral nucleosides of the base pair are indicated with open symbols; the protonated nucleoside is differentiated from the neutral nucleoside with a dot in the center of the symbol.

Figure 5. Noncovalent interaction maps superimposed on the B3LYP/6-311+G(d,p) optimized geometries of the ground conformers of the (dCyd)H^+(dCyd) and (m^5dCyd)H^+(m^5dCyd) base pairs. The NCI maps are shown at an isosurface of 0.5 a.u. of the B3LYP/6-311+G(2d,2p) reduced electron density gradients. Regions exhibiting strong attractive interactions such as ionic or strong neutral hydrogen-bonding interactions appear blue. Weaker attractive interactions such as London dispersion or noncanonical hydrogen-bonding interactions appear green. Highly repulsive interactions appear red.

Figure 6. B3LYP calculated vs. TCID measured BPEs of the (*x*Cyd)H^+(*x*Cyd) base pairs at 0 K (in kJ/mol), where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. Results for the 6-311+G(2d,2p) and def2-TZVPPD basis sets are compared. Calculated values include ZPE and BSSE corrections. The diagonal line indicates values for which calculated and measured BDEs are in perfect agreement.

Figure 7. TCID measured BPEs of the (*x*Cyd)H^+(*x*Cyd) protonated nucleoside base pairs at 0 K (in kJ/mol) versus the PBE0/6-311+G(2d,2p) computed isotropic molecular polarizability volumes (α in $Å^3$) of the neutral *x*Cyd nucleosides: Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The line is a linear regression fit to the data.

Figure 1.

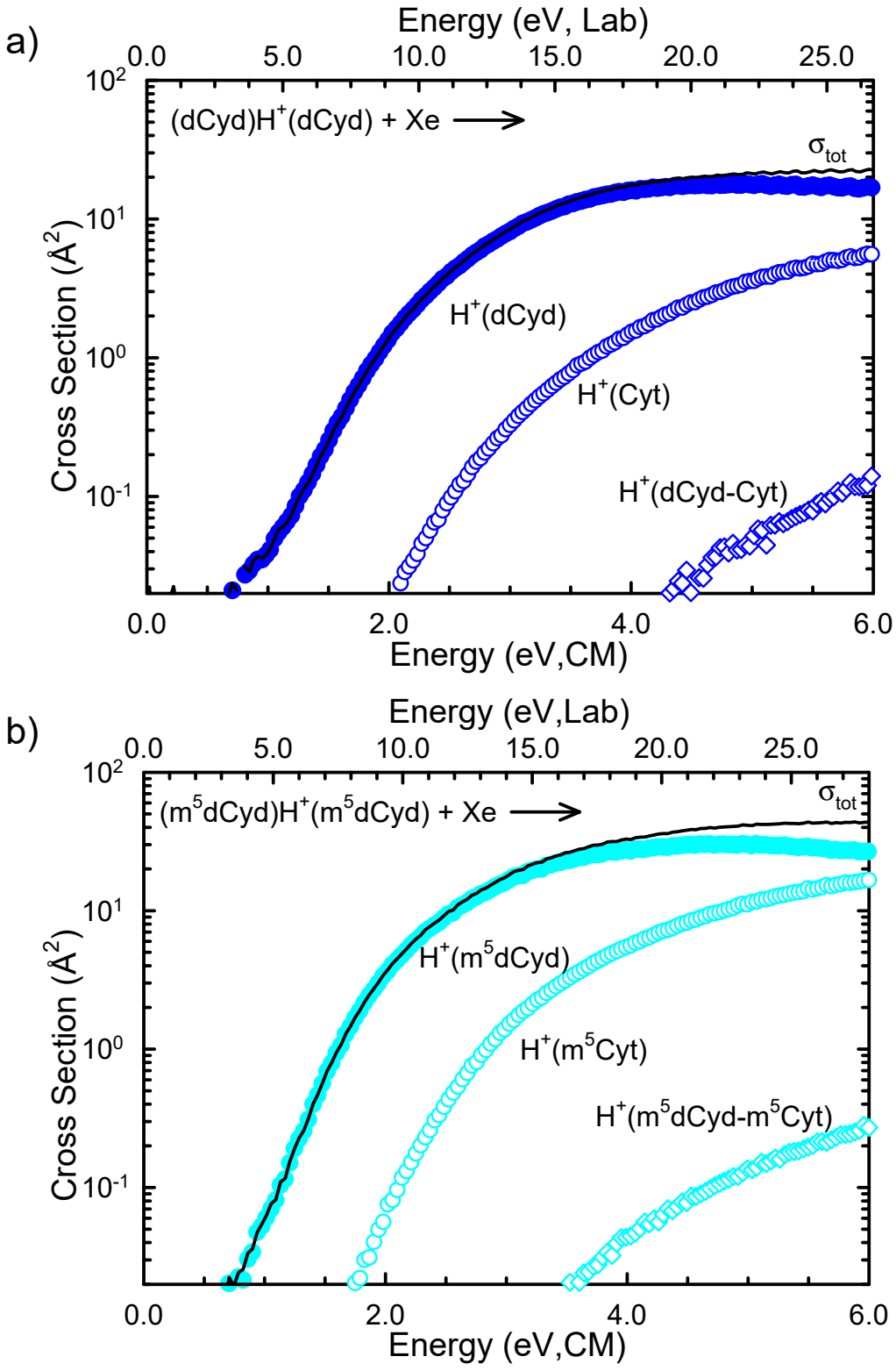

a)
Energy (eV, Lab)
0.0
5.0
10.0
15.0
20.0
25.0
10^2
10^1
10^0
10^{-1}
Cross Section (Å^2)
(dCyd)H$^+$(dCyd) + Xe
σ_{tot}
H$^+$(dCyd)
H$^+$(Cyt)
H$^+$(dCyd-Cyt)
0.0
2.0
4.0
6.0
Energy (eV,CM)
b)
Energy (eV,Lab)
0.0
5.0
10.0
15.0
20.0
25.0
10^2
10^1
10^0
10^{-1}
Cross Section (Å^2)
(m^{5}dCyd)H$^+$(m^{5}dCyd) + Xe
σ_{tot}
H$^+$(m^{5}dCyd)
H$^+$(m^{5}Cyt)
H$^+$(m^{5}dCyd-m^{5}Cyt)
0.0
2.0
4.0
6.0
Energy (eV,CM)

Figure 2.

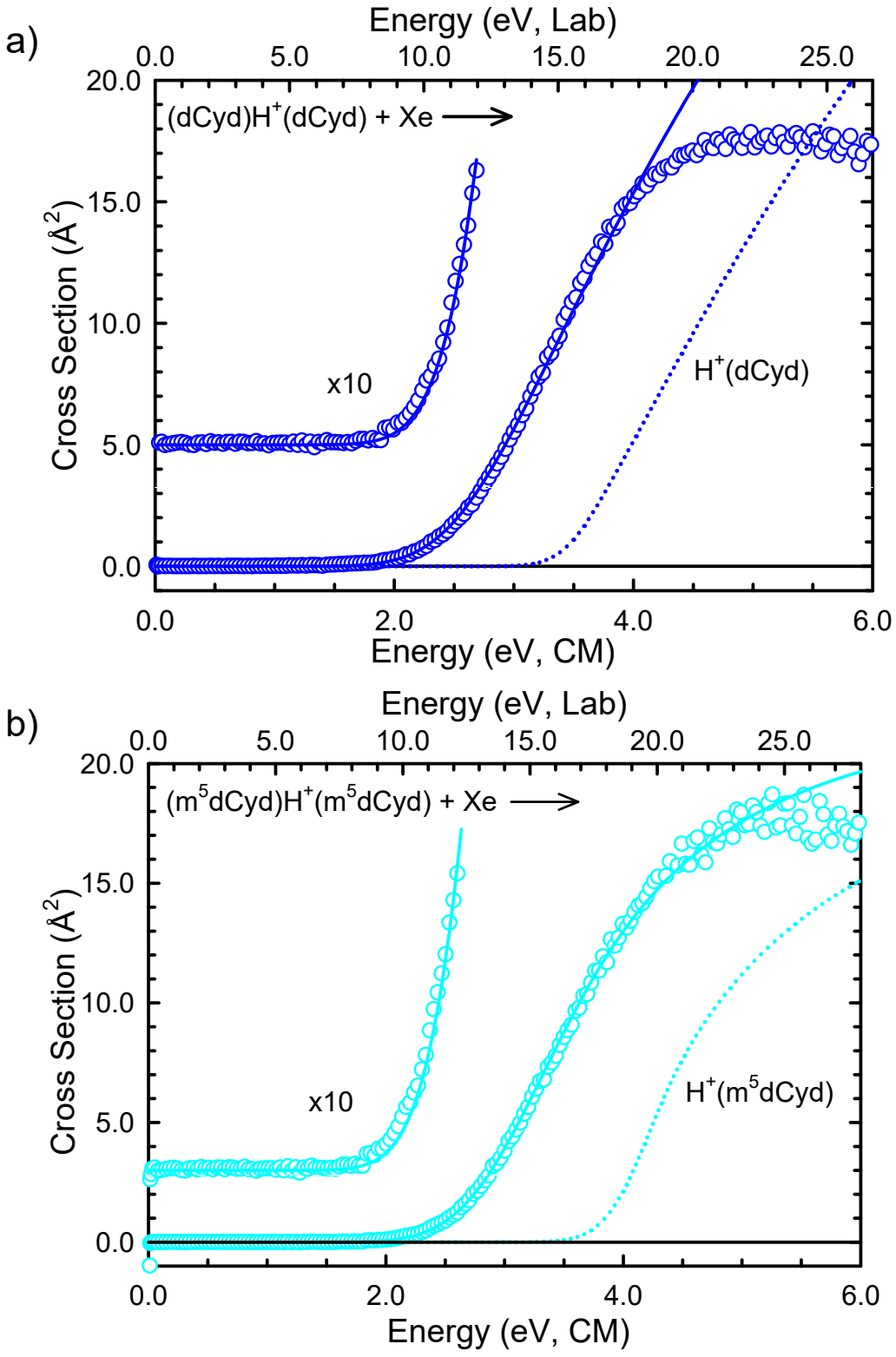
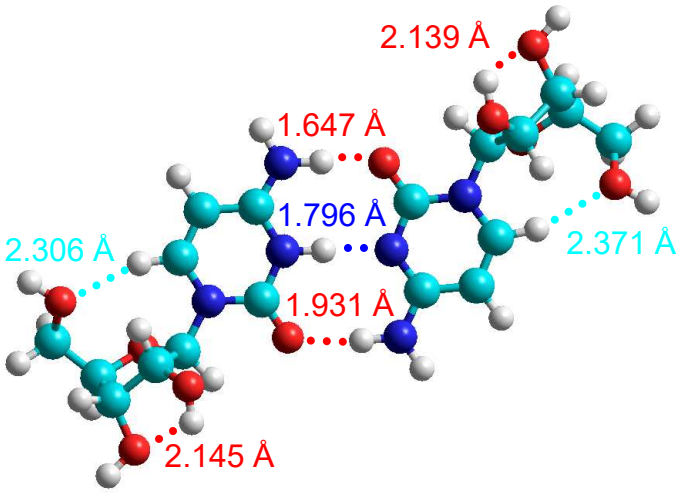

a)
Energy (eV, Lab)
0.0
5.0
10.0
15.0
20.0
25.0
(dCyd)H+(dCyd) + Xe
x10
H+(dCyd)
Cross Section (Å2)
20.0
15.0
10.0
5.0
0.0
0.0
2.0
4.0
6.0
Energy (eV, CM)
b)
Energy (eV, Lab)
(m5dCyd)H+(m5dCyd) + Xe
x10
H+(m5dCyd)
Cross Section (Å2)
Energy (eV, CM)

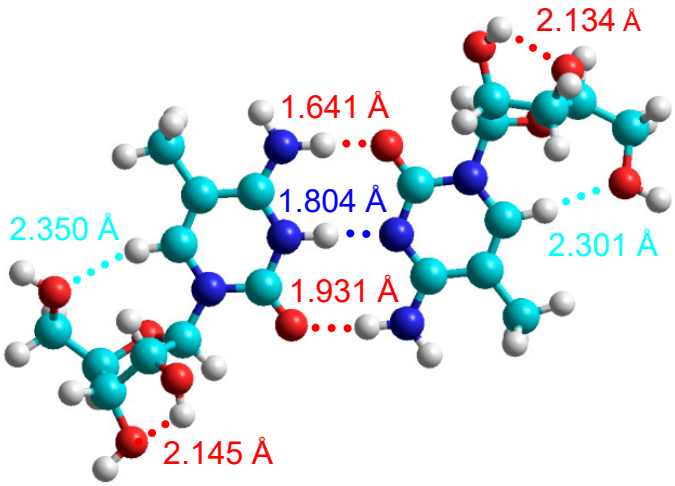
Figure 3.



(Cyd)H$^+$(Cyd)

***anti*, gauche$^+$·*anti*, gauche$^+$**

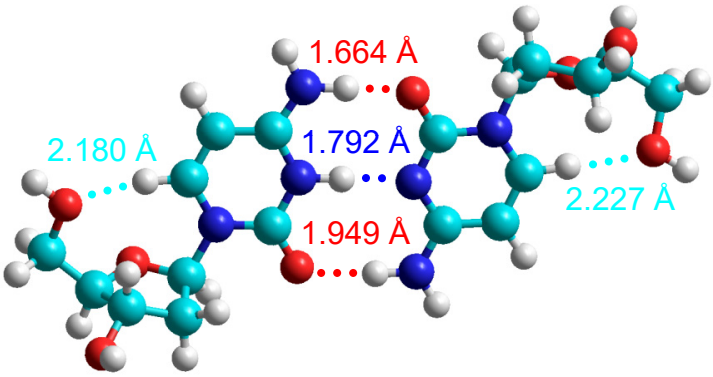
C2′-endo·C2′-endo (**$^3T_2$**·$^3T_2$)



(m^{5}Cyd)H$^+$(m^{5}Cyd)

***anti*, gauche$^+$·*anti*, gauche$^+$**

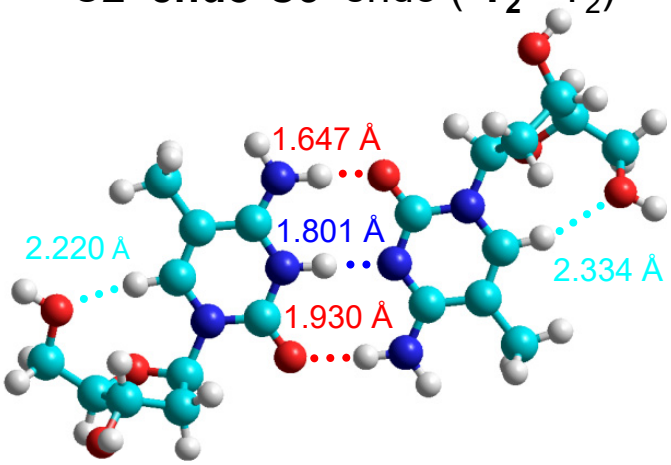
C2′-endo·C3′-endo (**$^3T_2$**·$^3T_2$)



(dCyd)H$^+$(dCyd)

***anti*, gauche$^+$·*anti*, gauche$^+$**

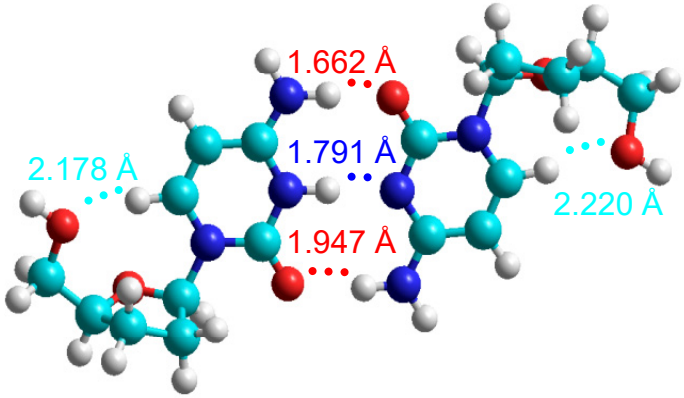
C3′-endo·C3′-endo (**$^3T_2$**·$^3T_2$)



(m^{5}dCyd)H$^+$(m^{5}dCyd)

***anti*, gauche$^+$·*anti*, gauche$^+$**

C3′-endo·C2′-endo (**$^3T_2$**·$^3T_2$)



(ddCyd)H$^+$(ddCyd)

***anti*, gauche$^+$·*anti*, gauche$^+$**

C3′-endo·C3′-endo ($^3T_2$·$^3T_2$)

Figure 4.

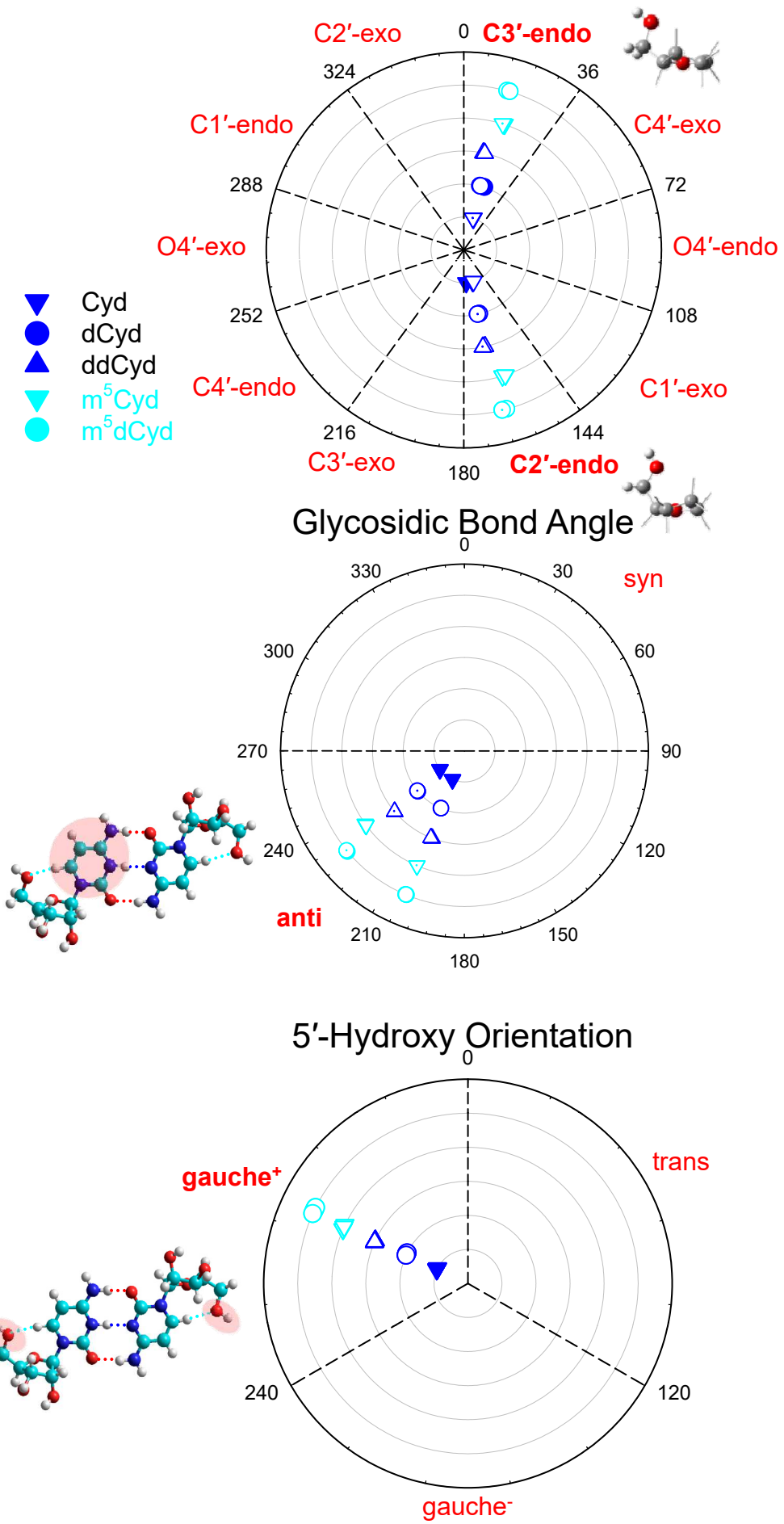
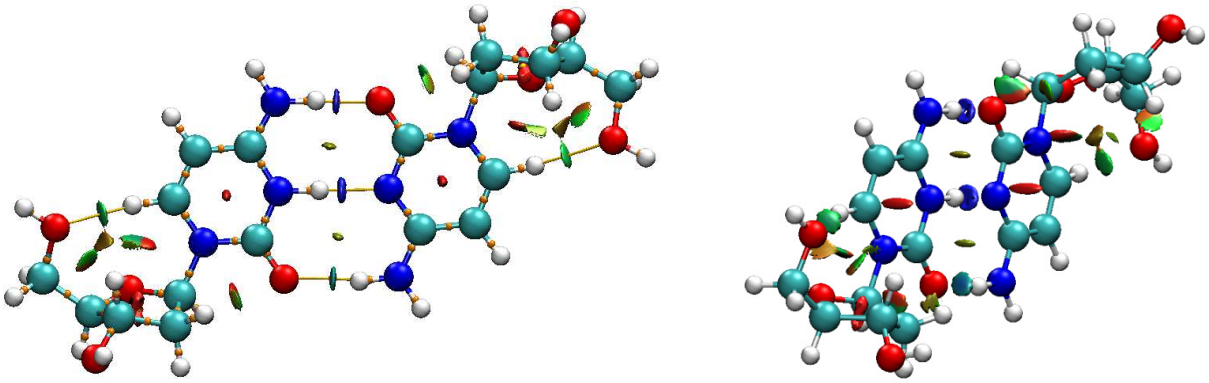
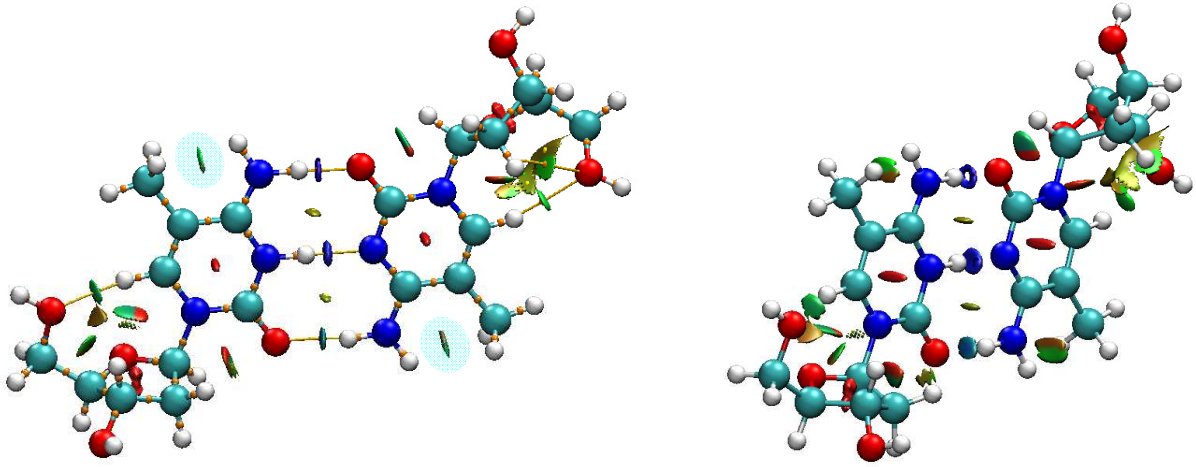

Pseudorotation Phase Angle (P)
C2′-exo
0
C3′-endo
36
C4′-exo
72
O4′-endo
108
C1′-exo
144
C2′-endo
180
C3′-exo
216
C4′-endo
252
O4′-exo
288
C1′-endo
324
Cyd
dCyd
ddCyd
m5Cyd
m5dCyd
Glycosidic Bond Angle
0
30
syn
60
90
120
150
180
anti
210
240
270
300
330
5′-Hydroxy Orientation
0
trans
120
gauche-
240
gauche+

Figure 5.



