GAS PHASE STUDIES OF HYPOXANTHINE

by

XUEJUN SUN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Chemistry and Chemical Biology

written under the direction of

Professor Jeehiun K. Lee

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

[January, 2010]

ABSTRACT OF THE DISSERTATION

GAS PHASE STUDIES OF HYPOXANTHINE

By XUEJUN SUN

Dissertation Director:
Professor Jeehiun K. Lee

Hypoxanthine is not only a naturally occurring nucleobase in tRNA but also a damaged one in DNA arising from oxidative deamination of adenine. This thesis describes studies of the intrinsic reactivity of hypoxanthine as a free base and the stability of DNA duplexes containing hypoxanthine in the gas phase versus in solution by calculation and mass spectrometry methods.

Firstly, the free base of hypoxanthine is studied as a damaged nucleobase. Hypoxanthine is one of the damaged nucleobases that can be excised by Alkyladenine DNA glycosylase (AAG) in humans. To understand the intrinsic properties of hypoxanthine, we examined the gas phase acidity and proton affinity using quantum mechanical calculations and gas-phase mass spectrometric experimental methods. We find that the N9-H of hypoxanthine is more acidic than that of adenine and guanine, pointing to a way by which AAG may discriminate damaged bases from normal bases. We hypothesize that AAG may cleave certain damaged nucleobases as anions and the active site may take advantage of a nonpolar environment to favor deprotonated hypoxanthine as a better leaving group than adenine and guanine. The acidities of AAG

substrates have been compared with those of normal bases by calculations and Cooks kinetics method. In addition, the acidity differences between damaged and normal bases are enhanced in the gas phase when we compare them with the acidities in solution. These results support our hypothesis.

Secondly, to understand the effect of hypoxanthine on DNA stability, we study a series of 9-mer DNA duplexes with the sequence 5'-d(GGTT**X**TTGG)-3'/3'-d(CCAA**Y**AACC)-5', where the central base X or Y = adenine (A), guanine (G), thymine (T), cytosine (C) and hypoxanthine (H). Comparison of the duplex stability in the gas phase versus in solution indicates that hypoxanthine has much less of a destabilizing effect in the gas phase versus in solution, relative to the normal complementary duplexes. The biological implications of these results are discussed in the context of hypoxanthine both as a universal base and as a damaged base.

DEDICATION

This Work is dedicated to my parents, Qichen Sun and Maozheng Xu, my husband,

Qianwei Han, and my son, Eric Bolin Han.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

Chapter 1 Background

1.1 Introduction

1.1.1 Nucleobases and their acidities and proton affinities in the gas phase and solution phase

Nucleobases are "the building blocks" of DNA，of which adenine (A), guanine (G), cytosine (C) and thymine (T) (**Figure 1.1**) are the four normal ones. The intrinsic acidity and basicity of nucleobases and their derivatives are of high importance to understand different issues in biological systems. They affect greatly the formation and stabilization of the DNA duplex. In DNA duplexes, two complementary nucleobases are held together by NH-O and NH-N hydrogen bonds to form a base pair, which is dependent on the intrinsic basicity of the acceptor atoms and on the acidity of the NH donor groups.[1,2] Also, the intrinsic acidity and basicity of nucleobases play significant roles in biosynthetic reactions for which nucleobases are substrates. The intrinsic reactivity of nucleobases is considerably influenced by their acidity and basicity.[3-5]
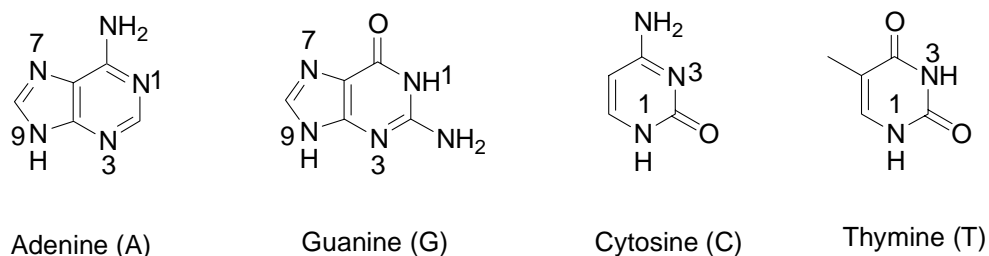
Figure 1.1 Structures of four nucleobases in DNA.

The gas phase is a unique medium for understanding intrinsic properties and reactivity of biological molecules without complications from solvent. Most biological media are not aqueous in Nature. For example, the interior of proteins is shown to be nonpolar, which causes shifts in acidity and basicity and changes in reactivity compared to behavior in aqueous solution.[6-8] The gas phase provides a 'clean' environment to examine intrinsic reactivity and to extrapolate the effects of media.[7,9] In addition, the gas phase data provide a bridge connecting calculations with condensed phase data.

Nucleobases have multiple acidic and basic sites. In the gas phase, the acidity is defined as the enthalpy change associated with deprotonation of a given chemical species HA to form H^+ and A^- ions (**Eq 1.1**). The proton affinity (PA) is defined as the negative value of the enthalpy change associated with protonation of a chemical species B to form HB^+ (**Eq 1.2**).

$$HA \longrightarrow H^+ + A^- \qquad \textbf{(Eq. 1.1)}$$

$$B + H^+ \longrightarrow HB^+ \qquad \textbf{(Eq. 1.2)}$$

Table 1.1 and **Table 1.2** list the acidities and proton affinities, respectively, of a series of prototypical compounds in the gas phase.[10] The gas phase acidities run from about 314 to 417 kcal mol^{-1}, with a higher ΔH°_{acid} value corresponding to a lower acidity (**Table 1.1**). The gas phase proton affinities range roughly from 130 to 291 kcal mol^{-1}, and a lower PA value indicates a lower proton affinity (**Table 1.2**).

Table 1.1 Representative gas-phase acidities.

Class of molecules	Specific example	ΔH^o_{acid} (kcal mol^{-1})
Alkanes	CH_4	416.7
Vinylic	C_2H_4	409.4
Amines	NH_3	403.6
Aromatic	C_6H_6	400.7
Allytic	CH_2CHCH_3	390.8
Water	H_2O	390.7
Alcohols	CH_3OH	381.7
Benzylic	$PhCH_3$	380.8
Acetylenes	C_2H_2	377.8
Nitriles	CH_3CN	372.9
α to Carbonyls	CH_3COCH_3	369.1
Thiols	CH_3SH	356.9
Phenols	$PhOH$	349.2
Carboxylic acids	CH_3COOH	348.6
Benzoic acids	$PhCOOH$	340.2
Strong acids	HI	314.4

Table 1.2 Representative gas-phase proton affinities.

Class of molecules	Specific example	PA (kcal mol^{-1})
Alkanes	CH_4	129.9
Strong acids	HI	150.0
Acetylene	C_2H_2	153.3
Vinylic	C_2H_4	162.6
Water	H_2O	165.2
Formaldehyde	CH_2O	170.4
Aromatic	C_6H_6	179.3
Alcohols	CH_3OH	180.3
Carboxylic acids	CH_3COOH	187.3
Thiols	C_2H_5SH	188.7
Nitriles	$CH_3CNC_2H_5CN$	189.8
Benzonitrile	C_6H_5CN	194.0
Ethers	$CH_3OCH_2CH_2CH_3$	194.8
Amides	C_2H_5NO	203.5
Amines	NH_3	204.0
Aniline	$C_6H_5NH_2$	210.9
Pyridine	C_5H_5N	222.3
Hydride	HNa	261.7
Monoxide	BaO	290.5

Our research group has pursued the acidities and proton affinities of normal nucleobases theoretically and experimentally for some years. Mary Ann Kurinovich calculated and measured the acidity of uracil and uracil analogs;[11-13] Seema Sharma studied the acidity of adenine and adenine derivatives by calculations and experiments.[14,15] The acidity and proton affinity of damaged bases, particularly those modified by deamination and alkylation, are also important and remain unknown. Base damage can occur during DNA replication. Damaged bases usually prefer to hydrogen bond to different nucleobases than do their normal counterparts. For example, hypoxanthine, which is an adenine mutation, prefers to bind cytosine instead of thymine. DNA glycosylases are enzymes involved in base excision repair. They recognize and remove the damaged bases from DNA. By measuring the acidities and proton affinities of normal and damaged nucleobases, we seek to discover: 1) why damaged bases are inherently different from normal nucleobases; 2) any differences in the acidity and proton affinity of certain sites on damaged bases versus their normal counterparts; 3) why they hydrogen bond differently and how they contribute to mutagenicity; 4) how the damaged bases are recognized by glycosylases. In Chapters 2 and 3, we describe our studies of the acidity and proton affinity of hypoxanthine in the gas phase and compare the acidity of hypoxanthine with other nucleobases in different media.

1.1.2 Stabilities of DNA duplexes in the gas phase and solution phase

Deoxyribonucleic acid (DNA) contains and transfers genetic information in all cell life. DNA is the so-called "molecule of heredity", and consists of paired strands that entwine to form a double helix. Each strand is a long chain polymer of linked nucleotides;

each nucleotide is composed of a nucleobase, a sugar ring and a phosphate. There are four nucleobases in DNA: adenine (A), cytosine (C), guanine (G) and thymine (T).[16] DNA sequences are expressed as the abbreviated attached bases (A, C, G and T), from the 5' to 3' end of the sugar rings. The two strands of DNA are held together by hydrogen bonds between nucleobases on each strand, and this is called base pairing. More than half a century ago, Watson and Crick first ascertained that each base pairs with one single predetermined base: adenine to thymine, guanine to cytosine.[17] Thus these two base pairs are called "Watson-Crick base pairs" (**Figure 1.2**).
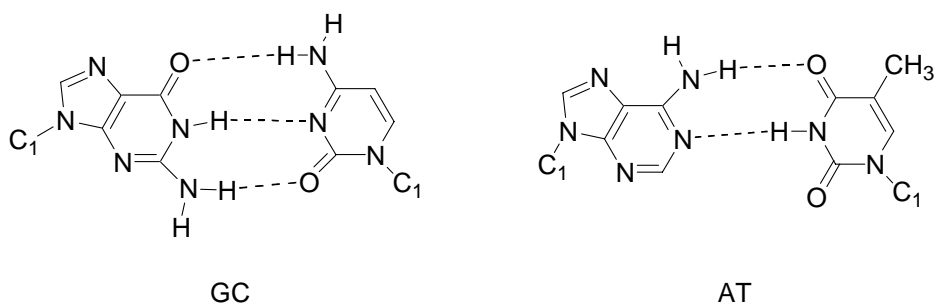
Figure 1.2 Watson-Crick GC and AT base pairs

Adjacent bases associate with each other electrostatically and these interactions are favored by the conformation of the DNA double helix. This sequence dependent interaction is called 'base stacking'.[18] During DNA replication, the cell's machinery is capable of unwinding a DNA double helix, and using each strand as a template to synthesize a new strand. Each new double helix is therefore identical to the previous one. Errors can occur in the course of replication and result in mutations when a base is incorrectly copied, inserted or skipped. Also, exposure to chemicals or radiation (UV light) can cause mutations.[19,20] Since mutations most often occur in DNA duplexes, it is of importance to study DNA duplex interactions and stabilities.

Gas phase methods have been extensively used to study DNA stabilities. Different interactions contribute to DNA stabilities, of which hydrogen bonding and base stacking are the most important ones. The involved forces include electrostatic interactions (Coulomb repulsion), van der Waals forces, and hydrophobic interactions.[21,22] In aqueous conditions, various solvent effects make the interactions even more complicated. Therefore the gas phase has been used to eliminate solvent effects and simplify the interactions contributing to DNA duplex stability. It is recognized that the main forces involved in the formation of DNA duplexes in the gas phase include charge repulsion, hydrogen bonding and base stacking.[23-25] The gas phase may also give us an idea of reactivity in the local nonpolar environment of protein-DNA interactions. Our group has successfully studied the stabilities of various series of complementary and mismatched DNA duplexes in the gas phase using LCQ with electrospray ionization (ESI).[26-28] In Chapters 4 and 5, we report our study of the stabilities of duplexes containing normal and damaged nucleobases in the gas phase and compare their stabilities with those in the solution phase.

1.2 Instrumentation

1.2.1 FTMS

Fourier transform ion cyclotron resonance mass spectrometry (FT-ICR MS or FTMS) was invented by Comisarow and Marshall in 1974.[29] Since then FTMS has become a popular instrument because of its high resolution and mass accuracy. Coupled with soft ionization methods, electrospray ionization (ESI)[30] and matrix-assisted laser desorption/ionization (MALDI),[31] FTMS has become a versatile instrument that can be

used to analyze both volatile and nonvolatile samples, including small molecules such as drug metabolites and large biomolecules such as peptides and proteins in the gas phase.[32,33]

The basic principle of FTMS is the cyclotron motion of charged ions in magnetic and electric fields. When a charged molecule enters a constant magnetic field *B* with a velocity *v*, the ion then can move on a circular trajectory in a plane perpendicular to the magnetic field. There are two forces placed on the ion, Lorentz force F_1 (magnetic force) (**Eq. 1.3**) and centrifugal force (**Eq. 1.4**):

Lorentz force $\quad\quad\quad\quad F_1 = qvB$ **Eq. 1.3**

Centrifugal force $\quad\quad\quad F_2 = mv^2/r$ **Eq. 1.4**

where q is the charge on the ion, m is the mass of the ion and r is the radius of the circular orbit. When the two forces are balanced by each other (**Eq. 1.5**), ions can be stabilized on a circular trajectory with a frequency of f ($f = v/2\pi r$) (**Eq. 1.6**):

$$qvB = mv^2 / r$$ **Eq. 1.5**

$$f = qB/2\pi m$$ **Eq. 1.6**

When magnetic field B is a constant, cyclotron motion frequency f is determined by the mass-to-charge ratio: for example, ions with the same mass-to-charge ratio will move with the same frequency in a fixed magnetic field, and the frequency is independent of the initial velocity of the ions.
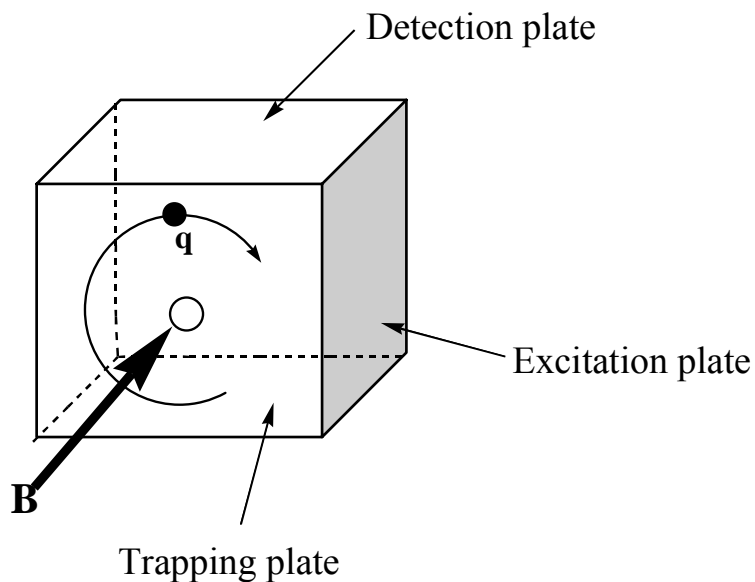
Figure 1.3 A cubic analyzer cell in FTMS

In FTMS, the analyzer cell is the essential component in which ions are trapped, analyzed and detected. Cubic and cylindrical open-end cells are two common types of analyzer cells. **Figure 1.3** shows the cubic analyzer cell which is composed of 6 plates. The pair of plates that are perpendicular to the applied magnetic field are called trapping plates. There is a hole in the center of the trapping plate from where ions or molecules enter the cell. When ions are moving parallel to the magnetic field, there is no magnetic field force applied on these ions. Therefore, ions can move along the magnetic field axis and drift out of the cell. However, if we apply a small potential (a positive potential for positive ions and a negative potential for negative ions) on the pair of trapping plates, ions can be trapped inside the cell.

There are two pairs of plates parallel to the magnetic field: one pair is called excitation plates and the other is called detection plates. In order to detect ions moving in the cell, those ions must be excited first. When a sinusoidal voltage is applied on the

excitation plates, an electric field perpendicular to the magnetic field is generated. If the ion cyclotron frequency is in resonance with the frequency of the applied field, the ions will continuously absorb the energy and convert it to kinetic energy which results in a continuous increase of radius and the ions move in a spiral trajectory.

Ions with the same mass-to-charge ratios are always grouped as "packets". When they are excited to a larger trajectory and get closer to the detection plates, they induce an image current on the detection plates. This induced current on the detection plates is digitized and amplified to a time domain signal, which is then converted to a frequency domain signal by Fourier transform. At last the frequency is translated to a mass-to-charge ratio by **Eq. 1.6**.

1.2.2 ESI and LCQ ion trap mass spectrometer

Electrospray ionization (ESI) is one of the most widely used ionization techniques in modern mass spectrometry, which ionizes analytes at atmospheric pressure.[34-37] It is especially useful in the analysis of polymer and biological macromolecules such as proteins/peptides and DNA/RNA. During the ESI process the analytes typically do not fragment, so it is a "soft" ionization technique. The electrospray process was first described by the Czech-American physicist John Zeleny in 1914.[38] He observed that charged droplets violently break into smaller droplets when the charge state reaches a critical point. In 1968, Malcolm Dole at Northwestern University proposed the concept of an ESI source.[39] In the 1980's Professor John Bennett Fenn at Yale University developed an ESI source for mass spectrometry and successfully applied it to protein analysis, a contribution that resulted in the 2002 Nobel Prize in Chemistry.[40]

Shown in **Figure 1.4** is a typical ESI process in positive mode[41]. The sample from HPLC or syringe is introduced into the source chamber through a capillary probe. At the tip of the probe, a strong electric potential (1 ~ 6kv) is applied to charge the analytes in the solvent. When the electric effect overcomes surface tension, the highly charged liquid drop at the tip takes a conical shape instead of a spherical shape.[42,43] The cone is called "Taylor Cone" after the name of Sir Geoffrey Ingram Taylor who first described the phenomena in 1964.[44] At onset of electrospray, small highly charged droplets are released as a fine jet from the tip of Taylor cone, and then quickly split up into a mist of smaller droplets which fly apart due to strong electric repulsion. At the same time, the solvent within the droplets evaporates very quickly and the droplets shrink, moving the ions within the droplets closer together, which increases Coulomb repulsion forces. At some point, the ions are close enough and their repulsion forces override the cohesion forces of the droplets. As a result, the droplets divide into finer droplets. The process repeats itself until a droplet becomes a single, charged molecule ion. The charge state of ions generated in ESI is greatly affected by the electrochemical process at the capillary tip and the charge accumulation process in the fine droplets. For example, many multiply charged ions of myoglobin at pH = 3 are formed in the droplets during the electrospray process.[45]
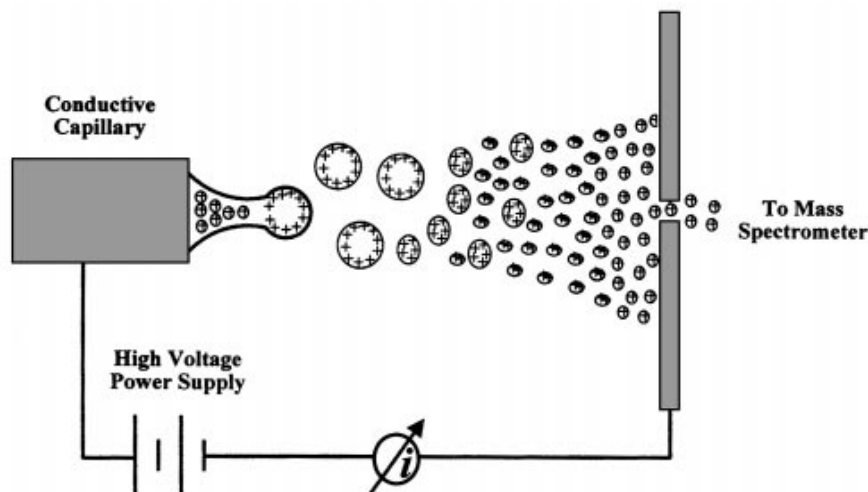
Figure 1.4 Diagram of Electrospray Ionization (Positive Mode)[41]

Though ESI has been used with mass spectrometry for over 20 years, there is still wide debate on the mechanism of the final formation of single ions. Two major mechanisms have been extensively discussed, the charged residue model and the ion evaporation model.[43] In the former model, the electrospray droplets undergo a cascade of division until the final droplets carry only one molecule or ion of analyte. After the final solvent evaporation, the single molecule or molecular ion forms with the charge of the droplet. In this model, a series of Coulomb fissions of droplets occurs until the final single ions form. In the latter ion evaporation model, the droplets first break up into finer droplets. When the radius of the daughter droplets reaches some critical point, the charge strength at the surface of the droplets is big enough to push a single, charged ion out. This releases the charge stress of the droplets and the evaporation of solvent continuous to the next critical point to release another single ion. This model proposes an ion evaporation process at certain critical points. Currently neither model can be proved exclusively. It is generally recognized that both kinds of ionization probably take place in most ESI applications.

ESI is probably the most popular ionization technique in modern mass spectrometry and is able to work efficiently in both positive and negative modes. Almost all polar, nonvolatile, large and small molecule analytes can be analyzed with an ESI source. ESI-MS can be easily automated and coupled with high performance liquid chromatography (HPLC) and capillary electrophoresis (CE). The ESI source has the following important characteristics which import various advantages. First, in an ESI process, the analytes remain intact and very little decomposition occurs even for very labile compounds. Second, ESI can generate multiply charged ions. This enables the analysis of compounds with high molecular weights using MS analyzers with regular mass ranges. Third, like most other sources, its sensitivity depends on sample concentration not on sample amount introduced to the source. Variants of ESI have been developed, including micro electrospray (μESI)[46] and nanospray (nESI)[47,48] using much lower flow rates and modified spray probes. Instruments with μESI and nESI source have been commercialized for the analysis where very little amount of sample is available.

The quadrupole ion trap (QIT) is a mass spectrometer with high sensitivity and specificity.[49] Decades ago, it was developed by Wolfgang Paul, a German physicist, and he received the Nobel Prize in 1989 for this work.[50] Now ion trap technology is moving fast, and a commercial ion trap can be made the size of a tennis ball.[51] The ion trap is extraordinary not only because it is an ion storage device for periods of up to hundreds of milliseconds, but also it has a tandem mass feature which allows for the study of ion chemistry and clarification of ion structure. The ESI-QIT has four main regions: ion generation, ion focusing, ion analysis, and ion detection (**Figure 1.5**). Ions are generated by ESI and then are focused and transferred by two octapole transmission systems.