(dCyd)H$^+$(dCyd)



(m^{5}dCyd)H$^+$(m^{5}dCyd)

Figure 6.

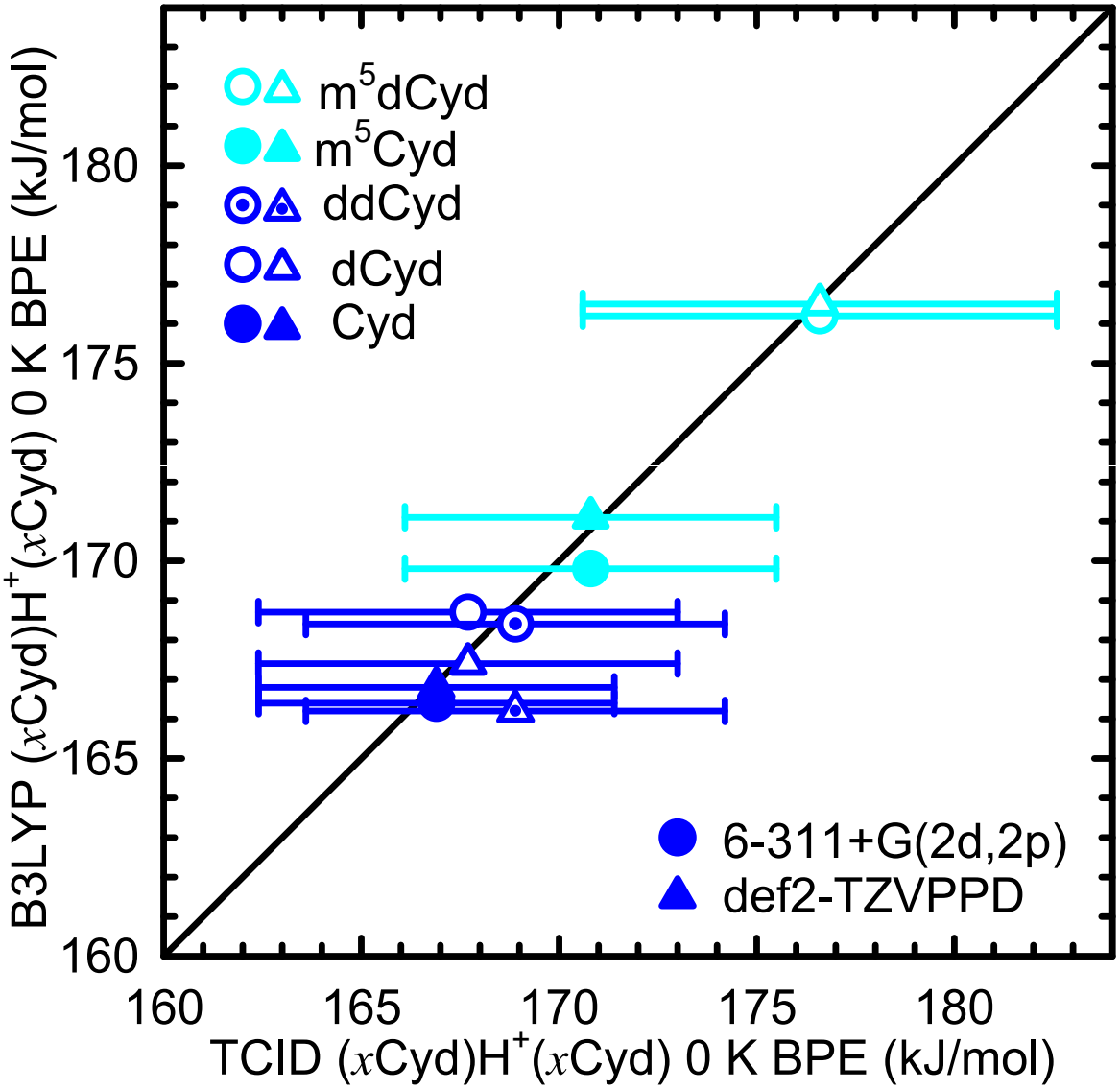
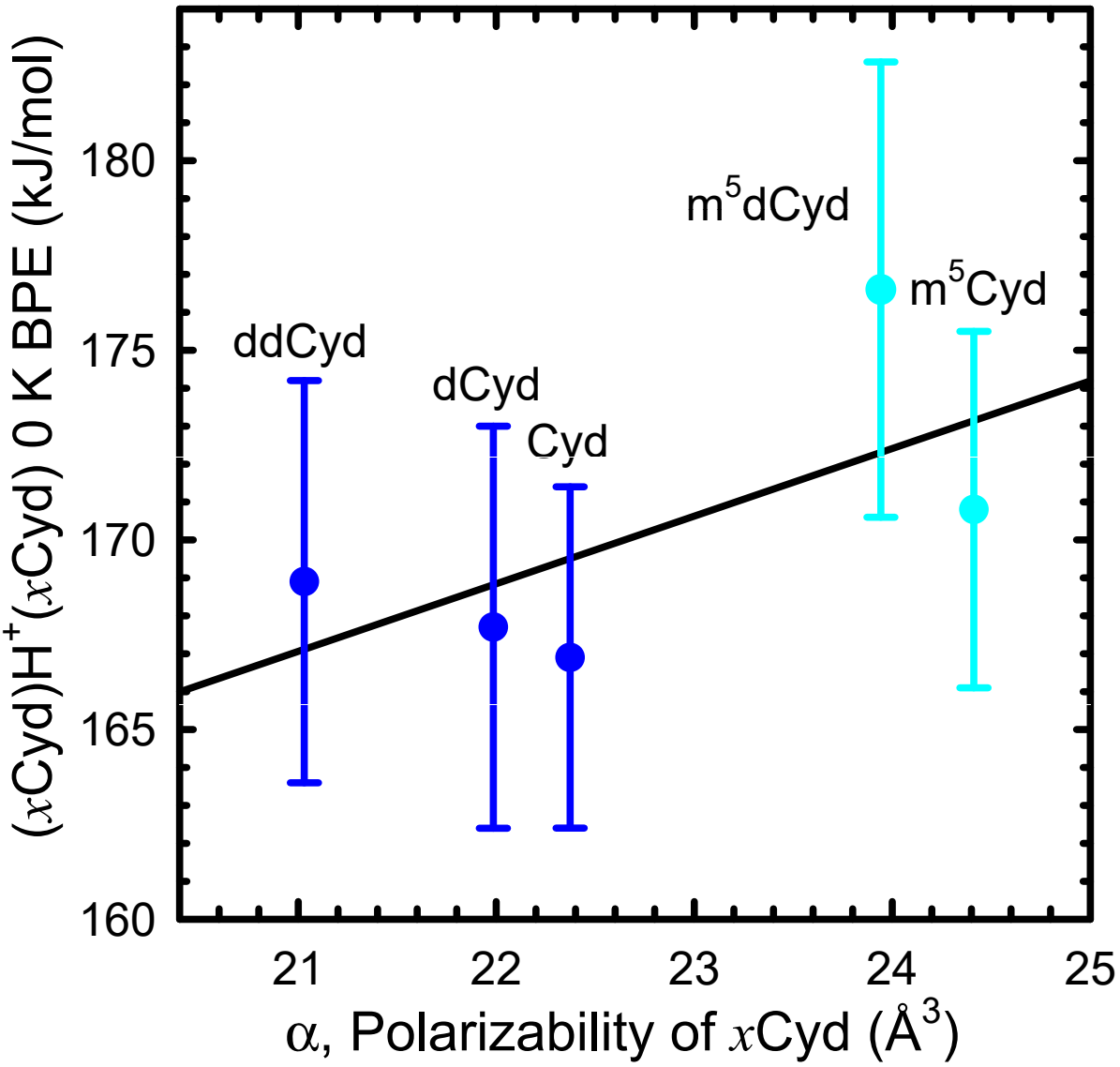
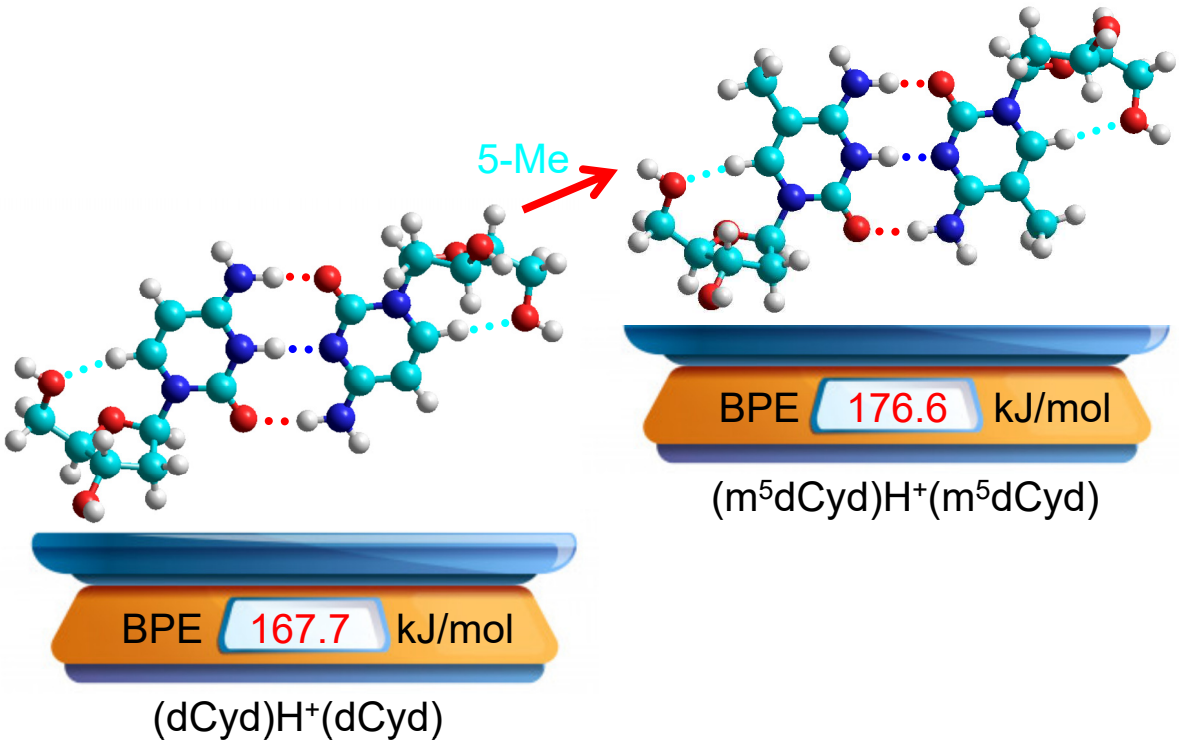

m5dCyd
m5Cyd
ddCyd
dCyd
Cyd
6-311+G(2d,2p)
def2-TZVPPD
B3LYP (xCyd)H+(xCyd) 0 K BPE (kJ/mol)
TCID (xCyd)H+(xCyd) 0 K BPE (kJ/mol)
160
165
170
175
180

Figure 7.


(xCyd)H+ (xCyd) 0 K BPE (kJ/mol)
180
175
170
165
160
21
22
23
24
25
α, Polarizability of xCyd (Å3)
ddCyd
dCyd
Cyd
m5dCyd
m5Cyd

TOC Graphic



Influence of 5-Methylation and the 2′- and 3′-Hydroxy Substituents on the Base Pairing Energies of Protonated DNA and RNA Cytidine Nucleoside Base Pairs: Implications for the Stabilities of *i*-Motif Structures

Yakubu S. Seidu, H.A. Roy, and M. T. Rodgers*
Department of Chemistry, Wayne State University, Detroit, MI, 48202, USA

Table S1. Vibrational Frequencies and Average Vibrational Energies of the (*x*Cyd)H$^+$(*x*Cyd) Base Pairs, N3-Protonated H$^+$(*x*Cyd) and Neutral *x*Cyd Nucleosides. [a]

Species	E_{int} (eV)[b]	Vibrational Frequencies (cm^{-1})[c]
(Cyd)H$^+$(Cyd)	0.88 (0.07)	11, 16, 27, 37, 49, 59, 66, 67, 85, 88, 93, 101, 125, 129, 139, 151, 161, 174, 181, 198, 212, 217, 229, 236, 246, 260, 277, 278, 300, 309, 328, 329, 341, 347, 351, 352, 403, 406, 414, 420, 431, 462, 468, 471, 479, 533, 537, 539, 549, 559, 585, 603, 615, 624, 649, 657, 696, 699, 708, 710, 738, 742, 765, 785, 787, 796, 798, 804, 869, 872, 873, 876, 904, 906, 925, 928, 981, 983, 1007, 1019, 1024, 1032, 1033, 1044, 1048, 1050, 1054, 1083, 1084, 1102, 1103, 1106, 1108, 1122, 1123, 1136, 1139, 1150, 1162, 1194, 1195, 1202, 1213, 1215, 1216, 1249, 1252, 1260, 1263, 1268, 1274, 1289, 1291, 1310, 1315, 1322, 1323, 1341, 1343, 1365, 1366, 1383, 1384, 1404, 1419, 1420, 1421, 1436, 1445, 1446, 1451, 1460, 1501, 1504, 1505, 1514, 1559, 1571, 1633, 1646, 1670, 1694, 1703, 1729, 1790, 2765, 2945, 3006, 3011, 3047, 3051, 3055, 3056, 3075, 3076, 3092, 3094, 3111, 3116, 3183, 3194, 3219, 3231, 3452, 3651, 3690, 3757, 3760, 3843, 3845, 3853, 3856
(dCyd)H$^+$(dCyd)	0.82 (0.07)	11, 20, 25, 43, 51, 53, 57, 68, 82, 85, 104, 108, 116, 126, 130, 146, 175, 181, 192, 200, 217, 220, 229, 232, 239, 243, 268, 274, 280, 281, 285, 299, 341, 344, 402, 409, 413, 425, 433, 442, 450, 479, 484, 527, 537, 539, 572, 582, 587, 601, 621, 632, 663, 665, 692, 711, 729, 736, 744, 749, 755, 773, 791, 796, 798, 803, 806, 842, 843, 881, 885, 919, 960(2), 974, 976, 1009, 1023, 1027, 1028, 1030, 1036, 1038, 1051, 1052, 1064, 1065, 1088(2), 1102, 1105, 1118, 1120, 1131, 1132, 1148, 1154, 1181, 1185, 1195(2), 1210, 1212, 1247, 1249, 1256, 1266, 1284, 1285, 1290, 1291, 1315, 1317, 1321, 1323, 1351(2), 1359, 1360, 1396, 1407, 1408, 1412, 1447, 1448, 1449, 1452, 1454, 1456, 1494(2), 1497, 1501, 1502, 1507, 1560, 1573, 1624, 1635, 1660, 1689, 1697, 1725, 1780, 2760, 2975, 3015, 3020, 3030, 3037, 3048, 3049, 3061, 3065, 3068, 3070, 3112, 3115, 3126, 3127, 3187, 3200, 3219, 3231, 3461, 3653, 3691, 3847, 3848, 3849, 3850
(ddCyd)H$^+$(ddCyd)	0.75 (0.07)	13, 25, 29, 49, 55, 62, 69, 72, 85, 93, 112, 127, 131, 143, 149, 157, 180, 192, 198, 220, 229, 251, 266, 269, 283, 288, 291, 301, 318, 322, 400, 404, 406, 410, 423, 454, 464, 527, 535, 538, 573, 584, 587, 599, 619, 626, 654, 657, 693, 709, 712, 728, 737, 757, 763, 775, 792, 797, 798, 803, 806, 832, 836, 858, 863, 915 (2), 921, 941, 942, 974, 975, 1010, 1013, 1014, 1025, 1026, 1037, 1039, 1055, 1056, 1078, 1080, 1094, 1095, 1112, 1115, 1125, 1126, 1147, 1153, 1181, 1183, 1194, 1196, 1215, 1216, 1237, 1238, 1256, 1265, 1282 (2), 1309 (2), 1317, 1321, 1329, 1331, 1344, 1345, 1371, 1373, 1395, 1409, 1416, 1418, 1447, 1449, 1450, 1456, 1492, 1493, 1497, 1502 (2), 1507, 1512, 1513, 1561, 1573, 1624, 1634, 1660, 1689, 1696, 1725, 1779, 2758, 2967, 3004, 3011, 3018, 3028, 3051, 3056, 3061, 3064, 3069 (2), 3112, 3114, 3116, 3117, 3128, 3129, 3191, 3206, 3219, 3230, 3460, 3653, 3692, 3847, 384
(m^{5}Cyd)H$^+$(m^{5}Cyd)	0.97 (0.08)	11, 15, 23, 38, 43, 46, 54, 60, 68, 85, 88, 90, 92, 111, 116, 125, 129, 134, 162, 173, 183, 187, 197, 205, 213, 216, 226, 234, 243, 246, 262, 274, 279, 280, 287, 302, 312, 321, 331, 336, 349, 351, 356, 404, 408, 421, 423, 437, 461, 468, 474, 483, 504, 516, 539, 540, 563, 573, 606, 617, 629, 645, 660, 684, 713, 726, 735, 736, 739, 748, 761, 771, 779, 788, 792, 799, 858, 874, 883, 891, 917, 923, 934, 946, 955, 977, 980, 986, 1002, 1024, 1027, 1032, 1034, 1047, 1056, 1060, 1072, 1075, 1081, 1085, 1098, 1103, 1107, 1108, 1126, 1127, 1138, 1141, 1153, 1161, 1191, 1195, 1206, 1216, 1228, 1234, 1251, 1252, 1264, 1272, 1276, 1284, 1290, 1303, 1311, 1320, 1323, 1324, 1343, 1344, 1366, 1372, 1375, 1384, 1386, 1397, 1406, 1420, 1422, 1426, 1435, 1436, 1446, 1450, 1454, 1486, 1488, 1491, 1493, 1503, 1504, 1506, 1510, 1547, 1563, 1620, 1660, 1673, 1696, 1727, 1786, 2788, 2924, 3011, 3015, 3021, 3035, 3040, 3054, 3055, 3059, 3060, 3070, 3074, 3076, 3092, 3109, 3117, 3123, 3124, 3182, 3202, 3447, 3656, 3697, 3751, 3757, 3844, 3848, 3851, 3854

Table S1. (continued) Vibrational Frequencies and Average Vibrational Energies of the (*x*Cyd)H$^+$(*x*Cyd) Base Pairs, N3-Protonated H$^+$(*x*Cyd) and Neutral *x*Cyd Nucleosides. [a]

Species	E_{int} (eV)[b]	Vibrational Frequencies (cm^{-1})[c]
(m^{5}dCyd)H$^+$(m^{5}dCyd)	0.91 (0.08)	11, 16, 23, 39, 42, 47, 60, 62, 74, 83, 92, 99, 106, 113, 123, 129, 138, 141, 173, 178, 186, 196, 213, 228, 236, 237, 260, 267, 274, 277, 279, 296, 303, 316, 330, 342, 353, 355, 378, 408, 417, 423, 424, 442, 446, 479, 489, 501, 516, 532, 549, 569, 586, 617, 631, 652, 665, 673, 688, 729, 733, 747, 748, 749, 768, 772, 784, 788, 795, 838, 867, 875, 884, 887, 889, 925, 932, 960, 962, 975, 981, 990, 991, 1001, 1020, 1028, 1030, 1053, 1056, 1072, 1073, 1076, 1088, 1091, 1103, 1104, 1112, 1119, 1120, 1133, 1149, 1160, 1188, 1200, 1205, 1209, 1225, 1233, 1236, 1247, 1264, 1272, 1274, 1286, 1290, 1313, 1314, 1320, 1323, 1328, 1339, 1351, 1361, 1364, 1387, 1389, 1390, 1407, 1419, 1423, 1425, 1442, 1446, 1448, 1449, 1452, 1480, 1487, 1488, 1492, 1493, 1503, 1504, 1505, 1512, 1547, 1564, 1623, 1653, 1674, 1691, 1700, 1724, 1776, 2776, 2944, 3004, 3015, 3019, 3020, 3033, 3036, 3046, 3047, 3051, 3061, 3064, 3068, 3081, 3110, 3115, 3125, 3125, 3141, 3189, 3444, 3658, 3696, 3831, 3846, 3847
H$^+$(Cyd)	0.41 (0.04)	40, 51, 81, 92, 126, 146, 170, 194, 220, 245, 257, 282, 289, 301, 314, 342, 362, 388, 422, 443, 471, 473, 544, 562, 583, 610, 637, 658, 704, 707, 748, 769, 789, 798, 832, 859, 884, 953, 998, 1007, 1027, 1042, 1059, 1066, 1093, 1103, 1110, 1127, 1145, 1188, 1192, 1201, 1210, 1246, 1284, 1301, 1311, 1322, 1341, 1370, 1379, 1386, 1399, 1413, 1447, 1449, 1456, 1504, 1557, 1607, 1654, 1692, 1830, 3024, 3025, 3059, 3069, 3070, 3116, 3167, 3234, 3565, 3585, 3697, 3750, 3844, 3846
H$^+$(dCyd)	0.38 (0.04)	43, 48, 92, 103, 128, 160, 183, 201, 237, 246, 274, 279, 292, 341, 362, 382, 421, 438, 468, 475, 524, 559, 585, 610, 657, 663, 705, 711, 748, 772, 794, 799, 842, 878, 958, 971, 1006, 1023, 1040, 1049, 1059, 1069, 1092, 1108, 1119, 1133, 1187, 1192, 1194, 1210, 1244, 1286, 1292, 1314, 1319, 1351, 1360, 1385, 1406, 1410, 1449, 1452, 1460, 1493, 1503, 1557, 1608, 1653, 1691, 1822, 3025, 3036, 3052, 3070, 3071, 3108, 3127, 3164, 3234, 3567, 3585, 3698, 3843, 3845
H$^+$(ddCyd)	0.34 (0.03)	5, 46, 63, 96, 126, 150, 163, 189, 228, 277, 283, 291, 319, 362, 377, 402, 444, 468, 524, 558, 585, 608, 652, 657, 694, 705, 754, 772, 796, 798, 829, 853, 916, 940, 972, 1006, 1014, 1039, 1052, 1060, 1079, 1101, 1115, 1126, 1186, 1192, 1193, 1214, 1236, 1283, 1309, 1317, 1331, 1344, 1373, 1385, 1409, 1417, 1450, 1460, 1492, 1503, 1510, 1556, 1609, 1652, 1691, 1822, 3018, 3039, 3062, 3066, 3070, 3111, 3118, 3130, 3168, 3234, 3567, 3586, 3698, 3845
H$^+$(m^{5}Cyd)	0.45 (0.04)	32, 50, 74, 88, 113, 126, 161, 174, 182, 218, 229, 244, 272, 274, 292, 296, 303, 317, 347, 375, 385, 419, 441, 473, 482, 497, 560, 594, 611, 636, 655, 706, 723, 729, 767, 771, 792, 855, 872, 904, 954, 994, 1002, 1018, 1026, 1050, 1067, 1077, 1092, 1102, 1110, 1130, 1144, 1184, 1197, 1210, 1245, 1250, 1285, 1300, 1308, 1322, 1340, 1370, 1374, 1382, 1388, 1406, 1429, 1440, 1447, 1450, 1489, 1499, 1503, 1556, 1603, 1657, 1687, 1825, 3022, 3025, 3026, 3057, 3068, 3069, 3072, 3114, 3130, 3173, 3567, 3588, 3700, 3752, 3844, 3846
H$^+$(m^{5}dCyd)	0.42 (0.04)	35, 48, 83, 96, 121, 137, 168, 174, 193, 227, 247, 255, 273, 284, 286, 302, 347, 375, 378, 418, 437, 470, 483, 494, 547, 587, 612, 654, 663, 682, 723, 748, 770, 782, 835, 854, 880, 958, 973, 999, 1011, 1022, 1046, 1054, 1074, 1077, 1092, 1108, 1121, 1136, 1185, 1194, 1208, 1244, 1249, 1287, 1292, 1311, 1319, 1351, 1360, 1379, 1386, 1407, 1430, 1447, 1449, 1452, 1490, 1493, 1500, 1503, 1555, 1604, 1656, 1687, 1817, 3022, 3025, 3035, 3050, 3070, 3071, 3072, 3107, 3127, 3130, 3171, 3567, 3588, 3701, 3843, 3845

Table S1. (continued) Vibrational Frequencies and Average Vibrational Energies of the (*x*Cyd)H^+(*x*Cyd) Base Pairs, N3-Protonated H^+(*x*Cyd) and Neutral *x*Cyd Nucleosides.[a]

Species	E_{int} (eV)[b]	Vibrational Frequencies (cm^{-1})[c]
Cyd	0.41 (0.04)	40, 48, 74, 95, 116, 138, 192, 203, 219, 240, 254, 259, 280, 289, 299, 308, 340, 370, 421, 438, 451, 522, 547, 571, 595, 619, 628, 702, 721, 754, 774, 778, 788, 848, 863, 889, 953, 969, 996, 1006, 1035, 1046, 1078, 1090, 1101, 1109, 1125, 1140, 1184, 1193, 1213, 1238, 1254, 1283, 1299, 1301, 1324, 1342, 1360, 1369, 1385, 1405, 1425, 1445, 1449, 1501, 1507, 1557, 1630, 1675, 1740, 3005, 3026, 3048, 3057, 3074, 3120, 3205, 3221, 3594, 3722, 3772, 3850, 3853
dCyd	0.38 (0.03)	44, 46, 73, 101, 118, 161, 191, 211, 225, 235, 258, 277, 283, 293, 337, 369, 421, 435, 476, 519, 526, 576, 591, 612, 660, 720, 741, 752, 774, 780, 792, 842, 886, 958, 971, 976, 1007, 1033, 1047, 1064, 1079, 1096, 1103, 1124, 1134, 1177, 1192, 1214, 1241, 1251, 1280, 1286, 1307, 1321, 1353, 1354, 1371, 1404, 1435, 1447, 1449, 1495, 1501, 1507, 1557, 1630, 1673, 1730, 2997, 3012, 3043, 3047, 3064, 3111, 3131, 3204, 3218, 3595, 3723, 3845, 3854
ddCyd	0.34 (0.03)	5, 50, 63, 93, 107, 147, 184, 198, 209, 226, 281, 295, 320, 369, 401, 463, 528, 536, 554, 573, 601, 609, 653, 716, 721, 754, 772, 775, 791, 833, 860, 900, 959, 967, 969, 982, 1036, 1061, 1067, 1078, 1101, 1129, 1153, 1192, 1209, 1237, 1249, 1261, 1301, 1318, 1327, 1331, 1351, 1370, 1385, 1408, 1438, 1452, 1488, 1497, 1506, 1513, 1557, 1632, 1682, 1728, 2978, 2991, 3008, 3048, 3055, 3073, 3117, 3140, 3189, 3216, 3600, 3675, 3731
m^5Cyd	0.45 (0.04)	30, 47, 68, 88, 113, 118, 177, 187, 201, 218, 233, 238, 256, 266, 271, 292, 303, 304, 315, 344, 387, 421, 442, 449, 508, 545, 562, 593, 622, 633, 701, 731, 744, 758, 785, 786, 860, 875, 909, 952, 962, 992, 1021, 1033, 1050, 1068, 1070, 1092, 1099, 1108, 1123, 1140, 1185, 1201, 1219, 1253, 1258, 1284, 1299, 1304, 1324, 1342, 1359, 1367, 1385, 1407, 1418, 1430, 1446, 1449, 1473, 1486, 1501, 1508, 1546, 1631, 1690, 1736, 3001, 3005, 3025, 3042, 3047, 3057, 3073, 3101, 3119, 3211, 3595, 3723, 3772, 3850, 3853
m^5dCyd	0.38 (0.04)	35, 45, 68, 92, 115, 137, 178, 189, 212, 220, 231, 251, 258, 271, 286, 303, 307, 345, 387, 421, 437, 479, 502, 543, 552, 583, 625, 663, 689, 740, 744, 771, 787, 836, 861, 888, 956, 966, 969, 1017, 1031, 1051, 1064, 1069, 1075, 1097, 1104, 1122, 1134, 1180, 1200, 1223, 1249, 1258, 1281, 1286, 1309, 1321, 1353, 1355, 1369, 1404, 1418, 1442, 1447, 1449, 1473, 1487, 1496, 1501, 1508, 1545, 1631, 1688, 1727, 2996, 3001, 3011, 3042, 3042, 3047, 3063, 3101, 3110, 3131, 3207, 3595, 3723, 3845, 3853

[a]Determined at the B3LYP/6-311G(d,p) level of theory and with frequencies scaled by 0.9887. [b]Uncertainties are listed in parentheses. [c]Degeneracies are listed in parentheses.

TableS2. Rotational Constants of (*x*Cyd)H^+(*x*Cyd) Protonated Nucleoside Base Pairs and PSL TSs for CID Determined at the B3LYP/6-311G(d,p) Level of Theory

Species	Ionic Product	Energized Molecule		Transition State	
		1D[a]	2D[b]	1D[c]	2D[c]
(Cyd)H^+(Cyd)	H^+(Cyd)	0.011	0.0013	0.026, 0.029	0.010, 0.010, 0.0009
(dCyd)H^+(dCyd)	H^+(dCyd)	0.013	0.0014	0.032, 0.033	0.010, 0.010, 0.0010
(ddCyd)H^+(ddCyd)	H^+(ddCyd)	0.016	0.0017	0.036, 0.037	0.010, 0.010, 0.0010
(m^5Cyd)H^+(m^5Cyd)	H^+(m^5Cyd)	0.0085	0.0012	0.023, 0.024	0.009, 0.008, 0.0007
(m^5dCyd)H^+(m^5dCyd)	H^+(m^5dCyd)	0.0098	0.0013	0.026, 0.026	0.009, 0.009, 0.0008

[a]Active external. [b]Inactive external. [c]Rotational constants of the TS treated as free internal rotors.

Table S3. Geometric Parameters of the B3LYP/6-311+G(d,p) Ground Conformers of the (*x*Cyd)H^+(*x*Cyd) Base Pairs[a].

(*x*Cyd)H^+(*x*Cyd)	C2–O2	C4–N4	∠O2C2N3	∠C2N3C4	∠N3C4N4	N3-H^+N4	H···O2	C6H···O5′	O2′H···O3′
(Cyd)H^+(Cyd)	**1.217**	**1.319**	**121.6**	**125.2**	**118.7**	**1.058**	**1.647**	**2.306**	**2.145**
	1.240	1.340	122.0	121.0	118.1	1.796	1.931	2.371	2.139
(dCyd)H^+(dCyd)	**1.22**	**1.321**	**122.1**	**124.8**	**118.5**	**1.059**	**1.664**	**2.180**	
	1.245	1.342	122.4	120.5	118.0	1.792	1.949	2.227	
(ddCyd)H^+(ddCyd)	**1.221**	**1.321**	**122.0**	**124.7**	**118.5**	**1.059**	**1.662**	**2.178**	
	1.245	1.342	122.4	120.5	118.0	1.791	1.947	2.220	
(m^5Cyd)H^+(m^5Cyd)	**1.218**	**1.319**	**121.7**	**125.4**	**118.3**	**1.056**	**1.641**	**2.250**	**2.145**
	1.245	1.342	122.6	120.8	117.5	1.804	1.931	2.301	2.134
(m^5dCyd)H^+(m^5dCyd)	**1.223**	**1.321**	**122.2**	**124.9**	**118.2**	**1.058**	**1.647**	**2.22**	
	1.243	1.342	122.1	121.0	117.6	1.801	1.930	2.334	

[a]Values for the protonated nucleoside of the base pair are indicated in boldface, whereas values describing the neutral nucleoside of the base pair are indicated in standard font. Bond distances are given in Angstroms and bond angles in degrees.

Table S4. Geometric Parameters of the B3LYP/6-311+G(d,p) Ground Conformers of the Neutral *x*Cyd Nucleosides[a]

*x*Cyd	C2–O2	C4–N4	∠OC2N3	∠C2N3C4	∠N3C4N4	O5′H···O2	O2′H···O2	O3′H···O2′
Cyd	1.233	1.359	123.9	120.4	116.8		1.820	2.036
dCyd	1.223	1.358	123.0	121.4	117.2	1.857		
ddCyd	1.222	1.361	124.8	120.7	117.0			
m^5Cyd	1.233	1.360	124.6	120.5	116.4		1.767	2.790
m^5dCyd	1.227	1.360	123.8	121.2	116.4	1.938		

[a]Bond distances are given in angstroms (Å) and bond angles in degrees (º).

Table S5. Geometric Parameters of the B3LYP/6-311+G(d,p) Ground N3-Protonated Conformers of the *x*Cyd Nucleosides[a]

H^+(*x*Cyd)	C2–O2	C4–N4	∠O2C2N3	∠C2N3C4	∠N3C4N4	N3–H^+	C6H···O5′
H^+(Cyd)	1.199	1.334	121.1	126.2	119.3	1.015	2.248
H^+(dCyd)	1.203	1.335	121.5	125.7	119.1	1.015	2.142
H^+(ddCyd)	1.203	1.335	121.5	125.7	119.1	1.015	2.133
H^+(m^5Cyd)	1.206	1.335	122.0	125.8	118.7	1.015	2.196
H^+(m^5dCyd)	1.204	1.336	121.8	125.9	118.8	1.015	2.183

[a]Bond distances are given in angstroms (Å) and bond angles in degrees (º).

Table S6. Geometric Parameters of the B3LYP/6-311+G(d,p) Ground O2-Protonated Conformers of of the *x*Cyd Nucleosides[a]

H^+(*x*Cyd)	C2–O2	C4–N4	∠O2C2N3	∠C2N3C4	∠N3C4N4	O2–H^+	C6H···O5′
H^+(Cyd)	1.321	1.334	120.1	118.9	117.1	0.970	2.252
H^+(dCyd)	1.326	1.335	120.3	118.4	117.0	0.970	2.139
H^+(ddCyd)	1.326	1.335	120.2	118.4	117.0	0.970	2.131
H^+(m^5Cyd)	1.324	1.336	120.7	118.7	116.7	0.970	2.227
H^+(m^5dCyd)	1.325	1.336	120.1	118.8	116.7	0.970	2.243

[a]Bond distances are given in angstroms (Å) and bond angles in degrees (º).

FIGURE CAPTIONS

Figure S1. Cross sections for CID of the (*x*Cyd)H^+(*x*Cyd) base pairs with Xe as a function of collision energy in the center-of-mass frame (lower *x*-axis) and laboratory frame (upper *x*-axis), where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd, parts a-e, respectively. Data are shown for a Xe pressure of ~0.2 mTorr.

Figure S2. Zero-pressure-extrapolated cross sections for CID of the (*x*Cyd)H^+(*x*Cyd) base pairs 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), where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd, parts a-e, respectively. The solid lines show the best fits to the data using the model of **eq 3** convoluted over the neutral and ion kinetic and internal energy distributions. The dotted lines show the model cross sections in the absence of experimental kinetic energy broadening for (*x*Cyd)H^+(*x*Cyd) base pairs with an internal temperature of 0 K.

Figure S3. B3LYP/6-311+G(d,p) optimized geometries of the ground and stable low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The puckering of the sugar moieties, nucleobase orientations, 5'-hydroxy orientations, and the B3LYP/6-311+G(2d,2p) relative Gibbs energies at 298 K (in kJ/mol) are also shown. Information regarding the protonated nucleoside is indicated first in boldface font, followed by that of the neutral nucleoside in standard font.

Figure S4. B3LYP/6-311+G(d,p) optimized geometries of the ground and stable low-energy conformers of the neutral *x*Cyd nucleosides, where *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The puckering of the sugar moieties, nucleobase orientations, 5'-hydroxy orientations, and the B3LYP/6-311+G(2d,2p) relative Gibbs energies at 298 K (in kJ/mol) are also shown.

Figure S5. B3LYP/6-311+G(d,p) optimized geometries of the ground and stable low-energy conformers of the N3- and O2-protonated H^+(*x*Cyd) nucleosides for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The puckering of the sugar moieties, nucleobase orientations, 5'-hydroxy orientations, and the B3LYP/6-311+G(2d,2p) relative Gibbs energies at 298 K (in kJ/mol) are also shown.

Figure S6. Polar plots depicting the pseudorotation phase angles (sugar puckering), 5'-hydroxy orientations, and glycosidic bond angles, of the B3LYP/6-311+G(d,p) optimized geometries of the ground and stable low-energy conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd.

Figure S7. Polar plots depicting the pseudorotation phase angles (sugar puckering), 5'-hydroxy orientations, and glycosidic bond angles of the B3LYP/6-311+G(d,p) optimized geometries of the ground and stable low-energy conformers of the neutral *x*Cyd and N3- and O2-protonated H^+(*x*Cyd) nucleosides for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd.

Figure S8. Electrostatic potential maps of the B3LYP/6-311+G(d,p) ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The ESP maps were generated at the B3LYP/6-311+G(2d,2p) level and are shown at an isosurface of 0.020 a.u. of the total SCF electron density. The Mülliken charges on the O2, N3, N4H atoms as well as the $N3H^+$ (excess proton) are labeled. The most electronegative regions are color-coded in red, regions of intermediate ESP are shown in green, and the most electropositive regions are color-coded in blue.

Figure S9. Noncovalent interaction maps superimposed on the B3LYP/6-311+G(d,p) optimized geometries of the ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The NCI maps are shown at an isosurface of 0.5 a.u. of the B3LYP/6-311+G(2d,2p) reduced electron density gradients. Regions exhibiting strong attractive interactions such as ionic hydrogen-bonding interactions and strong neutral hydrogen-bonding interactions appear blue. Weaker attractive interactions such as London dispersion or noncanonical hydrogen-bonding interactions appear green. Highly repulsive interactions appear red.

Figure S10. Reduced B3LYP/6-311+G(2d,2p) electron density gradient scatter maps of the BLYP/6-311+G(d,p) ground conformers of the (*x*Cyd)H^+(*x*Cyd) base pairs for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. Horizontal references lines indicate the values of the reduced electron density gradient at which the NCI maps of **Figure S8** are plotted.

Figure S11. Comparison of the B3P86/6-311+G(2d,2p) and M06-2X/6-311+G(2d,2p) computed 0 K BPEs versus measured threshold dissociation energies of the (*x*Cyd)H^+(*x*Cyd) base pairs for *x*Cyd = Cyd, dCyd, ddCyd, m^5Cyd, and m^5dCyd. The diagonal line indicates values for which the calculated and measured values are equal.

Figure S1.

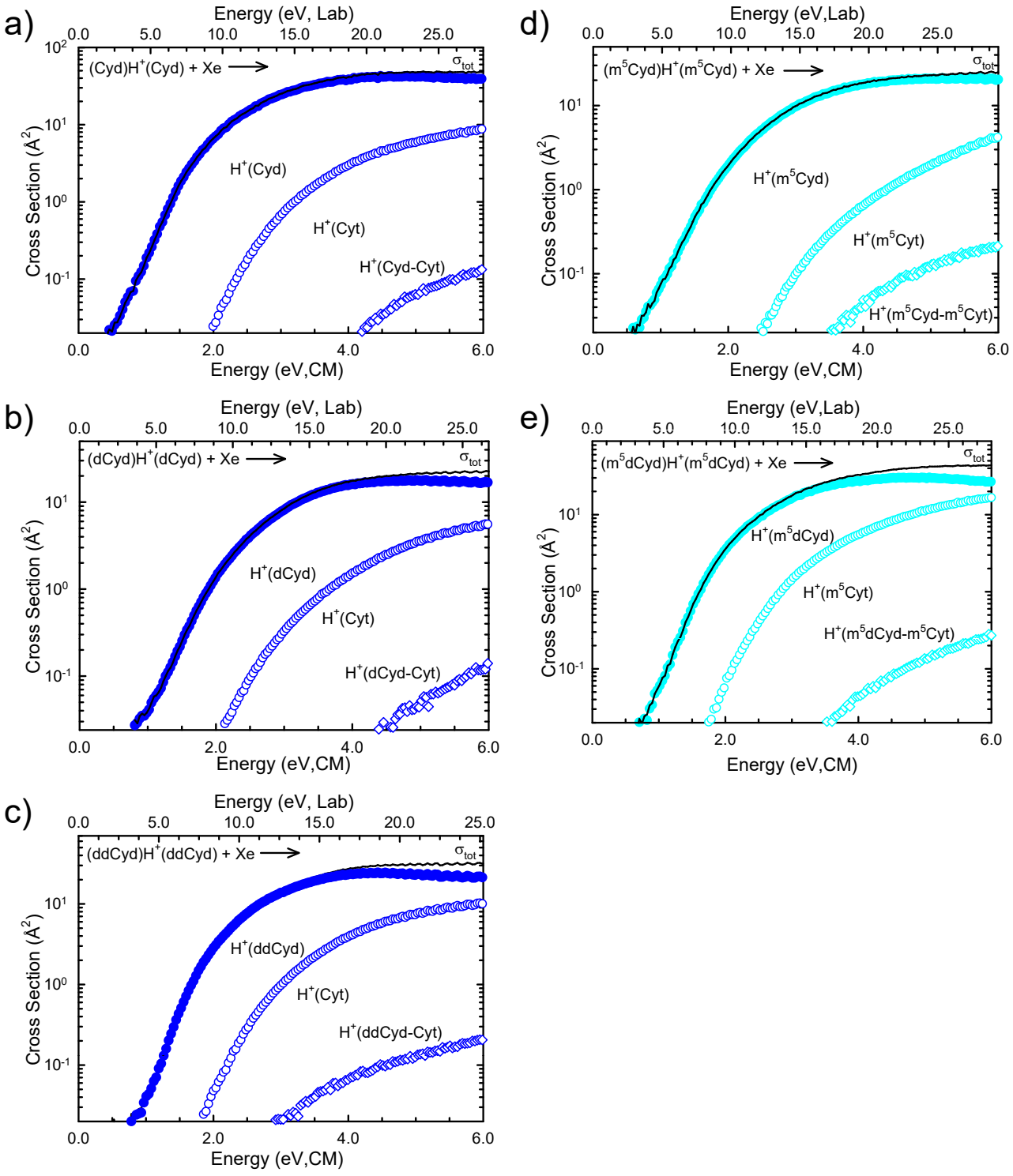

a)
Energy (eV, Lab)
(Cyd)H+(Cyd) + Xe
σtot
H+(Cyd)
H+(Cyt)
H+(Cyd-Cyt)
Cross Section (Å2)
Energy (eV,CM)
b)
Energy (eV, Lab)
(dCyd)H+(dCyd) + Xe
σtot
H+(dCyd)
H+(Cyt)
H+(dCyd-Cyt)
Cross Section (Å2)
Energy (eV,CM)
c)
Energy (eV, Lab)
(ddCyd)H+(ddCyd) + Xe
σtot
H+(ddCyd)
H+(Cyt)
H+(ddCyd-Cyt)
Cross Section (Å2)
Energy (eV,CM)
d)
Energy (eV,Lab)
(m5Cyd)H+(m5Cyd) + Xe
σtot
H+(m5Cyd)
H+(m5Cyt)
H+(m5Cyd-m5Cyt)
Cross Section (Å2)
Energy (eV,CM)
e)
Energy (eV,Lab)
(m5dCyd)H+(m5dCyd) + Xe
σtot
H+(m5dCyd)
H+(m5Cyt)
H+(m5dCyd-m5Cyt)
Cross Section (Å2)
Energy (eV,CM)

Figure S2.