Finally, the ions enter the ion trap. Three hyperbolic electrodes (one ring and two endcap electrodes) constitute the core structure of the ion trap, and make a cavity in the center, where the ions get trapped, isolated, excited and ejected. There is a hole in the center of each of the endcaps, through which ions could move in and out of the ion trap. The ESI-QIT mass spectrometer was extensively used in our nucleobase and DNA duplex gas phase studies.
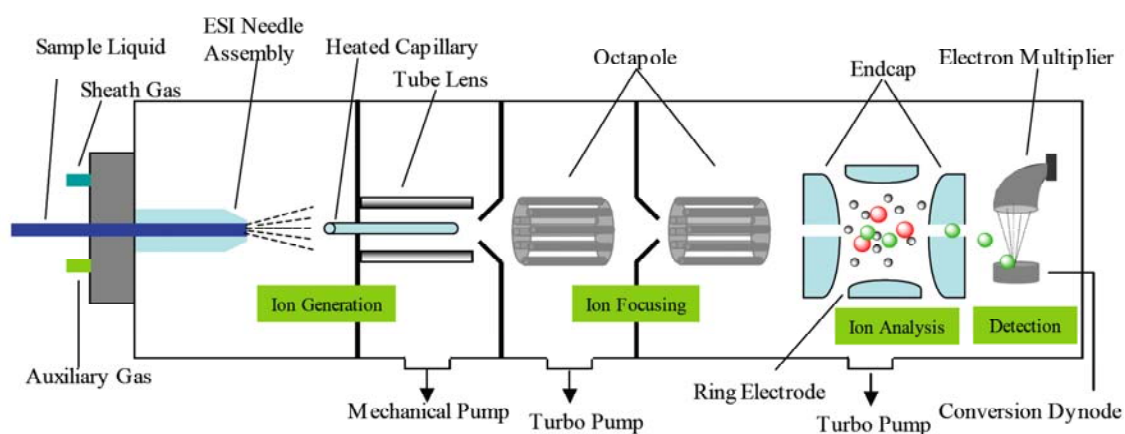
Figure 1.5 Schematic of an ESI-QIT mass spectrometer

1.3 Methodology

1.3.1 Bracketing method

The bracketing method was utilized to measure the gas phase acidities and proton affinities of nucleobases. All bracketing experiments are conducted on a dual-cell Finnigan 2001 Fourier transform mass spectrometer (FTMS). There are two "cells", which are each 2 cubic inches in size, and are connected, forming a "dual cell". The cells are essentially equivalent, but each can be used separately to trap ions, and ions can be

transferred from one cell to the other. One cell is called the "source cell" and the other is the "analyzer cell". Each side of the 2-inch cubic dual cell is pumped down to a baseline pressure of less than 1×10^{-9} Torr. The dual cell is positioned collinearly with the magnetic field produced by a 3.3 T superconducting magnet (**Figure 1.6**). Neutral samples were introduced into the FTMS by a heated batch inlet system (liquid sample) or a heated solids probe (solid sample). Most ions were produced by proton transfer to hydroxide (negative ions) or by proton transfer from hydronium (positive ions). Hydroxide and hydronium were generated by pulsing water vapor into the cell and sending an electron beam (8eV, 6μA to generate hydroxide ions; 20eV, 6μA to generate hydronium ions) through the center of the cell.
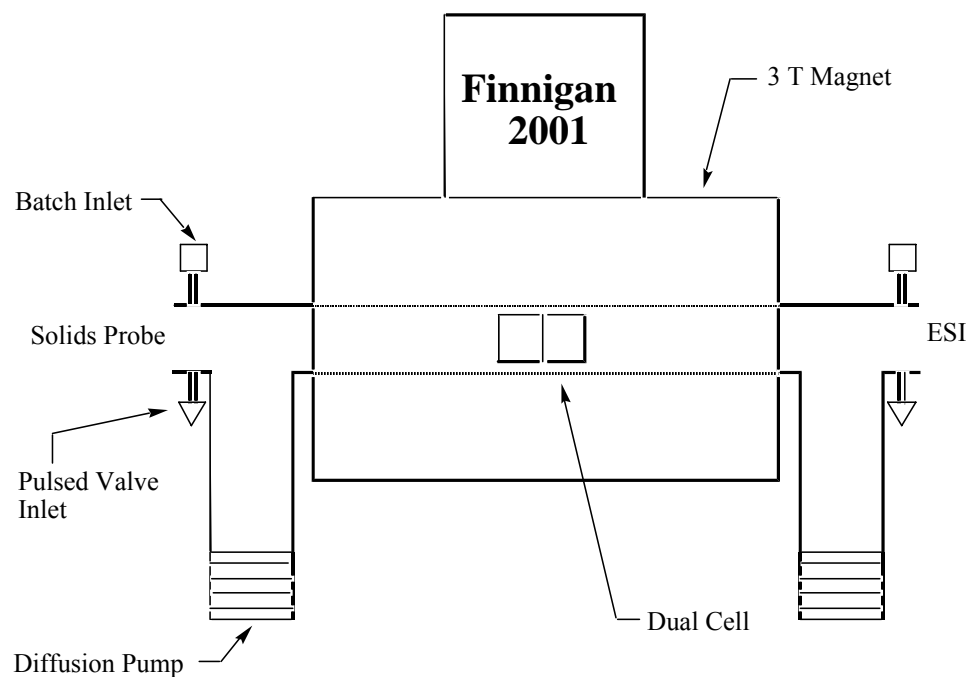
Figure 1.6 FTICR-MS

Another acid or base whose acidity or proton affinity is known is chosen as the reference, and the reactivity of the neutral reference with the ion of unknown substrate is measured. Where possible, the reverse reaction is also conducted (**Figure 1.7a** and **1.7b**).
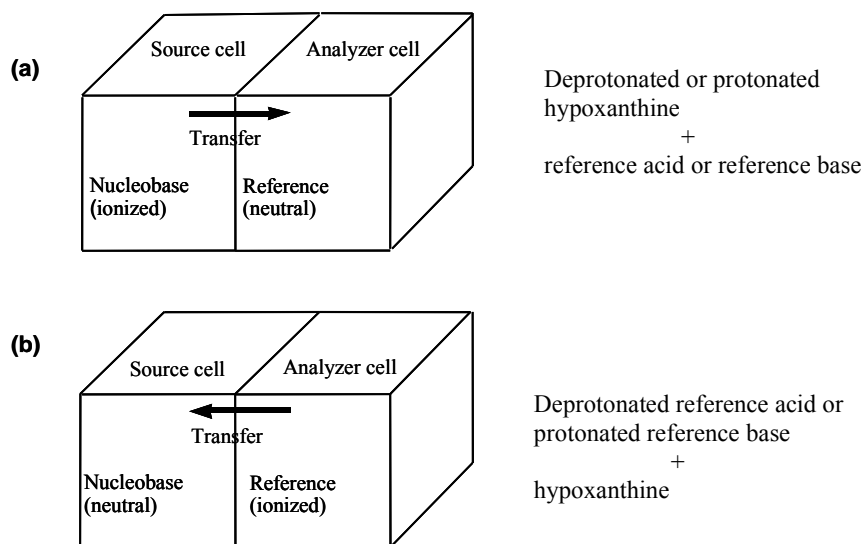
Figure 1.7 Schematic diagram of FTICR dual cell setup used in gas phase bracketing experiments

The nucleobases have multiple acidic and basic sites. Through use of different mass spectrometric (MS) protocols, we can measure more than one acidity (i.e. the most acidic and the second most acidic site) as well as multiple proton affinities. We will discuss acidity bracketing in detail as an example. Acetic acid is used as a simple example, as shown in **Scheme 1.1**. There are two acidic sites in acetic acid: the acidity of COO-H site is 348.1 ± 2.1 kcal mol^{-1} (more acidic site), and the acidity of H$_2$C-H site is 368.1 ± 3.1 kcal mol^{-1} (less acidic site).[52] First, we generate hydroxide ion; the hydroxide ions can then deprotonate acetic acid (HA), forming acetate (A1) and enolate (A2) in the source

cell. Because neutral acetic acid is present in the source cell, A2 can undergo isomerization (reaction with another molecule of neutral CH_3COOH) to form A1 (**Scheme 1.1**). Therefore, ultimately, the only ions in the cell will be the more acidic A1 ions. We can then transfer these A1 ions from the source cell to the analyzer cell to react with neutral reference acids whose acidities are known. There are two possible results: the reference acid can or cannot protonate the acetate A1. Using acids with known ΔH_{acid} values, we can bracket the proton affinity of acetate A1, which is equivalent to the acidity of the more acidic site of acetic acid. For the more acidic site bracketing, we can also use conjugate bases of known reference acids and test which are able to deprotonate acetic acid. For the less acidic site (enolate) measurement, the ions produced from reaction of acetic acid with hydroxide are quickly removed (within 0.5 sec) from the source cell which is rich in neutral acetic acid. In this way, we can isolate a mixture of A1 and A2 ions, before A2 isomerizes to A1. In this case, both the more acidic acetate ions A1 and less acidic enolate ions A2 are transferred to the analyzer cell. In the analyzer cell, the ions are allowed to react with reference acids. There are two possibilities: 1) the reference acid can protonate the enolate form A2; 2) the reference acid cannot react with either A1 nor A2. In this way, we are able to bracket the less acidic site.

Our plan for the proton affinity measurement is to follow the same protocol as for acidity bracketing, except to generate the protonated nucleobase using H_3O^+.

$$CH_3COOH + OH^{\ominus} \longrightarrow CH_3COO^{\ominus} + {}^{\ominus}CH_2COOH$$

HA   A1   A2

$$CH_3COOH$$

$$CH_3COO^{\ominus} + CH_3COOH$$

A1   HA
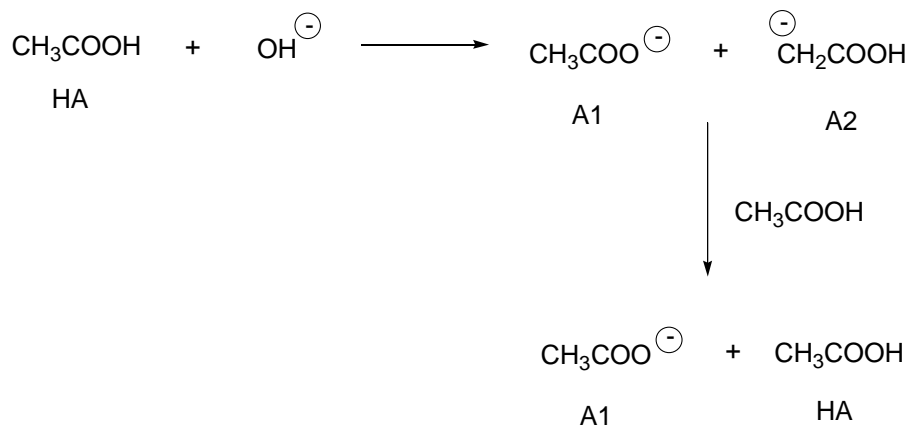
Scheme 1.1 Neutral-catalyzed isomerization

Rate constant calculation

   We conduct ion-molecule reactions in the gas phase to access their acidities and proton affinities. When ions collide with neutral molecules in the gas phase, proton transfer may occur (**Eq. 1.7**):

$$A^- + HB \rightarrow B^- + HA \qquad\qquad \textbf{Eq. 1.7}$$

Then the rate of this reaction can be expressed by **Eq. 1.8**:

$$Rate = k[HB] \qquad\qquad \textbf{Eq. 1.8}$$

Under FTMS conditions, neutral molecule HB is in excess, so $[HB] \gg [A^-]$ and can be used as a constant, then the reaction is a pseudo-first order reaction and the equation could be simplified as **Eq. 1.9**:

$$Rate = k_{obs}[A^-], \text{ where } k_{obs} = k\,[HB] \qquad\qquad \textbf{Eq. 1.9}$$

Also, $Rate = -d[A^-]/dt$ **Eq. 1.10**

Then, $\ln[A^-]_t - \ln[A^-]_0 = \ln([A^-]_t/[A^-]_0) = -k_{obs}t$ **Eq. 1.11**

Therefore, if we plot the natural log of [A⁻] versus reaction time, we should get a straight line with a slope of k_{obs}. Since $k_{obs} = k_{exp}$ [HB], if [HB] is known, then the actual rate constant k (k_{exp}) could be determined.

The theoretical ion-molecule collision rate constant k_{coll} is calculated by the Average Dipole Orientation (ADO) program.[53,54] This program estimates the collision rate constant based on the dipole moment, polarizability, mass of the neutral molecule and the mass of the ion, see **Eq. 1.12**:

$$k_{coll} = (2\pi q / \mu^{1/2})[\alpha^{1/2} + C\mu_D (2/\pi kT)^{1/2}]$$
Eq. 1.12

In **Eq. 1.12** μ_D is the dipole moment of the neutral molecule, μ is the reduced mass of the ion-molecule system, q is the charge of the ion, C is the dipole locking constant, and α is the polarizability of the neutral.[55-57]

Reaction efficiency is defined by the ratio of k_{exp} and k_{coll}. 10% is used as a cutoff to characterize proton transfer. If the efficiency is higher than 10%, then the reaction is treated as a proton transfer reaction (denoted as a positive sign); if it is lower than 10%, then the proton transfer reaction does not happen (denoted as a negative sign).

In order to get the neutral molecule concentration, another control reaction is conducted where hydroxide ion reacts with neutral molecule HB. Here, since OH⁻ is much more basic than HB, we can assume for this reaction proton transfer happens at the collisional rate: $k_{exp} = k_{coll}$. When k_{exp} is known, the slope yielded from a plotting of natural log of [OH⁻] versus reaction time could be used to calculate the neutral concentration.

1.3.2 Cooks Kinetic Method

We also utilize the Cooks kinetic method in a Finnigan-LCQ quadrupole ion trap mass spectrometer to measure the proton affinity of the more basic site or more acidic site of nucleobases and compare acidity among different nucleobases.[58-61] In this method, a proton bound complex or dimer of an unknown compound and a reference base with known proton affinity is formed. Collision-induced dissociation (CID) of this complex leads to products of protonated unknown and protonated reference base and the ratio of these two products reflects the relative proton affinities of the compounds.

$$[AHB_i]^+ \xrightarrow{\ k_1\ } AH^+ + B_i \qquad\qquad \textbf{(Eq. 1.13)}$$

$$\xrightarrow{\ k_2\ } B_iH^+ + A \qquad\qquad \textbf{(Eq. 1.14)}$$

$$B_iH^+ + A \underset{\longleftarrow}{\overset{K}{\longrightarrow}} AH^+ + B_i \qquad\qquad \textbf{(Eq. 1.15)}$$
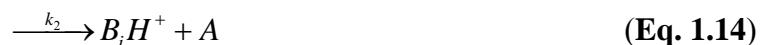
$$K \approx k_1/k_2 \qquad\qquad \textbf{(Eq. 1.16)}$$

$$PA(A) - PA(B_i) \approx RT_{eff} \ln K \qquad\qquad \textbf{(Eq. 1.17)}$$

$$\ln(k_1/k_2) = \frac{1}{RT_{eff}}(PA(A) - PA(B_i)) \qquad\qquad \textbf{(Eq. 1.18)}$$

In the above equations, A represents the nucleobase in our case and B_i represents a series of reference bases with known proton affinities. T_{eff} is the effective temperature of the proton bound complex in Kelvin. We could find the T_{eff} from the slope of a plot of the $[\ln(k_1/k_2)]$ versus the proton affinities of a series reference bases (**Eq. 1.18**). There are two assumptions involved in above equations: the dissociation does not have a reverse activation barrier and the transition state is late and represents the stabilities of the

products[61-63]. In addition, entropy effects should be negligible, which means the two bases bound by the proton should have similar structures. Since the number of reference bases resembling the structure of hypoxanthine is limited, we chose a series of reference bases that are similar to each other to minimize the entropy effects.

1.3.3 Measuring DNA duplex stability in the gas phase by LCQ

There are two major scan modes that we use in our study of DNA duplexes by ESI-MS – full scan and MS/MS. Full scan yields a simple spectrum of what is in solution. It is a useful tool to probe the composition and relative abundance of the complexes. In MS/MS, the target ions from full scan are isolated first and then slightly excited by collision with helium gas. The extra kinetic energy is converted to internal energy, leading to dissociation. This process is called collision-induced dissociation (CID). Armed with these two tools, our group studied DNA duplex stabilities in the gas phase. This provides us with more information about intrinsic stability which is related to the nonpolar environment. Many factors, for example, the size and sequence of the duplex, the charge level of the duplex ion, and hydrogen bonding and base stacking between the two single strands are involved in stabilizing DNA duplexes in the gas phase. Our group studied a set of 9-mer DNA duplexes in the gas phase by ESI-MS and compared their stabilities in the gas phase with those in the solution phase.[26] We found that gas phase stability tracks with solution phase stability generally, although there are key outliers that do not follow the correlation. Our group also studied a systematic series of duplexes with different length, different charge level and different sequence.[27]

1.3.4 Computational method

Theoretical calculations are conducted to predict the tautomerism, acidities and proton affinities of nucleobases.

The GAUSSIAN03 program was used.[64] All the gas phase calculations were conducted using density functional theory (DFT) with the B3LYP method and 6-31+G* basis set. The structures were fully optimized and then frequency calculations were performed on the optimized structures at the same level. The relative stabilities of each tautomer, the acidities ΔH^o_{acid} and PA's are reported at 298 K. The calculations provide important information on the structures of nucleobases in the gas phase and help us to choose proper reference acids and bases for the experiments.

Dielectric calculations of acidities in different solvent media are also performed using the conductor-like polarizable continuum model (CPCM).[65] This model places molecules in a cavity surrounded by an infinite polarizable dielectric and simulates the structure and free energy. The results are used to elucidate the acidity changes of nucleobases in different environments ranging from a polar medium such as water ($\varepsilon = 78$) to a non polar solvent such as cyclohexane ($\varepsilon = 2.0$).

1.3.5 Measurement of nucleobase acidity in different media

It is widely recognized that both reaction rates and acidities are greatly affected when the solvent changes from water to dipolar aprotic solvents.[66-71] In our study of the damaged nucleobase hypoxanthine, we found that hypoxanthine is considerably more acidic than normal nucleobases in the gas phase. However, the acidity differences between hypoxanthine and normal nucleobases are much smaller in water. Hence the

acidity differences are probably enhanced by the nonpolar environment, and the solvent polarities influence the acidity differences. To determine the relationship of nucleobase acidity and solvent polarity, the acidity of damaged and normal nucleobases were measured in different media (dielectric constant changes from 78 (water) to 47 (DMSO)). In solution phase, the acidities of substrates are expressed by pK values and usually measured by three methods: potentiometric, spectrophotometric, and conductometric methods.[70,72-76] We used a spectrometric method for our study. The method and relevant results will be discussed in detail in Chapter 3.

1.4 Reference:

(1)   Nguyen, M. T.; Chandra, A. K.; Zeegers-Huyskens, T. *J. Chem. Soc.* 1998, *94*, 1277-1280.

(2)   Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *J. Phys. Chem. A* 1999, *103*, 8853-8860.

(3)   Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.

(4)   Lee, J. K.; Houk, K. N. *Science* 1997, *276*, 942-945.

(5)   Lundegaard, C.; Jensen, K. F. *Biochemistry* 1999, *38*, 3327-3334.

(6)   Gilson, M. K.; Honig, B. H. *Biopolymers* 1986, *25*, 2097-2119.

(7)   Jordan, F.; Li, H.; Brown, A. *Biochemistry* 1999, *38*, 6369-6373.

(8)   Simonson, T.; Brooks, C. L. I. *J. Am. Chem. Soc.* 1996, *118*, 8452-8458.

(9)   McEwen, W. K. *J. Am. Chem. Soc.* 1936, *58*, 1124-1129.

(10)   Mallard, W. G.; Linstrom, P. J., Eds.; National Institute of Standards and Technology: Gaithersburg, MD 20899, 2003, 2003.

(11)   Kurinovich, M. A.; Lee, J. K. *J. Am. Chem. Soc.* 2000, *122*, 6258-6262.

(12)   Kurinovich, M. A.; Lee, J. K. *J. Am. Soc. Mass Spectrom.* 2002, *13*, 985-995.

(13)   Kurinovich, M. A.; Phillips, L. M.; Sharma, S.; Lee, J. K. *Chem. Commun.* 2002, 2354-2355.

(14)   Sharma, S.; Lee, J. K. *J. Org. Chem.* 2002, *67*, 8360-8365.

(15)   Sharma, S.; Lee, J. K. *J. Org. Chem.* 2004, *69*, 7018-7025.

(16)   Saenger, W. *Principles of Nucleic Acid Structure*; Strange-Verlag: New York, 1984.

(17)   Watson, J. D.; Crick, F. H. C. *Nature* 1953, *171*, 737-738.

(18)   Sponer, J.; Laszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* 1996, *14*, 117.

(19)   Brown, T. *Aldrichimica Acta* 1995, *28*, 15-20.

(20)   Brown, T.; Hunter, W. N. *Biopolymers* 1997, *44*, 91-103.

(21)   Schneider, H. J. *Angew. Chem., Int. Ed. Engl.* 1991, *30*, 1417.

(22)   Lehn, J. M. *Supramolecular Chemistry*; VCH: Weinheim, 1995.

(23)   Sharp, K. A.; Honig, B. *Current Opinion in Structural Biology* 1995, *5*, 323-328.

(24)   Turner, D. H. *Current Opinion in Structural Biology* 1996, *6*, 299-304.

(25)   Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. *Biopolymers* 1997, *44*, 217-239.

(26)   Pan, S.; Sun, X.; Lee, J. K. *Int. J. Mass Spectrom.* 2006, *253*, 238-248.

(27)   Pan, S.; Sun, X.; Lee, J. K. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 1383-1395.

(28)   Pan, S.; Verhoeven, K.; Lee, J. K. *J. Am. Soc. Mass Spectrom.* 2005, *16*, 1853-1865.

(29)   Comisarow, M. B.; Marshall, A. G. *Chem. Phys. Lett.* 1974, *25*, 282-283.

(30)   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* 1989, *246*, 64-71.

(31)   Hillenkamp, F. *Anal. Chem.* 1991, *63:A*, 1193-1203.

(32)   Hudgins, R. R. *Biophys. J.* 1999, *76*, 1591-1597.

(33)   McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* 1992, *3*, 599-614.

(34)   Hoffmann, E. d.; Stroobant, V. *Mass Spectrometry, Principles and Applications*; 2 ed.; John Wiley & Sons: Chichester, 2002.

(35)   Cole, R. B. *Electrospray Ionization Mass Spectrometry*; Wiley: Chichester, 1997.

(36)   Mora, J. F. d. l.; Berkel, G. J. V.; Enke, C. G.; Cole, R. B.; Martinez-Sanchez, M.; Fenn, J. B. *JOURNAL OF MASS SPECTROMETRY* 2000, *35*, 939-952.

(37)   Bruins, A. P. *Mass Spectrometry Reviews* 1991, *10*, 53-77.

(38)   Zeleny, J. *Phys. Rev.* 1914, *3*, 69-91.

(39)   Dole, M.; Mack, L. L.; Hines, R. L.; Mobley, R. C.; Ferguson, L. D.; Alice, M. B. *J. Chem. Phys.* 1968, *49*, 2240-2249.

(40)   Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. *Science* 1989, *246*, 64-71.

(41)   Sanz-Medel, A. *Analyst* 2000, *125*, 35-43.

(42)   Gomez, A.; Tang, K. *Phys. Fluids* 1994, *6*, 404-414.

(43)   Kebarle, P. *J. Mass Spectrom.* 2000, *35*, 804-817.

(44)   Taylor, G. *Proc. Roy. Soc. London Ser. A* 1964, *280*, 383-397.

(45)   Kelly, M. A.; Vestling, M. M.; Fenselau, C. C.; Smith, P. B. *Org. Mass Spectrom.* 1992, *27*, 1143-1147.

(46)   Emmett, M. R.; Caprioli, R. M. *J. Am. Soc. Mass. Spectrom.* 1994, *5*, 605-613.

(47)   Wilm, M.; Shevchenko, A.; Houthaeve, T.; Breit, S.; Schweigerer, L.; Fotsis, T.; Mann, M. *Nature* 1996, *379*, 466-469.

(48)   Wilm, M.; Mann, M. *Anal. Chem.* 1996, *68*, 1-8.

(49)   March, R. E. *Quadrupole Ion Trap Mass Spectrometer*; John Wiley & Sons Ltd: chichester, 2000.

(50)   Paul, W. *Angew. Chem.* 1990, *102*, 780-789.

(51)   Cook, R. G.; Glish, G. L.; McLuckey, S. A.; Kaiser, R. E. *Chem. Eng. News* 1991, *69*, 26-30.

(52)   Grabowski, J. J.; Cheng, X. *J. Am. Chem. Soc* 1989, *111*, 3106-3108.

(53)   Chesnavich, W. J.; Su, T.; Bowers, M. T. *J. Chem. Phys.* 1980, *72*, 2641-2655.

(54)   Su, T.; Chesnavich, W. J. *J. Chem. Phys.* 1982, *76*, 5183-5185.

(55)   Bowers, M. T.; Laudenslager, J. B. *J. Chem. Phys.* 1972, *56*, 4711.

(56)   Gupta, S. K. *Can. J. Chem.* 1967, *45*, 3107.

(57)   Moran, T. F.; Hamil, W. H. *J. Chem. Phys.* 1963, *39*, 1413.

(58)   McLuckey, S. A.; Cameron, D.; Cooks, R. G. *J. Am. Chem. Soc* 1981, *103*, 1313-1317.

(59)   McLuckey, S. A.; Cooks, R. G.; Fulford, J. E. *Int. J. Mass Spectrom. Ion Proc.* 1983, *52*, 165-174.

(60)   Brodbelt-Lustig, J. S.; Cooks, R. G. *Talanta* 1989, *36*, 255-260.

(61)   Green-Church, K. B.; Limbach, P. A. *J. Am. Soc. Mass Spectrom.* 2000, *11*, 24-32.

(62)   Ervin, K. M. *Chem. Rev.* 2001, *101*, 391-444 and references therein.

(63)   Gronert, S.; Feng, W. Y.; Chew, F.; Wu, W. *Int. J. Mass Spectrom.* 2000, *196*, 251-258.

(64)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; 03 ed.; Gaussian, Inc.: Pittsburgh, PA, 2003.

(65)   Barone, V. C.; Tomasi, J. J. *J. Comput. Chem.* 1998, *19*, 404-417.

(66)   Bordwell, F. G.; McCallum, R. J.; Olmstead, W. N. *J. Org. Chem.* 1984, *49*, 1424-1427.

(67)   Buncel, E.; Wilson, H. *Adv. Phys. Org. Chem* 1977, *14*, 133-202.

(68)   Haberfield, P.; Friedman, J.; fPinkston, M. F. *J. Am. Chem. Soc.* 1972, *94*, 71-75.

(69)   Hughes, E. O.; Ingold, C. K. *J. Chem. Soc.* 1935, 244-252.

(70)   Kolthoff, I. M.; Chantooni, M. K.; Bhowmik, S. *J. Am. Chem. Soc.* 1968, *90*, 23-28.

(71)   Parker, A. J. *Chem. Rev.* 1969, *69*, 1-32.

(72)   Courtot-Coupez, J.; LeDemezet, M. *Bull. Soc. Chim. Fr.* 1969.

(73)   Kolthoff, I. M.; Chantooni, M. K. *J. Am. Chem. Soc.* 1971, *93*, 3843-3849.

(74)   Ritchie, C. D.; Uschold, R. E. *J. Am. Chem. Soc.* 1967, *89*, 1721-1725.

(75)   Ritchie, C. D.; Uschold, R. E. *J. Am. Chem. Soc.* 1967, *89*, 2752-2753.

(76)   Ritchie, C. D.; Uschold, R. E. *J. Am. Chem. Soc.* 1968, *90*, 2821-2824.