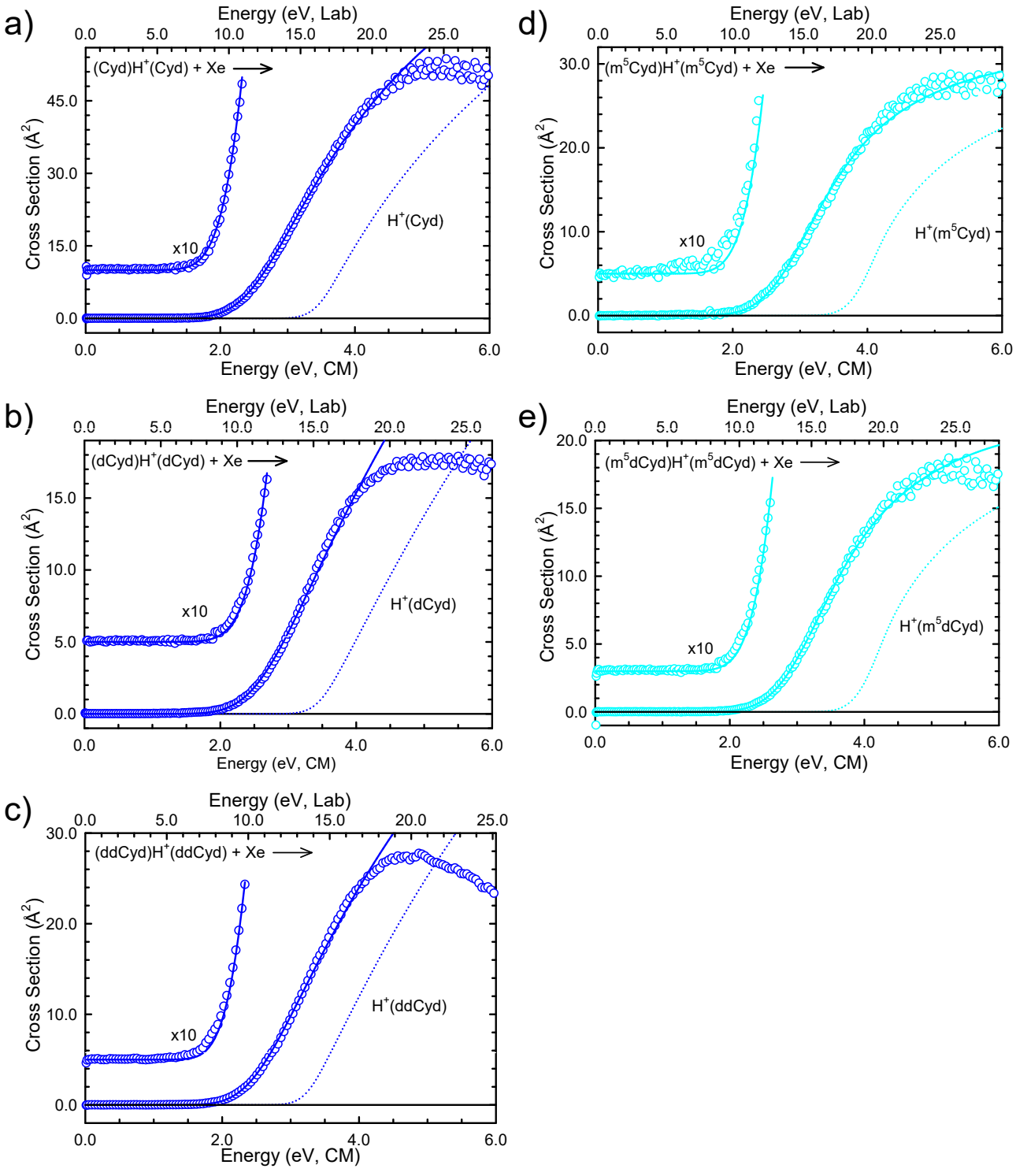
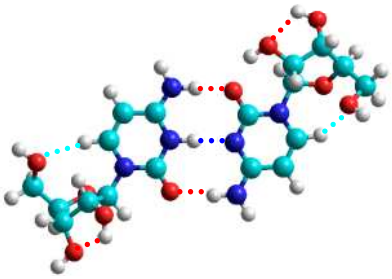
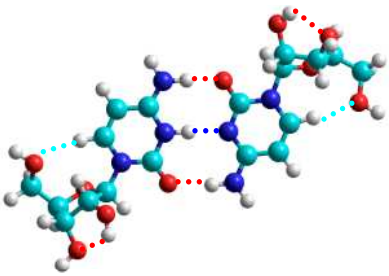

a)
Energy (eV, Lab)
(Cyd)H+(Cyd) + Xe
Cross Section (Å2)
x10
H+(Cyd)
Energy (eV, CM)
d)
(m5Cyd)H+(m5Cyd) + Xe
H+(m5Cyd)
b)
(dCyd)H+(dCyd) + Xe
H+(dCyd)
e)
(m5dCyd)H+(m5dCyd) + Xe
H+(m5dCyd)
c)
(ddCyd)H+(ddCyd) + Xe
H+(ddCyd)

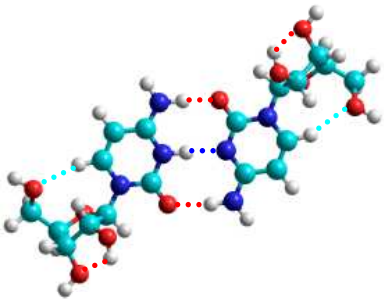
Figure S3.



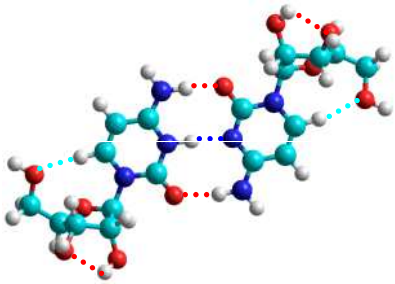
C2′-endo·C2′-endo (1)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.0 kJ/mol



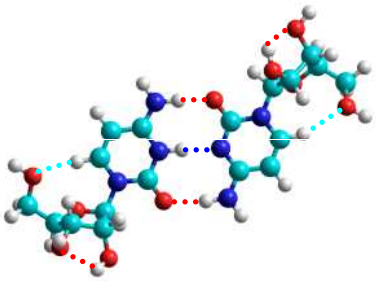
C2′-endo·C3′-endo (2)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.1 kJ/mol



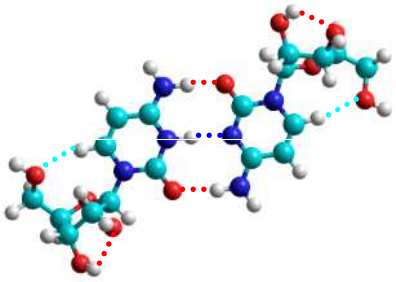
C2′-endo·C2′-endo (3)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.7 kJ/mol



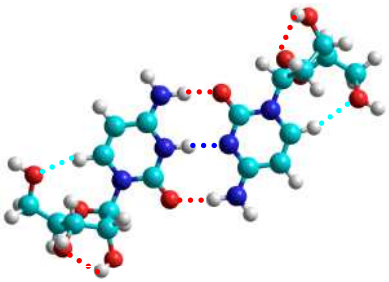
C3′-endo·C3′-endo (4)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.9 kJ/mol



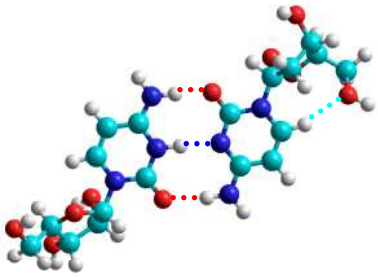
C3′-endo·C2′-endo (5)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.9 kJ/mol



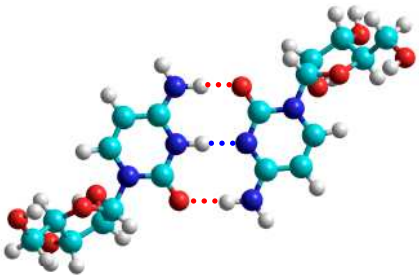
C2′-endo·C3′-endo (6)
***anti*, gauche$^+$**·*anti*,gauche$^+$
5.2 kJ/mol



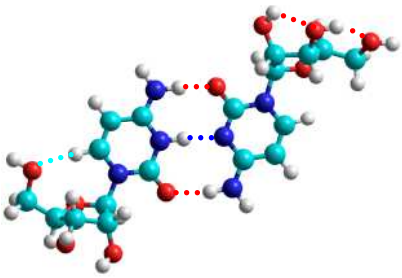
C3′-endo·C2′-endo (7)
***anti*, gauche$^+$**·*anti*,gauche$^+$
5.5 kJ/mol



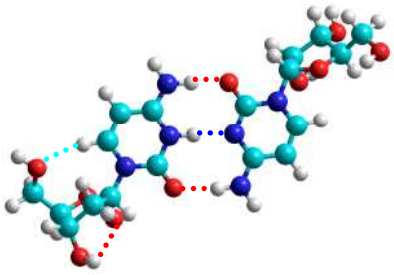
C3′-endo·C2′-endo (9)
***anti*, trans**·*anti*,gauche$^+$
7.3 kJ/mol



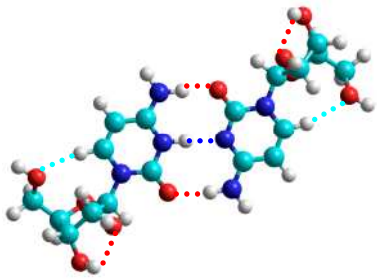
C3′-endo·C3′-endo (8)
***anti*, trans**·*anti*,trans
5.8 kJ/mol



C3′-endo·C3′-endo (10)
***anti*, gauche$^+$**·*anti*,gauche$^-$
7.6 kJ/mol



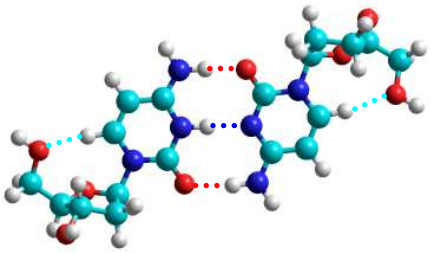
C2′-endo·C3′-endo (11)
***anti*, gauche$^+$**·*anti*,trans
7.8 kJ/mol



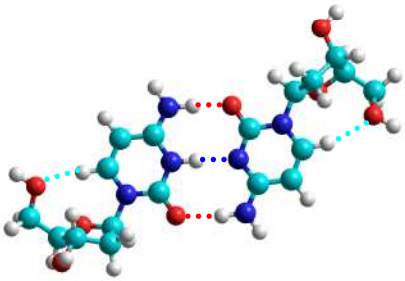
C2′-endo·C2′-endo (12)
***anti*, gauche$^+$**·*anti*,gauche$^+$
8.9 kJ/mol

(Cyd)H$^+$(Cyd)

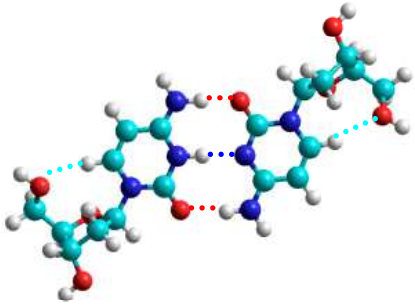
Figure S3.



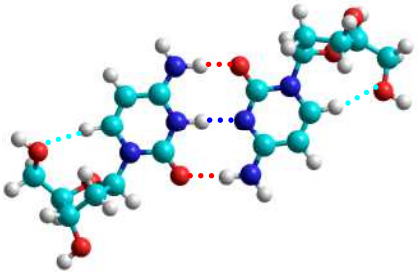
C3′-endo·C3′-endo (1)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.0 kJ/mol



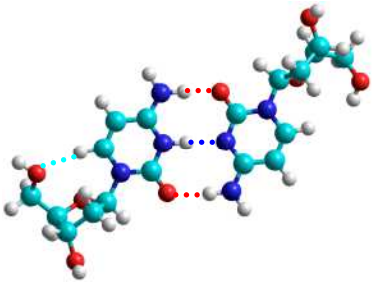
C3′-endo·C2′-endo (2)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.3 kJ/mol



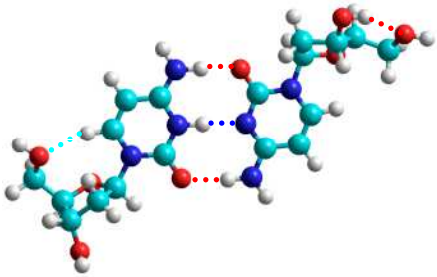
C2′-endo·C2′-endo (3)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.0 kJ/mol



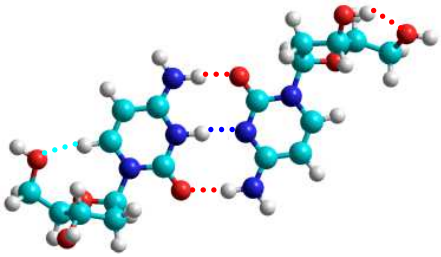
C2′-endo·C3′-endo (4)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.0 kJ/mol



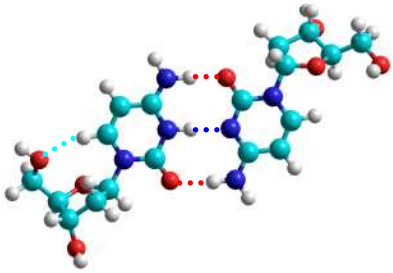
C2′-endo·C1′-exo (5)
***anti*, gauche$^+$**·*anti*,gauche$^+$
4.4 kJ/mol



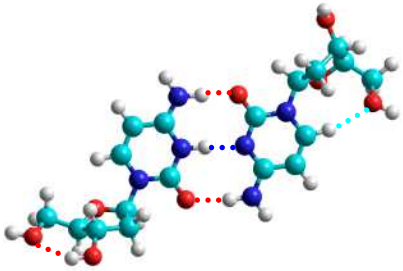
C2′-endo·C3′-endo (6)
***anti*, gauche$^+$**·*anti*,gauche$^-$
6.1 kJ/mol



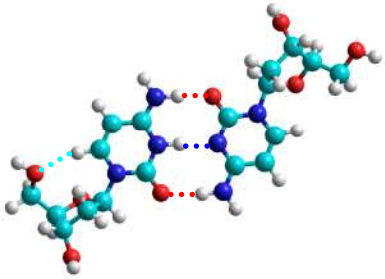
C3′-endo·C3′-endo (7)
***anti*, gauche$^+$**·*anti*,gauche$^-$
6.6 kJ/mol



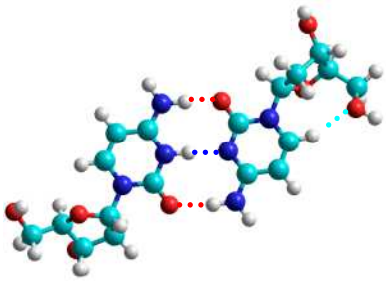
C2′-endo·C3′-endo (8)
***anti*, gauche$^+$**·*anti*,trans
7.5 kJ/mol



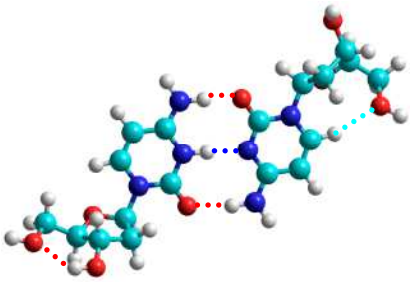
C4′-exo·C2′-endo (9)
***anti*, gauche$^-$**·*anti*,gauche$^+$
8.7 kJ/mol



C2′-endo·C2′-endo (10)
***anti*, gauche$^+$**·*anti*,gauche$^-$
8.9 kJ/mol



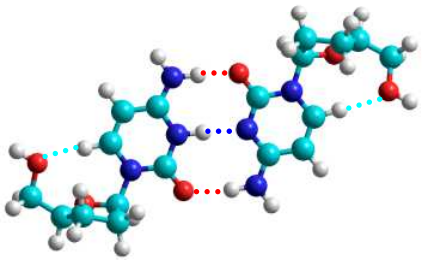
C2′-endo·C4′-exo (11)
***anti*, gauche$^+$**·*anti*, **gauche$^+$**
9.5 kJ/mol



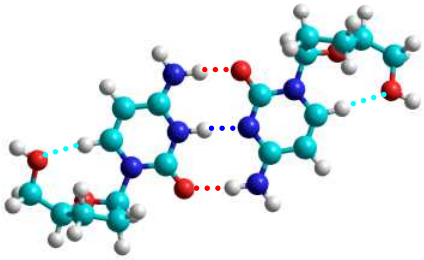
C3′-endo·C2′-endo (12)
***anti*, gauche$^-$**·*anti*,gauche$^+$
9.9 kJ/mol

(dCyd)H$^+$(dCyd)

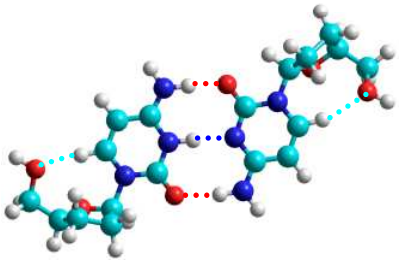
Figure S3.



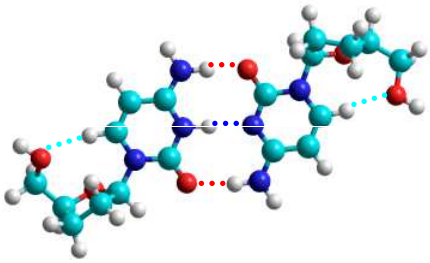
C3′-endo·C3′-endo (1)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.0 kJ/mol



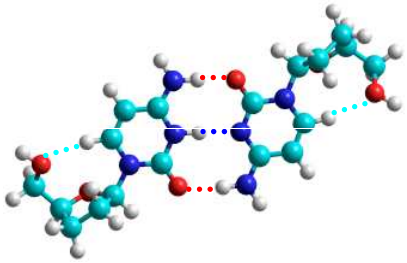
C3′-endo·C3′-endo (2)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.9 kJ/mol



C3′-endo·C2′-endo (3)
***anti*, gauche$^+$**·*anti*,gauche$^+$
4.1 kJ/mol



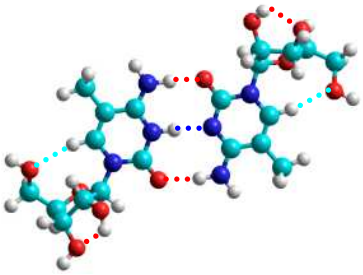
C2′-endo·C3′-endo (4)
***anti*, gauche$^+$**·*anti*,gauche$^+$
5.0 kJ/mol



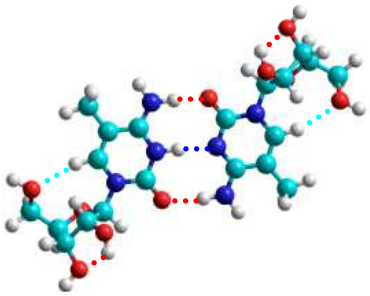
C2′-endo·C2′-endo (5)
***anti*, gauche$^+$**·*anti*,gauche$^+$
9.0 kJ/mol

(ddCyd)H$^+$(ddCyd)

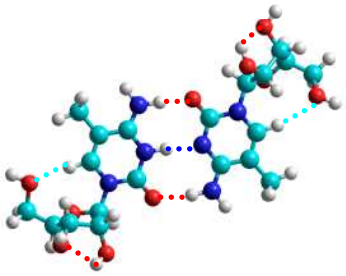
Figure S3.



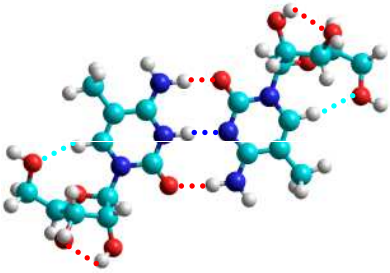
C2′-endo·C3′-endo (1)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.0 kJ/mol



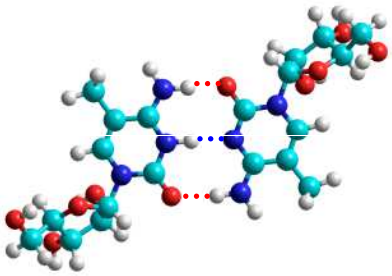
C2′-endo·C2′-endo (2)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.5 kJ/mol



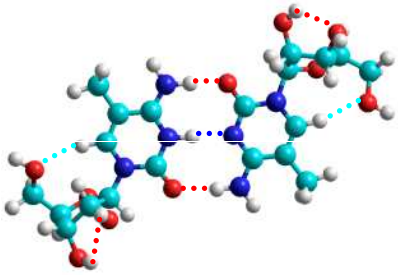
C3′-endo·C2′-endo (3)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.3 kJ/mol



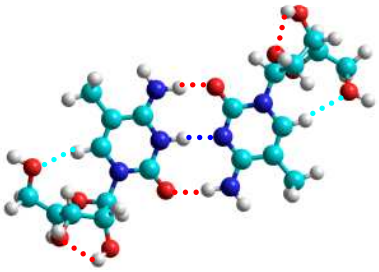
C3′-endo·C3′-endo (4)
***anti*, gauche$^+$**·*anti*,gauche$^+$
1.4 kJ/mol



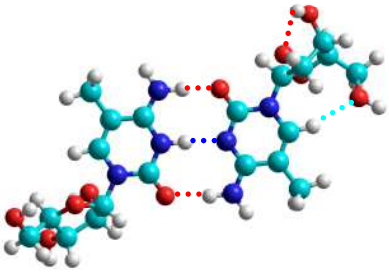
C3′-endo·C3′-endo (5)
***anti*, trans**·*anti*,trans
3.7 kJ/mol



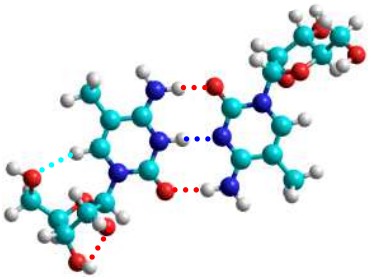
C2′-endo·C3′-endo (6)
***anti*, gauche$^+$**·*anti*,gauche$^+$
4.3 kJ/mol



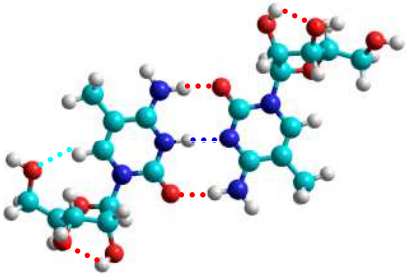
C2′-endo·C3′-endo (7)
***anti*, gauche$^+$**·*anti*,gauche$^+$
5.1 kJ/mol



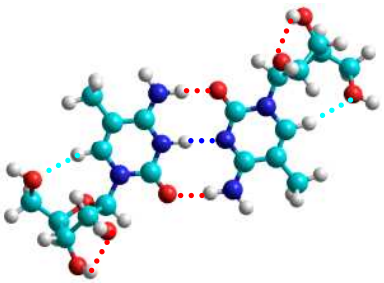
C3′-endo·C2′-endo (8)
***anti*, gauche$^+$**·*anti*,gauche$^+$
5.7 kJ/mol



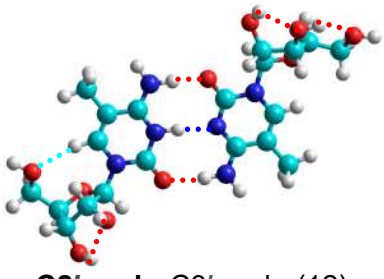
C2′-endo·C3′-endo (9)
***anti*, gauche$^+$**·*anti*,gauche$^+$
6.1 kJ/mol



C3′-endo·C3′-endo (10)
***anti*, gauche$^+$**·*anti*,gauche$^+$
6.4 kJ/mol



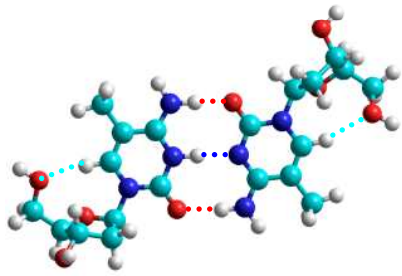
C2′-endo·C2′-endo (11)
***anti*, gauche$^+$**·*anti*,gauche$^+$
7.9 kJ/mol



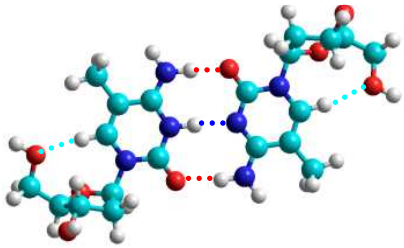
C2′-endo·C3′-endo (12)
***anti*, gauche$^+$**·*anti*,gauche$^+$
9.7 kJ/mol

$(m^5Cyd)H^+(m^5Cyd)$

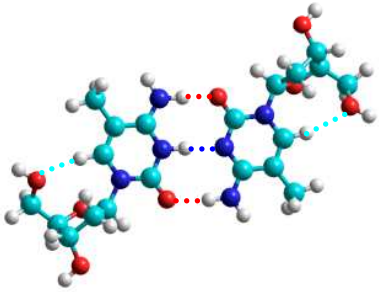
Figure S3.



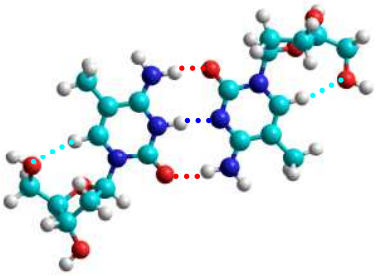
C3′-endo·C2′-endo (1)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.0 kJ/mol



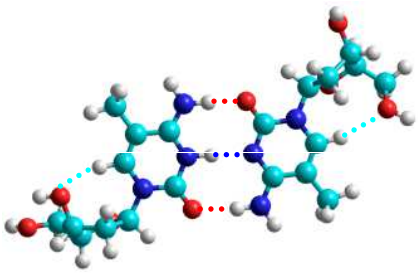
C3′-endo·C3′-endo (2)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.1 kJ/mol



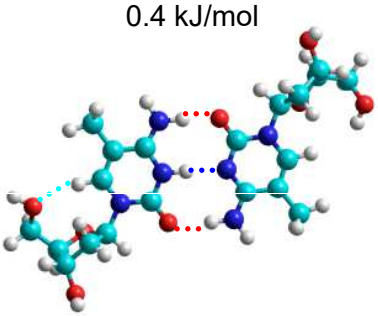
C2′-endo·C2′-endo (3)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.4 kJ/mol



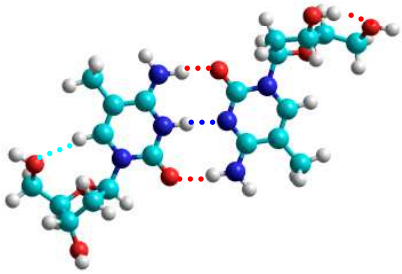
C2′-endo·C3′-endo (4)
***anti*, gauche$^+$**·*anti*,gauche$^+$
0.4 kJ/mol



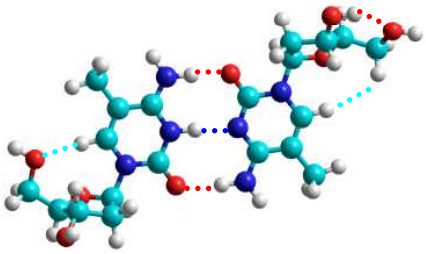
C4′-exo·C2′-endo (5)
***anti*, gauche$^-$**·*anti*,gauche$^+$
0.7 kJ/mol



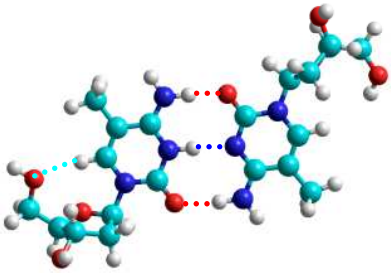
C2′-endo·C1′-exo (6)
***anti*, gauche$^+$**·*anti*,gauche$^+$
4.9 kJ/mol



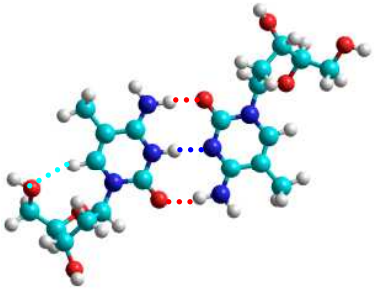
C2′-endo·C3′-endo (7)
***anti*, gauche$^+$**·*anti*,gauche$^-$
5.0 kJ/mol



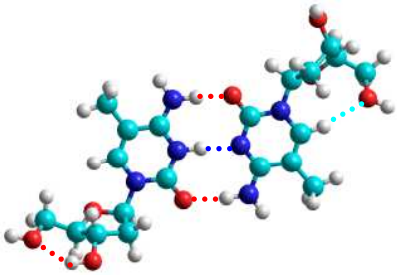
C3′-endo·C3′-endo (8)
***anti*, gauche$^+$**·*anti*,gauche$^-$
5.6 kJ/mol



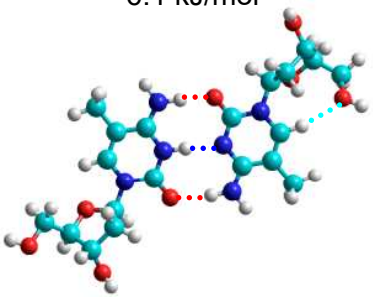
C3′-endo·C1′-exo (9)
***anti*, gauche$^+$**·*anti*,gauche$^+$
6.1 kJ/mol



C2′-endo·C2′-endo (10)
***anti*, gauche$^+$**·*anti*,gauche$^+$
8.1 kJ/mol



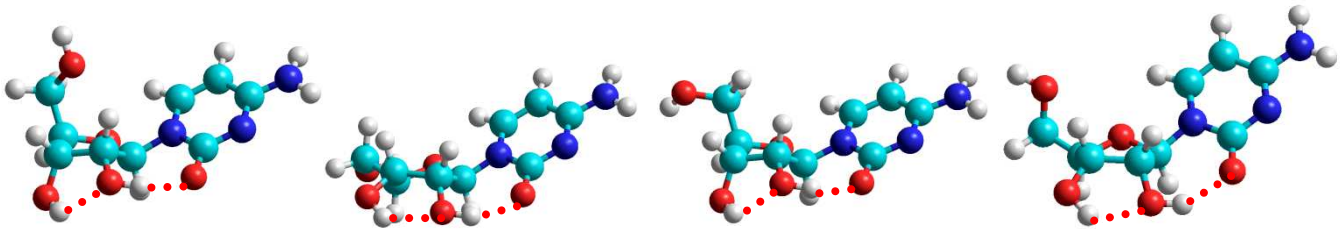
C3′-endo·C2′-endo (11)
***anti*, gauche$^-$**·*anti*,gauche$^+$
8.3 kJ/mol



C2′-endo·C2′-endo (12)
***anti*, gauche$^+$**·*anti*,gauche$^+$
10.3 kJ/mol

(m^{5}dCyd)H$^+$(m^{5}dCyd)

Figure S4.

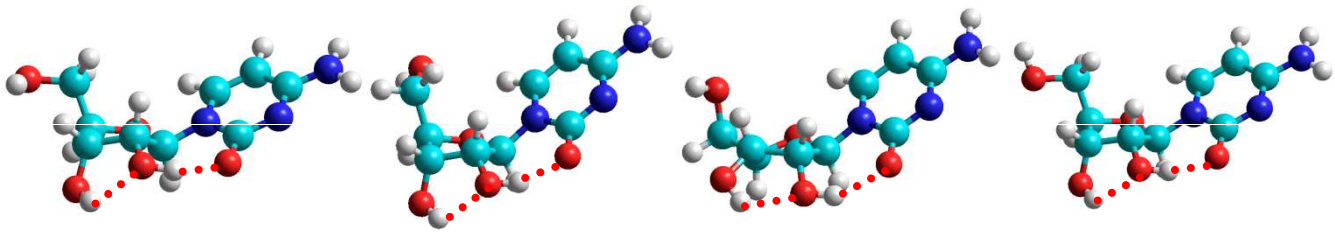


Cyd1, C2′-endo
anti, gauche$^+$
0.0 kJ/mol

Cyd2, O4′-endo
anti, trans
1.0 kJ/mol

Cyd3, C2′-endo
anti, gauche$^-$
3.8 kJ/mol

Cyd4, O4′-endo
anti, gauche$^+$
4.7 kJ/mol

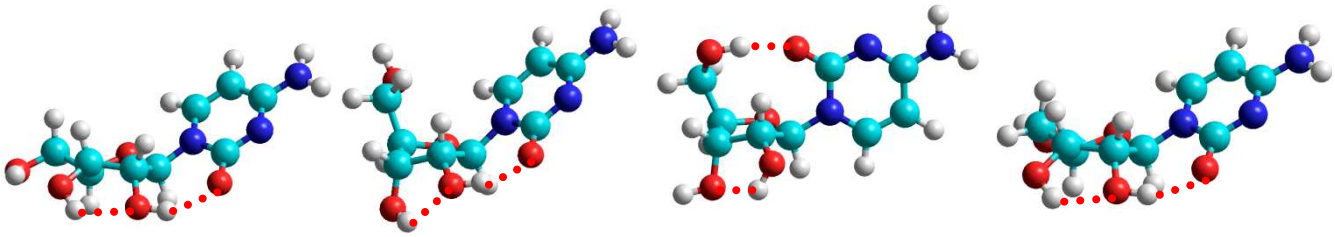


Cyd5, C2′-endo
anti, trans
5.0 kJ/mol

Cyd6, C3′-exo
anti, gauche$^-$
5.0 kJ/mol

Cyd7, C4′-endo
anti, gauche$^+$
5.1 kJ/mol

Cyd8, C2′-endo
anti, gauche$^-$
5.4 kJ/mol

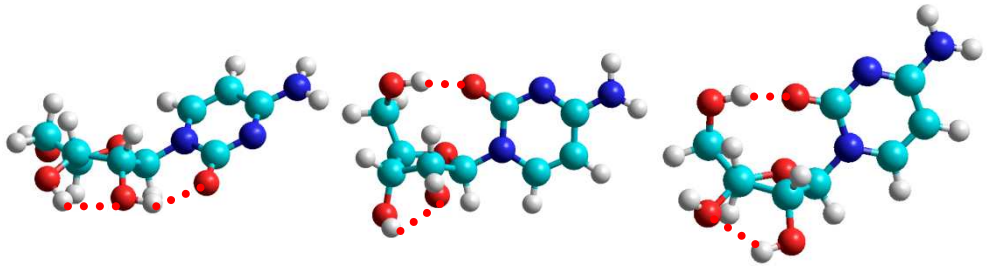


Cyd9, C4′-exo
anti, gauche$^-$
6.7 kJ/mol

Cyd10, C3′-exo
anti, trans
7.0 kJ/mol

Cyd11, C2′-endo
syn, gauche$^+$
7.2 kJ/mol

Cyd12, C4′-endo
anti, gauche$^+$
7.3 kJ/mol



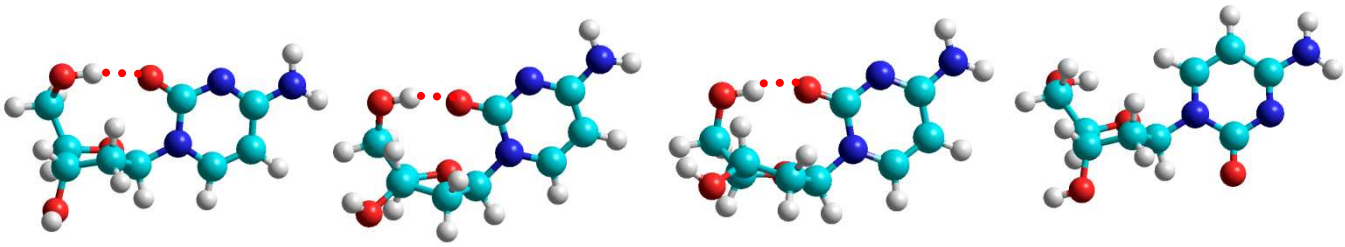
Cyd13, O4′-endo
anti, gauche$^+$
7.5 kJ/mol

Cyd14, C2′-endo
syn, gauche$^+$
8.6 kJ/mol

Cyd15, C4′-exo
syn, gauche$^+$
9.5 kJ/mol

Cyd

Figure S4.

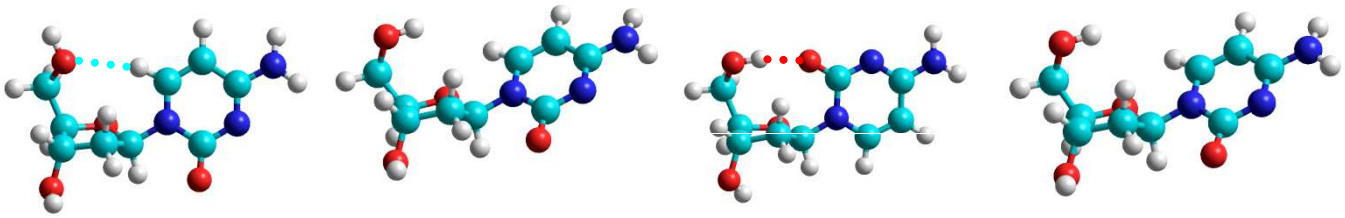


dCyd1, C2′-endo
syn, gauche$^+$
0.0 kJ/mol

dCyd2, C4′-exo
syn, gauche$^+$
0.3 kJ/mol

dCyd3, C3′-endo
syn, gauche$^+$
2.5 kJ/mol

dCyd4, C1′-exo
anti, trans
4.3 kJ/mol

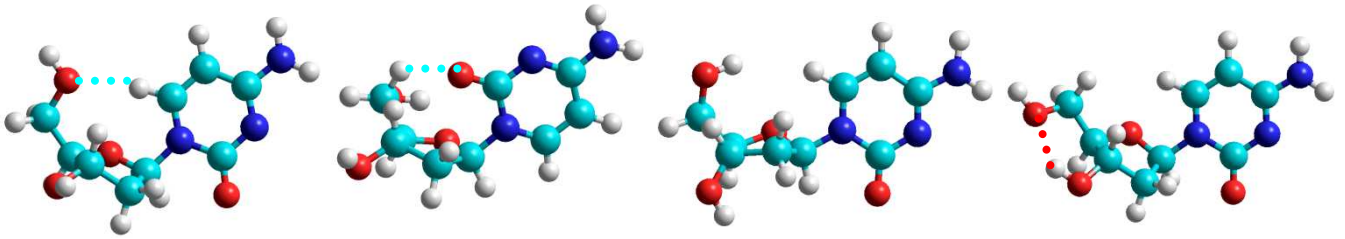


dCyd5, C2′-endo
anti, gauche$^+$
5.0 kJ/mol

dCyd6, C1′-exo
anti, gauche$^+$
5.1 kJ/mol

dCyd7, C2′-endo
syn, gauche$^+$
5.8 kJ/mol

dCyd8, C1′-exo
anti, gauche$^+$
5.9 kJ/mol

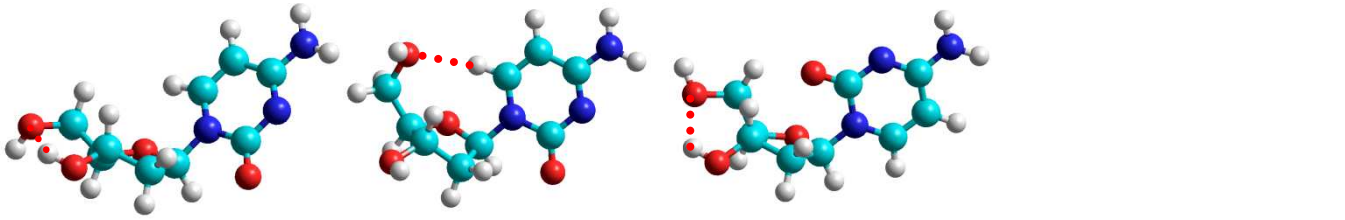


dCyd9, C3′-endo
anti, gauche$^+$
6.2 kJ/mol

dCyd10, C3′-endo
syn, gauche$^+$
6.7 kJ/mol

dCyd11, C4′-endo
anti, trans
6.7 kJ/mol

dCyd12, C3′-endo
anti, gauche$^-$
7.4 kJ/mol



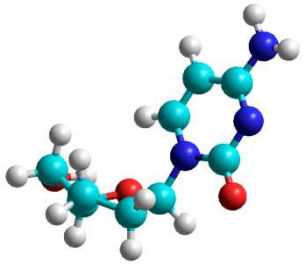
dCyd13, C3′-endo
anti, gauche$^-$
7.7 kJ/mol

dCyd14, C3′-endo
syn, gauche$^+$
8.4 kJ/mol

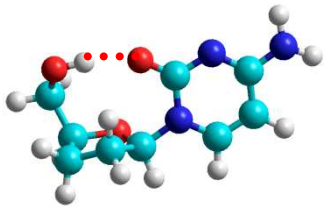
dCyd15, C3′-endo
syn, gauche$^-$
8.7 kJ/mol

dCyd

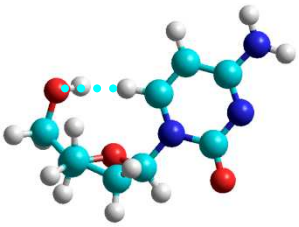
Figure S4.



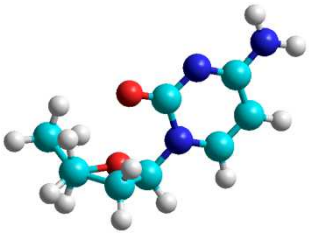
ddCyd1, C3′-endo
anti, trans
0.0 kJ/mol



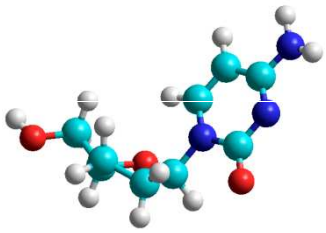
ddCyd2, C2′-endo
syn, gauche$^+$
2.0 kJ/mol



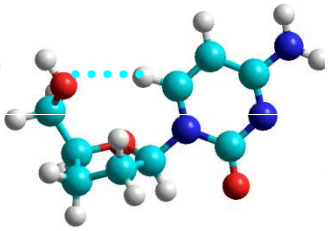
ddCyd3, C3′-endo
anti, gauche$^+$
2.3 kJ/mol



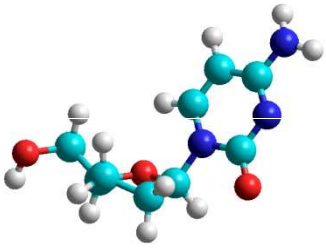
ddCyd4, C3′-endo
syn, trans
3.4 kJ/mol



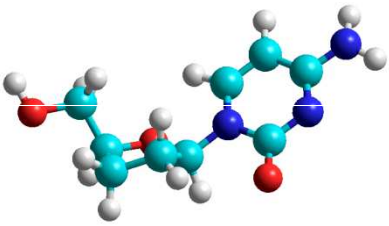
ddCyd5, C3′-endo
anti, gauche$^-$
5.4 kJ/mol



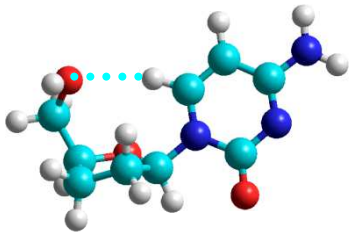
ddCyd6, C2′-endo
anti, gauche$^+$
5.8 kJ/mol



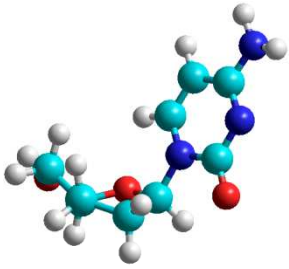
ddCyd7, C3′-endo
anti, gauche$^-$
6.6 kJ/mol



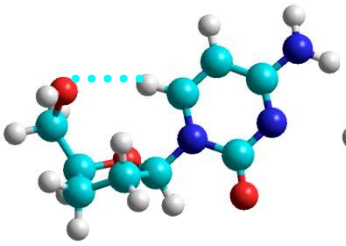
ddCyd8, C2′-endo
anti, gauche$^-$
7.0 kJ/mol



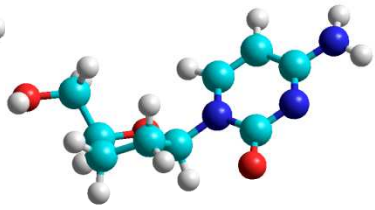
ddCyd9, C2′-endo
anti, gauche$^+$
7.6 kJ/mol



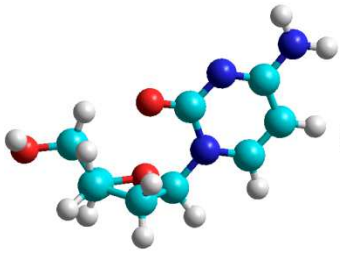
ddCyd10, C3′-endo
anti, trans
8.0 kJ/mol



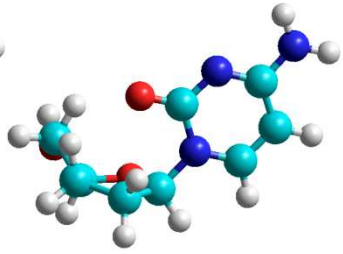
ddCyd11, C2′-endo
anti, gauche$^+$
8.3 kJ/mol



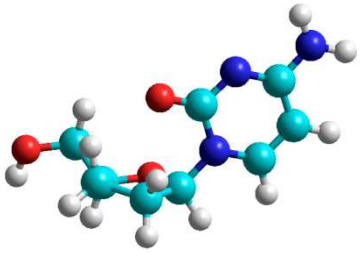
ddCyd12, C2′-endo
anti, gauche$^-$
9.0 kJ/mol



ddCyd13, C3′-endo
syn, gauche$^-$
10.8 kJ/mol



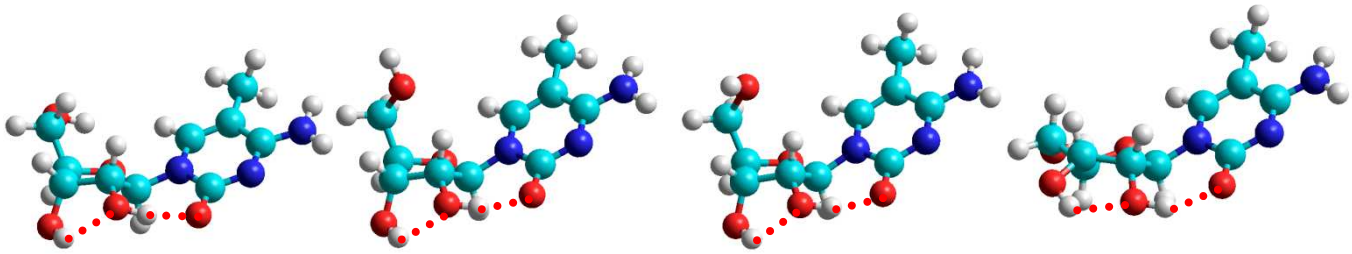
ddCyd14, C3′-endo
syn, trans
11.4 kJ/mol



ddCyd15, C3′-endo
syn, gauche$^-$
13.0 kJ/mol

ddCyd

Figure S4.