Chapter 2 The Acidity and Proton Affinity of Hypoxanthine in the Gas Phase versus in Solution: Intrinsic Reactivity and Biological Implications

2.1 Introduction

As we addressed in the first chapter, the acidities and the basicities of nucleobases and nucleobase derivatives are germane to several biological issues. We study these properties of nucleobases in the gas phase because the gas phase is the 'ultimate' nonpolar environment, where intrinsic reactivity can be explored and extrapolated to other media.[1-9]

Recently, our studies of nucleobases have focused on the mutated purine base hypoxanthine (**1**).[10] While hypoxanthine is in fact a naturally occurring base -- in its nucleotide form (called "inosine" (**2**)), it is a key intermediate in the *de novo* biosynthesis of purine nucleotides -- it is also a damaged base that is formed in DNA when adenine (**3**) undergoes oxidative deamination.[10-15] The mutagenicity of hypoxanthine is believed to arise from the fact that unlike adenine, which hydrogen bonds to thymine in the double helix, hypoxanthine prefers to hydrogen bond with cytosine.[13-17] This mispairing when propagated is deleterious to the genome and unless repaired, could lead to cancer or cell death.[18]

1

2   **a** R = H   (nucleoside)
      b R = PO_3^{2-} (nucleotide)

3

4
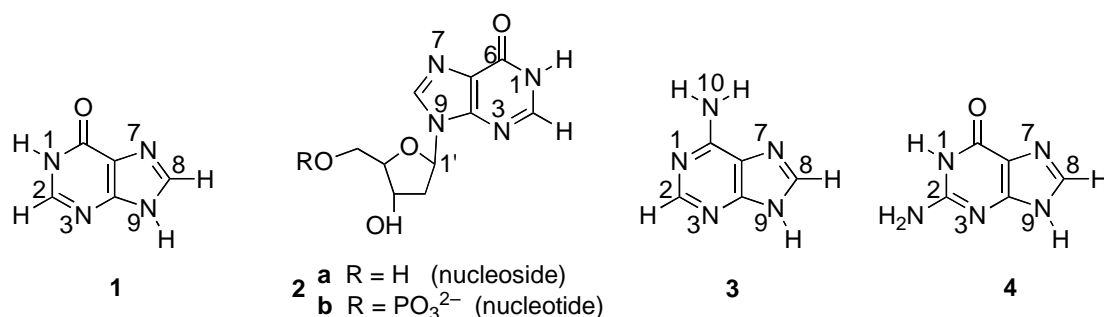
As with many damaged bases, hypoxanthine is cleaved from DNA via a genome-protecting reaction catalyzed by an enzyme called glycosylase.[15,17,19-24] In human cells, alkyladenine DNA glycosylase (AAG) effects this cleavage.[19-22,24-31] AAG is particularly interesting because it cleaves a broad range of bases yet does not cleave normal bases; the conundrum is how the enzyme achieves this "broad specificity".[18,19,21,22] Most glycosylases, like uracil DNA glycosylase (UDGase), are specific to one nucleobase (in the case of UDGase, the enzyme will cleave only uracil).[18,19,21,22] However, AAG is specific yet broad, cleaving a variety of alkylated purines, including 3-methyladenine, 7-methylguanine, and 1, N^6-ethenoadenine in addition to hypoxanthine, yet leaving normal bases (adenine (**3**) and guanine (**4**)) untouched.[18,21,22,32]

In terms of mechanism, one can imagine cleavage with "deprotonated hypoxanthine" as the leaving group (**Scheme 2.1A**), or protonation of the hypoxanthine first to facilitate cleavage (**Scheme 2.1B**).[18,22,33] Little is known about the catalytic mechanism of AAG, but pH rate profiles imply that AAG-catalyzed excision of hypoxanthine requires both a general base (Glu 125) and a general acid. Mutation experiments have thus far failed to identify which residue is the general acid catalyst.[18,20,22]

Scheme 2.1
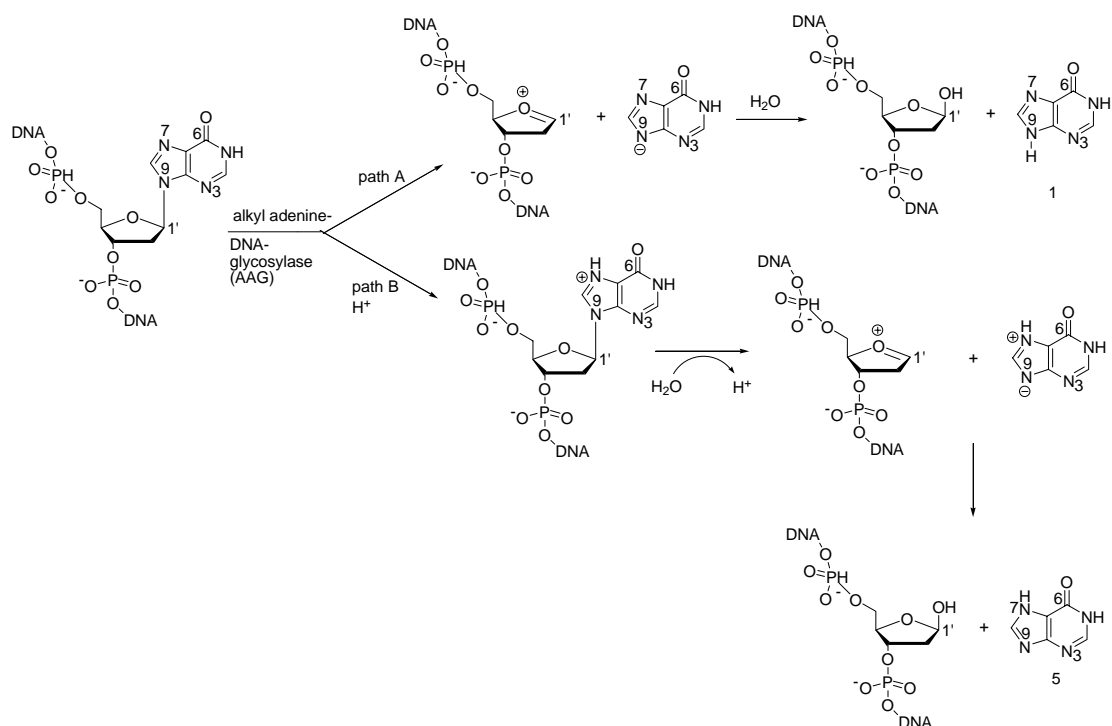
  We have found that gas-phase reactivity can be extremely intriguing and lend insight into glycosylase mechanisms.[4,5,8,9] In previous studies, we showed that the gas phase acidities of uracil (a mutation when present in DNA) and 3-methyladenine are very high, leading to a prediction that relatively speaking, the deprotonated anions of these bases would be good leaving groups and the nonpolar enzyme active sites could take advantage of this property for facile cleavage.[4,8] We also found that relative acidities change in the gas phase versus solution; for example, while the two "NH" groups in uracil have the same pK_a in solution, they have very different acidities in the gas phase, with the N1-H being 15 kcal mol^{-1} more acidic than the N3-H.[4] This raises the interesting possibility that a nonpolar environment could be used to enhance discrimination between sites or substrates.

This leads to the question of the intrinsic, gas phase reactivity of hypoxanthine. The gas phase acidity of hypoxanthine at N9 should correlate to its leaving group ability in a nonpolar active site; if the hypoxanthine N9 proton is unusually acidic, this could be a path by which AAG might favor cleavage of damaged bases over normal bases. We were interested to probe whether hypoxanthine would have enhanced acidity relative to normal nucleobases (adenine and guanine) and whether the relative acidities change in solution versus the gas phase.[4,5,8,9,34] Secondly, the gas phase proton affinity of hypoxanthine is also of interest, since proton transfer may precede cleavage; again, we sought to uncover the differences in the gas phase versus solution. In this chapter, we describe a thorough gas phase computational and experimental examination of hypoxanthine, providing the first measurements of the acidic and basic properties of this damaged base, followed by a discussion of both the intrinsic properties of this nucleobase and also how those properties relate to the biological mechanism of the enzyme human alkyladenine DNA glycosylase (AAG).

2.2 Experimental

All chemicals are commercially available and were used as received. Experiments were conducted on a Fourier Transform mass spectrometer (FTMS) with a dual cell setup which has been described previously.[4,8] Bracketing methods were utilized to measure multiple acidities and basicities (proton affinities) in the gas phase. Rapid proton transfer (i.e., near the collision rate) was taken as evidence that the reaction was exothermic and is indicated by a "+" in **Tables 2.1** and **2.2**.[35] In our experiments, the ions are the reactants and the neutrals are in excess, creating pseudo-first order conditions. Because the

nucleobase is introduced via an external solids probe which, when introduced to the high

vacuum, typically causes the pressure of the cell to rise to about 10^{-7} Torr, we utilize the

following procedure (which has also been described previously) to ascertain the pressure

of the neutral.[9] First, we obtain the pseudo-first order rate constant for the reaction of

hydroxide with the relevant neutral. We then assume that this reaction between hydroxide,

which is very basic, and the neutral proceeds at the theoretical collisional rate.[35] We then

use that calculated collisional rate constant to "back out" the neutral pressure. This

procedure also precludes the errors which can arise from ion gauge issues (such as remote

location and varying sensitivity for different substrates).[36]

$$[AHB_i]^+ \xrightarrow{k_1} AH^+ + B_i \qquad \textbf{Eq. 2 .1}$$
$$[AHB_i]^+ \xrightarrow{k_2} B_iH^+ + A$$
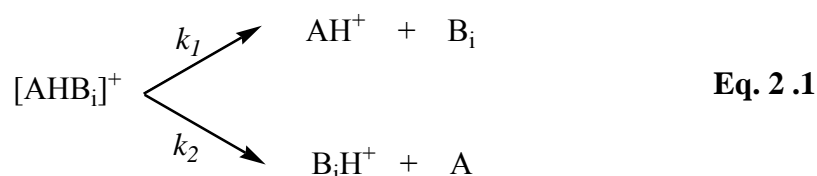
$$\ln(k_1/k_2) = [PA(A)/(RT_{eff}) - \Delta(\Delta S)/R] - PA(B_i)/(RT_{eff}) \qquad \textbf{Eq. 2.2}$$

$$\ln(k_1/k_2) = \ln([AH^+]/[B_iH^+]) \qquad \textbf{Eq. 2.3}$$

$$GB^{app}(A)/RT_{eff} = PA(A)/(RT_{eff}) - \Delta(\Delta S)/R \qquad \textbf{Eq. 2.4}$$

For our proton affinity studies of hypoxanthine in a quadrupole ion trap mass

spectrometer, we utilized four reference bases and measured the product ion distributions

three separate times to ensure reproducibility. We also conducted the Cooks kinetic

method experiments using the "extended" method.[37-41] This method has been well-

described and involves acquiring ion abundance ratios at different collision energies (and

therefore different effective temperatures (vide infra)), which allows for deconvolution of

the enthalpic and entropic contributions. **Eq. 2.2-2.4** summarize the data analysis. Briefly,

T_{eff} is the effective temperature of the dissociating proton bound complex in Kelvin. The term "$\Delta(\Delta S)$" is the difference in the ΔS associated with the two channels in **Eq. 2.1**. A plot of $\ln(k_1/k_2)$ (which is equal to $\ln(/[B_iH^+])$, **Eq. 2.3**) versus $PA(B_i)$ yields the T_{eff} from the slope (**Eq. 2.2**) and the "$GB^{app}(A)$" from the intercept (**Eq. 2.2** and **Eq. 2.4**). Plotting **Eq. 2.4** at different values of T_{eff} yields the proton affinity of hypoxanthine. We find that the standard deviation of our measurements is ± 3 kcal mol^{-1}.

For the kinetic method experiments, solutions of hypoxanthine and the reference bases were subjected to electrospray ionization (10^{-3} to 10^{-4} M solutions in methanol; a small amount of acetic acid is also added). The typical flow rate is 25μL/min. An electrospray needle voltage of ~ 4500 V was used. The proton-bound complexes of hypoxanthine and the reference base were isolated and subjected to collision-induced dissociation (CID); the complexes were activated for about 30 ms. About 40 scans were averaged for product ions.

Gas phase calculations were conducted at B3LYP/6-31+G* using Gaussian03.[42-47] All gas phase structures, including those in the Supporting Information, were fully optimized; frequency calculations confirmed true minima, with no negative frequencies. Acidity and proton affinity values are reported as ΔH at 298 K, which allows for direct comparison with the experimentally measured values. Solvation studies were conducted using the CPCM-SCRF method (full optimizations at B3LYP/6-31+G*; UAKS cavity) as implemented in Gaussian03.[48,49] This method has been shown to be reliable for neutral and ionic organic molecules.[50] A dielectric constant of 78.4 was used in order to simulate an aqueous environment; we did not conduct frequency calculations so we report ΔE values.

2.3 Results

2.3.1 Computational results: tautomers

The structure of nucleobases is such that several tautomers are often possible, and hypoxanthine is no exception. We calculated the relative stability of the possible tautomers of hypoxanthine. The canonical structure **1**, which we refer to as the "H19" tautomer (since the protons reside on the N1 and the N9) is calculated to be *less* stable than the "H17" tautomer (**5**, where the protons reside on the N1 and the N7), by 0.8 kcal mol^{-1} (**Figure 2.1**). The next most stable tautomer is **6**; at 5.4 kcal mol^{-1} less stable than the H17 tautomer, it is not likely to be present in any appreciable amount in the gas phase. The remaining possible tautomers are even less stable (Figure S1, Supporting Information).

1 (H19)	5 (H17)	6
0.8 kcal mol^{-1}	0.0 kcal mol^{-1}	5.4 kcal mol^{-1}
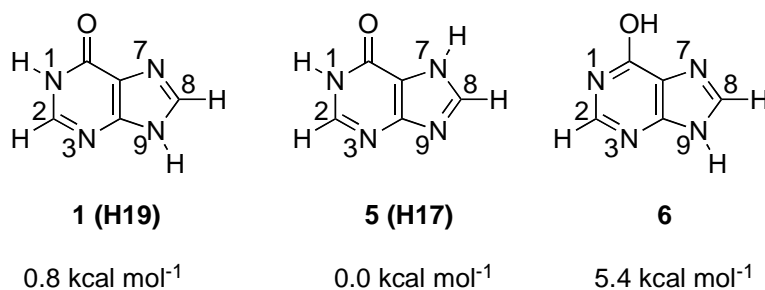
Figure 2.1. Relative energies (ΔH) of the three most stable tautomers of hypoxanthine, calculated at B3LYP/6-31+G*, at 298 K.

2.3.2 Computational results: acidity

We focused our acidity studies on the two most stable tautomers, H19 (**1**) and H17 (**5**). The H19 tautomer is the "canonical" structure and the most relevant biologically. The

H17 structure is of importance intrinsically, since its high stability is such that it might be present in our gas phase studies.

The gas phase acidities for all the protons of hypoxanthine calculated at B3LYP/6-31+G* are summarized in **Figure 2.2**. The most acidic site of the H19 tautomer **1** is the N9-H, at 330.5 kcal mol^{-1}. The N1-H is slightly less acidic, at 337.0 kcal mol^{-1}. The C-H protons are the least acidic; the C2-H calculates to 368.7 kcal mol^{-1} and the C8-H to 373.2 kcal mol^{-1}.
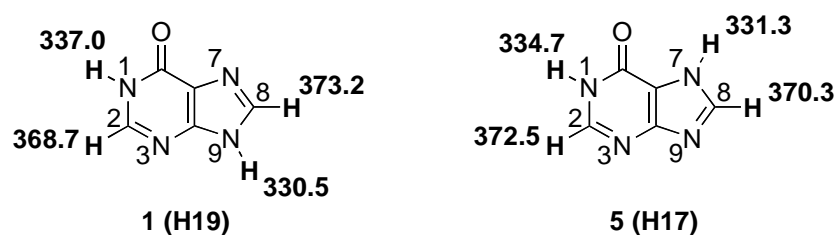
Figure 2.2. Calculated acidities (ΔH_{acid}) of the two most stable tautomers of hypoxanthine using B3LYP/6-31+G*, at 298 K. Acidic protons are in bold.

The H17 tautomer **5** is also quite acidic. The N7-H calculates to 331.3 kcal mol^{-1} while the N1-H is slightly less acidic, at 334.7 kcal mol^{-1}. The least acidic sites are the C-H protons; the C2-H calculates to 372.5 kcal mol^{-1} and the C8-H to 370.3 kcal mol^{-1}.

2.3.3 Computational results: proton affinity

The proton affinities of the tautomers H19 and H17 were also explored computationally (**Figure 2.3**). The H19 tautomer has several heteroatoms that could accept a proton; we find the most basic site to be N7, at 219.6 kcal mol^{-1}. The O6 site is the next most basic, with a PA of 213.9 kcal mol^{-1} (for the protonated structure where the proton is "pointing toward" the imidazolic ring). Protonation of O6 to form the structure

where the proton is on the same side as the N1-H is exothermic by 206.2 kcal mol^{-1}, and the least basic heteroatom on the H19 tautomer is N3, at 204.5 kcal mol^{-1}. The H17 tautomer is of similar overall basicity (as compared to the H19); the most basic site is N9, at 218.8 kcal mol^{-1}. The remaining proton affinities are 215.6 (N3), 201.9 (O6, imidazole (N7) side) and 201.5 (O6, N1 side) kcal mol^{-1}.
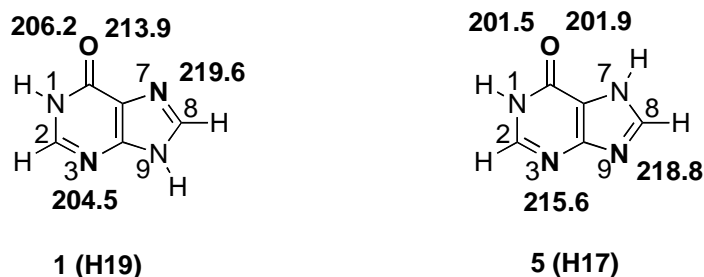
Figure 2.3. Calculated proton affinities (ΔH) of the two most stable tautomers of hypoxanthine using B3LYP/6-31+G*, at 298 K. Basic sites are in bold.

2.3.4 Experimental results: acidity

Our first step was to bracket the most acidic site of hypoxanthine (**Table 2.1**). We find that the conjugate base of hypoxanthine deprotonates $(CF_3)_3COH$ (ΔH_{acid} = 331.6 $\pm$ 2.2 kcal mol^{-1}), but not HCl ((ΔH_{acid} = 333.4 $\pm$ 0.1); also, Cl$^-$ deprotonates hypoxanthine but $(CF_3)_3CO^-$ does not. We therefore bracket the most acidic site of hypoxanthine to be 332 $\pm$ 2 kcal mol^{-1}. We next used our "less acidic" method (**Table 2.2**). We find that the conjugate base of hypoxanthine does not deprotonate acetone (CH_3COCH_3, ΔH_{acid} = 368.8 $\pm$ 2.0) but it does deprotonate methyl ethyl ketone ($CH_3COCH_2CH_3$, ΔH_{acid} = 367.2

$\pm$ 2.8). Therefore, we bracket the gas phase acidity of the less acidic site in hypoxanthine

to be 368 $\pm$ 3 kcal mol^{-1}. At this point, we do not know which tautomer(s) or which sites

we are bracketing; we simply were interested as a first step to bracket the experimental

acidity values.

Table 2.1　　Summary of results for acidity bracketing of more acidic site of
hypoxanthine**.**

reference compound	$\Delta H_{acid}{}^{a}$	proton transfer[b]	
		ref. acid	conjugate base
HCOOH	346.2 ± 1.2	–	+
$CH_3COCH_2COCH_3$	343.8 ± 2.1	–	+
m-CF$_3$PhOH	339.3 ± 2.1	–	+
CH$_3$CHBrCOOH	336.8 ± 2.1	–	+
CNCH$_2$CN	335.8 ± 2.1	–	+
HCl	333.4 ± 0.1	–	+
(CF$_3$)$_3$COH	331.6 ± 2.2	+	–
CHF$_2$COOH	331.0 ± 2.2	+	–
(CF$_3$)$_2$C$_6$H$_3$OH	329.8 ± 2.1	+	–
CF$_3$COCH$_2$COCH$_3$	328.3 ± 2.9	+	–

[a] Acidities are in kcal mol^{-1} and come from reference 44. [b] A "+" indicates

the occurrence and a "–" denotes the absence of proton transfer.

Table 2.2 Summary of results for acidity bracketing of less acidic site of hypoxanthine.

reference compound	$\Delta H_{acid}{}^a$	proton transfer[b] ref. acid
$CH_3CH_2CH_2OH$	375.7 ± 1.3	–
CH_3CN	372.9 ± 2.1	–
$PhCH_2OH$	370.0 ± 2.1	–
CH_3COCH_3	368.8 ± 2.0	–
$CH_3COCH_2CH_3$	367.2 ± 2.8	+
$C_6H_5NH_2$	366.4 ± 2.1	+
m-F-$PhNH_2$	362.6 ± 2.2	+
C_4NH_5	359.5 ± 0.3	+
CH_3COOH	348.1 ± 2.2	+
$CH_3CH_2CH_2CH_2COOH$	346.2 ± 2.1	+
HCOOH	346.2 ± 1.2	+

[a] Acidities are in kcal mol^{-1} and come from reference 44. [b] A "+" indicates the occurrence and a "–" denotes the absence of proton transfer.

2.3.5 Experimental results: proton affinity

We next measured the most basic site of hypoxanthine. We find that 3-bromopyridine (PA = 217.5 ± 2.0 kcal mol^{-1}) cannot deprotonate protonated hypoxanthine, but that hypoxanthine can protonate 3-bromopyridine. 1-Butylamine (PA = 220.2 ± 2.0 kcal mol^{-1}) deprotonates protonated hypoxanthine, while the reverse reaction does not occur. In between these two reference bases, we obtain some conflicting results. For example, while the reference base benzylamine (PA = 218.3 ± 2.0 kcal mol^{-1}) deprotonates protonated hypoxanthine, N-methylaniline (PA = 219.1 ± 2.0 kcal mol^{-1}) *cannot*

deprotonate protonated hypoxanthine. The ambiguity of our bracketing results may be due to the error (about ± 2 kcal mol^{-1}) in any given PA value; from our data, we only can say that the PA of hypoxanthine is between 217.5 (±2) and 220.2 (±2) kcal mol^{-1}.[51]

Given the large range of our bracketing result, we decided to measure the PA using a second method, the Cooks extended kinetic method (see Experimental section for details). Four reference bases were used: 1-octanamine (PA = 222.0 kcal mol^{-1}), isobutylamine (PA = 221.0 kcal mol^{-1}), 1-butylamine (220.2 kcal mol^{-1}), and 1-propylamine (219.4 kcal mol^{-1}), yielding a proton affinity of 222 ± 3 kcal mol^{-1} for hypoxanthine, which is consistent (though on the high end) with the window that we find via bracketing.[52-55]

The measurement of the less basic site of hypoxanthine had experimental issues that we could not overcome. In essence, under the "less basic" FTMS conditions, we always see proton transfer from protonated hypoxanthine to the reference base, even if the reference base is very weak (such as acetone, PA = 194.1 kcal mol^{-1}). We have experienced this problem in the past, and it appears to be due to the presence of protonated water, which is very acidic.[56] We tried various methods to rid the system of H_3O^+, but to no avail, and therefore could not measure the less basic site.

2.4 Discussion

In aqueous solution and in the solid state, the canonical H19 (**1**) tautomer of hypoxanthine predominates.[57-59] Furthermore, this particular tautomer is the biologically relevant one; when hypoxanthine is in nucleotide form ("inosine" (**2**)), the deoxyribose moiety is attached to the N9; therefore, the H19 tautomer is a model for the nucleobase portion of inosine.

Unlike in solution and in the solid state, more than one tautomer is accessible in the gas phase. Our calculations at B3LYP/6-31+G* indicate that the H19 structure (**1**) is actually less stable than the H17 structure (**5**) by 0.8 kcal mol^{-1} (**Figure 2.1**), which is consistent with other calculations on hypoxanthine tautomers.[58-64] In this Discussion section, we will first discuss the intrinsic gas phase properties of hypoxanthine, in which we consider both the H19 (**1**) and H17 (**5**) tautomers. We will then move on to discuss the biological relevance of our experimental and computational results, which will focus on the H19 canonical tautomer **1**.

2.4.1 Can we differentiate between tautomers?

The acidity and proton affinity calculations indicate that measurements of these properties are not likely to reveal which tautomer(s) is (are) present. That is, sometimes two tautomers have very differing properties; one tautomer might be much more acidic than the other, and the measurement of the acidity can thus reveal which tautomer is present. In the case of hypoxanthine, however, the most acidic site of the H19 tautomer is calculated to be 330.5 kcal mol^{-1}, while the most acidic site of the H17 tautomer calculates to 331.3 kcal mol^{-1}. These acidities are so close that any measured value could be attributable to either tautomer. For proton affinity (PA), the most basic site of the H19 tautomer has a calculated PA of 219.6 kcal mol^{-1}, while the most basic site of the H17 tautomer has a calculated PA of 218.8 kcal mol^{-1}. Again, these values are sufficiently close that the actual measurement will not divulge which tautomer has been probed. It may be that the more stable H17 tautomer is prevalent and that our measurements pertain to that structure or a mix of H17 and H19.

We measured two acidic sites on hypoxanthine, one at 332 ± 2 kcal mol^{-1} and the other at 368 ± 3 kcal mol^{-1} (**Tables 2.1, 2.2**). The more acidic site is consistent with the calculated values for the most acidic site of both the H19 and the H17 tautomers (330.5 kcal mol^{-1} (N9-H of H19 tautomer) and 331.3 kcal mol^{-1} (N7-H of H17 tautomer), **Figure 2.2**). Either or both tautomers may be present; calculations would indicate that in the gas phase, we have a mixture of both tautomers.[58,61]

The "less acidic" bracketing experiment yields a value of 368 ± 3 kcal mol^{-1}. By calculations, however, the next most acidic site on the H19 (**1**) tautomer is the N1-H, at 337.0 kcal mol^{-1} (**Figure 2.2**). On the H17 (**5**) tautomer, the second most acidic site is also the N1-H, at 334.7 kcal mol^{-1} (**Figure 2.2**). Our "less acidic" bracketing method is such that if we have a mixture of [M-H]$^-$ ions deprotonated at N1 as well as at C2 and C8, we will only see the acidity value corresponding to the most basic anions (in this case, the carbanions).[4,5,8,9,65] Therefore, we may have ions resulting from the N1-H of the H17 tautomer and/or the N1-H of the H19 tautomer, but we can only bracket the most basic ion due to the nature of the experiment.[4-6,8,9] Given how close in energy both tautomers are calculated to be, we believe that we are likely to have a tautomer mixture.[58-64] Therefore, the less acidic site of 368 ± 3 kcal mol^{-1} could correspond to any of the C-H sites in the H19 (**1**) and/or the H17 (**5**) tautomers (**Figure 2.2**).