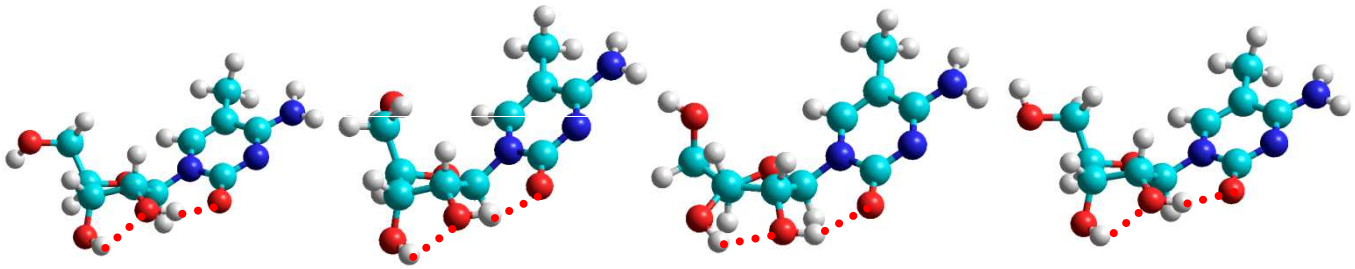


m^{5}Cyd1, C2′-endo
anti, trans
0.0 kJ/mol

m^{5}Cyd2, C2′-endo
anti, gauche$^+$
0.6 kJ/mol

m^{5}Cyd3, C2′-endo
anti, gauche$^+$
1.1 kJ/mol

m^{5}Cyd4, O4′-endo
trans, gauche$^+$
2.4 kJ/mol

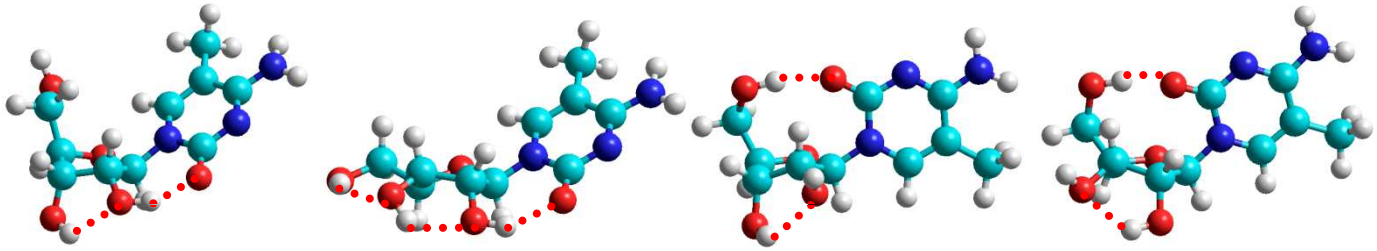


m^{5}Cyd5, C2′-endo
anti, gauche$^-$
3.7 kJ/mol

m^{5}Cyd6, C3′-exo
trans, gauche$^+$
4.8 kJ/mol

m^{5}Cyd7, O4′-endo
anti, gauche$^-$
5.0 kJ/mol

m^{5}Cyd8, C2′-endo
anti, gauche$^+$
5.7 kJ/mol

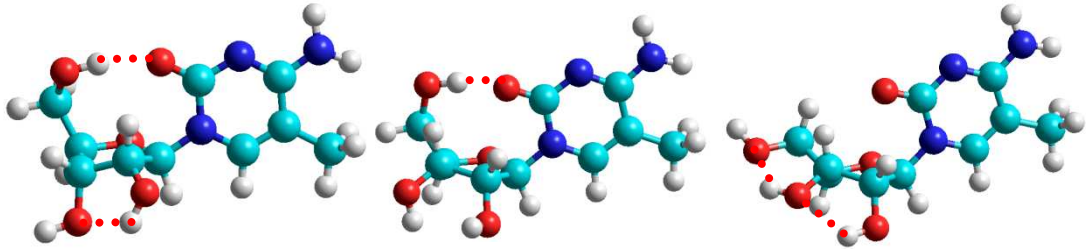


m^{5}Cyd9, C3′-exo
anti, trans
6.8 kJ/mol

m^{5}Cyd10, C4′-exo
anti, gauche$^-$
7.1 kJ/mol

m^{5}Cyd11, C2′-endo
syn, gauche$^+$
8.3 kJ/mol

m^{5}Cyd12, C3′-endo
syn, gauche$^+$
9.3 kJ/mol



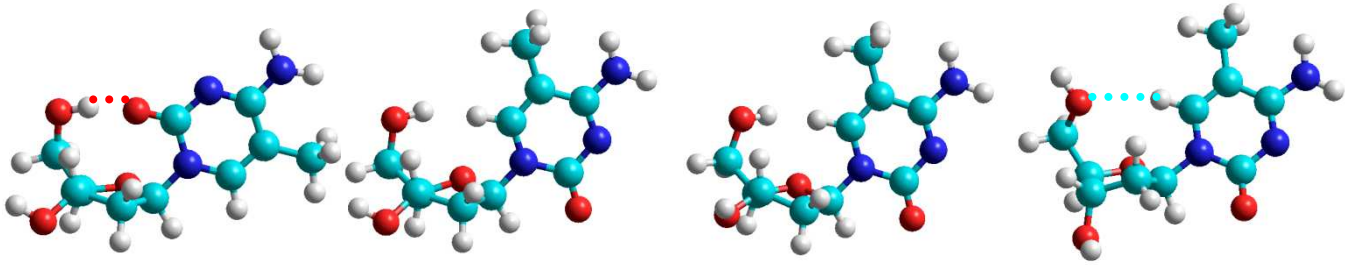
m^{5}Cyd13, C2′-endo
syn, gauche$^+$
9.5 kJ/mol

m^{5}Cyd14, C4′-exo
syn, gauche$^+$
14.6 kJ/mol

m^{5}Cyd15, C3′-endo
anti, gauche$^-$
16.2 kJ/mol

m^{5}Cyd

Figure S4.

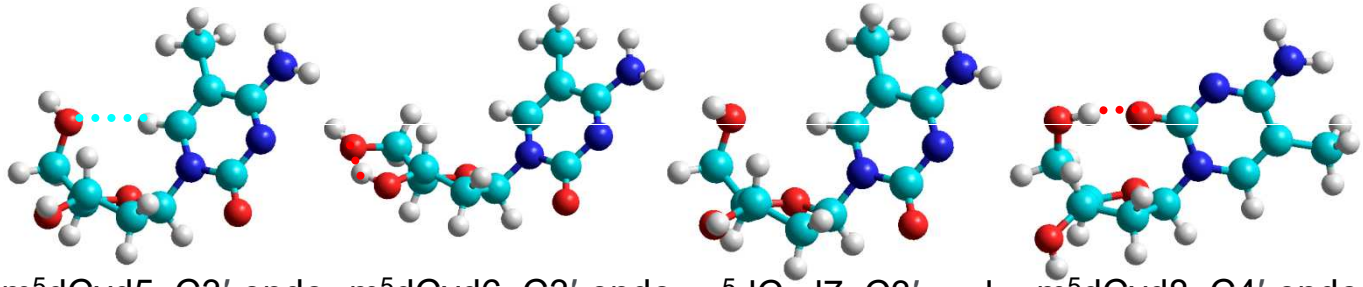


m^{5}dCyd1, C3′-endo
syn, gauche$^+$
0.0 kJ/mol

m^{5}dCyd2, C4′-exo
anti, gauche$^+$
2.5 kJ/mol

m^{5}dCyd3, C3′-endo
anti, gauche$^+$
2.6 kJ/mol

m^{5}dCyd4, C2′-endo
anti, gauche$^+$
3.5 kJ/mol

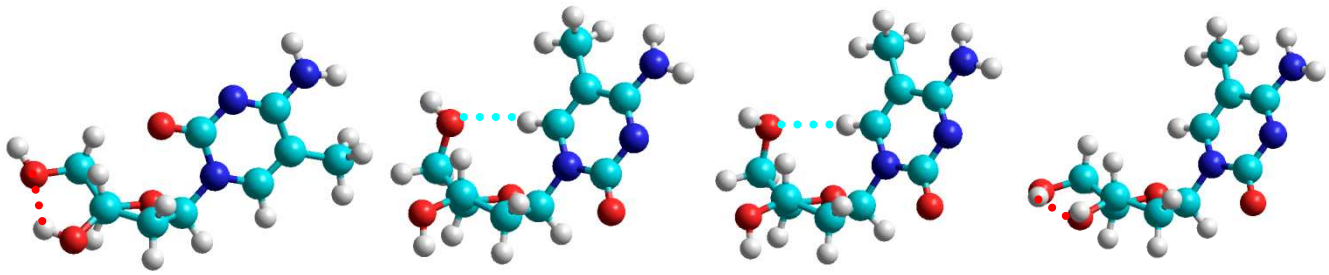


m^{5}dCyd5, C3′-endo
anti, gauche$^+$
4.9 kJ/mol

m^{5}dCyd6, C3′-endo
anti, gauche$^-$
5.9 kJ/mol

m^{5}dCyd7, C3′-endo
anti, gauche$^+$
6.3 kJ/mol

m^{5}dCyd8, C4′-endo
syn, gauche$^+$
6.9 kJ/mol

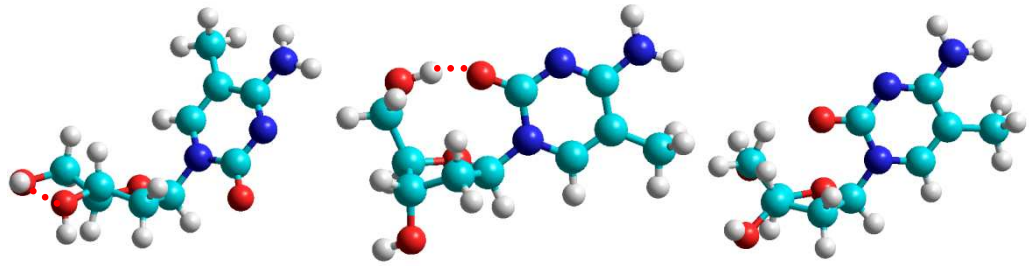


m^{5}dCyd9, C3′-endo
syn, gauche$^-$
7.0 kJ/mol

m^{5}dCyd10, C3′-endo
anti, gauche$^+$
7.5 kJ/mol

m^{5}dCyd11, C3′-endo
anti, gauche$^+$
8.3 kJ/mol

m^{5}dCyd12, C3′-endo
anti, gauche$^-$
9.0 kJ/mol



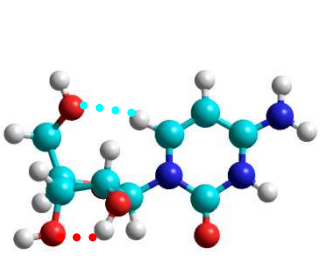
m^{5}dCyd13, C3′-endo
anti, gauche$^-$
9.7 kJ/mol

m^{5}dCyd14, C3′-exo
syn, trans
10.3 kJ/mol

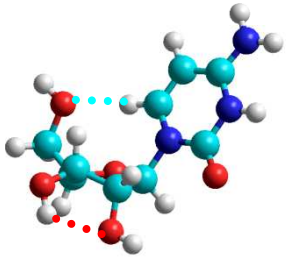
m^{5}dCyd15, C3′-endo
syn, trans
11.3 kJ/mol

m^{5}dCyd

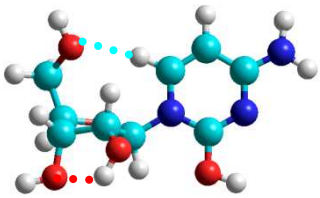
Figure S5.



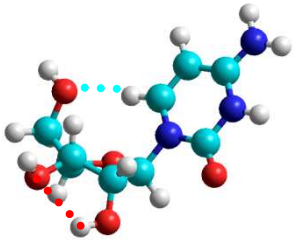
N3H(1), C2′-endo
anti, gauche$^+$
0.0 kJ/mol



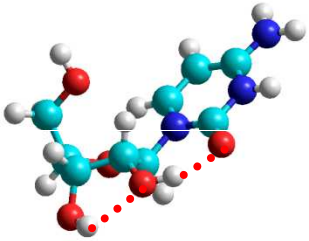
N3H(2), C3′-endo
anti, gauche$^+$
0.6 kJ/mol



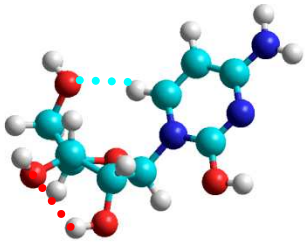
O2H(3), C2′-endo
anti, gauche$^+$
1.7 kJ/mol



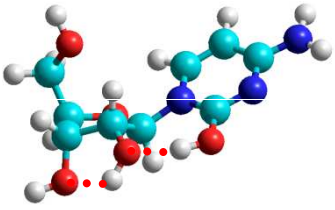
N3H(4), C3′-endo
anti, gauche$^+$
2.1 kJ/mol



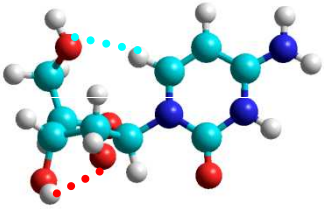
N2H(5), C2′-endo
anti, gauche$^+$
3.7 kJ/mol



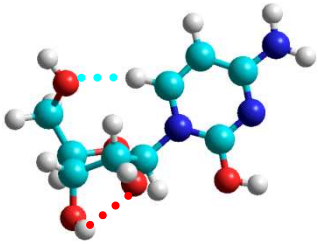
O2H(6), C3′-endo
anti, gauche$^+$
4.1 kJ/mol



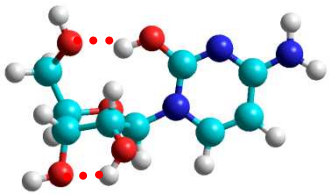
O2H(7), C2′-endo
anti, gauche$^+$
5.4 kJ/mol



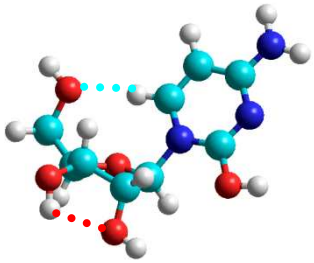
N3H(8), C2′-endo
anti, gauche$^+$
6.3 kJ/mol



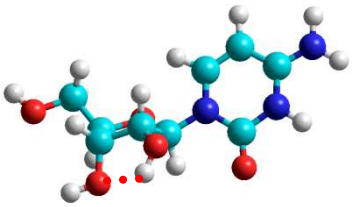
O2H(9), C2′-endo
anti, gauche$^+$
6.8 kJ/mol



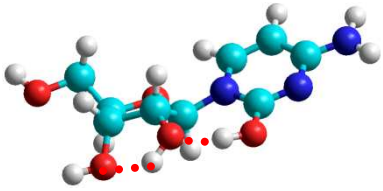
O2H(10), C2′-endo
syn, gauche$^+$
7.3 kJ/mol



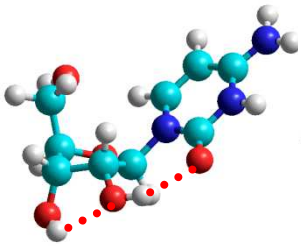
O2H(11), C3′-endo
anti, gauche$^+$
8.7 kJ/mol



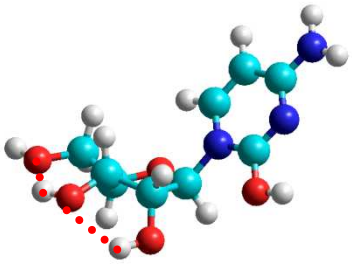
N3H(12), C1′-exo
anti, gauche$^+$
12.7 kJ/mol



O2H(13), C1′-exo
anti, gauche$^+$
13.9 kJ/mol



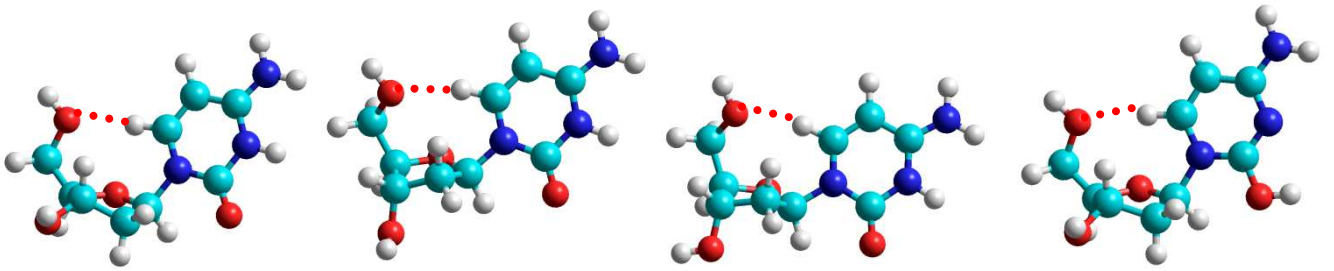
N3H(14), C3′-exo
anti, gauche$^+$
14.6 kJ/mol



O2H(15), C4′-exo
anti, gauche$^-$
16.6 kJ/mol

H$^+$(Cyd)

Figure S5.

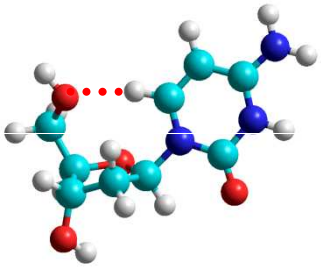


N3H(1), C3′-endo
anti, gauche$^+$
0.0 kJ/mol

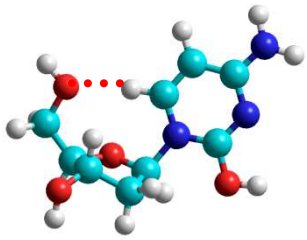
N3H(2), C2′-endo
anti, gauche$^+$
0.5 kJ/mol

N3H(3), C2′-endo
anti, gauche$^+$
1.7 kJ/mol

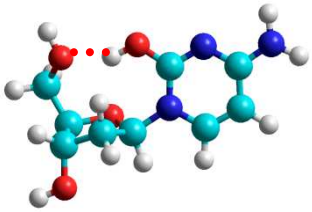
O2H(4), C3′-endo
anti, gauche$^+$
4.8 kJ/mol



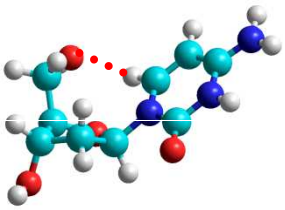
N3H(5), C2′-endo
anti, gauche$^+$
10.3 kJ/mol



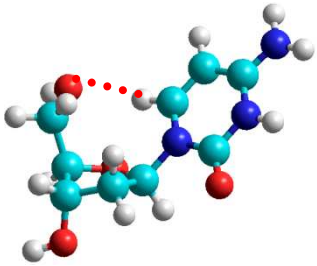
O2H(6), C3′-endo
anti, gauche$^+$
10.6 kJ/mol



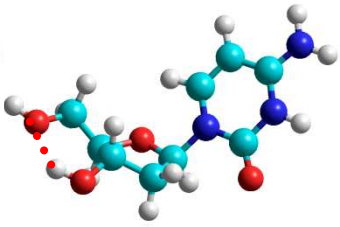
O2H(7), C2′-endo
syn, gauche$^+$
11.1 kJ/mol



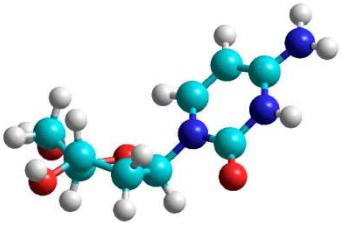
N3H(8), C3′-exo
anti, trans
12.2 kJ/mol



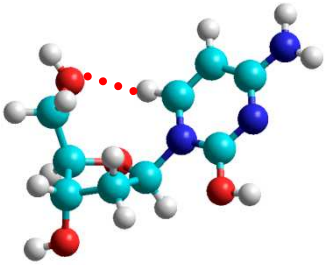
N3H(9), C3′-exo
anti, trans
12.7 kJ/mol



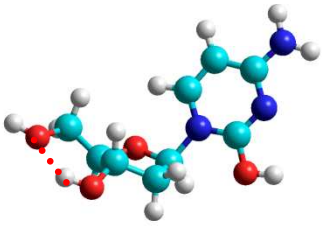
N3H(10), C3′-endo
anti, gauche$^-$
15.5 kJ/mol



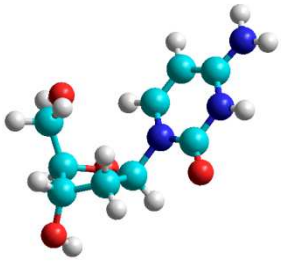
N3H(11), C4′-exo
anti, trans
16.6 kJ/mol



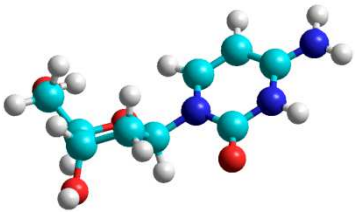
O2H(12), C3′-exo
anti, trans
18.0 kJ/mol



O2H(13), C3′-endo
anti, gauche$^-$
19.0 kJ/mol



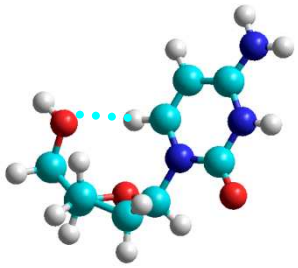
N3H(14), C3′-exo
anti, trans
20.1 kJ/mol



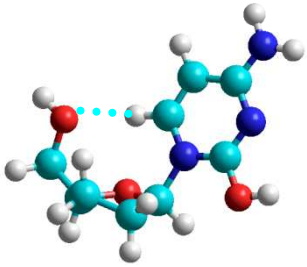
N3H(15), C1′-exo
anti, trans
21.8 kJ/mol

H^+(dCyd)

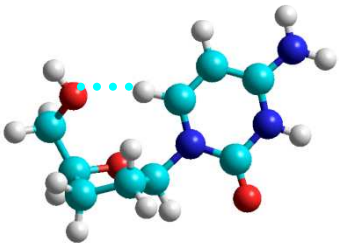
Figure S5.



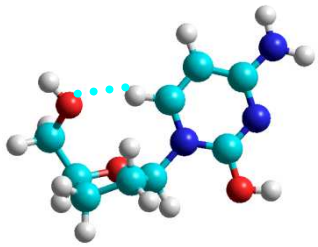
N3H(1), C3′-endo
anti, gauche$^+$
0.0 kJ/mol



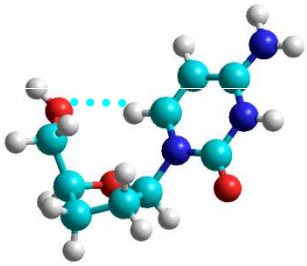
O2H(2), C3′-endo
anti, gauche$^+$
5.0 kJ/mol



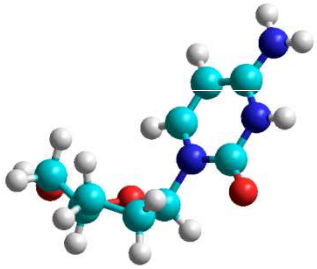
N3H(3), C2′-endo
anti, gauche$^+$
6.4 kJ/mol



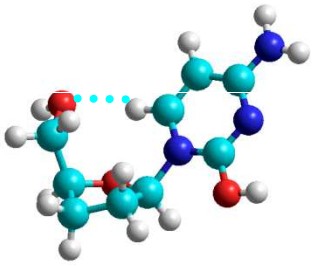
O2H(4), C2′-endo
anti, gauche$^+$
10.4 kJ/mol



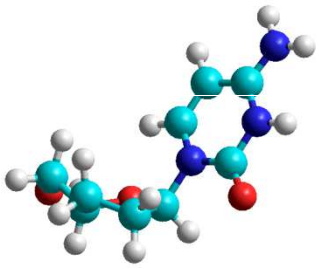
N2H(5), C3′-exo
anti, trans
16.1 kJ/mol



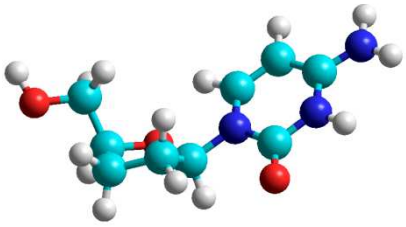
N3H(6), C3′-endo
anti, trans
18.0 kJ/mol



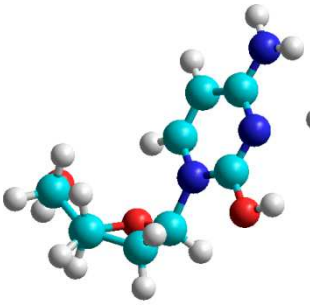
O2H(7), C3′-exo
anti, trans
20.9 kJ/mol



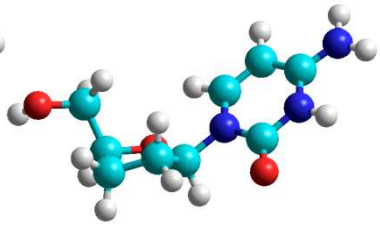
N3H(8), C3′-endo
anti, trans
21.7 kJ/mol



N3H(9), C2′-endo
anti, gauche$^-$
21.9 kJ/mol



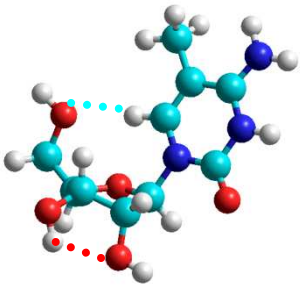
O2H(10), C3′-endo
anti, trans
22.3 kJ/mol



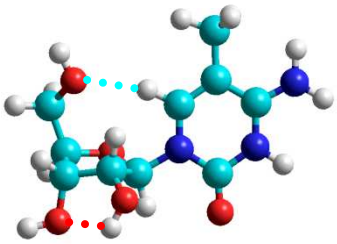
N3H(11), C2′-endo
anti, gauche$^-$
23.3 kJ/mol

H^+(ddCyd)

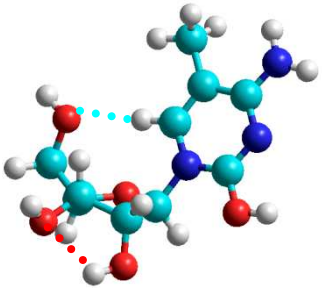
Figure S5.



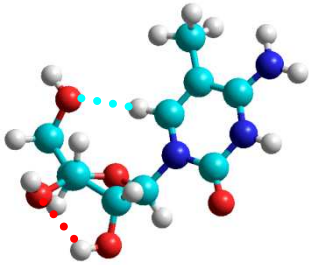
N3H(1), C3′-endo
anti, gauche$^+$
0.0 kJ/mol



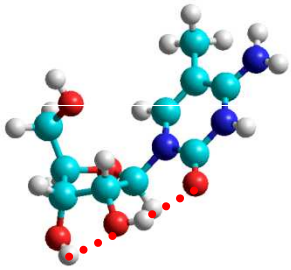
N3H(2), C2′-endo
anti, gauche$^+$
0.0 kJ/mol



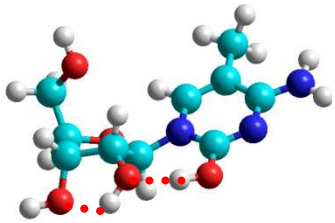
O2H(3), C3′-endo
anti, gauche$^+$
1.0 kJ/mol



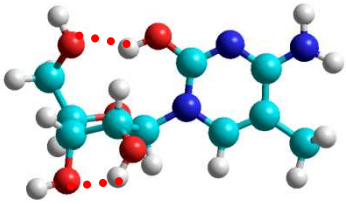
N3H(4), C3′-endo
anti, gauche$^+$
2.3 kJ/mol



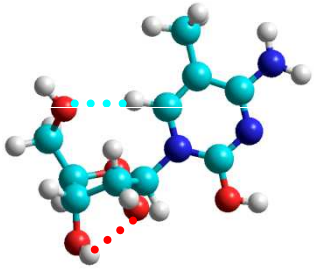
N3H(5), C3′-exo
anti, gauche$^+$
2.9 kJ/mol



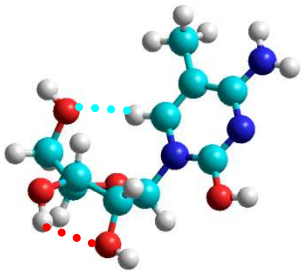
O2H(6), C2′-endo
anti, gauche$^+$
2.9 kJ/mol



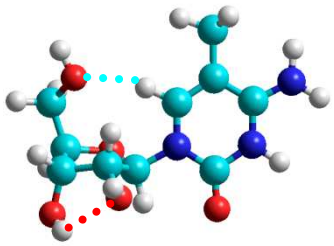
O2H(7), C2′-endo
syn, gauche$^+$
3.4 kJ/mol



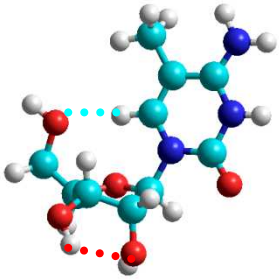
O2H(8), C2′-endo
anti, gauche$^+$
3.6 kJ/mol



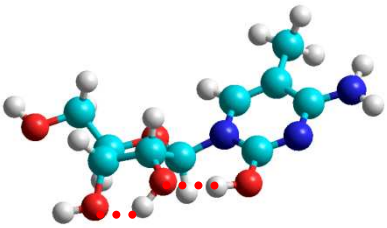
O2H(9), C3′-endo
anti, gauche$^+$
4.9 kJ/mol



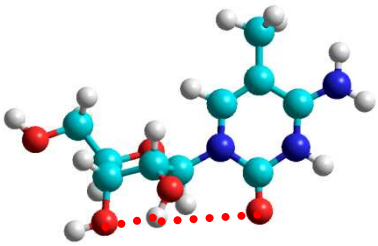
N3H(10), C2′-endo
anti, gauche$^+$
6.4 kJ/mol



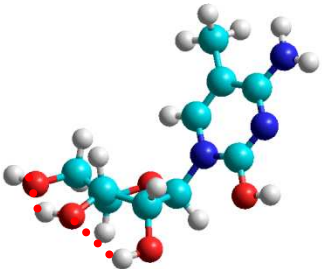
N3H(11), C2′-endo
anti, gauche$^+$
9.2 kJ/mol



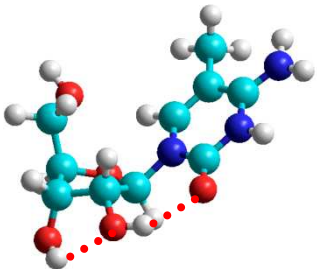
O2H(12), C1′-exo
anti, gauche$^-$
10.9 kJ/mol



N3H(13), C1′-exo
anti, gauche$^-$
12.3 kJ/mol



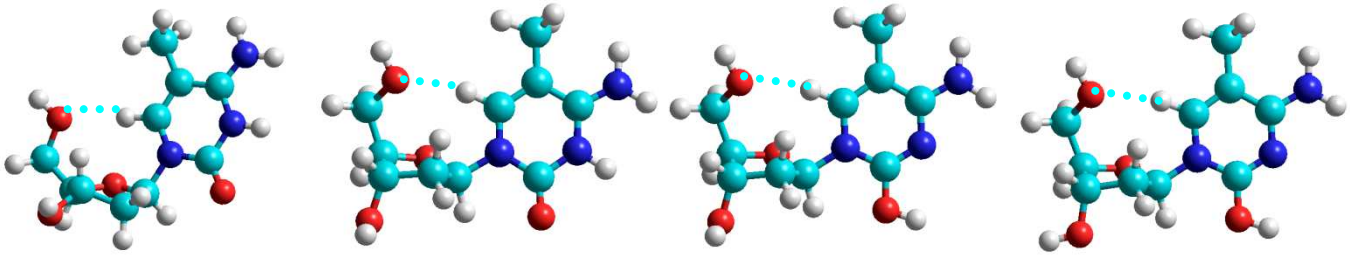
O2H(14), C4′-exo
anti, gauche$^-$
12.7 kJ/mol



N3H(15), C3′-exo
anti, trans
13.3 kJ/mol

$H^+(m^5Cyd)$

Figure S5.

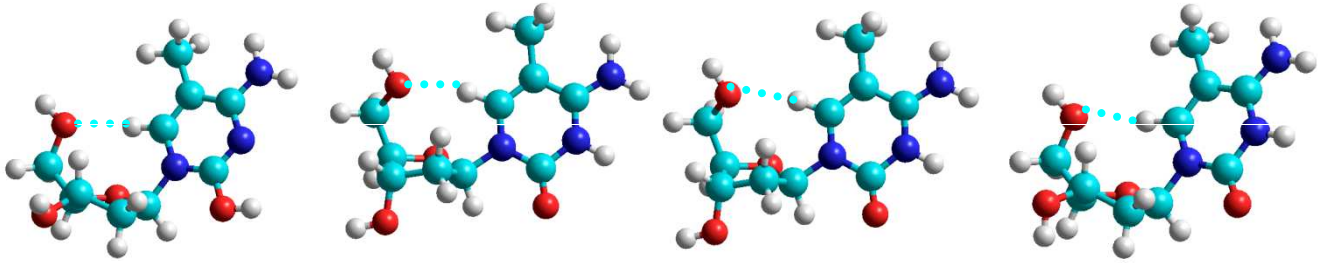


N3H(1), C3′-endo *anti*, gauche$^+$ 0.0 kJ/mol

N3H(2), C2′-endo *anti*, gauche$^+$ 0.1 kJ/mol

O2H(3), C2′-endo *anti*, gauche$^+$ 0.6 kJ/mol

O2H(4), C2′-endo *anti*, gauche$^+$ 1.2 kJ/mol

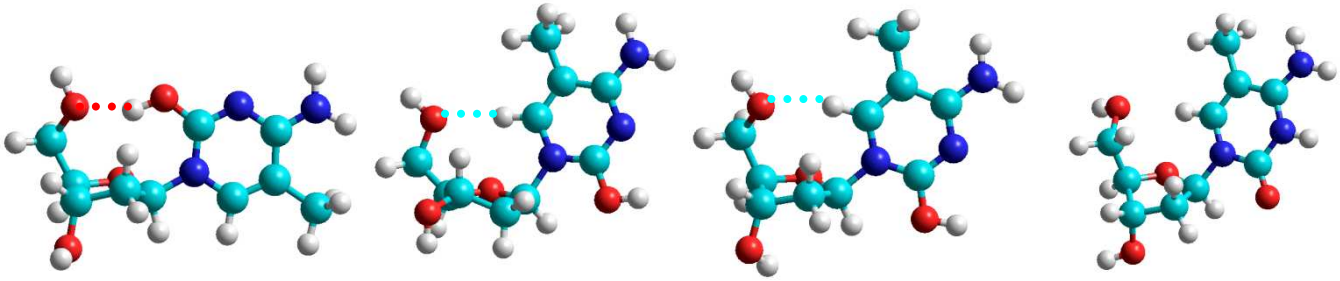


O2H(5), C3′-endo *anti*, gauche$^+$ 1.5 kJ/mol

N3H(6), C2′-endo *anti*, gauche$^+$ 1.6 kJ/mol

N3H(7), C2′-endo *anti*, gauche$^+$ 1.6 kJ/mol

N3H(8), C3′-endo *anti*, gauche$^+$ 4.6 kJ/mol

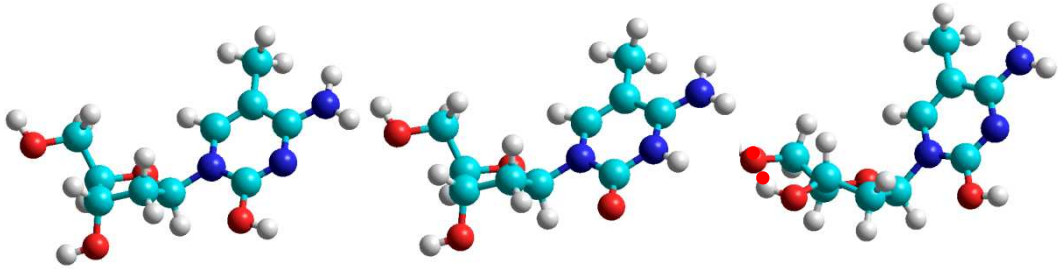


O2H(9), C2′-endo *syn*, gauche$^+$ 6.5 kJ/mol

O2H(10), C3′-endo *anti*, gauche$^+$ 7.0 kJ/mol

O2H(11), C2′-endo *anti*, gauche$^+$ 11.3 kJ/mol

N3H(12), C3′-exo *anti*, trans 11.8 kJ/mol



O2H(13), C2′-endo *anti*, gauche$^-$ 12.3 kJ/mol

N3H(14), C2′-endo *anti*, gauche$^-$ 13.0 kJ/mol

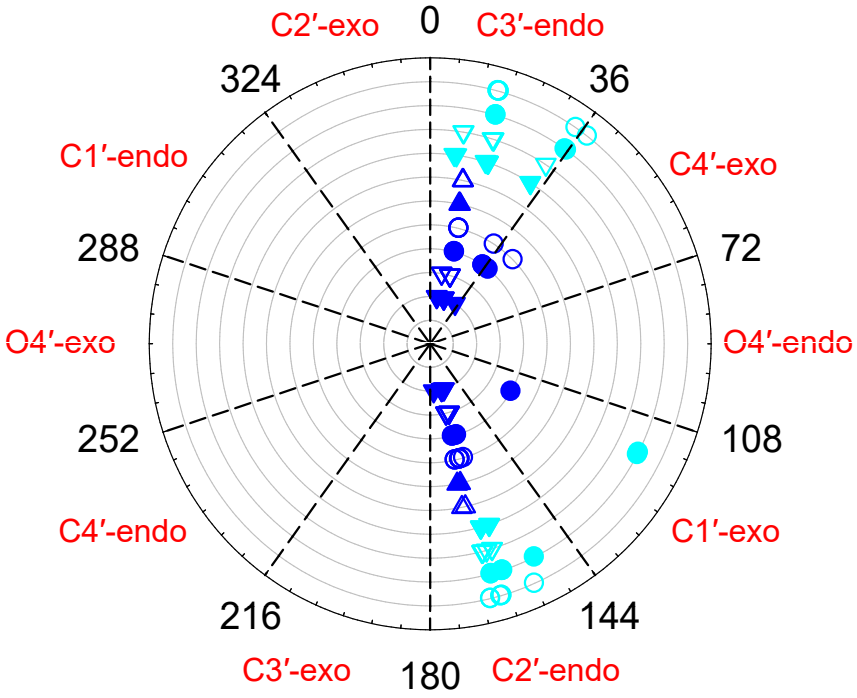
O2H(15), C3′-endo *anti*, gauche$^-$ 14.6 kJ/mol

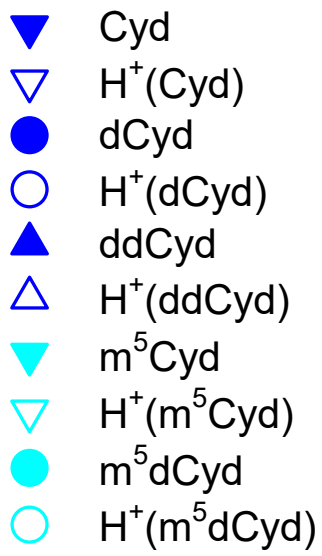
$H^+(m^5dCyd)$

Figure S6.

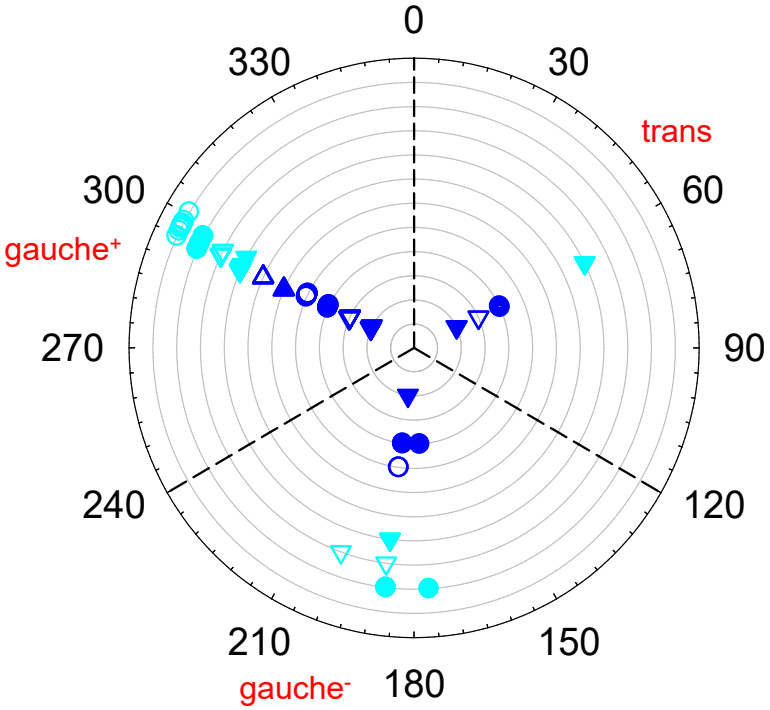
(*x*Cyd)H⁺(*x*Cyd)

Pseudorotation Phase Angle





5′-Hydroxy Orientation



Glycosidic Bond Angle

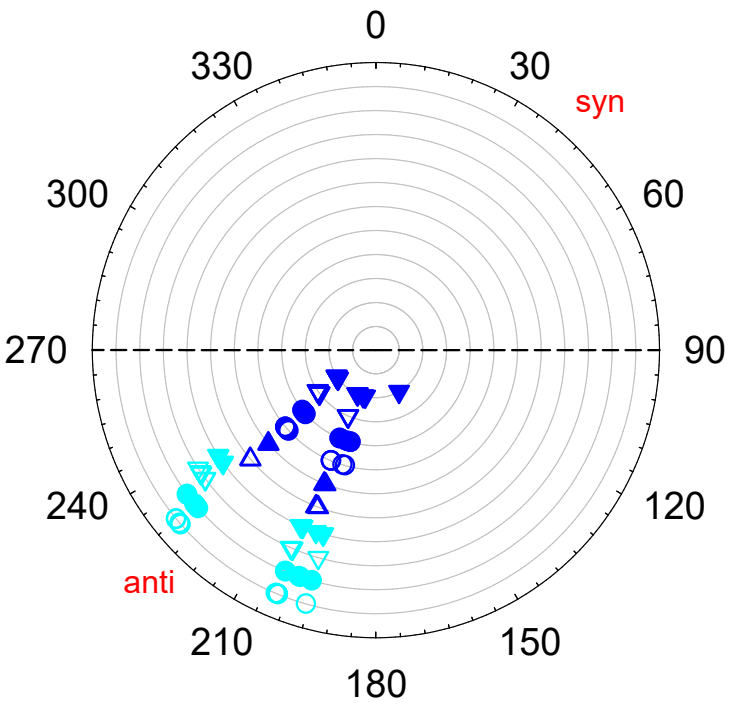


Figure S7.

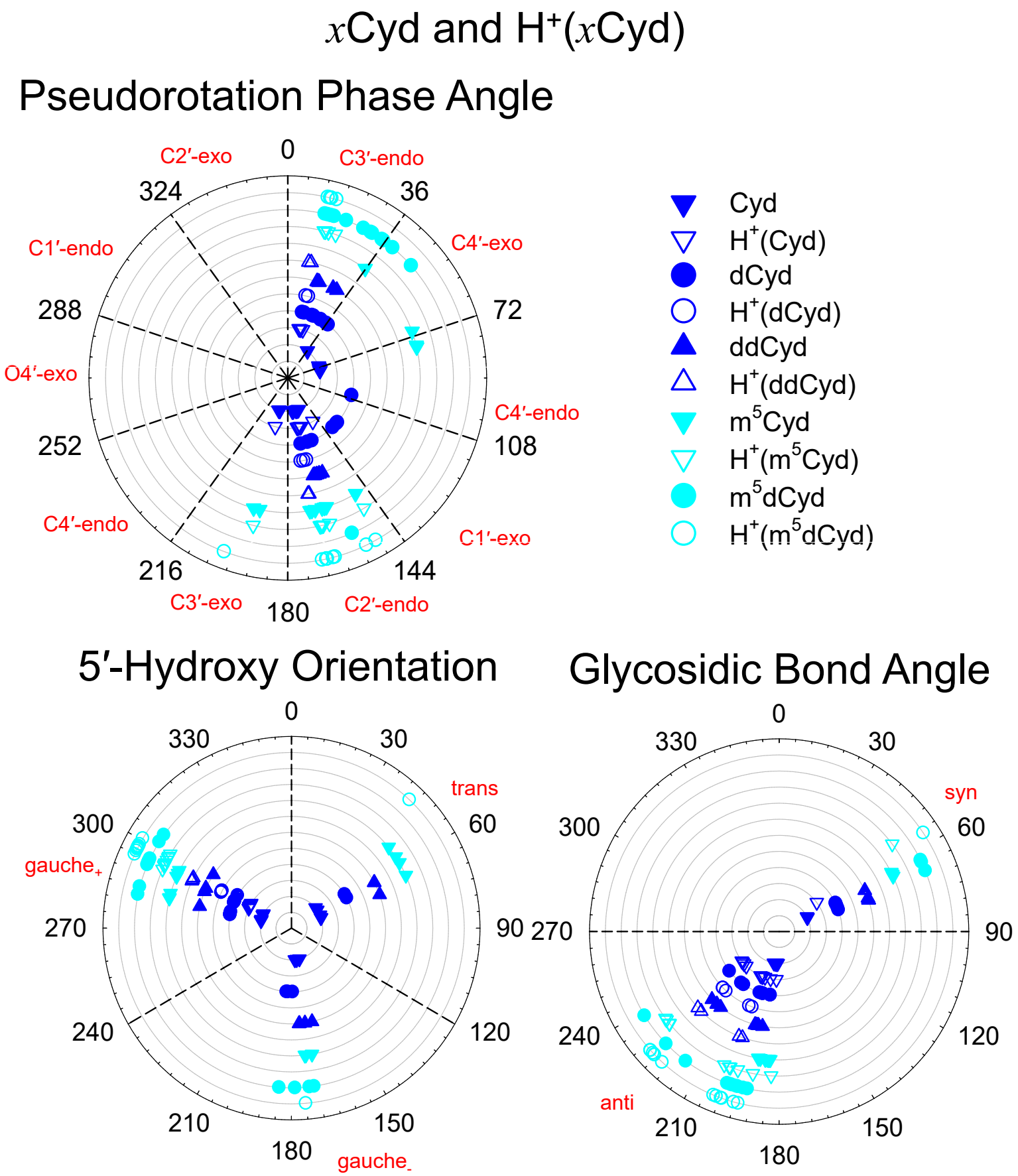
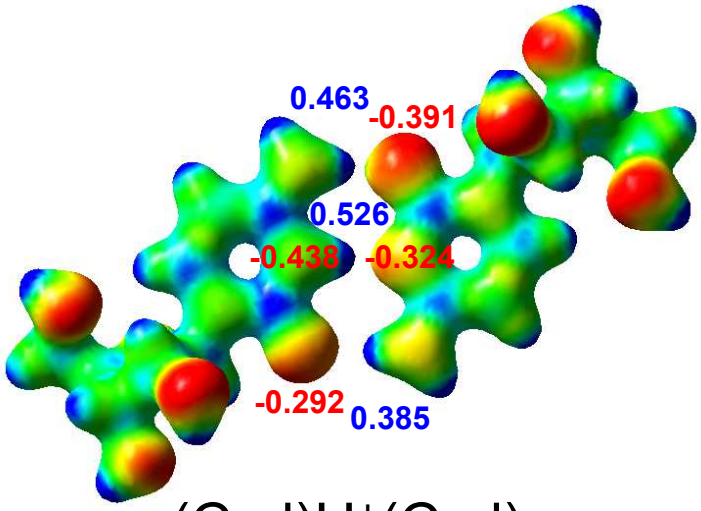
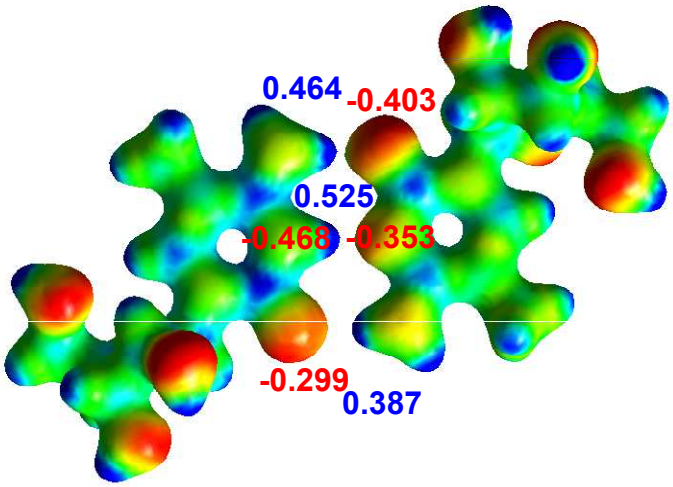

xCyd and H+(xCyd)
Pseudorotation Phase Angle
0
C3′-endo
36
C4′-exo
72
C4′-endo
108
C1′-exo
144
C2′-endo
180
C3′-exo
216
C4′-endo
252
O4′-exo
288
C1′-endo
324
C2′-exo
Cyd
H+(Cyd)
dCyd
H+(dCyd)
ddCyd
H+(ddCyd)
m5Cyd
H+(m5Cyd)
m5dCyd
H+(m5dCyd)
5′-Hydroxy Orientation
0
30
trans
60
90
120
gauche-
150
180
210
240
270
gauche+
300
330
Glycosidic Bond Angle
0
30
syn
60
90
120
150
180
anti
210
240
270
300
330

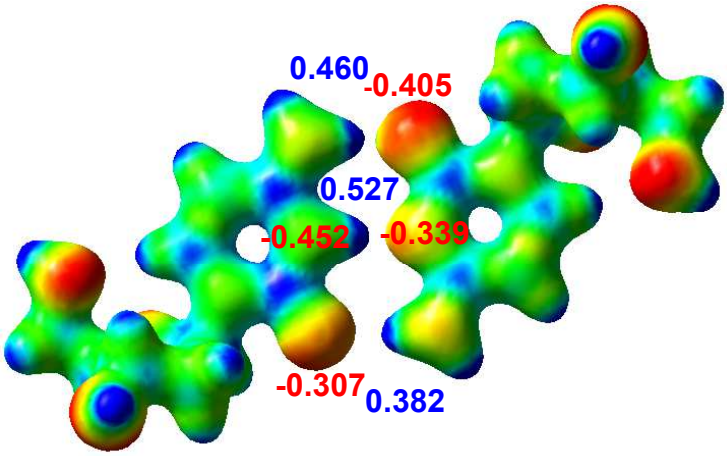
Figure S8.