For proton affinities, we find the more basic site to have a PA of 222 ± 3 kcal mol^{-1}. This value is consistent with the most basic calculated sites (**Figure 2.3**) of the H19 (**1**) tautomer (N7, 219.6 kcal mol^{-1}), and the H17 (**5**) tautomer (N9, 218.8 kcal mol^{-1}). The PA results thus also correlate with the computational results.

These acidities and proton affinities of hypoxanthine are the first such measurements and serve to both benchmark the calculations as well as to provide insights into the intrinsic reactivity of this damaged base.

2.4.2 Biological implications

As noted earlier, one of our interests in hypoxanthine is that it is a mutagenic base. When incorporated into DNA, hypoxanthine is called "inosine" (**2**). As shown in **Scheme 2.1**, hypoxanthine is excised from DNA via scission of the N9-C1' bond; the relevant tautomer is therefore the H19 (**1**), and the discussions herein will focus solely on this canonical tautomer. Our genome is protected by an enzyme called alkyladenine DNA glycosylase (AAG), which cleaves hypoxanthine from DNA. AAG is a particularly intriguing enzyme because it cleaves a wide range of damaged bases, thus achieving a "broad specificity".[18,19,21,22] One of the puzzles is how an enzyme could cleave *so many* different bases -- in this case, a variety of alkylated purines, including 3-methyladenine, 7-methylguanine, and 1, N^6-ethenoadenine in addition to hypoxanthine -- yet leave normal bases untouched. The exact mechanism by which AAG excises hypoxanthine is unknown.[19-22,24-31] Two possibilities are that the hypoxanthine is excised in an anionic state or that it is protonated prior to departing as a neutral (**Scheme 2.1**). Our results lend insight into both possibilities.

2.4.2.1 Deprotonated hypoxanthine as a leaving group

Should hypoxanthine undergo cleavage without prior protonation, the N9-deprotonated hypoxanthine would be the leaving group (**Scheme 2.1A**). The more acidic the N9-H, the more easily deprotonated hypoxanthine should be cleaved. Our prior studies have shown

that the *gas phase* acidities of uracil (an undesirable base in DNA) and 3-methyladenine (a mutated base) are very high, leading to a prediction that relatively speaking, the deprotonated anions of these damaged bases would be good leaving groups in nonpolar enzyme active sites (since the gas phase is the ultimate nonpolar environment), which in turn translates to facile cleavage.[4,5,8,9,66] Furthermore, recently, Drohat and coworkers published a thorough study of a *pyrimidine* glycosylase that also has broad specificity, human thymine DNA glycosylase (TDG).[34] Their studies show a relationship between the solution phase acidity (pK$_a$) of a series of nucleobase substrates and the ease of excision of these substrates by the enzyme TDG (the more acidic the nucleobase, the more facile the excision), indicating the nucleobase is cleaved in its deprotonated form. In addition, the differences in acidity among the substrates appear to be much greater in a nonpolar gas phase environment than in solution. That is, the nonpolarity of the active site of TDG appears to contribute to the enzyme's specificity by enhancing the differences in acidity of the various substrates and therefore favoring cleavage of those nucleobases that have high acidities (and are therefore good deprotonated leaving groups) in the gas phase.[34]

The gas phase acidity of hypoxanthine at N9 should likewise indicate its leaving group ability in a nonpolar active site. The calculated acidity of the N9-H of hypoxanthine (**1**) at B3LYP/6-31+G* is 330.5 kcal mol^{-1}. The acidity of the N9-H of guanine (**4**) calculated at the same level is 334.3 kcal mol^{-1}.[67-70] That is, hypoxanthine is ~4 kcal mol^{-1} more acidic than guanine in the gas phase. This enhanced acidity of hypoxanthine would translate to its ease of excision; AAG, like TDG, may capitalize on a nonpolar enzyme active site to favor cleavage of damaged bases like hypoxanthine due to their better leaving group ability. The gas phase acidity of hypoxanthine is also calculated to be 4 kcal mol^{-1} less

(i.e., more acidic) than that of adenine (hypoxanthine (**1**), 330.5 kcal mol^{-1} versus adenine (**3**), 334.8 kcal mol^{-1}).[8,9,67-69,71] Thus, the enhanced acidity of the damaged base hypoxanthine relative to the normal bases adenine and guanine would render it more easily excised by AAG; this provides an explanation for why AAG might favor hypoxanthine for removal.

We were also interested in ascertaining whether the acidity of hypoxanthine is enhanced, relative to adenine and guanine, in the gas phase versus in solution.[34] The solution phase acidities (pK$_a$) of the N9-H of adenine, guanine, and hypoxanthine are reported to be 9.8, 10.0 and 8.9, respectively.[59,72-75] These pK$_a$ values translate to deprotonation at 298 K being more favorable for hypoxanthine by 1.2 kcal mol^{-1} over adenine, and 1.5 kcal mol^{-1} over guanine. This preference is enhanced in the gas phase, where the acidity of hypoxanthine is calculated to be 4.3 and 3.8 kcal mol^{-1} greater than that of adenine and guanine, respectively. A nonpolar environment would therefore serve to enhance the differences among the substrates adenine, guanine and hypoxanthine, and favor cleavage of the damaged base hypoxanthine. We hypothesize that one of the ways AAG provides selectivity is by targeting hypoxanthine and other damaged bases and cleaving them as deprotonated, anionic nucleobases.

We also conducted dielectric medium calculations on the acidities of hypoxanthine, adenine and guanine to determine how the gas phase acidities change in a medium of dielectric 78.4 (aqueous solution). We find that the results are consistent with trends seen with the experimental pK$_a$ values. The calculated acidity of hypoxanthine in an aqueous continuum is 295.7 kcal mol^{-1}; that of adenine and guanine are 297.8 and 298.4 kcal mol^{-1}. Therefore, hypoxanthine is still the most acidic, but less so than in the gas phase (more

acidic than adenine and guanine, respectively, by only 2.1 and 2.7 kcal mol[-1] in a water dielectric versus 4.3 and 3.8 kcal mol[-1] in the gas phase). Therefore, the gas phase and solvation calculations and the experimental pK_a data all indicate that hypoxanthine is more acidic than adenine and guanine, and that the differences in acidity are greatest in the gas phase.

2.4.2.2 Neutral hypoxanthine as a leaving group

Should hypoxanthine leave as a neutral, the acidity of the N9 site after protonation is relevant: the more acidic the protonated hypoxanthine is, the more easily it will be cleaved as a neutral (**Scheme 2.1B**). Assuming that protonation is barrierless, cleavage at the N-glycosidic bond would be a rate-determining step. There are many acidic residues in the AAG active site, including Glu125 that can protonate the leaving nucleobase and hence activate the hydrolysis of N-glycosidic bond.[76-79] We used protonation at N7 as a model because when N7 is protonated, it increases C8-N9 π-bonding, which compensates the loss of C1'-N9 bond, thus promotes the purine departure.[80] Calculation methods were used to simulate this process. The acidity of the N9 site of the H19 tautomer of hypoxanthine after protonation at N7 (**1**, which is relevant to the biological compound inosine (**2**) since the ribose is attached to N9, **Scheme 2.1B**) is calculated to be 218.8 kcal mol[-1], at B3LYP/6-31+G*. We calculate the acidity of N7 protonated adenine (**3**) and guanine (**4**) to be, using the same method, respectively, 223.7 and 226.9 kcal mol[-1].[68,69,81-83] Therefore, N7-protonated hypoxanthine appears to be *more acidic* than N7-protonated adenine and guanine. A mechanism involving protonation prior to cleavage (**Scheme 2.1B**) would also energetically favor hypoxanthine cleavage. However, we also, incidentally, explored the acidity of N7- protonated 1, N[6]-ethenoadenine, another

damaged base which is also excised by AAG, and found the ΔH_{acid} to be 223.0 kcal mol^{-1}. This value is comparable to that of protonated adenine, and makes the protonation mechanism less likely in terms of discrimination between damaged and normal bases.

In addition to the N7 site, we also explored the acidity of N9H after protonation at the most basic site of each nucleobase (N1 for adenine, N7 for guanine, N7 for hypoxanthine, and N10 for 1, N^6-ethenoadenine). However, the acidities of N9H of these protonated bases (242.2, 226.9, 218.8 and 244.7 kcal mol^{-1} for adenine, guanine, hypoxanthine, and 1, N^6-ethenoadenine, respectively) do not show any trend.

Of course, AAG might also differentiate hypoxanthine from normal bases through other means (binding energetics, size selectivity),[18,21,22] but our results show that energetically speaking, a mechanism involving initial proton transfer does not favor all the damaged bases. We therefore propose the possibility that like thymine DNA glycosylase (TDG), hypoxanthine is excised by AAG as an anion, and that hypoxanthine should be a better leaving group (N9-H is more acidic) than the normal bases adenine and guanine.[4,5,8,9,34] The acidity of the N9 site in 1, N^6-ethenoadenine was measured by our group to be 332 kcal mol^{-1} (close to the acidity of hypoxanthine, and more acidic than adenine and guanine), which is consistent with this mechanism.[84] Furthermore, this trend is enhanced in the gas phase, and would therefore be enhanced in a nonpolar active site. Further studies of the acidity difference between damaged and normal bases in different media will be discussed in the next chapter.

2.5 Conclusions

We have established the acidities and proton affinity of the damaged nucleobase hypoxanthine. The experimental studies allow us to confirm the accuracy of the ab initio calculations. The acidic and basic properties of hypoxanthine are compared to those of the normal nucleobases adenine and guanine, both in the gas phase and in solution, to provide insights into understanding how the enzyme alkyladenine DNA glycosylase (AAG) might discriminate damaged bases from normal bases. The results highlight the possibility that AAG cleaves damaged nucleobases as anions and that the active site may take advantage of a nonpolar environment to favor deprotonated hypoxanthine as a leaving group versus deprotonated adenine or guanine. Future studies of the gas phase properties of other damaged nucleobases that serve as substrates for AAG are underway.

2.6 References

(1)   Brauman, J. I.; Blair, L. K. *Journal of the American Chemical Society* **1970**, *92*, 5986-5992.

(2)   Lowry, T. H.; Richardson, K. S. *Mechanism and Theory in Organic Chemistry*; Third ed.; Harper & Row, Publishers: New York, 1987.

(3)   McEwen, W. K. *Journal of the American Chemical Society* **1936**, *58*, 1124-1129.

(4)   Kurinovich, M. A.; Lee, J. K. *Journal of the American Chemical Society* **2000**, *122*, 6258-6262.

(5)   Kurinovich, M. A.; Lee, J. K. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 985-995.

(6)   Kurinovich, M. A.; Phillips, L. M.; Sharma, S.; Lee, J. K. *Chemical Communications* **2002**, 2354-2355.

(7)   Phillips, L. M.; Lee, J. K. *Journal of the American Chemical Society* **2001**, *123*, 12067-12073.

(8)   Sharma, S.; Lee, J. K. *Journal of Organic Chemistry* **2002**, *67*, 8360-8365.

(9)   Sharma, S.; Lee, J. K. *Journal of Organic Chemistry* **2004**, *69*, 7018-7025.

(10)   Voet, D.; Voet, J. G. *Biochemistry*; Second ed.; John Wiley & Sons, Inc.: New York, 1995.

(11)   Crick, F. H. C. *Journal of Molecular Biology* **1966**, *19*, 548-555.

(12)   Munns, A. R. I.; Tollin, P. *Acta Crystallographica, Section B: Structural Crystallography and Crystal Chemistry* **1970**, *B26*, 1101-1113.

(13)   Miao, F.; Bouziane, M.; O'Connor, T. R. *Nucleic Acids Research* **1998**, *26*, 4034-4041.

(14)   Lindahl, T. *Nature* **1993**, *362*, 709-715.

(15)   Karran, P.; Lindahl, T. *Biochemistry* **1980**, *19*, 6005-6011.

(16)   Hang, B.; Singer, B.; Margison, G. P.; Elder, R. H. *Proceedings of the National Academy of the Sciences, USA* **1997**, *94*, 12869-12874.

(17)   Saparbaev, M.; Laval, J. *Proceedings of the National Academy of the Sciences, USA* **1994**, *91*, 5873-5877.

(18)   O'Brien, P. J.; Ellenberger, T. *Journal of Biological Chemistry* **2004**, *279*, 9750-9757.

(19)   Stivers, J. T.; Jiang, Y. L. *Chemical Reviews* **2003**, *103*, 2729-2759.

(20)   Berti, P. J.; McCann, J. A. B. *Chemical Reviews* **2006**, *106*, 506-555 and references therein.

(21)   O'Brien, P. J.; Ellenberger, T. *Journal of Biological Chemistry* **2004**, *279*, 26876-26884.

(22)   O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418-12429.

(23)   Singer, B.; Hang, B. *Chem. Res. Toxicol.* **1997**, *10*, 713-732.

(24)   Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. *BioEssays* **1999**, *21*, 668-676.

(25)   *Repair of Alkylation Damage to DNA*; Seeberg, E.; Berdal, K. G., Eds.; Landes Bioscience: New York, 1997.

(26)   Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. *Proceedings of the National Academy of the Sciences, USA* **2000**, *97*, 13573-13578.

(27)   In E. coli, 3-methyladenine DNA glycosylase II (AlkA) excises hypoxanthine. For reviews, see: references 19 and 20.

(28)   Lau, A. Y.; Schärer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, T. *Cell* **1998**, *95*, 249-258.

(29)   Vallur, A. C.; Feller, J. A.; Abner, C. W.; Tran, R. K.; Bloom, L. B. *Journal of Biological Chemistry* **2002**, *277*, 31673-31678.

(30)   Xia, L.; Zheng, L.; Lee, H.-W.; Bates, S. E.; Federico, L.; Shen, B.; O'Connor, T. R. *Journal of Molecular Biology* **2005**, *346*, 1259-1274.

(31)   Guliaev, A. B.; Hang, B.; Singer, B. *Nucleic Acids Research* **2002**, *30*, 3778-3787.

(32)   Current understanding is that when attached to the deoxyribose in DNA, 3-methyladenine and 7-methylguanine are positively charged substrates while hypoxanthine and ethenoadenine are neutral substrates. This paper focuses on damaged bases such as hypoxanthine and ethenoadenine, which are neutral substrates. See references 18-22.

(33)   Previous work indicates that glycosidic bond hydrolyses proceed via a highly dissociative S_N2 or a stepwise S_N1 mechanism; to simplify Scheme 1 we show the mechanism as S_N1. Furthermore, the mechanism for enzyme cleavage also involves a

"base-flipping" step, whereby the nucleobase is "flipped" into the enzyme active site; this step may also differ for different substrates. It is not known whether base flipping is a relatively fast or slow step in AAG, and herein we focus on the steps involving actual nucleobase excision. See references 18-20.

(34)   Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C. *Journal of the American Chemical Society* **2006**, *128*, 12510-12519.

(35)   Su, T.; Chesnavich, W. J. *Journal of Chemical Physics* **1982**, *76*, 5183-5185.

(36)   Bartmess, J. E.; Georgiadis, R. M. *Vacuum* **1983**, *33*, 149-153.

(37)   Williams, T. I.; Denault, J. W.; Cooks, R. G. *Internat. Journal of Mass Spectrom* **2001**, *210/211*, 133-146.

(38)   Armentrout, P. B. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 371-379.

(39)   Cheng, X. H.; Wu, Z. C.; Fenselau, C. *Journal of the American Chemical Society* **1993**, *115*, 4844-4848.

(40)   Cerda, B. A.; Wesdemiotis, C. *Journal of the American Chemical Society* **1996**, *118*, 11884-11892.

(41)   Nold, M. J.; Cerda, B. A.; Wesdemiotis, C. *Journal of the American Society for Mass Spectrometry* **1999**, *10*, 1-8.

(42)   GAUSSIAN03, R. C.; Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; C., B. J.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; O. Kitao; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A., Gaussian, Inc., Wallingford CT, 2004.

(43)   Kohn, W.; Becke, A. D.; Parr, R. G. *Journal of Physical Chemistry* **1996**, *100*, 12974-12980.

(44)   Becke, A. D. *Journal of Chemical Physics* **1993**, *98*, 1372-1377.

(45)   Becke, A. D. *Journal of Chemical Physics* **1993**, *98*, 5648-5652.

(46)   Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

(47)   Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *Journal of Physical Chemistry* **1994**, *98*, 11623-11627.

(48)   Barone, V.; Cossi, M. *Journal of Physical Chemistry A* **1998**, *102*, 1995-2001.

(49)   Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *Journal of Computational Chemistry* **2003**, *24*, 669-681.

(50)   Takano, Y.; Houk, K. N. *J. Chem. Theory Comp.* **2005**, *1*, 71-78.

(51)   *NIST Chemistry WebBook, NIST Standard Reference Database Number 69, June 2005*; Linstrom, P. J.; Mallard, W. G., Eds.; National Institute of Standards and Technology: Gaithersburg, MD 20899, 2005; Vol. http://webbook.nist.gov.

(52)   The $\Delta\Delta S$ value obtained for hypoxanthine from our Cooks extended kinetic method experiments is 5 cal K^{-1} mol^{-1}. It has been noted that the $\Delta\Delta S$ value is related to the accuracy of the PA value obtained by the Cooks extended method. Ideally, the actual $\Delta\Delta S$ value should be less than or equal to about 5 cal K^{-1} mol^{-1}; otherwise, the extended kinetic method may underestimate the PA. Another caveat is that the $\Delta\Delta S$ value obtained from the extended method is often underestimated. There is therefore a possibility that the PA value of 222 kcal mol^{-1} obtained via our Cooks extended method experiment is underestimated; however, given that the bracketing result gives a PA of 217.5-220.2 kcal mol^{-1}, we are inclined to believe that the extended kinetic method value is not too low.

(53)   Wesdemiotis, C. *Journal of Mass Spectrometry* **2004**, *39*, 998-1003.

(54)   Bouchoux, G.; Sablier, M.; Berruyer-Penaud, F. *Journal of Mass Spectrometry* **2004**, *39*, 986-997.

(55)   Ervin, K. M.; Armentrout, P. B. *Journal of Mass Spectrometry* **2004**, *39*, 1004-1015.

(56)   Kurinovich, M. A.; Phillips, L. M.; Sharma, S.; Lee, J. K. *Chem. Commun.* **2002**, 2354-2355.

(57)   Schmalle, H. W.; Hänggi, G.; Dubler, E. *Acta Crystallographica* **1988**, *C44*, 732-736.

(58)   Ramaekers, R.; Maes, G.; Adamowicz, L.; Dkhissi, A. *Journal of Molecular Structure* **2001**, *560*, 205-221.

(59)   Lichtenberg, D.; Bergmann, F.; Neiman, Z. *Isr. J. Chem.* **1972**, *10*, 805-817.

(60)   Shukla, M. K.; Leszcynski, J. *Journal of Physical Chemistry A* **2003**, *107*, 5538-5543.

(61)   Sheina, G. G.; Stepanian, S. G.; Radchenko, E. D.; Blagoi, Y. P. *Journal of Molecular Structure* **1987**, *158*, 275-292.

(62)   Costas, M. E.; Acevedo-Chávez, R. *Journal of Physical Chemistry A* **1997**, *101*, 8309-8318.

(63)   Lin, J.; Yu, C.; Peng, S.; Akiyama, I.; Li, K.; Lee, L. K.; LeBreton, P. R. *Journal of Physical Chemistry* **1980**, *84*, 1006-1012.

(64)   Chenon, M.-T.; Pugmire, R. J.; Grant, D. M.; Panzica, R. P.; Townsend, L. B. *Journal of the American Chemical Society* **1975**, *97*, 4636-4642.

(65)   Grabowski, J. J.; Cheng, X. *Journal of the American Chemical Society* **1989**, *111*, 3106-3108.

(66)   Drohat, A. C.; Stivers, J. T. *Journal of the American Chemical Society* **2000**, *122*, 1840-1841.

(67)   Huang, Y.; Kenttämaa, H. *Journal of Physical Chemistry A* **2004**, *108*, 4485-4490.

(68)   Chandra, A. K.; Nguyen, M. T.; Uchimaru, T.; Zeegers-Huyskens, T. *Journal of Physical Chemistry A* **1999**, *103*, 8853-8860.

(69)   Chen, E. C. M.; Herder, C.; Chen, E. S. *Journal of Molecular Structure* **2006**, *798*, 126-133 and references therein.

(70)   Note that in reference 67, the calculated guanine acidity at B3LYP/6-31+G* is reported as 334.8 kcal mol^{-1}. We are not sure why their results differ slightly from ours, but the difference is very small and does not change the conclusions herein.

(71)   Lee, J. K. *Internat. Journal of Mass Spectrom* **2005**, *240*, 261-272.

(72)   Taylor, H. F. W. *Journal of the Chemical Society* **1948**, 765-766.

(73)   Jang, Y. H.; Goddard III, W. A.; Noyes, K. T.; Sowers, L. C.; Hwang, S.; Chung, D. S. *Journal of Physical Chemistry B* **2003**, *107*, 344-357.

(74)   Albert, A.; Brown, D. J. *Journal of the Chemical Society* **1954**, 2060-2071.

(75)   Langman, S. R.; Shohoji, M. C. B. L.; Telo, J. P.; Vieira, A. J. S. C.; Novais, H. M. *Journal of the Chemical Society, Perkin Transactions 2* **1996**, 1461-1465.

(76)   Horenstein, B. A.; Parkin, D. W.; Estupinan, B.; Schramm, V. L. *Biochemistry* **1991**, *30*, 10788-10795.

(77)   O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418-12429.

(78)   Rios-Font, R.; Rodriguez-Santiago, L.; Bertran, J.; Sodupe, M. *J. Phys. Chem. B* **2007**, *111*, 6071-6077.

(79)   Zoltewicz, J. A.; Clark, D. F.; Sharpless, T. W.; Grahe, G. *J. Am. Chem. Soc.* **1970**, *92*, 1741-1749.

(80)   Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106*, 506-555 and reference cited herein.

(81)   Gas phase calculations indicate that the most basic site of hypoxanthine and guanine is N7; for adenine it is N1. The general acid catalyst for AAG has not been identified so it is as yet unknown whether an acid is proximal enough to both the N7 of hypoxanthine and guanine as well as to the N1 of adenine to effect protonation. See references 18 and 22.

(82)   Meot-Ner (Mautner), M. *Journal of the American Chemical Society* **1979**, *101*, 2396-2403.

(83)   Turecek, F.; Chen, X. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1713-1726.

(84)   Liu, M.; Xu, M.; Lee, J. K. *J. Org. Chem.* **2008**, *73*, 5907-5914.

Chapter 3 Acidity Comparison among Nucleobases

3.1 Biological Introduction

DNA damage is ubiquitous in cells and may be caused by endogenous and exogenous agents, including chemicals from cell metabolism, inflammation, ultraviolet light, ionizing radiation, and environmental toxins. The damage to nucleobases results in abnormal base pairing, and when propagated by replication the damage is linked to aging and disease. Base excision repair (BER) is one of the important repair pathways that counteract the damage to maintain the fidelity of the genome. In the BER process, DNA glycosylases recognize and cleave the damaged DNA base by catalyzing the hydrolysis of the glycosidic bond, generating an apurinic site. The remaining abasic lesion (phosphate and sugar) is subsequently removed by an AP lyase and a gap forms at the abasic site of the DNA strand. This gap is then filled with the correct nucleobase by DNA polymerase and finally rejoined by DNA ligase.[1,2] The human alkyladenine DNA glycosylase (AAG) is one of the most important enzymes that initiate the BER repair process.

It is not clear how DNA glycosylases differentiate damaged nucleobases from normal bases. Those that only cleave specific damaged nucleobases are believed to form a highly discriminating active site to accommodate a certain damaged base. For example, uracil DNA glycosylase (UDG) accommodates only uracil and inhibits other bases in its active pocket by specific electrostatic interactions and steric effects.[3-7] In contrast, other DNA glycosylases, including AAG and thymine DNA glycosylase (TDG), cleave a broad range of damaged nucleobases. For example, AAG cleaves 3-methyladenine, 7-

methylguanine, hypoxanthine, and 1, N^6-ethenoadenine, while leaving guanine and adenine untouched.[3-5] These damaged nucleobases are often very similar in structure and size. It is not well established how these DNA glycosylases can recognize such a diverse group of substrates against normal nucleobases. It shows that neither the exocyclic amino group nor the steric effect prevents the normal purines from docking in the active site of AAG, because 3-methyladenine and 7-methylguanine also have exocyclic amino group and 1, N^6-ethenoadenine is much more bulky (see **Figure 3.1**).