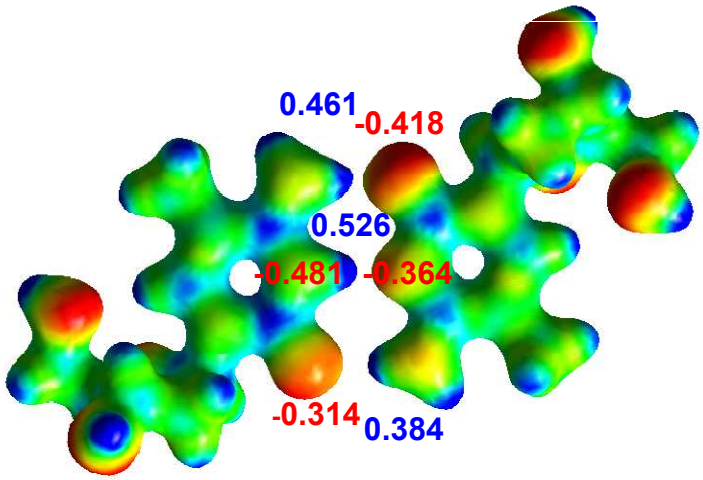
(Cyd)H$^+$(Cyd)



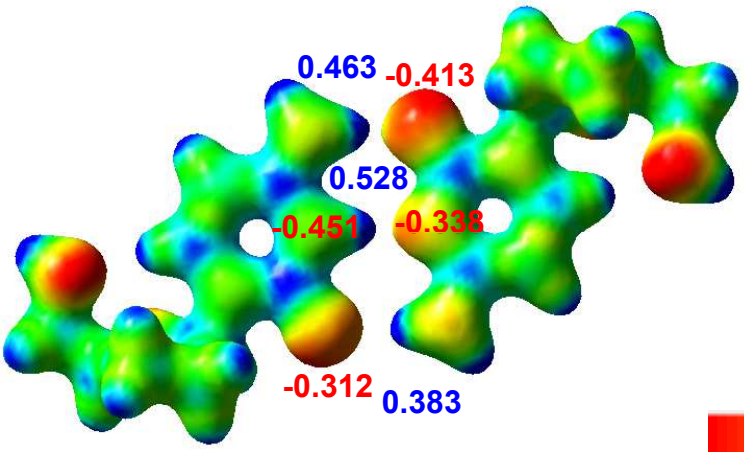
(m^{5}Cyd)H$^+$(m^{5}Cyd)



(dCyd)H$^+$(dCyd)



(m^{5}dCyd)H$^+$(m^{5}dCyd)



(ddCyd)H$^+$(ddCyd)

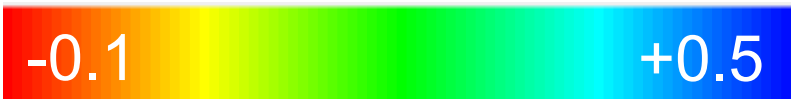
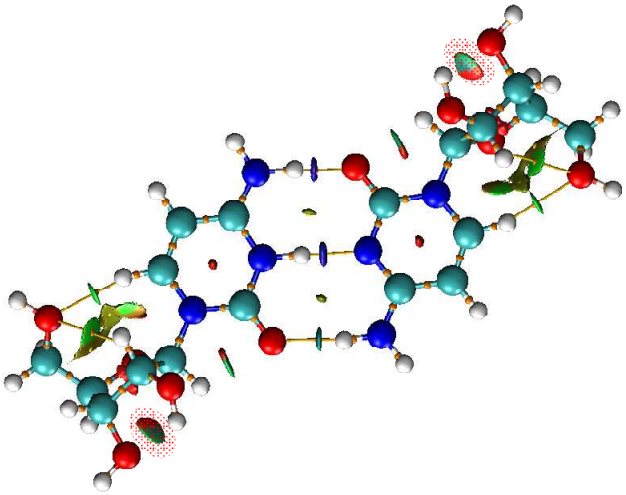
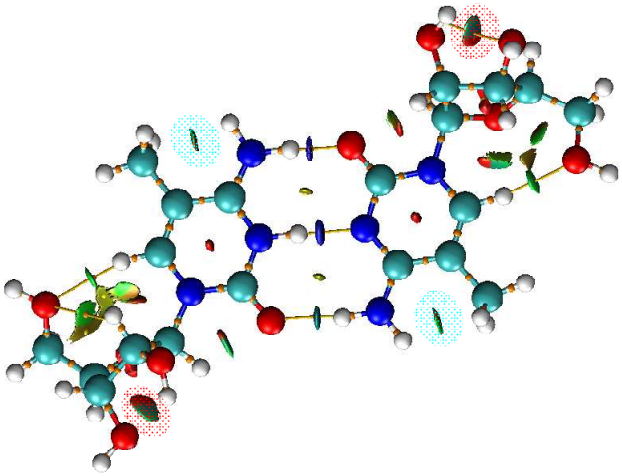


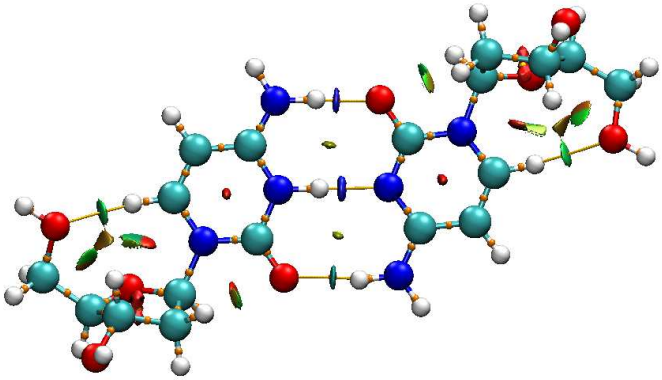
Figure S9.



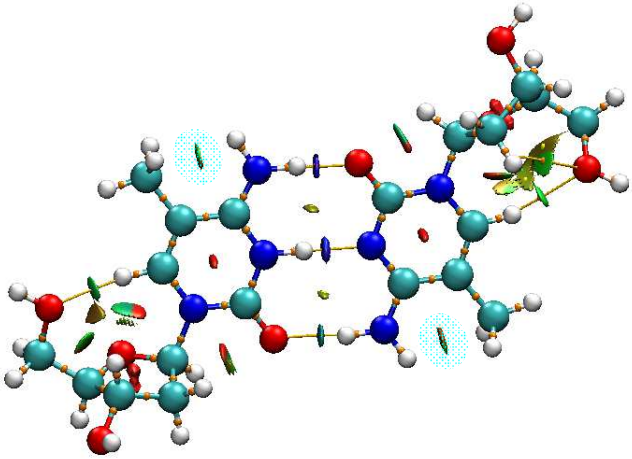
(Cyd)H$^+$(Cyd)



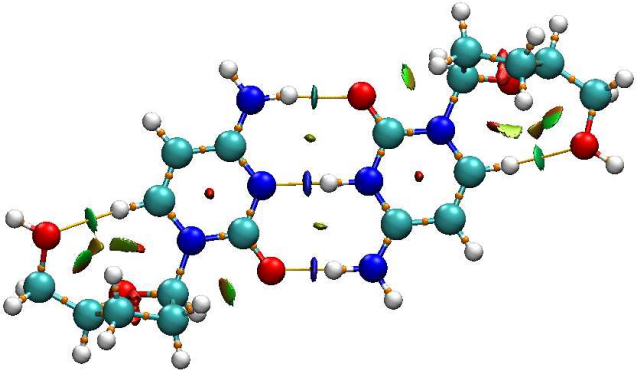
(m^{5}Cyd)H$^+$(m^{5}Cyd)



(dCyd)H$^+$(dCyd)

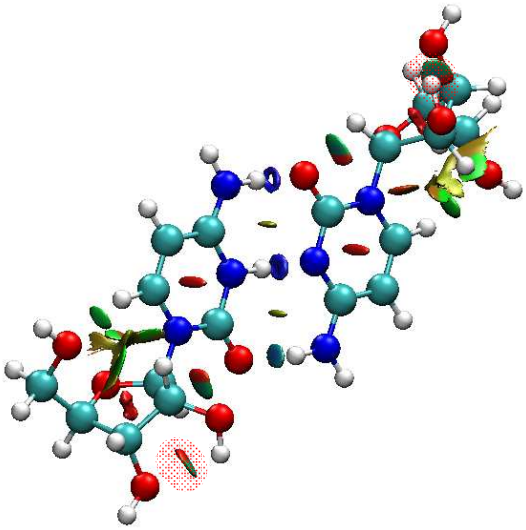


(m^{5}dCyd)H$^+$(m^{5}dCyd)

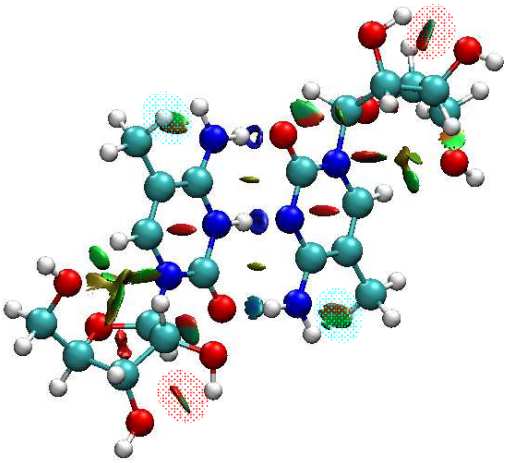


(ddCyd)H$^+$(ddCyd)

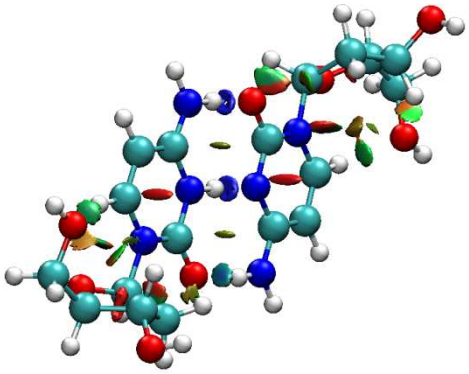
Figure S9.



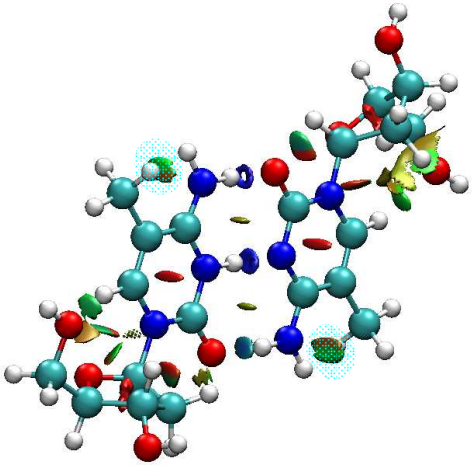
(Cyd)H$^+$(Cyd)



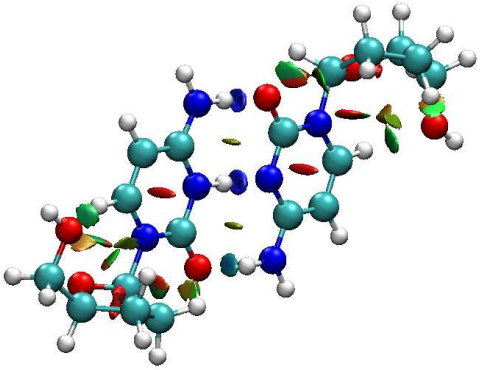
(m^{5}Cyd)H$^+$(m^{5}Cyd)



(dCyd)H$^+$(dCyd)

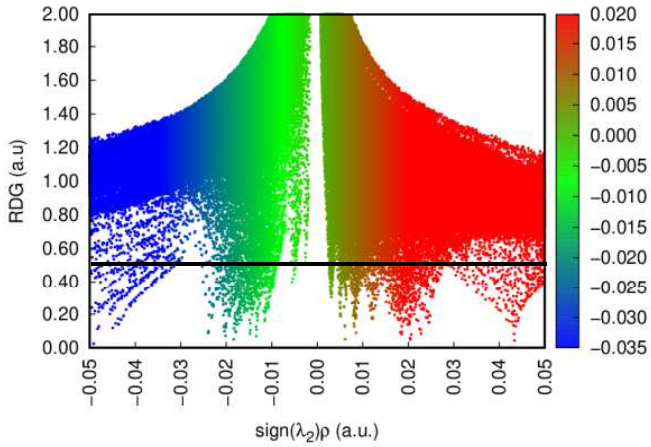


(m^{5}dCyd)H$^+$(m^{5}dCyd)

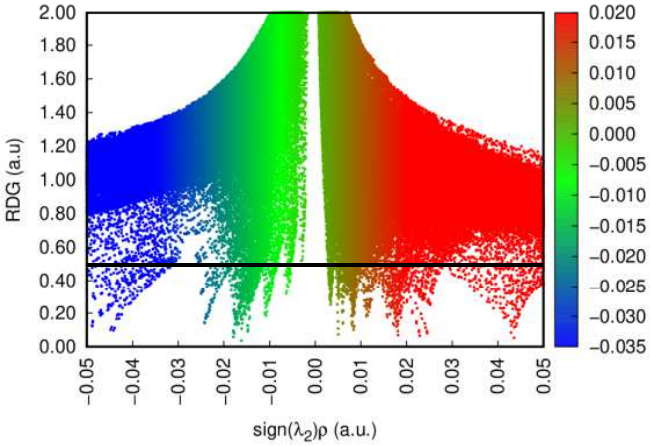


(ddCyd)H$^+$(ddCyd)

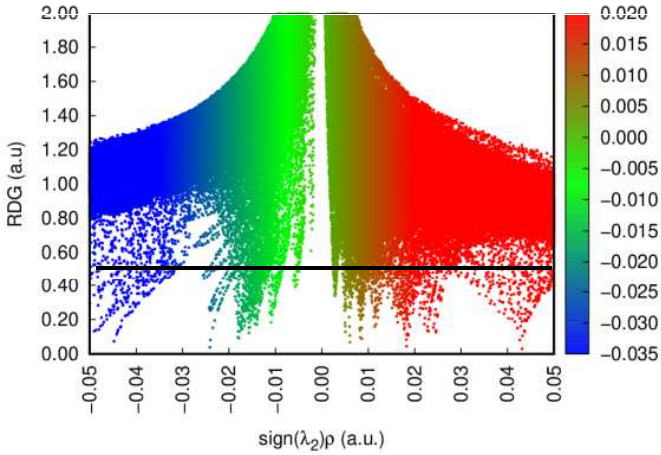
Figure S10.



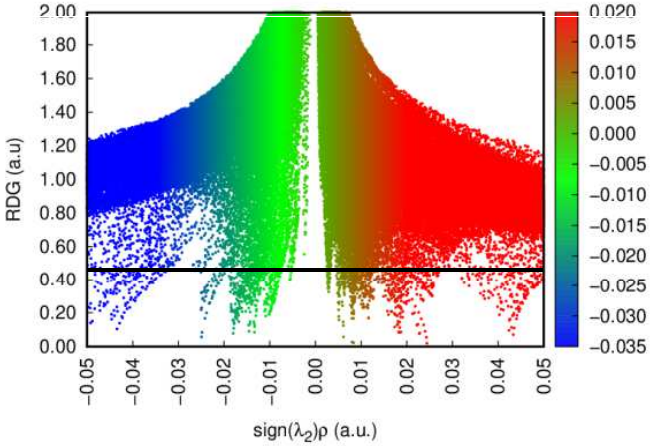
(Cyd)H⁺(Cyd)



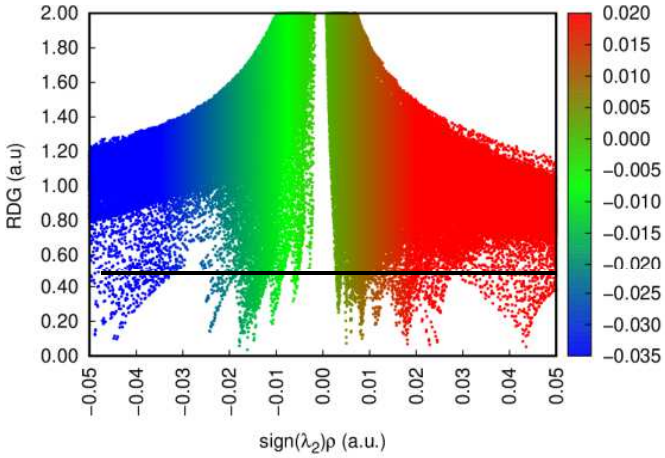
(m⁵Cyd)H⁺(m⁵Cyd)



(dCyd)H⁺(dCyd)

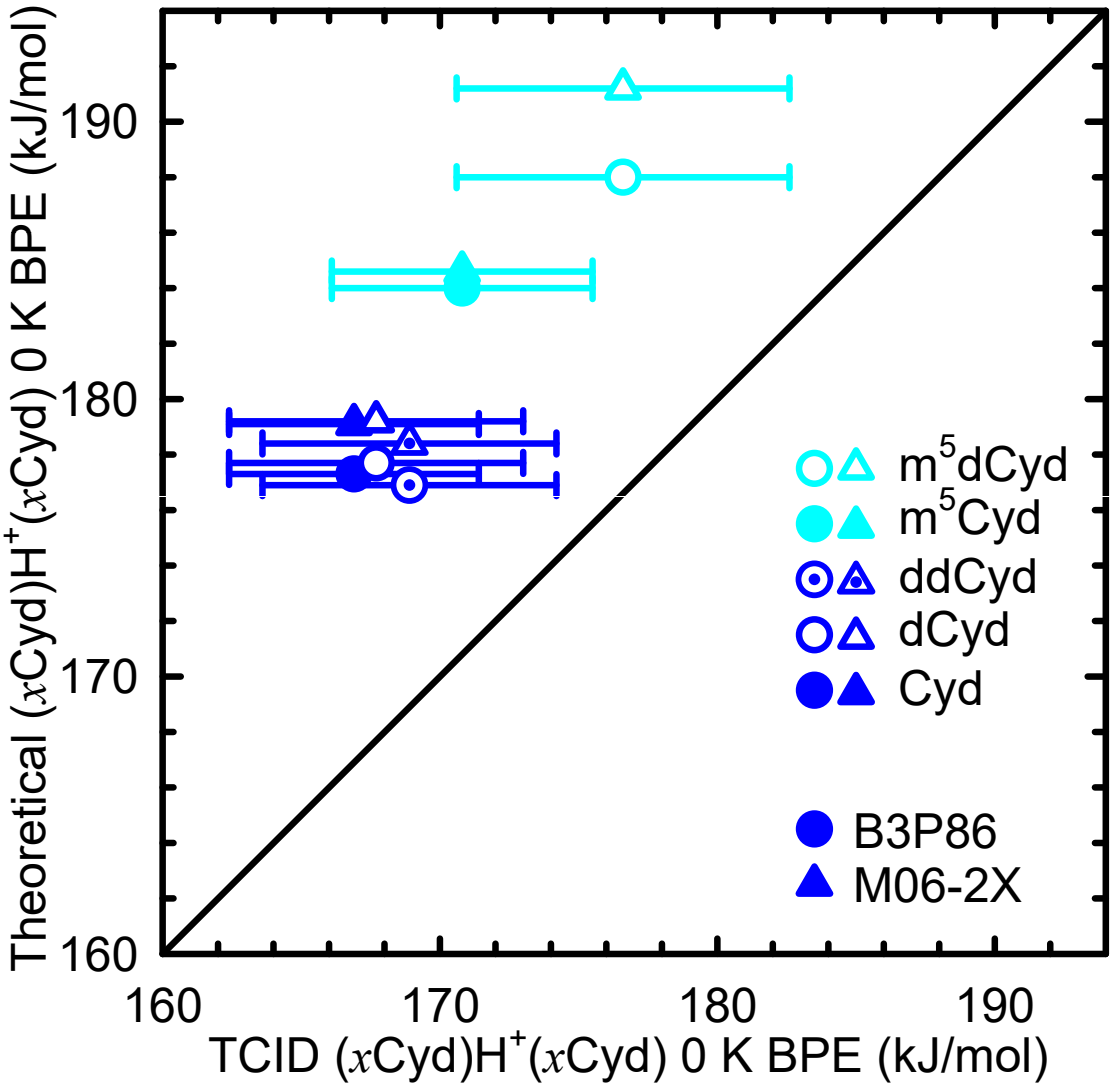


(m⁵dCyd)H⁺(m⁵dCyd)



(ddCyd)H⁺(ddCyd)

Figure S11.


Theoretical (xCyd)H$^+$(xCyd) 0 K BPE (kJ/mol)
TCID (xCyd)H$^+$(xCyd) 0 K BPE (kJ/mol)
160
170
180
190
m^{5}dCyd
m^{5}Cyd
ddCyd
dCyd
Cyd
B3P86
M06-2X