Adenine   Guanine

3-methyladenine   7-methylguanine   hypoxanthine   1,N^6-ethenoadenine
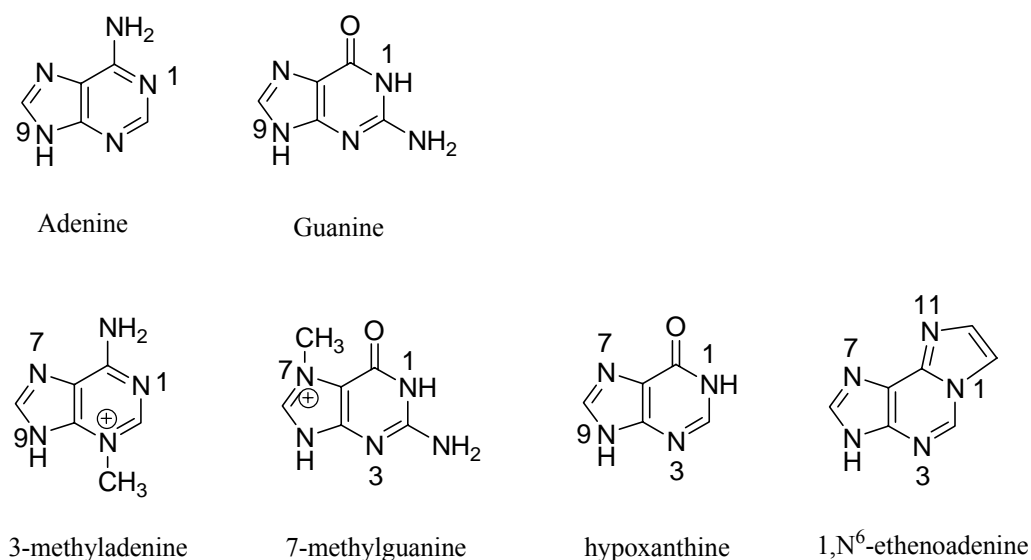
Figure 3.1 Structures of normal nucleobases and damaged nucleobases.

The mechanism underlying this "broad specificity" of AAG is unknown. The study of pH rate profiles showed that AAG-catalyzed excision of hypoxanthine involves both a general base (Glu 125) and general acids.[3,4,8] It can be imagined that hypoxanthine is cleaved as an anion or a cation. The acidity and proton affinity study in our group demonstrated that hypoxanthine is more acidic than guanine and adenine in the gas phase, which is an extreme mimic for the nonpolar environment at the active site of enzymes. The results indicate that the deprotonated hypoxanthine is a better leaving group than

normal bases.[9] The study of another AAG substrate, 1, N^6-ethenoadenine gave similar results.[10] Therefore AAG substrates seem to be more easily cleaved by enzymes than the normal purines. The hydrolysis of the glycosidic bond may proceed via a highly dissociative S_N2 or a stepwise S_N1 mechanism.[4,6,8] To make it simple, we only show the S_N1 mechanism in **Scheme 3.1**. In addition, the base excision repair initiated by DNA glycosylases involves a "base-flipping" step, whereby the nucleobase is "flipped" out of the duplexes and docked into the enzyme active site. The rate of flipping may be different for different substrates due to variable hydrogen bonding and base stacking abilities. It is not clear whether the flipping step is a fast or slow process. In our study, we focus on the steps where only actual base excision is involved.[4,6,8] A similar mechanism has been proposed for TDG, which cleaves a diverse group of pyrimidine nucleobases.[11] TDG has been shown to discriminate the substrates by their intrinsic leaving ability, and the hydrophobic active site of TDG enhances the differences in the substrates' leaving group abilities. Herein, we would like to study the leaving group abilities of the AAG substrates and compare them with the normal nucleobases to elucidate the excision mechanism.
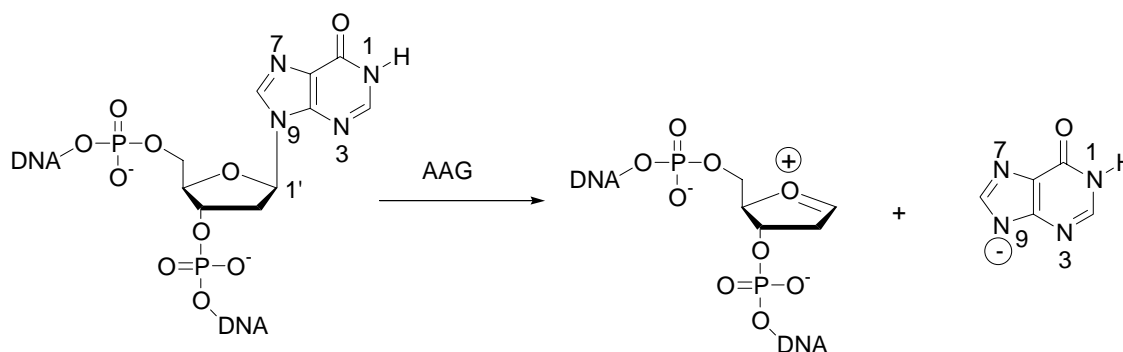
Scheme 3.1 Proposed S_N1 mechanism for AAG excising hypoxanthine.

O^6-Methylguanine (OMG) is a damaged base from alkylation of guanine and it is highly mutagenic since if not repaired, it can cause "GC to AT" transitions.[12-21] O^6-Methylguanine can be repaired by methylguanine methyl transferase (MGMT), which dealkylates OMG.[22-24] O^6-Methylguanine has been reported not to be cleaved by AAG within a period of 160 minutes in a study of DNA enzyme kinetics.[40] There are no obvious structural features that exclude O^6-methylguanine from the active site of AAG. The acidity of O^6-methylguanine might be the reason why AAG does not recognize and cleave it as we proposed. The comparison of O^6-methylguanine with normal bases and hypoxanthine may shed light on AAG cleavage mechanism.

3.2 Experimental Methods

3.2.1 Computational method

Theoretical calculations have been widely used to predict acidities in the gas phase and in different solvents. Gaussian03 was employed in our study.[25-30] All the gas phase calculations were carried out by density functional theory (DFT) at B3LYP and 6-31+G* basis set. The structures were fully optimized and then frequency calculations were performed on the optimized structures at the same level. The acidities ΔH^o_{acid} were calculated at 298K. The calculations provide important information on the structures and energies of nucleobases in the gas phase where the experimental measurements are hard to carry out.

Solvation studies were performed using the CPCM-SCRF method (full optimizations at B3LYP/6-31+G*; UAKS cavity) as implemented in Gaussian03.[31,32] This method has

been shown to be reliable for neutral and ionic organic molecules.[33] Molecules are placed in a cavity surrounded by an infinite polarizable dielectric, and their structures and free energies can be simulated. The results were used to determine the acidity changes of nucleobases in different environments ranging from very polar solvents such as water (ε = 78) to nonpolar solvents such as cyclohexane (ε = 2.0).

3.2.2 Cooks kinetic method

The Cooks kinetic method was used to compare the acidities in a quadrupole ion trap mass spectrometer.[34-37] This method involves the formation of a proton bound dimer of the two deprotonated nucleobases, denoted by A and B in **Equations 3.1-3.6**. Collision-induced dissociation (CID) of this dimer leads to the formation of deprotonated nucleobases. The ratio of these two deprotonated products yields the relative acidities of the two compounds. **Equations 3.1-3.6** summarize the data analysis process. The method assumes that the dissociation has no reverse activation energy barrier, and that the dissociation transition structure is late and therefore indicates the stability of the two deprotonated products. Both these assumptions are generally true for proton bound systems.[37-39]

$$[AHB]^- \xrightarrow{k_1} AH + B^- \qquad \textbf{(Eq. 3.1)}$$

$$\xrightarrow{k_2} BH + A^- \qquad \textbf{(Eq. 3.2)}$$

$$BH + A^- \underset{}{\overset{K}{\rightleftharpoons}} AH + B^- \qquad \textbf{(Eq. 3.3)}$$
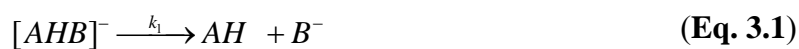
$$K \approx k_1 / k_2 \qquad \textbf{(Eq. 3.4)}$$

$$\Delta H_{acid1} - \Delta H_{acid2} \approx RT_{eff} \ln K \qquad \textbf{(Eq. 3.5)}$$

$$\ln(k_1 / k_2) = \frac{1}{RT_{eff}} (\Delta H_{acid1} - \Delta H_{acid2}) \qquad \textbf{(Eq. 3.6)}$$

The acidity of hypoxanthine with five other nucleobases was compared. The product ion distributions were measured three times to ensure reproducibility. As shown in **Eq. 3.1-3.6**, T_{eff} is the effective temperature of dissociation of the proton bound complex in Kelvin. Here T_{eff} was obtained from the measurement of proton affinity of hypoxanthine by Cooks Kinetic Method. The $\ln(k_1/k_2)$ is equal to $\ln([A^-]/[B^-]$ which was given by the measured ion abundance of A and B (**Eq. 3.1**). Then the acidity difference between two nucleobases A and B, $\Delta H_{acid1} - \Delta H_{acid2}$, was calculated. The standard deviation of our measurements is ± 3 kcal mol^{-1}.

In the kinetic method experiments, solutions of nucleobases were subjected to electrospray ionization (10^{-3} to 10^{-4} M solutions in methanol; a small amount of ammonia is added). Guanine could not be completely dissolved in the solvent, and its supernatant was used. The typical infusion flow rate was 25µL/min and the electrospray needle voltage was ~ 4500 V. The proton-bound complexes of the two nucleobases were isolated and subjected to collision-induced dissociation (CID) for about 30ms. The isolation window for the parent ions is 3 Da. About 40 scans were averaged to generate the spectra of product ions.

3.2.3 pK$_a$ measurement using spectrophotometric method

A spectrophotometric method was also used to measure the pK$_a$ values. Since pK$_a$ values are solvent dependent, the same substance shows different pK$_a$ values in different solvents. Thymol blue was chosen as an indicator to measure pK$_a$. A thymol blue solution of certain concentration was first prepared by mixing with excess tertabutylammonium hydroxide and the UV absorbance was measured. Under these conditions, thymol blue is totally deprotonated. The total concentration of the indicator was calculated using the extinction coefficient at 622nm. A second thymol blue solution of the same concentration was then prepared by mixing with tertabutylammonium hydroxide, but this time thymol blue is in excess. The UV absorbance was measured and the initial concentration of deprotonated thymol blue present in the cell was determined. A small aliquot of prepared nucleobase solutions was then added to the second solution and the UV absorbance was measured. This was repeated three times to the same indicator solution. From the change of UV absorbance after addition of each aliquot of nucleobase solutions under study, the equilibrium constant between indicator and the nucleobase added was calculated. Finally the pK$_a$ of the nucleobase was determined from the equilibrium constant and known pK$_a$ value of the indicator (see **Equation 3.7**). For each set of the aliquots added, several pK$_a$ values were calculated and they varied by about 0.2 pK unit. Three indicator solutions were prepared and three parallel runs were made for reproducibility.

$$I^- + HA \rightleftharpoons A^- + HI$$

$$K_{eq} = \frac{[A^-][HI]}{[I^-][HA]} = \frac{[A^-][H^+]}{[HA]} \frac{[HI]}{[I^-][H^+]} = K_a^{HA} / K_a^{HI} \qquad \textbf{(Eq. 3.7)}$$
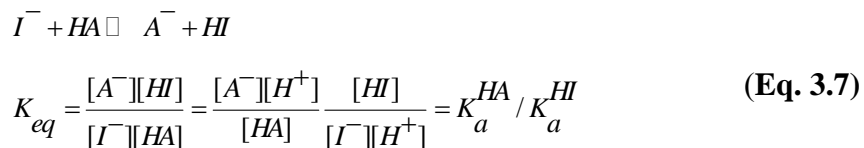
3.3 Results and Discussion

3.3.1 Calculated acidity in different media

Although there are several tautomers for each nucleobase, herein we only calculate the biologically relevant structures (canonical structures, with a proton attached to N9 site) in each medium. Neutral and deprotonated (at N9 site) nucleobases were placed in environments with different dielectric constants ranging from cyclohexane ($\varepsilon = 2.0$) to water ($\varepsilon = 78.4$). Their structures were optimized at B3LYP/6-31+G* level in CPCM model. Their acidities were calculated from the difference of free energy between neutral and deprotonated molecules. The results are summarized in **Table 3.1**. The acidities relative to hypoxanthine are listed in parentheses.

Table 3.1 Summary of results for acidity calculations in different media

Media	hypoxanthine	1, N^6-ethenoadenine	Adenine	Guanine	O^6-methyl guanine
Gas phase ($\varepsilon = 1.0$)	330.5 (0.0)	330.7 (0.2)	334.8 (4.3)	334.3 (3.8)	335.3 (4.8)
Cyclohexane ($\varepsilon = 2.0$)	315.1 (0.0)	315.9 (0.8)	318.5 (3.4)	318.7 (3.6)	319.5 (4.4)
Methylene chloride ($\varepsilon = 9.0$)	297.7 (0.0)	298.7 (1.0)	300.5 (2.8)	300.8 (3.1)	301.7 (4.0)
Dimethyl sulfoxide ($\varepsilon = 46.7$)	293.5 (0.0)	294.5 (1.0)	296.1 (2.6)	296.5 (3.0)	297.3 (3.8)
Water ($\varepsilon = 78.4$)	292.6 (0.0)	294.1 (1.5)	294.8 (2.2)	294.6 (2.0)	295.8 (3.3)

In each medium, hypoxanthine is the most acidic nucleobase. The acidities of adenine and guanine are very close. The calculated acidity of each nucleobase increases as the polarity of solvents increases. The acidity differences between hypoxanthine and adenine, hypoxanthine and guanine (the values in parentheses) increase as the polarity of the

solvents decreases. The acidity differences show the biggest values in the gas phase, allowing AAG to easily differentiate the damaged nucleobases from the natural ones. The results are consistent with the proposed mechanism of deprotonation prior to cleavage.

3.3.2 Acidity comparison among nucleobases by Cooks kinetics method

The absolute acidity of hypoxanthine in the gas phase has been measured in Chapter 2.[9] Other members in our group have measured gas phase acidities of adenine, adenine derivatives, guanine, 1, N^6-ethenoadenine and O^6-methylguanine.[10,41-43] The measurements were carried out by bracketing method on FTMS or Cooks Kinetic method on LCQ. Each of these results associates with a 2-3kcal mol^{-1} error bar, which causes considerable uncertainty for comparison. To solve this problem, the acidities of the damaged and normal bases were directly compared by Cooks kinetic method. In the sample preparation, two nucleobases under study were added to form a proton bound complex. By applying collision energy, the complex was dissociated into nucleobase anions and their ion intensities in the gas phase were compared. The nucleobase with higher ion intensity has the higher acidity, and their ion abundance ratio reflects their acidity difference. The comparison results are listed in **Table 3.2**.

The T_{eff} of 447K, obtained from Cooks kinetic measurements of proton affinity of hypoxanthine, was used in the calculation of acidity difference. The masses of hypoxanthine and adenine are similar, 135 and 136. With an isolation window of 3, the dimer of adenine (*m/z* 269) was isolated and subjected to CID together with the complex of hypoxanthine and adenine (*m/z* 270). In CID process, both adenine dimer and targeted complexes dissociate, and the ions from adenine dimer interfere with the determination of

ion ratio for acidity difference. To solve this problem, dimethyladenine was introduced to compare with adenine and hypoxanthine, respectively. Moreover, the result usually associates with a large error bar if the acidity difference is larger than 3 kcal mol^{-1} in **Table 3.2.** That is because the big difference in acidities of the nucleobases results in a very low ion abundance of one of the two peaks, which cannot be measured accurately. This may lead to a very large uncertainty of the ion ratio and thus a large error bar of acidity difference. Generally, it is better to keep the acidity difference within 3 kcal mol^{-1} when comparing different nucleobases.

Table 3.2 Experimental acidity differences among nucleobases (first listed is the more acidic one).

	Acidity difference (kcal mol^{-1})	STDEV
A-DMA*	1.70	0.19
H-DMA	4.99	0.72
G-A	0.67	0.08
H-G	2.55	0.10
H-eA	0.04	0.01
H-OMG	3.83	0.30
G-OMG	3.07	0.10
A-OMG	2.46	0.10
DMA-OMG	1.00	0.17

*In this table, A stands for adenine, G for guanine, H for hypoxanthine, DMA for dimethyladenine, eA for 1, N^6-ethenoadenine, and OMG for O^6-methylguanine.

The acidities of nucleobases from different methods are summarized in **Table 3.3**. Prepared in aqueous solution, all of the samples have canonical structures and the most acidic sites are all at the N9 site. Presumably, the electrospray is a fast process, and would not change the structure conformation. It ensures that the structures we study in the gas phase are still the ones we prepare in solution. Hypoxanthine and 1, N^6-ethenoadenine are the two most acidic compounds. They are both substrates of AAG and

the acidity difference between them is about 0.04 kcal mol^{-1}. Both of them are about 2-3 kcal mol^{-1} more acidic than normal bases (guanine by 2.6 kcal mol^{-1} and adenine by 3.2 kcal mol^{-1}). O^6-methylguanine is much less acidic than normal bases. These results are consistent with other groups' enzyme kinetics experiments and support the cleavage mechanism we proposed. AAG recognizes the damaged nucleobases by their acidities and cleaves them as anions. The more acidic the nucleobases, the better the leaving groups. O^6-methylguanine is much less acidic and therefore can not be cleaved by AAG.

Table 3.3 Summary of the gas phase acidities and relative acidities of nucleobases (in kcal mol-1).

Nucleobase	Calculated acidity	FTMS (bracketing)	LCQ (Cooks)	Relative acidity
hypoxanthine	330.5	332		0.0
1,N^6-ethenoadenine	330.7	332		0.04 (±0.01)
Guanine	334.3		335	2.55 (±0.10)
Adenine	334.8	333	335	3.22 (±0.13)
6-dimethyladenine	334.0	333		4.99 (±0.72)
O^6-methylguanine	337.4	338	337	5.62 (±0.14)

3.3.3 pK$_a$ measurement using spectrophotometric method

Our calculations have already shown that the calculated acidity difference between the damaged nucleobases and normal ones increases as the dielectric constant decreases. The experimental measurements of the acidities were carried out in different solvents using spectrophotometric method to confirm the calculation results. However, the

solubility of the nucleobases caused problems. With decrease of the dielectric constant of the solvent, nucleobases become less soluble in nonpolar solvents, making it impossible to accurately measure their acidities. Only pK_a of these nucleobases in DMSO were measured and the results are summarized in **Table 3.4**. The pK_a of nucleobases in DMSO are higher than those in water, with hypoxanthine as the most acidic one. This is consistent with the computational results. The acidity differences of nucleobases in DMSO are similar to those in water. The dielectric constant of DMSO is 46 and lies between water (dielectric constant = 78) and the gas phase (dielectric constant = 1), but the acidity differences are not as what we predicted. Maybe the acidity differences change dramatically when the environment changes from a solvent to a total solvent-free medium, although they change a little when we compare them among different solvent media.

Table 3.4 Acidity of nucleobases in the gas phase and different solvents

nucleobases	Gas phase (kcal/mol) calculated	Water (pK_a) Difference in kcal/mol	DMSO (pK_a) Difference in kcal/mol
guanine	334.3 (3.8)	10.0 (1.5)	14.67±0.09 (1.2)
adenine	334.8 (4.3)	9.8 (1.2)	14.66±0.04 (1.2)
1,N^6-ethenoadenine	330.7 (0.2)	9.9 (1.4)	14.75±0.04 (1.3)
hypoxanthine	330.5 (0.0)	8.9 (0.0)	13.81±0.13 (0.0)

3.4 Conclusions

The mechanism of damaged nucleobase recognition and cleavage by repair enzyme is of high importance to understand functions of enzyme in biosystems. We proposed that AAG differentiates damaged nucleobases from normal nucleobases by their acidity

difference. Damaged nucleobases are more acidic (better leaving groups) than normal bases, thus it supports the mechanism that the damaged bases are recognized and cleaved by AAG as an anion. In addition, the nonpolar environment at the enzyme active site could enhance the acidity difference and cleavage selectivity between damaged and normal bases.

In this chapter, the acidities of damaged and normal bases in different environments were calculated. The results showed that hypoxanthine and 1, N^6-ethenoadenine, the substrates of AAG, are more acidic than normal bases in the gas phase and all solvents studied. O^6-Methylguanine, which could not be cleaved by AAG, was calculated to be the least acidic one in all media among the bases studied here, including the normal bases. The acidity differences are enlarged with the decrease of polarity of the solvents, especially when the polarity of solvents is close to gas phase. The results support the above AAG cleavage mechanism. However, measuring the acidities of bases in different solvents was limited by the solubility of the nucleobases. We only obtained partial experimental data, which makes it hard to compare these experimental results with those of calculations.

The acidity difference between damaged and normal nucleobases in the gas phase was experimentally measured by Cooks kinetic method. This method enables the direct and accurate acidity comparison of these nucleobases. As expected, hypoxanthine and 1, N^6-ethenoadenine are more acidic than normal bases and O^6-methylguanine, which is much less acidic than normal bases and cannot be cleaved by AAG. These results are consistent with the AAG cleavage mechanism we proposed.

3.5 Reference:

(1)   Cunningham, R. P. *DNA Glycosylases. Mutat. Res.* **1997**, *383*, 189-196.

(2)   Friedberg, E. C.; Walker, G. C.; Siede, W. *DNA Repair and Mutagenesis*; ASM Press: Washington, DC, 1995.

(3)   O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418-12429.

(4)   O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *2004*, 9750-9757.

(5)   O'Brien, P. J.; Ellenberger, T. *J. Biol. Chem.* **2004**, *279*, 26876-26884.

(6)   Stivers, J. T.; Jiang, Y. L. *Chem. Rev.* **2003**, *103*, 2729-2759.

(7)   Parikh, S. S.; Walcher, G.; Jones, G. D.; Slupphaug, G.; Krokan, H. E.; Blackburn, G. M.; Tainer, J. A. *Proc. Nati. Acad. Sci. US* **2000**, *97*, 5083-5088.

(8)   Berti, P. J.; McCann, J. A. B. *Chem. Rev.* **2006**, *106*, 506-555 and reference cited herein.

(9)   Sun, X.; Lee, J. K. *J. Org. Chem.* **2007**, *72*, 6548-6555.

(10)   Liu, M.; Xu, M.; Lee, J. K. *J. Org. Chem.* **2008**, *73*, 5907-5914.

(11)   Bennett, M. T.; Rodgers, M. T.; Hebert, A. S.; Ruslander, L. E.; Eisele, L.; Drohat, A. C. *J. Am. Chem. Soc.* **2006**, *128*, 12510-12519.

(12)   Beranek, D. T. *Mutat. Res.* **1990**, *231*, 11-30.

(13)   De Bont, R.; van Larabeke, N. *Nutagenesis* **2004**, *19*, 169-185.

(14)   Dimitri, A.; Burns, J. A.; Broyde, S.; Scicchitano, D. A. *Nucleic Acids Res.* **2008**, *36*, 6459-6471.

(15)   Liu, J. G.; Li, M. H. *Carcinogenesis* **1989**, *10*, 617-620.

(16)   Loechler, E. L. *Chem. Res. Toxicol.* **1994**, *7*, 277-280.

(17)   Rydberg, B.; Lindahl, T. *EMBO J.* **1982**, *1*.

(18)   Sedgwick, B. *Carcinogenesis* **1997**, *18*, 1561-1567.

(19)   Souliotis, V. L.; Kaila, S.; Boussiotis, V. A.; Pangalis, G. A.; Kyrtopoulos, S. A. *Cancer Res.* **1990**, *50*, 2759-2764.

(20)   Taverna, P.; Catapano, C. V.; Citti, L.; Bonfanti, M.; D'Incalci, M. *Anticancer Drugs* **1992**, *3*, 401-405.

(21)   Wyatt, M. D.; Pittman, D. L. *Chem. Res. Toxicol.* **2006**, *19*, 1580-1594.

(22)   Chen, J. M.; Zhang, Y. P.; Wang, C.; Sun, Y.; Fujimoto, J.; Ikenaga, M. *Carcinogenesis* **1992**, *13*, 1503-1507.

(23)   Pegg, A. E.; Dolan, M. E.; Scicchitano, D. A.; Moriomoto, K. *Environ. health Perspect.* **1985**, *62*, 109-114.

(24)   Pegg, A. E.; Fang, Q.; Loktionova, N. A. *DNA Repair* **2007**, *6*, 1071-1078.

(25)   Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372-1377.

(26)   Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.

(27)   Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A. J.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A.; 03 ed.; Gaussian, Inc.: Pittsburgh, PA, 2003.

(28)   Kohn, W.; Becke, A. D.; Parr, R. G. *J. Phys. Chem.* **1996**, *100*, 12974-12980.

(29)   Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785-789.

(30)   stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.

(31)   Barone, V.; Cossi, M. *Journal of Physical Chemistry A* **1998**, *102*, 1995-2001.

(32)   Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. *Journal of Computational Chemistry* **2003**, *24*, 669-681.

(33)   Takano, Y.; Houk, K. N. *J. Chem. Theory Comp.* **2005**, *1*, 71-78.

(34)   McLuckey, S. A.; Cameron, D.; Cooks, R. G. *Journal of the American Chemical Society* **1981**, *103*, 1313-1317.

(35)   McLuckey, S. A.; Cooks, R. G.; Fulford, J. E. *Internat. Journal of Mass Spectrom & Ion Processes* **1983**, *52*, 165-174.

(36)   Brodbelt-Lustig, J. S.; Cooks, R. G. *Talanta* **1989**, *36*, 255-260.

(37)   Green-Church, K. B.; Limbach, P. A. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 24-32.

(38)   Ervin, K. M. *Chemical Reviews* **2001**, *101*, 391-444 and references therein.

(39)   Gronert, S.; Feng, W. Y.; Chew, F.; Wu, W. *Internat. Journal of Mass Spectrom* **2000**, *196*, 251-258.

(40)   Abner, C. W.; Lau, A. Y.; Ellenberger, T.; Bloom, L. B. *J. Biol. Chem.* **2001**, *276*, 13379-13387.

(41)   Sharma, S.; Lee, J. K. *J. Org. Chem.* **2002**, *67*, 8360-8365.

(42)   Sharma, S.; Lee, J. K. *J. Org. Chem.* **2004**, *69*, 7018-7025.

(43)   Zhachkina, A.; Liu, M.; Sun, X.; Amegayibor, F. S.; Lee, J. K. *J. Org. Chem.* **2009**.

Chapter 4 Hydrogen Bonding and Base Stacking

4.1 Biological Introduction

Deoxyribonucleic acid (DNA) contains and transfers genetic information in cellular lives. Hydrogen bonding and base stacking of DNA bases are the two major forces to form and stabilize the DNA double helix. The two complementary strands are associated by hydrogen bonds, which play important roles in stabilizing the DNA duplex structure and provide the specificity for genetic information transfer. Adjacent bases also associate with each other by 'base stacking',[1] which is an electrostatic interaction and favored by conformation of the DNA double helix. Two major interactions involved in base stacking are London dispersion forces[2] and interactions between partial charges within adjacent rings (**Figure 4.1**).[3]
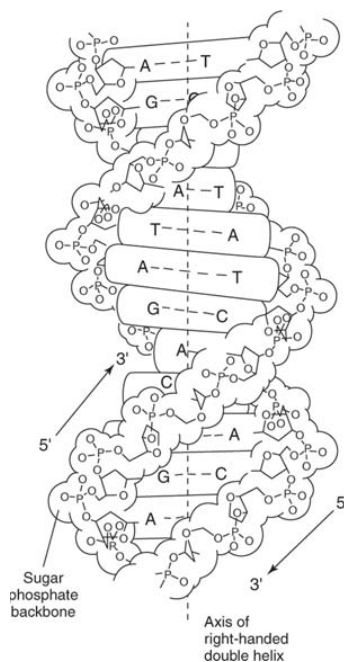
Figure 4.1 DNA duplex.

There are many methods used to study the DNA double helix. Experimental methods, such as X-ray crystallography and NMR, may yield some structural information but can not simultaneously determine the geometries and stabilization energies of the DNA duplexes. Quantum mechanical (QM) calculations, including the Hartree-Fock (HF) and density functional theory (DFT) methods, give reasonable structural and energetic information for hydrogen bonding but fail to yield accurate predictions for London dispersion energy. As a result the calculations and experiment results are not always consistent.

Mass spectrometry is a very useful tool to investigate biomolecular interactions. Without interference of solvents, the intrinsic reactivity of molecules can be studied. For example, DNA in aqueous solution is stabilized by hydrogen bonding, base stacking, electrostatic forces (Coulombic repulsion), van der Waals forces, hydrophobic interactions (solvent dependent) and some nonclassical interactions. Furthermore, these interactions are sensitive to the local environment. Examining the DNA duplexes in the gas phase, which is a solvent free environment, reduces the complexity and reveals the intrinsic interactions.[4,5]

Herein, our objective is to study hydrogen bonding and base stacking, the two major interactions, in terms of stabilizing the DNA duplexes in the gas phase, and evaluate their relative significance. First, we have learned from previous work that hydrogen bonding is dominant in the gas phase and important in solution: a higher GC content (more hydrogen bonds) leads to higher solution and gas-phase stabilities of duplexes.[6-9] Therefore, the ranking of the gas phase stability should track with solution phase for the duplexes with the same length but different GC content. If GC content is kept constant within a series of

DNA duplexes, which means the hydrogen bonding effects are similar, the base stacking should account for the stability differences in the gas phase and in solution. To test this, a set of 6-mer (6-mer-A-D), a set of 8-mer (8-mer-G-L) complementary DNA duplexes with the same GC content and a set of 8-mer (8-mer-A-F) with varying GC content are designed for this study. Second, to further study the effect of base stacking itself, we also designed some series of DNA duplexes (8-mers and 9-mers, for instance, 9-mer with GXG/CYC moiety in the center, 8-mer 14-17xy, etc.) with the same GC composition, leaving base stacking to be the only major factor that results in the stability difference in the gas phase (assume other interactions are negligible).

4.2 Experimental Methods

4.2.1 Sample Preparation

Oligodeoxynucleotide single strands were purchased from Sigma Aldrich (The Woodlands, TX). These single strands were pre-desalted and used without further purification. Two single strands as designed were put together to make duplex stock solutions containing 62.5 μM of each strand in 40 mM NH$_4$OAc aqueous solution at pH 7.0. The stock solutions were annealed at 90 °C for 10 minutes and then slowly cooled down to 0 °C. Before injecting to ESI-MS, the stock solutions were diluted to 12.5 μM in 40 mM NH$_4$OAc mixed with 20% methanol.

4.2.2 Melting Temperature in Solution

An Aviv 14D-S spectrophotometer was utilized to obtain the UV melting curves. The concentration of the duplex sample was 5 μM in 40 mM NH$_4$OAc and kept constant for

all the melting measurements. The quartz cells were heated from 0 °C to 80 °C and absorbance was measured at 260 nm. Data were processed by KaleidaGraph software. The error bar for this measurement was estimated to be 0.3 °C.[10]

The T_m's of the DNA duplexes was predicted by the program "MELTING".[11] The settings are as follows: (1) hybridization type: dnadna (for a DNA duplex); (2) nearest-neighbor parameters set: all97a.nn;[12] (3) salt concentration: 0.04 M; (4) nucleic acid concentration (total): 10 µM; (5) nucleic acid correction factor: 4 (for non-self-complementary duplex[13]); (6) salt correction: san98a.[14] The error for estimating T_m's by MELTING program is ±1.6 °C.[14]

4.2.3 ESI-Quadrupole Ion Trap Mass Spectrometer

Mass spectrometric measurements were performed on a Finnigan LCQ mass spectrometer (San Jose, CA). The 0°C solution was injected into the Mass spectrometer by direct infusion at 25 µL/min. The spray voltage was –4.0 kV (negative ion mode) and the capillary was heated to 175 °C. Duplex abundance is calculated as follows (**Equation 4.1**): [6]

$$\text{Duplex(\%)} = \frac{2 \times [\text{all duplexes}]}{[\text{all single strands}] + 2 \times [\text{all duplexes}]} \times 100 \qquad (\textbf{Eq. 4.1})$$

Collision-induced dissociation (CID) experiments were conducted in ion trap by varying the relative collision energy with a default activation time of 30ms and a q value of 0.25. The isolation width under CID mode is 5 Da for all the duplex ions. Experimental MS conditions were tuned using the –3 charged duplex ions of 8-mer-II (m/z =1605). Then the optimized parameters were utilized to all the duplexes.

In CID process, a normalized collision energy (in %) is applied to the parent ions based on the different m/z ratio. The normalized collision energy is scaled across the entire mass range with relatively less energy to lower m/z ions and more energy to higher m/z ions.[15] In the measurement of the gas phase stability, proper duplex ions are chosen as the parent ions and subjected to increasing collision energies. E_{50} is used to characterize the gas-phase stability and defined as the applied collision energy at which 50% of the duplexes are dissociated into single strands.[9] The larger the E_{50} value, the more stable the parent duplex ions in the gas phase. Plot of the duplex ion abundance to collision energy by sigmoid equation is similar to a solution phase melting curve of double-strand DNA. The corresponding E_{50} values were derived by Origin 6.0 software.[16] It is noteworthy that comparison between E_{50} values can only be made under the same experimental conditions.[17,18]

4.3 Results and Discussion

In previous work, our group studied a series of complementary and mismatched 9-mer duplexes[18] and a series of duplexes varying in length and sequence.[19] The gas-phase stability does not always track with the solution phase stability. To further study the intrinsic interactions, we designed a series of 6-mers and two sets of 8-mer non-self-complementary duplexes. The 6-mers were chosen because we saw the gas/solution phase exception within this set when we tested different length duplexes: the melting temperature of 6-mer-I is higher than 6-mer-II in the solution phase, whereas, 6-mer-II has a higher E_{50}, which indicates a more stable structure in the gas phase.[19] We chose the 8-mers because these are extremely well behaved and have a strong signal at the -4

charge state among the original duplexes examined in previous study of duplexes with different length.[19] The 6-mer DNA duplexes (6-mer-A, B, C, D, I and II) have the same composition but different sequence. One set of 8-mer DNA duplexes (8-mer-A-F) has varying GC content, whereas the other 8-mer set (8-mer-G-L) has the same GC content and different sequence. The term "ssX" and "ssY" are used to refer to the two single strands that comprise each duplex. In a given duplex, "ssX" is always the strand with higher mass.

4.3.1 Solution Phase Stability of DNA Duplex

T_m (melting temperature) is a conventional parameter to characterize the stability of DNA in the solution phase. The ordered helix double strand dissociates into two single strands when a DNA duplex is slowly heated. The single strands and the duplexes show different UV absorption properties. Therefore, the dissociation results in an increase of UV absorption, which is called "hyperchromicity".[20] T_m is the midpoint of a UV absorption transition. The melting temperature ($T_{m,calc}$) can also be calculated by the program MELTING.[11] Both T_m's and $T_{m,calc}$'s of the 8-mer DNA duplexes were determined and listed in **Table 4.1**. To evaluate how good the prediction is, $T_{m,calc}$ was plotted versus $T_{m,expt}$ (**Figure 4.2**). The obtained line with good linearity indicates that the calculated T_m value is a reliable predictor of duplex stability. Since it is time-consuming to measure the T_m experimentally, $T_{m,calc}$ can be conveniently used to study DNA stability in the solution phase.

Table 4.1 $T_{m, calc}$ and $T_{m,expt}$ for 8-mer duplexes

Duplex Name	Sequence	$T_{m,calc}$ (° C)	$T_{m,expt}$ (° C)
SERIES 1			
8-mer-A	5'-AAAAAAAA-3'	6.8	5.2
	3'-TTTTTTTT-5'		
8-mer-B	5'-CAAAAAAG-3'	13.1	17.2
	3'-GTTTTTTC-5'		
8-mer-C	5'-ACTGGATT-3'	18.0	18.5
	3'-TGACCTAA-5'		
8-mer-D	5'-AAGCGTAG-3'	24.1	24.2
	3'-TTCGCATC-5'		
8-mer-E	5'-GAGGTCGT-3'	28.0	32.3
	3'-CTCCAGCA-5'		
8-mer-F	5'-GAGCCGTG-3'	33.4	38.8
	3'-CTCGGCAC-5'		
SERIES 2			
8-mer-G	5'-AGAGAGAG-3'	18.3	n/a
	3'-TCTCTCTC-5'		
8-mer-H	5'-GACTAGGT-3'	18.9	20.1
	3'-CTGATCCA-5'		
8-mer-J	5'-AGTCCAGA-3'	22.1	24.0
	3'-TCAGGTCT-5'		
8-mer-D	5'-AAGCGTAG-3'	24.1	24.2
	3'-TTCGCATC-5'		
8-mer-K	5'-ATACAGCG-3'	24.2	n/a
	3'-TATGTCGC-5'		
8-mer-L	5'-AAGCGCTA-3'	27.6	33.8
	3'-TTCGCGAT-5'		

Figure 4.2 Correlation of experimental T_m with calculated T_m for 8-mers.

4.3.2 Full-Scan Mass Spectrometry of DNA Duplexes

Full scan can be used to establish the relationship between solution phase stability and mass spectrometric ion abundance. It is generally recognized that the ion abundance in the gas phase reflects the strength of non-covalent interactions for non-covalent complexes, such as enzyme-inhibitor complexes.[6,18,21-23] Moreover, GC composition (hydrogen bonding) is not the only factor that contributes to solution and gas-phase stability of duplexes. Base stacking also plays an important role in stabilizing the DNA double helical structures. To study the importance of hydrogen bonding and base stacking effects, two series of 8-mers (series 1 and 2) were studied. The first set, 8mer A-F (series 1), varies in GC% (0-75%), while the second set of 8-mers, G, H, J, D, K and L (series 2) has a constant GC% which is 50%. To assess the correlation between full-scan duplex ion abundance and solution phase stability, the normalized duplex percentages (DS%) were plotted versus solution phase $T_{m,calc}$ values (**Table 4.2** and **Figure 4.3**).

Table 4.2 GC%, $T_{m, calc}$, DS% and CID data for 8-mer duplexes

Duplex Name	Sequence	GC%	$T_{m,calc}$ (°C)[*]	DS%	E_{50} (%)
SERIES 1					
8-mer-A	5'-AAAAAAAA-3' 3'-TTTTTTTT-5'	0	7.5	25.6±0.6	n/a
8-mer-B	5'-CAAAAAAG-3' 3'-GTTTTTTC-5'	25	13.8	25.7±1.4	7.27±0.07
8-mer-C	5'-ACTGGATT-3' 3'-TGACCTAA-5'	37.5	18.8	35.6±0.4	9.02±0.06
8-mer-D	5'-AAGCGTAG-3' 3'-TTCGCATC-5'	50	24.7	51.9±1.5	9.86±0.04
8-mer-E	5'-GAGGTCGT-3' 3'-CTCCAGCA-5'	62.5	28.7	70.5±0.4	10.19±0.03
8-mer-F	5'-GAGCCGTG-3' 3'-CTCGGCAC-5'	75	34.0	81.6±0.2	10.41±0.06
SERIES 2					
8-mer-G	5'-AGAGAGAG-3' 3'-TCTCTCTC-5'	50	19.0	44.9±0.7	9.73±0.01
8-mer-H	5'-GACTAGGT-3' 3'-CTGATCCA-5'	50	19.7	45.7±1.4	9.14±0.08
8-mer-J	5'-AGTCCAGA-3' 3'-TCAGGTCT-5'	50	22.9	47.3±0.8	10.03±0.07
8-mer-D	5'-AAGCGTAG-3' 3'-TTCGCATC-5'	50	24.7	51.9±1.5	9.86±0.04
8-mer-K	5'-ATACAGCG-3' 3'-TATGTCGC-5'	50	24.9	55.1±1.7	9.81±0.06
8-mer-L	5'-AAGCGCTA-3' 3'-TTCGCGAT-5'	50	28.3	61.1±0.2	10.27±0.06

[*] T_m is calculated at ESI condition, whose concentration is 12.5 μM.
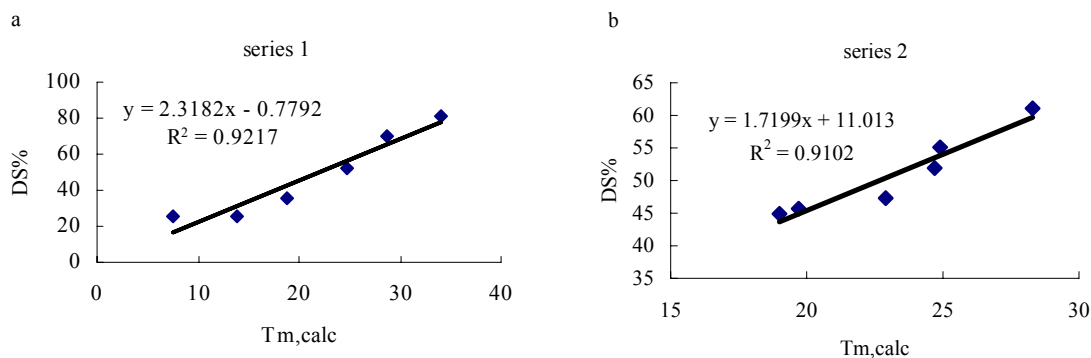
Figure 4.3 Correlation of calculated T_m values ($T_{m,calc}$) with normalized duplex percentage (DS%) of 8-mers: (a) Series 1; (b) Series 2.

For series 1, from 8-mer-A to F, the $T_{m,calc}$ increases with increasing of the GC content, and so do the DS% values. For series 2, although the GC content is a constant (GC% = 50%), the DS% values still roughly track with $T_{m,calc}$ (see **Table 4.2**). Nevertheless, while taking a whole look at **Table 4.2**, the overall DS% values are not consistent with $T_{m,calc}$. The complicating factor is GC content. For instance, 8-mer-C and 8-mer-G have nearly the same $T_{m, calc}$ (18.8 versus 19.0° C), but the 8-mer-G, which has a higher GC%, shows a higher DS% (44.9% versus 35.6%). Similarly, 8-mer-E has a comparable $T_{m,calc}$ as 8-mer-L (28.7 versus 28.3° C), but 8-mer-E (higher GC content) shows a higher DS% (70.5% versus 61.1%). Thus, the duplexes with the same melting temperature may exhibit different relative ion abundances in the ESI-MS. The possible reason is that the duplexes with higher GC content may survive better during the ESI process and therefore show stronger ion signals.

4.3.3 Gas-Phase Stability of DNA Duplexes Examined by CID

Our interest here is to compare the gas and solution phase stabilities of the DNA duplexes. As a general rule, a higher E_{50} means a more stable duplex in the gas phase, and a higher $T_{m,calc}$ indicates a more stable duplex in the solution phase. The comparison between the gas phase and solution phase stabilities can be carried out by comparing E_{50} and $T_{m,calc}$. In previous studies, we found that the gas phase stabilities of 6-mer-I and II (sequences in **Table 4.3,** GC% = 50%) do not track with their solution phase stabilities. The gas phase stability of 6-mer-I (E_{50} = 12.57%) is less than that of 6-mer-II (E_{50} = 12.75%), whereas, the $T_{m,calc}$ of 6-mer-I is higher, meaning that 6-mer-I is more stable in the solution phase. In a study of four 16-mer duplexes with 50% GC content, Gabelica and coworkers found one duplex with unparallel gas/solution phase stabilities.[7] They explained that the four 'A's in a row in the sequence of the "misbehaving" duplex might be so bulky that distort the local B-DNA structure in solution and cause problems to predict solution phase stability by T_m. We also found some outliers when we compare T_m versus E_{50} in our study of a series of 9-mers.[18] All these results demonstrate that gas phase and solution phase stabilities do not always track with each other even though they have the same number of hydrogen bonds. The possible reason is that the factors affecting the stabilities are different in the gas phase from in solution, and base stacking may play a more important role in the gas phase than in solution.[18]

Up to now, there is no systematic study reported to investigate the effect of base stacking on the duplexes stabilities in the gas phase and the solution phase.[7,11-14,24,25] In the series of duplexes with varying GC content, the gas phase stabilities track with the increasing GC%, and also parallel the stabilities in solution. This correlation may be

because the hydrogen bonding effect dominates in the gas phase (lack of distraction from the solvent) and is still a strong force in the solution phase. When the GC content is constant, the hydrogen bonding effect is similar, and the differences of stability would result from base stacking. To test this, we studied the gas phase stabilities of these three series of complementary duplexes including one series of 6-mers and two series of 8-mers (series 1 and 2) by CID, and the results were compared with those in solution (**Table 4.3** and **Table 4.2**).

Table 4.3 CID data for 6-mer duplexes

Duplex Name	Sequence	GC%	$T_{m,calc}$ (° C)	E_{50} (%)
6-mer-A	5'-AGAGAG-3' 3'-TCTCTC-5'	50	-0.6	12.83 ± 0.06
6-mer-B	5'-GTCTGT-3' 3'-CAGACA-5'	50	3.7	12.12 ± 0.05
6-mer-II	5'-GTGTCA-3' 3'-CACAGT-5'	50	5.3	12.75 ± 0.10
6-mer-C	5'-GTGTGT-3' 3'-CACACA-5'	50	6.3	12.26 ± 0.05
6-mer-D	5'-TGTGTG-3' 3'-ACACAC-5'	50	6.4	12.19 ± 0.06
6-mer-I	5'-GCGAAT-3' 3'-CGCTTA-5'	50	9.3	12.57 ± 0.10

The 6-mers were designed to have 50% GC content, and their calculated melting temperatures and E_{50}'s are listed in **Table 4.3**. From the results, the values of E_{50} do not increase with $T_{m,calc}$, and a scattered pattern showed up in the plot of E_{50} versus T_m (see **Figure 4.4**). The stability difference can be attributed to base stacking because the GC percentage is the same and hydrogen bonding effects can be cancelled out. Base stacking

contributes differently to the stabilities of the DNA duplexes in the gas phase than the solution phase due to the impact of media properties.
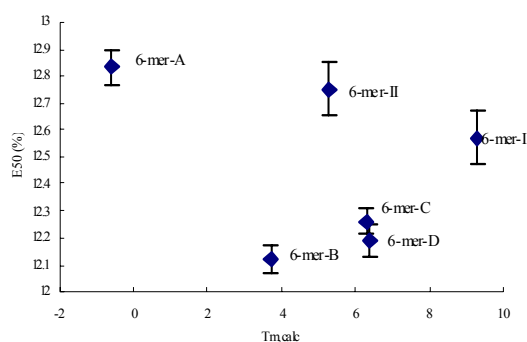
Figure 4.4 Plot of gas-phase stabilities (E_{50}) versus solution phase stabilities ($T_{m,calc}$) for six 6-mer duplexes.

To see whether the results are affected by length, we also tested two series of 8-mers to assess the effect of hydrogen bonding and base stacking. Results of the 8-mers are listed in **Table 4.2**. The series 1, 8-mers (A-F), are listed with increasing GC content, while the series 2 with a constant GC content of 50%, 8-mers (G, H, J, D, K and L), are listed with increasing T_m. From **Table 4.2**, E_{50} values increase with $T_{m,calc}$ for the first series of 8mers, and also increase with GC content (8-mer-A produced too weak duplex signal to obtain reproducible gas phase data). In the plot of E_{50} versus $T_{m,calc}$, a clear gas-solution phase correlation shows up, though it is not a perfect linear correlation (**Figure 4.5 (a)**). When the GC content is held constant (series 2), E_{50} values do not increase with $T_{m,calc}$ and the plotted data is scattered (**Figure 4.5 (b)**). Hence, we see a reproducible result as the 6-mers that the gas phase stability does not correlate to the solution stability

for the duplexes with the same GC content. This is probably because base stacking becomes more important when hydrogen bond effects are cancelled out.
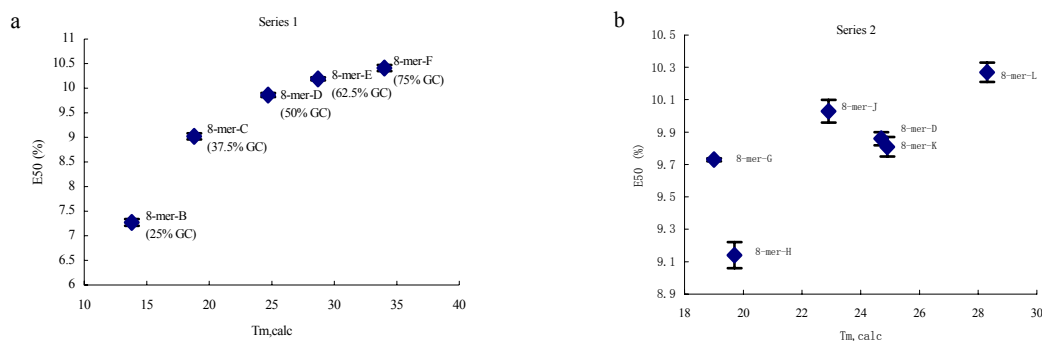
Figure 4.5 Plot of gas-phase stabilities (E_{50}) versus solution phase stabilities ($T_{m,calc}$) for 8-mers: (a) Series 1; (b) Series 2.

The above results show that base stacking also plays an important role in terms of stabilizing the duplexes in the gas phase. Our next step is to evaluate the effect of base stacking. In our previous study of 9-mer duplexes (5'-GGTTXTTGG-3'/3'-CCAAYAACC-5'), it was discovered that the stacking ability of the X base in the sequence GGT<u>TXT</u>TGG follows the order of T>A>G>C, whereas that of Y base in the sequence CCA<u>AYA</u>ACC follows the order of C>G>A>T in the gas phase.[18] Based on the next neighbor theory, this conclusion could be generalized to any sequence with a central TXT/AYA moiety.

To complement the study of base stacking, we chose another 9-mer duplex (5'-GTT<u>GXG</u>TTG-3'/3'-CAA<u>CYC</u>AAC-5') as a template with the central base varied (X = A, C, G, and T; Y = A, C, G, and T). Herein, we still use the 9-mer because the duplexes at this length show good CID spectra, and symmetric structure is not affected by varying the middle base. In addition, we want to keep terminal GC base pairs to keep the behavior

of the duplexes to a two-state model (all-or-none, the two strands of DNA either stay together as a double strand or dissociate to two single strands without intermediate state).[26] In this template, we also keep the same GC% as the previous 9-mer template (5'-GGT<u>TXT</u>TGG-3'/3'-CCA<u>AYA</u>ACC-5'). If the GC% is too high, T_m's will be too high and a fair amount of undesired covalent fragmentation instead of the desired noncovalent dissociation will take place when performing CID on the –4 charge duplex. The designed duplexes and their calculated T_m's and experimental E_{50}'s are listed in **Table 4.4**.

Table 4.4 $T_{m, calc}$ and E_{50} data for 9-mer duplexes

Duplex Name	Sequence	$T_{m,calc}$ (° C)	E_{50} (%)
AA	5'-GTTGAGTTG-3' 3'-CAACACAAC-5'	8.5	10.48 ± 0.11
AC	5'-GTTGAGTTG-3' 3'-CAACCCAAC-5'	-2.8	10.85 ± 0.17
AG	5'-GTTGAGTTG-3' 3'-CAACGCAAC-5'	12.5	10.88 ± 0.06
AT	5'-GTTGAGTTG-3' 3'-CAACTCAAC-5'	26.8	10.48 ± 0.10
CA	5'-GTTGCGTTG-3' 3'-CAACACAAC-5'	2.3	10.41 ± 0.13
CC	5'-GTTGCGTTG-3' 3'-CAACCCAAC-5'	0.0	11.01 ± 0.09
CG	5'-GTTGCGTTG-3' 3'-CAACGCAAC-5'	35.1	10.83 ± 0.11
CT	5'-GTTGCGTTG-3' 3'-CAACTCAAC-5'	3.7	10.50 ± 0.13
GA	5'-GTTGGGTTG-3' 3'-CAACACAAC-5'	12.0	10.75 ± 0.09
GC	5'-GTTGGGTTG-3' 3'-CAACCCAAC-5'	31.5	11.38 ± 0.10
GG	5'-GTTGGGTTG-3' 3'-CAACGCAAC-5'	19.6	11.30 ± 0.10
GT	5'-GTTGGGTTG-3' 3'-CAACTCAAC-5'	10.4	10.63 ± 0.09
TA	5'-GTTGTGTTG-3' 3'-CAACACAAC-5'	28.3	10.57 ± 0.07
TC	5'-GTTGTGTTG-3' 3'-CAACCCAAC-5'	-1.8	11.28 ± 0.08
TG	5'-GTTGTGTTG-3' 3'-CAACGCAAC-5'	18.0	10.93 ± 0.09
TT	5'-GTTGTGTTG-3' 3'-CAACTCAAC-5'	11.2	10.69 ± 0.06

As stated above, duplexes with the same composition have the same number of hydrogen bonds, thus any difference in stability in the gas phase could be attributed to base stacking. Among the duplexes studied in **Table 4.4**, there are 6 isomeric duplex pairs, GC versus CG, AT versus TA, AC versus CA, AG versus GA, CT versus TC, and GT versus TG. Their differences in E_{50}'s (ΔE_{50}) reflect the base stacking stability of the

duplexes.[6] The ΔE_{50}'s for these 6 pairs of isomers are listed in **Table 4.5**. The more stable duplex is listed first in each pair, for instance, the first entry "GC versus CG" means that GC is more stable than CG. The ΔE_{50} is calculated by subtracting the E_{50} of the less stable duplex from that of the more stable one in each isomeric pair. E_{50}'s of these duplexes range from 10.41% (CA) to 11.38% (GC), and the values cover a range of 0.97%. The largest ΔE_{50} is 0.78 between TC versus CT, which is a big portion of the total range. The rank of ΔE_{50} listed from the highest to the lowest is (TC versus CT) > (GC versus CG) > (AC versus CA) > (TG versus GT) > (AG versus GA) > (TA versus AT). The ΔE_{50}'s for AG versus GA and TG versus GT are 0.09% and 0.13%, which are comparable to the error associated with the measurement. These two pairs can be treated as to have the same gas phase stabilities.

Table 4.5 Gas phase ΔE_{50} values for selected isomeric XY duplexes

Isomeric Pair	ΔE_{50} (%)
TC vs. CT	0.78
GC vs. CG	0.55
AC vs. CA	0.44
TG vs. GT	0.30
AG vs. GA	0.11
TA vs. AT	0.09

The ΔE_{50}'s of **Table 4.5** were reorganized in **Table 4.6** to rank the order of the intrinsic stacking ability of each central base in each sequence. Each entry in **Table 4.6** is the ΔE_{50} of the corresponding XY duplex and its isomeric pair. For example, the first entry "0.78" reflects the E_{50} of the XY duplex (TC) minus that of the YX duplex (CT), with the positive sign indicating that the XY duplex is more stable than the YX duplex in the gas phase. Another example is AT/TA (second row, fourth column). The value "-

0.09" in this entry indicates that the E_{50} of the XY (AT) duplex is lower than that of the YX duplex (TA) and AT is less stable than TA in the gas phase. It is necessary to emphasize that the base stacking is sequence-dependent. Since the sequence studied here is 5'-GTTGXGTTG-3'/3'-CAACYCAAY-5', the stacking ability of X in GXG/AYA moiety can be examined. For the relative stacking abilities of variable base X in 5'-GTTGXGTTG-3', signs of each row in **Table 4.6** were evaluated. For the first row (X = T), which represents the stacking ability of T, there are two positive signs and one zero; for the second row, which represents the stacking ability of A, there is one positive sign and two zeros; the third row, which represents the stacking ability of G, has one positive sign, one zero and one negative sign; the last row represents the stacking ability of C, and the three signs are all negative. Therefore, the stacking ability of X in the sequence 5'-GTTGXGTTG-3' follows the order T > A > G > C. Similarly, signs of each column in **Table 4.6** were evaluated for the stacking ability of Y in the sequence 3'-CAACYCAAC-5'. The first column represents the stacking ability of C and there are 3 positive signs; the second column represents G, and has one positive, one negative and one zero; the third one represents A, and has two zeros and one negative sign; the last column represents T, and has one zero and two negative signs. Therefore, the overall relative stacking ability of Y in the sequence CAACYCAAC follows the order of C > G > A > T in the gas phase. Based on the next neighbor interactions, GXG/CYC motif should show the same base stacking trend in any sequence.

Table 4.6 Base stacking analysis for isomeric XY duplexes (5'-GTTGXGTTG-3'/3'-CAACYCAAC-5') in the gas phase[a]

	Y = C	Y = G	Y = A	Y = T	Rank
X = T	0.78 +	0.30 +	0.09 0	n/a	1
X = A	0.44 +	0.13 0	n/a	-0.09 0	2
X = G	0.55 +	n/a	-0.13 0	-0.30 -	3
X = C	n/a	-0.55 -	-0.44 -	-0.78 -	4
Rank	1	2	3	4	

[a] A positive sign "+" indicates that an XY duplex has a higher E_{50} than its isomer YX duplex and a negative sign "-" denotes the opposite trend. A "0" is used for the cases in which two E_{50} values are the same within experimental error. The values are listed for each entry.

Up to now, we have studied the base stacking effect in two 3-membered sequences GXG/CYC and TXT/AYA. It would be more versatile if we could generalize the base stacking effect in a "next neighbor" fashion, for example, what is the base stacking difference between G stacking on G and G stacking on C. Therefore, we designed several series of complementary duplexes with the same length, the same GC contents, the same sequence, but with different last two base pairs, for example, duplex 8mer14xy has GG/CC base pairs on the 3' end, while 8mer15xy has GC/CG base pairs on the 3' end. Here GG refers to GG/CC base pairs, and GC refers to GC/CG base pairs. From MS CID results, 8mer14xy is more stable than 8mer15xy in the gas phase by about 0.77%. Since the only difference between these two sequences is the last two base pairs, the stability difference comes from the contribution of the base stacking effect (GG vs. GC). For 8mer16xy with 8mer17xy, they have similar gas phase stabilities (8.13% vs. 8.16%), so CC stacking is as strong as CG stacking. Since GG (GG/CC) is the same stacking as CC (CC/GG), GG stacking is obviously as stable as CG stacking. The overall stabilities in

this series is CG ≥ GG (CC) >> GC. However, the order of predicted melting temperature is inconsistent with this result. The stability of 8mer14xy is lower than that of 8mer15xy, while 8mer16xy is lower than 8mer17xy. This is probably because a lot of other factors, such as hydrophobic effect and solvent effect, influence the stabilities of duplexes in solution besides base stacking and hydrogen bonding.

Table 4.7 8mer14-17xy: 5'-GGTTTTXX-3' (XX = GG, GC, CC and CG)

DNA oligos	Sequence	$T_{m, calc}$ (° C)	E_{50} (%)
8mer-14xy	5'-GGTTTTGG-3' 3'-CCAAAACC-5'	22.79	8.70
8mer-15xy	5'-GGTTTTGC-3' 3'-CCAAAACG-5'	25.19	7.93
8mer-16xy	5'-GGTTTTCC-3' 3'-CCAAAAGG-5'	21.99	8.13
8mer-17xy	5'-GGTTTTCG-3' 3'-CCAAAAGC-5'	24.25	8.16

To further test whether the ranking is generally true for all the DNA duplexes, we studied another 4 series of duplexes with the same base pairs on the 3' end but different length and sequences (see **Table 4.8** to **Table 4.11**). From **Table 4.8** and **Table 4.9**, similar results were obtained as **Table 4.7**: CG ≥ or > GG (CC) >> GC.

Table 4.8 9mer20-23xy:5'-GTTGTATXX-3' (XX = GG, GC, CC and CG)

DNA oligos	Sequence	$T_{m, calc}$ (° C)	E_{50}
9mer-20xy	5'-GTTGTATGG-3' 3'-CAACATACC-5'	25.08	11.55 ± 0.08
9mer-21xy	5'-GTTGTATGC-3' 3'-CAACATACG-5'	27.17	10.95 ± 0.08
9mer-22xy	5'-GTTGTATCC-3' 3'-CAACATAGG-5'	24.37	10.78 ± 0.10
9mer-23xy	5'-GTTGTATCG-3' 3'-CAACATAGC-5'	26.31	10.84 ± 0.07

CG (0.66) ≥ GG (CC) (0.60)> GC (0.00);

Table 4.9 9mer24-27xy: 5'-GTTGTTTXX-3' (XX = GG, GC, CC and CG)

DNA oligos	Sequence	$T_{m, calc}$ (° C)	E_{50}
9mer-24xy	5'-GTTGTTTGG-3' 3'-CAACAAACC-5'	27.82	10.87 ± 0.07
9mer-25xy	5'-GTTGTTTGC-3' 3'-CAACAAACG-5'	29.83	10.16 ± 0.09
9mer-26xy	5'-GTTGTTTCC-3' 3'-CAACAAAGG-5'	27.13	10.00 ± 0.14
9mer-27xy	5'-GTTGTTTCG-3' 3'-CAACAAAGC-5'	28.95	10.11 ± 0.13

CG (0.82) > GG (CC) (0.71) > GC (0.00);

However, this stability order does not occur for the duplexes in **Table 4.10** and **Table 4.11**. The duplexes 9mer14xy and 9mer15xy were initially designed as 5'-GGTTGTTCC-3'/3'-CCAACAAGG-5' and 5'-GGTTGTTCG-3'/3'-CCAACAAGC-5', respectively. But when CID was performed on 9mer14xy, the daughter ions (double charges) were not differentiable from the parent ions (four charges) since ssX has similar mass with ssY (the single strands and duplex show up as one peak in spectra, making it complicated to calculate the DS% at different collision energies). To solve the problem, 9mer14xy and 9mer15xy were redesigned and are shown together with 9mer1xy (GG) and 9mer11xy (GC) in **Table 4.10**. Surprisingly, this series showed an order of base stacking stability different from **Table 4.7**: GG (CC) > GC > CG. Apparently, 9mer14xy has a higher E_{50} than 9mer15xy, which indicates that CC stacking is more stable than CG. This is opposite from the earlier result that CG is slightly higher than or comparable to CC stacking. In the study of another series of duplexes, a similar trend was obtained as listed in **Table 4.11**: GG (CC) > GC $\cong$ CG. Why we get these inconsistent results is still a puzzle, and it may be due to the charge on phosphate and other factors that affect the stability, for instance, the location of the charges on the phosphate might be slightly

different in different sequences thus this charge repulsion results in the stability difference.

Table 4.10 9mer1xy and 11xy: 5'-GGTTGTTGX-3' (X = G or C)
9mer14xy and 15xy: 5'-CCTTGTTCY-3' (Y = C or G)

DNA oligos	Sequence	$T_{m, calc}$ (° C)	E_{50}
9mer-1xy	5'-GGTTGTTGG-3' 3'-CCAACAACC-5'	31.51	12.15 ± 0.13
9mer-11xy	5'-GGTTGTTGC-3' 3'-CCAACAACG-5'	33.47	11.68 ± 0.12
9mer-14xy	5'-CCTTGTTCC-3' 3'-GGAACAAGG-5'	30.00	11.63 ± 0.11
9mer-15xy	5'-CCTTGTTCG-3' 3'-GGAACAAGC-5'	31.76	11.28 ± 0.11

GG (CC) (0.47) > GC (0.12) > CG (0.00)

Table 4.11 8mer14xy and 15xy: 5'-GGTTTTGX-3' (X = G or C)
8mer30xy and 31xy: 5' -CCTTTTCY-3' (Y = C or G)

DNA oligos	Sequence	$T_{m, calc}$ (° C)	E_{50}
8mer-14xy	5'-GGTTTTGG-3' 3'-CCAAAACC-5'	22.79	9.06 ± 0.08
8mer-15xy	5'-GGTTTTGC-3' 3'-CCAAAACG-5'	25.19	8.49 ± 0.05
8mer-30xy	5'-CCTTTTCC-3' 3'-GGAAAAGG-5'	21.00	9.12 ± 0.16
8mer-31xy	5'-CCTTTTCG-3' 3'-GGAAAAGC-5'	23.31	8.53 ± 0.18

GG (CC) (0.59) > GC (0.02) ≅ CG (0.00)

4.4 Conclusions

In this chapter, several series of duplexes were examined to study the effects of their hydrogen bonding and base stacking abilities on DNA duplexes in the gas phase versus solution phase.

It was found that ion abundances (DS%) do not always track with solution phase stability (T_m). Two duplexes with the same solution phase stability (T_m) may yield different ion abundance (DS%) if they have different GC content. The higher the GC

content, the higher the DS%. It indicates that the duplex with a higher GC content can survive better during ESI.

Two series of duplexes (8merA-L, 6merA-D and I, II) were used to study the effect of GC content on gas-phase stabilities. When the GC content varies within the series, their gas-phase stabilities tend to track with their solution-phase stabilities. When the GC content are the same, base stacking becomes a more important factor for the stabilities in the gas phase, which may cause different duplex stabilities in the gas phase and in the solution phase.

In a further study of the base stacking ability in the gas phase, several series of 9mers and 8mers were investigated. The base stacking abilities were evaluated in GXG/CYC context of 16 XY duplexes, and they were also studied with a nearest neighbor setup in several series of 8-mer and 9-mer duplexes which only vary the sequence at the end. The results are not consistent. Developing a robust method to quantitatively study the base stacking and compare it with hydrogen bonding by mass spec would be our future plan.

4.5 Reference:

(1)   Sponer, J.; Laszczynski, J.; Hobza, P. *J. Biomol. Struct. Dyn.* 1996, *14*, 117.

(2)   Hanlon, S. *Biochem. Biophys. Res. Commun.* 1966, *23*, 861-867.

(3)   Sarai, A.; Mazur, J.; Nussinov, R.; Jernigan, R. L. *Biochem.* 1988, *27*, 8498-8502.

(4)   Schneider, H. J. *Angew. Chem., Int. Ed. Engl.* 1991, *30*, 1417.

(5)   Lehn, J. M. *Supramolecular Chemistry*; VCH: Weinheim, 1995.

(6)   Gabelica, V.; De Pauw, E. *J. Mass Spectrom.* 2001, *36*, 397-402.

(7)   Gabelica, V.; De Pauw, E. *Int. J. Mass Spectrom.* 2002, *219*, 151-159.

(8)   Gabelica, V.; De Pauw, E. *J. Am. Soc. Mass Spectrom.* 2002, *13*, 91-98.

(9)   Wan, K. X.; Gross, M. L.; Shibue, T. *J. Am. Soc. Mass Spectrom.* 2000, *11*, 450-457.

(10)   Gaffney, B. L.; Jones, R. A. *Biochemistry* 1989, *28*, 5881-5889.

(11)   Le Novere, N. *Bioinformatics* 2001, *17*.

(12)   SantaLucia, J., Jr.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* 1996, *35*, 3555-3562.

(13)   SantaLucia, J., Jr; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* 2004, *33*, 415-440.

(14)   SantaLucia, J., Jr *Proc. Natl. Acad. Sci. U.S.A.* 1998, *95*, 1460-1465.

(15)   Lopez, L. L.; Tiller, P. R.; Senko, M. W.; Schwartz, J. C. *Rapid Commun. Mass Spectrom.* 1999, *13*, 663-668.

(16)   Deschenes, L. A.; Vanden Bout, D. A. *J. Am. Chem. Soc* 2000, *122*, 9567-9568.

(17)   Pan, S.; Lee, J. K. In *Proceedings of the 50th ASMS Conference on Mass Spectrometry and Allied Topics* Orlando, FL, 2002.

(18)   Pan, S.; Sun, X.; Lee, J. K. *Int. J. Mass Spectrom.* 2006, *253*, 238-248.

(19)   Pan, S.; Sun, X.; Lee, J. K. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 1383-1395.

(20)   Michelson, A. M. *Biochim. Biophys. Acta* 1962, *55*, 841-848.

(21)   Ganem, B.; Li, Y. T.; Henion, J. D. *Tetrahedron Lett.* 1993, *34*, 1445-1448.

(22)   Ding, J.; Anderegg, R. J. *J. Am. Soc. Mass Spectrom.* 1995, *6*, 159-164.

(23)   Greig, M. J.; Gaus, H. J.; Griffey, R. H. *Rapid Commun. Mass Spectrom.* 1996, *10*, 47-50.

(24)   Breslauer, K. J.; Frank, R.; Blocker, H.; Marky, L. A. *Proc. Natl. Acad. Sci. U.S.A.* 1986, *83*, 3746-3750.

(25)   Sugimoto, N.; Nakano, S.; Yoneyama, M.; Honda, K. *Nucleic Acids Res.* 1996, *24*, 4501-4505.

(26)   Poland, D.; Scheraga, H. A. *Theory of Helix-Coil Transitions in Biopolymers*; Academic Press: New York, 1970.

Chapter 5 The Stability of DNA Duplexes Containing Hypoxanthine (Inosine): Gas versus Solution Phase and Biological Implications

5.1 Introduction

Hypoxanthine (**1a**) is a nucleobase that occurs naturally in tRNA and is a key intermediate in the *de novo* biosynthesis of purine nucleotides; it is also a mutation that occurs in DNA when adenine is deaminated.[1-6]
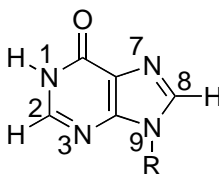
$$1 \quad \begin{array}{l} \textbf{a} \ \ R = H \quad \text{(hypoxanthine)} \\ \textbf{b} \ \ R = \text{(deoxy)ribosemonophosphate} \ \ \text{(inosine)} \end{array}$$

Hypoxanthine (also called "inosine" in its nucleotide form (**1b**)) is often referred to as a "universal base". This ability was first recognized after the discovery that inosine was often present in the first anticodon position in various tRNA sequences.[2] That first anticodon position pairs to the third codon position on mRNA, which is proposed to have some "play" that allows for non-Watson-Crick base pairs (such as hypoxanthine•adenine); this is the well-known Crick "wobble hypothesis".[2] As a potentially useful universal base, hypoxanthine can form base-pair structures with all the normal nucleobases (**Figure 5.1**).[7-15]
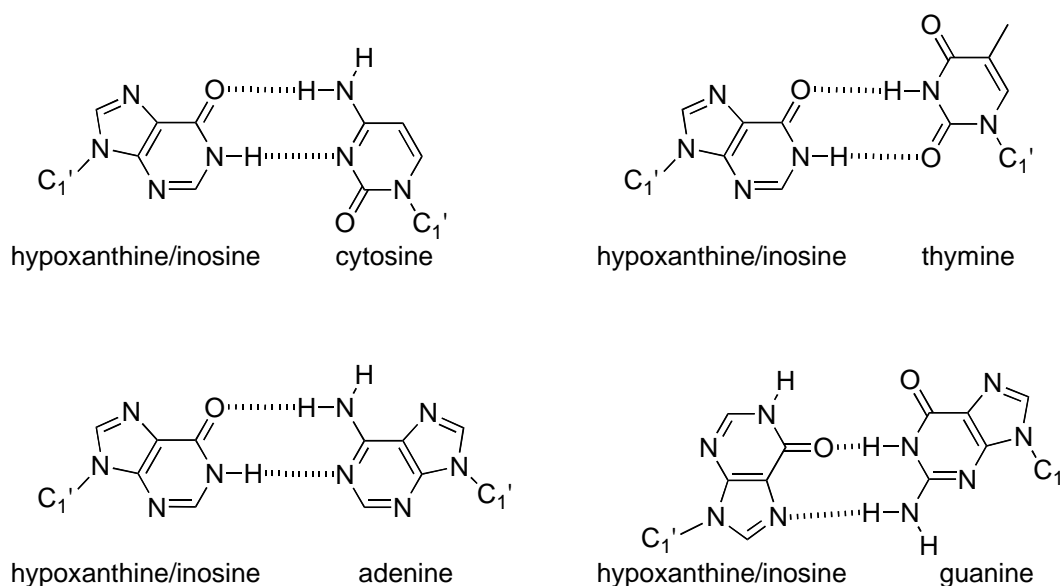
Figure 5.1 Possible structures of hypoxanthine•normal base pairs.

Universal bases have a myriad of potential uses in molecular biology.[16-18] In many applications, a needed oligonucleotide sequence target may not be exactly known and the universal base can act as a "wild card" that will bind to any nucleobase indiscriminately.[9,17,19-22] For example, hypoxanthine residues have been placed at ambiguous points in oligonucleotide probes that screen genomic DNA libraries.[22-26] Universal bases are of particular utility when probes and primers are designed based on the amino acid sequence of a protein, which can be complicated by codon degeneracy or fragmentary peptide sequence data.[17,19,22-24,27-31]

Although hypoxanthine has been used in primers and probes, the fundamental examination of the stability of base pairs containing it has been limited; the major experimental study was conducted by Watkins and SantaLucia in 2005.[16,18,19,26,32,33] They examined the solution phase stability (melting temperature) of 84 dimers containing hypoxanthine; combining their data with that of 13 additional oligonucleotide dimers that were examined previously in the literature, they were able to characterize the hypoxanthine nearest-neighbor parameters, which allows for accurate prediction of the

stability of oligonucleotides containing hypoxanthine. Overall, in solution, melting temperatures of duplexes containing hypoxanthine vary widely, and are on average lower than those for complementary duplexes containing only adenine, guanine, cytosine and thymine.[16,18,19,26,32,33]   Also, hypoxanthine•cytosine is more stable as a base pair than other combinations; hypoxanthine is therefore not so universal that it binds equally to each natural nucleobase in solution.[16,17,19]

In the gas phase, numerous computational studies and limited experimental studies have explored the tautomeric, acidic and basic properties of hypoxanthine, as well as the various hydrogen-bonded dimers.[34-46]   As far as we know, however, no gas phase experimental study of the effect of hypoxanthine on duplex stability has been carried out.

In our previous studies, we have found that gas phase studies can be useful for elucidating the intrinsic reactivity of biological molecules.[46-55] The interior of ribosomes, polymerases, the cell plasma, and other biological media in which universal bases may have a role vary in polarity and are not fully aqueous; gas phase studies yield intrinsic data that can be extrapolated to other media.[46-55]

In this work, we examine the effect of hypoxanthine on the gas phase stability of a series of DNA duplexes using mass spectrometry, and compare those values to solution, to evaluate the effect of hypoxanthine on duplex stability in non-aqueous environments. We discuss our results in the context of hypoxanthine as a possible universal base, and also in its role as a damaged, mutagenic base.

5.2 Experimental Methods

5.2.1 Sample preparation

Oligodeoxynucleotide single strands were purchased from Sigma Aldrich. These single strands were pre-desalted and used without further purification. Stock solutions of

62.5 μM duplex were prepared in 40 mM NH$_4$OAc aqueous solution at pH 7.0. Then the stock solutions were annealed at 90 °C for 10 minutes and cooled down slowly to 0 °C. Before injecting into the ESI-MS, the stock solutions were diluted to 12.5 μM in 40 mM NH$_4$OAc mixed with 20% methanol.

5.2.2 Melting calculations

The T$_m$ values of the DNA duplexes are predicted by the program "MELTING".[56] The settings are as follows: (1) hybridization type: dnadna (for a DNA duplex); (2) nearest-neighbor parameters set: all97a.nn; (3) salt concentration: 0.04 M; (4) nucleic acid concentration (total): 25.0 μM; (5) nucleic acid correction factor: 4 (for non-self-complementary duplex); (6) salt correction: san98a; (7) nearest neighbor parameters for inosine mismatches: san05a. The error for estimating T$_m$ values by the MELTING program is ±1.6 °C.[16,19,56-59]

5.2.3 ESI-quadrupole ion trap mass spectrometer and "E$_{50}$" experiments

Negative ion ESI-MS spectra were obtained with the Finnigan LCQ mass spectrometer (San Jose, CA). The 0ºC solution was infused at 25 μL/min directly into the mass spectrometer. The spray voltage was -4.0 kV while the capillary temperature was 175 °C. Collision induced dissociation (CID) was performed in the mass analyzer by varying the relative collision energy with a default activation time of 30 ms and a *q* value of 0.25. The applied collision energy is a normalized collision energy (in %) that corrects for the *m/z* dependence of the activation voltage required for ions of different *m/z* ratios.[60] The relative mass difference between the base pair with the highest *m/z* ratio and that with the lowest *m/z* ratio is less than 1%. Given the similar sequences and *m/z* ratio, we assume that the energies deposited into the ions are the same when the same normalized

collision energies are applied. The gas phase stability of the duplexes is measured in a relative way by subjecting the duplex parent ions to increasing collision energies during the CID event in an ion trap. E_{50} is defined as the collision energy at which 50% of the duplexes are dissociated into single strands, and is used to characterize the gas phase stability.[52,61-65] A higher E_{50} corresponds to a more stable duplex in the gas phase. Although CID is a kinetic experiment, the dissociation is assumed to be endothermic enough such that the barrier and the endothermicity are similar.[52,62,65-69] The internal energy distribution of the parent ion is poorly defined due to the multiple collision events in the ion trap. We therefore do not intend to report absolute duplex dissociation energies, but rather relative gas phase stabilities as reflected by the E_{50}'s, a method established previously by our group and others'.[52,60-65,70-72]

Experimental conditions were tuned by optimizing the –4 charged duplex ions of the TA duplex (5'-d(GGTT**T**TTGG)-3'/3'-(CCAA**A**AACC)-5') (*m/z* = 1358); the conditions thus obtained were applied to all the duplexes. Duplex abundance is normalized by using the equation: % Duplex (DS%) = (2 × [all duplexes]) / ([all single strands] + 2 × [all duplexes]), where the values in brackets are absolute ion abundances.[52,61-65] The reported duplex abundance is an average of six full-scan measurements; the average standard deviation is 1.69%. Duplex dissociation profiles were fitted with sigmoid equations, and the corresponding E_{50} values were derived using Origin 6.0 software.[73] Each CID experiment was performed under a parent ion isolation width of w = 5; in previous studies we show that changing the isolation width does not change relative E_{50} values.[52,65] The reported E_{50} value for each XY duplex is an average of six measurements. The average standard deviations for the measurement of E_{50} for all the XY duplexes are 0.06%.

5.3 Results and Discussion

We examined the 9-mer 5'-d(GGTTXTTGG)-3'/3'-d(CCAAYAACC)-5', where the central base X or Y = adenine (A), guanine (G), thymine (T), cytosine (C) and hypoxanthine (H). This particular sequence, where X or Y = all possible combination of normal bases, has been well-characterized both in the gas phase (in our lab) and in the solution phase, and therefore is an ideal choice for these hypoxanthine-substitution studies.[52,65,74,75] We have found that this 9-mer is long enough to form helical structures but also manifests measurable changes in stability when just the central base pair is changed.[52,65,74,75] Furthermore, these duplexes, because of the terminal GC base pairs that help maintain helical structure during dissociation, are particularly well-suited to the traditional two-state dissociation model that allows for accurate theoretical prediction of melting temperatures.[57]

To simplify the nomenclature of each duplex in this thesis, we use only the variable central base of each strand to represent the whole duplex. For example, a duplex called "GH" refers to the duplex 5'-d(GGTTGTTGG)-3'/3'-d(CCAAHAACC)-5', where X = guanine (G), Y = hypoxanthine (H). The various combinations where X or Y is an H can form nine duplexes (AH, CH, GH, TH, HH, HA, HC, HG and HT); we studied all of these plus the four complementary duplexes GC, CG, AT and TA.

5.3.1. Solution phase stability

The traditional method for assessing the solution phase stability of a DNA duplex is via melting temperature (T_m). When the temperature of a DNA duplex solution is slowly increased, the ordered double helical structures dissociate into single strands. The

midpoint of this transition is the T_m. The higher the T_m, the more stable the duplex. The calculated solution T_m's for our XY duplexes are shown in **Table 5.1**.[16,19,56] The normal duplexes (GC, CG, AT and TA) are significantly more stable than those containing hypoxanthine, with the exception of CH and HC, which are comparable in stability to AT. As expected, the duplexes containing G and C are more stable than those containing A and T (the former has three hydrogen bonds while the latter has only two). Among the hypoxanthine duplexes, CH is the most stable, and HG is the least stable.

Table 5.1. Calculated Tm values for XY duplexes.

XY	T_m (°C)
GC	33.48
CG	31.98
TA	29.32
CH	26.97
AT	26.54
HC	26.40
AH	20.60
HT	18.87
HA	15.88
GH	13.21
HH	10.31
TH	8.68
HG	8.66

In earlier work, we showed that when X and Y are normal nucleobases, the duplex ion abundance resulting from electrospray (which we term DS%) reflects the solution phase stability.[52,65] That is, if the electrospray process should volatilize the duplexes with

relative integrity, then the resultant mass spectrum should be a "snapshot" of the solution phase composition, and the relative ion abundances among a series of duplexes with differing stability should reflect their relative solution phase stabilities.[66,76-82] In order to assess whether this method of using ion abundances to measure relative solution phase stabilities can be generalizable to DNA duplexes containing a damaged base such as hypoxanthine, we plotted the duplex ion abundances versus the solution phase T_m values (**Figure 5.2**). A reasonable linear relationship is seen; thus, for this series of duplexes, monitoring the duplex ion signal is a fast and efficient method for assessing relative solution phase stabilities. This is the first study establishing a correlation between solution phase stability and mass spectrometric signal abundance for mismatched duplexes containing a damaged base, and is extremely valuable. The ability to quickly assess, by mass spectrometry, the solution phase noncovalent complex abundance has implications for the development of efficient screening assays for potential DNA binders.
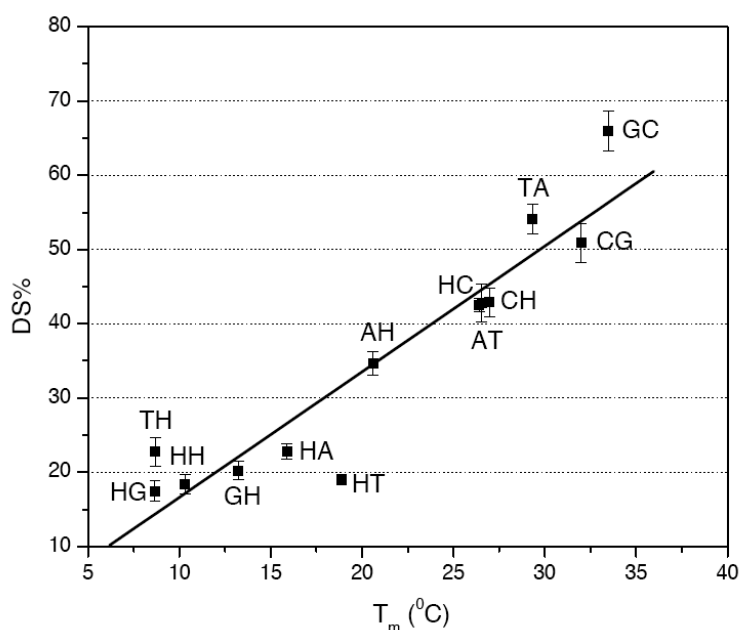

Figure 5.2. Plot of DS% versus T_m for complementary duplexes and duplexes containing hypoxanthine (5'-d(GGTT**X**TTGG)-3'/3'-d(CCAA**Y**AACC)-5').

5.3.2. Gas phase stability

Whereas ion abundance yields the solution phase stability, collision-induced dissociation (CID) to dissociate a duplex into its single strand components yields gas phase stability.[52,61-65,70,71] Electrospray of these 9-mer sequences yields duplex ions with both –3 and –4 charges. Under gentle CID conditions, we find that dissociation of the –4 duplex into its constitutive single strands (–2 charge on each) is the major fragmentation pathway (**Figure 5.3**). The dissociation of the –3 charged duplexes preferentially yields extensive cleavage of covalent bonds, with little noncovalent dissociation. The effect of charge state on duplex stability has been discussed previously, and stability comparison among different ions is valid only if all the duplexes have the same charge states.[60,70,71,83-86] Therefore, we report the CID experiments on the –4 charged duplexes only.
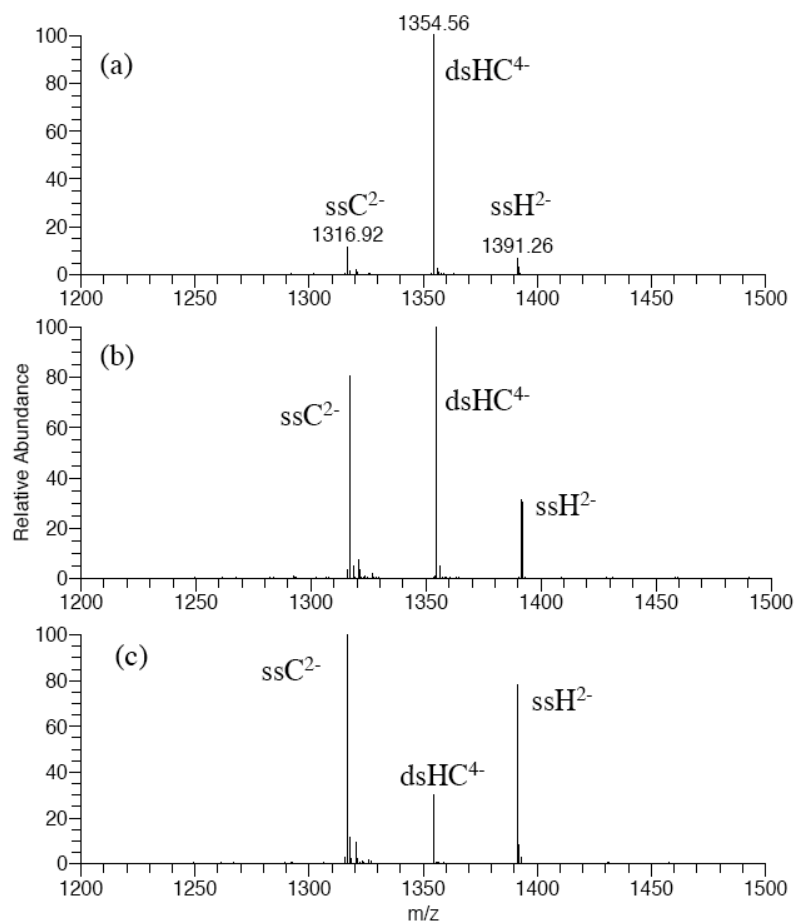
Figure 5.3. CID spectra of the duplex HC^{4-} ions at relative collision energies of (a) 10.4%. (b) 11.2% and (c) 12.0%; "ds" indicates double strand; "ss" indicates single strand.

The dissociation of the parent duplex ion is monitored by the disappearance of the duplex signal and the appearance of the single strands. The collision energy at which 50% of the duplexes are dissociated into single strands ("E$_{50}$"; more details in Experimental section) is used to characterize the gas phase stability.[52,61-65,70,71] The dissociation profiles of four duplexes (GC, AT, HC and HT) are displayed in Figure 4. To achieve the same degree of dissociation among all these duplexes (as indicated by the E$_{50}$ value), the GC duplex (squares) requires the highest collision energy (as indicated by the largest E$_{50}$ value), followed by HC then HT (circles and upright triangles), with AT (upside down triangles) having the lowest E$_{50}$ and therefore being the least stable. This

would indicate that the gas phase stability order of these four duplexes is GC > HC > HT > AT.
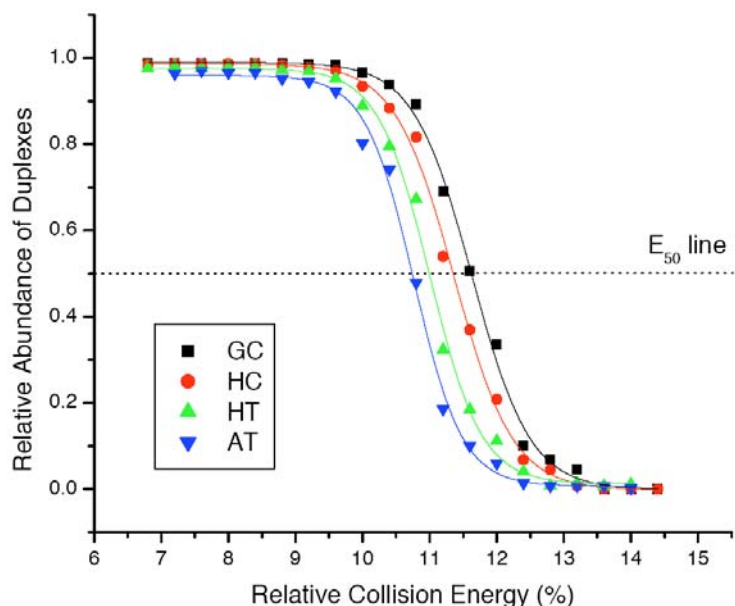
Figure 5.4. Gas phase dissociation profiles of four XY duplexes: GC, HC, HT and AT.

Table 5.2 lists all the E_{50} values for the 9 XY duplexes containing H, and the four complementary duplexes, together with the predicted melting temperatures, to allow for direct comparison between gas phase and solution phase stabilities. The XY duplexes are listed in decreasing order of their E_{50}'s. The E_{50} difference between the most stable GC duplex and the least stable AH duplex is only 0.98%. However, these differences are significant because the average standard deviation is only 0.06%. Immediately from the table, one can see that the solution and gas phase stabilities do not track. Many of the hypoxanthine-containing duplexes (HC, HG, GH, HT, TH, HH, CH) are more stable than the normal duplexes AT and TA. A plot of the E_{50} values versus melting temperature

reveals that there seems to be no straightforward correlation between the solution and gas phase data (**Figure 5.5**).

Table 5.2. T_m and E_{50} values for XY duplexes.

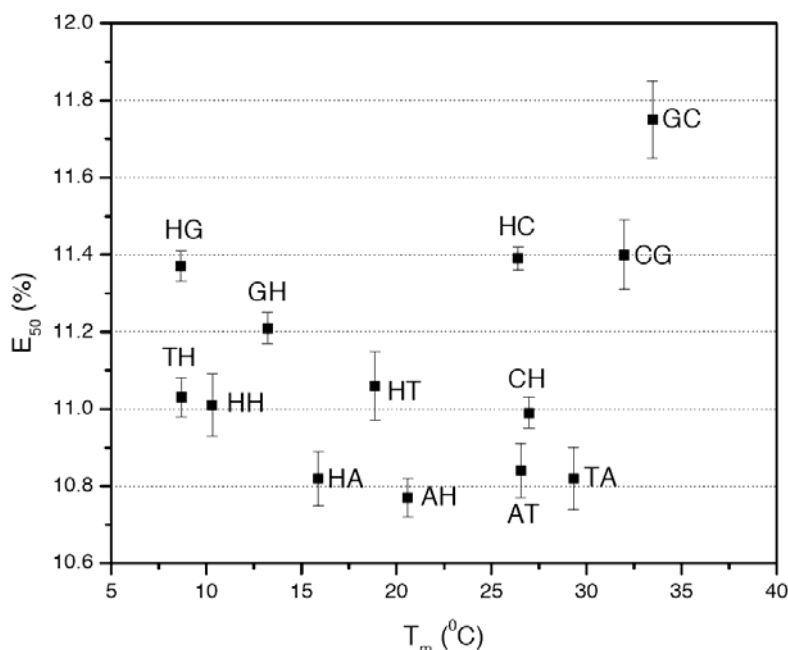| XY	T_m (°C)	E_{50} (%)
GC	33.48	11.75 ± 0.10
CG	31.98	11.40 ± 0.09
HC	26.40	11.39 ± 0.03
HG	8.66	11.37 ± 0.04
GH	13.21	11.21 ± 0.04
HT	18.87	11.06 ± 0.09
TH	8.68	11.03 ± 0.05
HH	10.31	11.01 ± 0.08
CH	26.97	10.99 ± 0.04
AT	26.54	10.84 ± 0.07
TA	29.32	10.82 ± 0.08
HA	15.88	10.82 ± 0.07
AH	20.60	10.77 ± 0.05

Figure 5.5 Comparison of gas phase stability (E_{50}) and solution phase stability (T_m) of the 13 XY duplexes. Error bars indicate the standard deviation for each E_{50} value.

5.3.3 Biological implications

Hypoxanthine as a universal base. The most striking aspect of our data is the disparity between the solution and gas phase stabilities. Although the two do not track, there are some clear trends (Table 3). In **Table 5.3**, we arrange the data so that the duplexes that do and do not contain H can be more easily compared. In solution, with the exception of CH and HC, all duplexes containing hypoxanthine are less stable (have lower melting temperatures) than the normal GC (T_m = 33.48 ˚C), CG (31.98 ˚C), AT (26.54 ˚C) and TA (29.32 ˚C) duplexes. This is consistent with earlier studies pointing to a wide range in stability in hypoxanthine duplexes, with hypoxanthine clearly favoring cytosine for base pairing.[16,18,19,26,32,33] In the gas phase, however, the hypoxanthine-containing duplexes are *not* less stable than the normal duplexes. In fact, the E_{50} values for those duplexes containing hypoxanthine fall, for the most part, between the E_{50} values of TA

(E_{50} = 10.82%) and GC (11.75%). (AH is the only duplex that may be less stable than TA, with an E_{50} of 10.77 ± 0.05). Therefore, in the nonpolar environment of the gas phase, the hypoxanthine-containing duplexes are not unusually unstable but rather fall into a range between the complementary GC and TA duplexes. This difference in hypoxanthine's effect on duplex stability (relative to normal complementary duplexes) in a nonpolar versus a polar environment could be useful in applications utilizing hypoxanthine as a universal base. The differences we see could explain why although hypoxanthine (inosine) is a key component in the anticodon-codon wobble pairing of tRNA and mRNA (which occurs in the ribosome), it has had mixed success as a universal base in *in vitro* applications such as PCR primers and hybridization probes (where its wide-ranging effects on duplex stability make it not-so-universal).[2,16,17,19] In applications that might occur in media that are not aqueous solution (for example, in inhibitory antisense strands designed to bind to mRNA *in vivo*), hypoxanthine may be a better universal base.[1,30,31] Of course, our fundamental studies are at the extreme (the gas phase), but these results point to interesting differences in the intrinsic stability of duplexes versus those that are in solution.[17,30,31]

Table 5.3 T_m and E_{50} data for XY duplexes (5'-d(GGTTXTTGG)-3'/3'-d(CCAAYAACC)-5').

XY	T_m (°C)		E_{50} (%)	
GC or CG	33.48 (GC)	31.98 (CG)	11.75 ± 0.10 (GC)	11.40 ± 0.09 (CG)
AT or TA	26.54 (AT)	29.32 (TA)	10.84 ± 0.07 (AT)	10.82 ± 0.08 (TA)
HC or CH	26.40 (HC)	26.97 (CH)	11.39 ± 0.03 (HC)	10.99 ± 0.04 (CH)
HG or GH	8.66 (HG)	13.21 (GH)	11.37 ± 0.04 (HG)	11.21 ± 0.04 (GH)
HT or TH	18.87 (HT)	8.68 (TH)	11.06 ± 0.09 (HT)	11.03 ± 0.05 (TH)
HA or AH	15.88 (HA)	20.60 (AH)	10.82 ± 0.07 (HA)	10.77 ± 0.05 (AH)
HH	10.31 (HH)		11.01 ± 0.08 (HH)	

Hypoxanthine as a damaged base. Hypoxanthine can result from the deamination of adenine, and as such is a damaged base in DNA, and a mutagenic agent.[1-6] The mutagenicity of hypoxanthine lies in its ability to cause an A•T to G•C transition.[4-6,87,88] If the adenine in an A•T base pair is deaminated, the base pair becomes an H•T base pair. In replication, H prefers C, so once the duplex unwinds and replicates, an H•C base pair is formed. When that duplex replicates, the C will base pair with a G; this is the A•T to G•C transition. Since the specific sequence of the human genome is responsible for coding proteins, signaling, and a myriad of other important functions, the hypoxanthine mutation can be deleterious.

The human genome is protected by an enzyme, alkyladenine DNA glycosylase (AAG), which excises hypoxanthine from DNA.[6,88-102] AAG will excise hypoxanthine from both H•T and H•C base pairs, but is more efficient at excising hypoxanthine when it is base paired to T.[103,104] Biologically, this makes sense -- when hypoxanthine is first formed from adenine, the target is a hypoxanthine hydrogen bonded to thymine, and it is better to excise hypoxanthine before replication occurs. Mechanistically, the hypoxanthine "flips" out of the duplex and into the active site of AAG before excision. One might expect that hypoxanthine would be easier to "flip" out if its base pair were *weak*. In solution, our results indicate that the HT and TH base pairs are *less* stable than HC and CH base pairs, which is consistent with the enzymatic preference for hypoxanthine hydrogen bonded to thymine.[103,104] In the gas phase, HT is less stable than HC, but TH and CH are comparable in stability. This difference from the solution phase results is intriguing and is almost certainly based on differences in base stacking rules in solution versus the gas phase, which we have started to explore.[52] This may in fact have relevance to AAG. It is found that excision of hypoxanthine is much lower when hypoxanthine is flanked by a 5'G and a 3'C versus a 5'T and a 3'A.[104] In our studies, hypoxanthine was either flanked by 5' and 3' adenines, or 5' and 3' thymines. In the future, we would like to ascertain how gas phase stabilities of duplexes containing

hypoxanthine with different nearest neighbors compare to solution phase stabilities. Such data will be helpful both in the context of AAG and also for further assessing the utility of hypoxanthine as a universal base.

5.4 Conclusions

Comparison of the gas phase and solution phase stabilities of the 13 XY duplexes (5'-d(GGTT**X**TTGG)-3'/3'-d(CCAA**Y**AACC)-5') where X, Y = A, G, T, C and H indicates that although hypoxanthine has a fairly consistent destabilizing effect in aqueous solution, in the gas phase, those DNA duplexes with hypoxanthine are for the most part as stable or more stable than the normal AT and TA duplexes (and less stable than the GC and CG duplexes). The relatively higher stability of the hypoxanthine-containing duplexes in vacuo could mean that hypoxanthine, which has limitations as a universal base *in vitro*, might prove useful in *in vivo* applications where the environment may not be aqueous.[17] Our results are also consistent with hypoxanthine's role as a damaged base; when hypoxanthine arises from deamination of adenine, it is excised by the enzyme AAG. AAG cleaves hypoxanthine from hypoxanthine•thymine base pairs more readily than from hypoxanthine•cytosine base pairs; our results are consistent with this observation in that the HC base pair is the most stable (and therefore it is less facile for the hypoxanthine to flip "out" of the duplex and "into" the AAG active site). Future studies will probe nearest neighbor effects on the stability of duplexes containing hypoxanthine in the gas phase versus in solution.

5.5 Reference and Footnote

(1)   Voet, D.; Voet, J. G. *Biochemistry*; Second ed.; John Wiley & Sons, Inc.: New York, 1995.

(2)   Crick, F. H. C. *Journal of Molecular Biology* **1966**, *19*, 548-555.

(3)   Munns, A. R. I.; Tollin, P. *Acta Crystallographica, Section B: Structural Crystallography and Crystal Chemistry* **1970**, *B26*, 1101-1113.

(4)   Miao, F.; Bouziane, M.; O'Connor, T. R. *Nucleic Acids Research* **1998**, *26*, 4034-4041.

(5)   Lindahl, T. *Nature* **1993**, *362*, 709-715.

(6)   Karran, P.; Lindahl, T. *Biochemistry* **1980**, *19*, 6005-6011.

(7)   Oda, Y.; Uesugi, S.; Ikehara, M.; Kawase, Y.; Ohtsuka, E. *Nucleic Acids Research* **1991**, *19*, 5263-5267.

(8)   Leonard, G. A.; Booth, E. D.; Hunter, W. N.; Brown, T. *Nucleic Acids Research* **1992**, *20*, 4753-4759.

(9)   Corfield, P. W. R.; Hunter, W. N.; Brown, T.; Robinson, P.; Kennard, O. *Nucleic Acids Research* **1987**, *15*, 7935-7949.

(10)   Carbonnaux, C.; V., F. G.; Sowers, L. C. *Nucleic Acids Research* **1990**, *18*, 4075-4081.

(11)   Cruse, W. B. T.; Aymani, J.; Kennard, O.; Brown, T.; Jack, A. G. C.; Leonard, G. A. *Nucleic Acids Research* **1989**, *17*, 55-72.

(12)   Kumar, V. D.; Harrison, R. W.; Andrews, L. C.; Weber, I. T. *Biochemistry* **1992**, *31*, 1541-1550.

(13)   Xuan, J.-C.; Weber, I. T. *Nucleic Acids Research* **1992**, *20*, 5457-5464.

(14)   Lipanov, A.; Kopka, M. L.; Kaczor-Grzeskowiak, M.; Quintana, J.; Dickerson, R. E. *Biochemistry* **1993**, *32*, 1373-1389.

(15)   Base pairs shown are those that prevail at pH 7.0 in crystal and NMR structures; see preceding references.

(16)   Watkins, N. E. J.; SantaLucia, J. J. *Nucleic Acids Research* **2005**, *33*, 6258-6267.

(17)   Loakes, D. *Nucleic Acids Research* **2001**, *29*, 2437-2447 and references therein.

(18)   Case-Green, S. C.; Southern, E. M. *Nucleic Acids Research* **1994**, *22*, 131-136.

(19)   Martin, F. H.; Castro, M. M.; Aboul-ela, F.; Tinoco, I. J. *Nucleic Acids Research* **1985**, *13*, 8927-8938.

(20)   Nichols, R.; Andrews, P. C.; Zhang, P.; Bergstrom, D. E. *Nature* **1994**, *369*, 492-493.

(21)   Loakes, D.; Brown, D. M.; Linde, S.; Hill, F. *Nucleic Acids Research* **1995**, *23*, 2361-2366.

(22)   Frey, K. A.; Woski, S. A. *Chemical Communications* **2002**, 2206-2207.

(23)   Takahashi, Y.; Kato, K.; Hayashizaki, Y.; Wakabayashi, T.; Ohtsuka, E. *Proceedings of the National Academy of the Sciences, USA* **1985**, *82*, 1931-1935.

(24)   Ohtsuka, E.; Matsuki, S.; Ikehara, M.; Takahashi, Y.; Matsubara, K. *Journal of Biological Chemistry* **1985**, *260*, 2605-2608.

(25)   Miura, N.; Ohtsuka, E.; Yamaberi, N.; Ikehara, M.; Uchida, T.; Okada, Y. *Gene* **1985**, *38*, 271-274.

(26)   Kawase, Y.; Iwai, S.; Miura, K.; Ohtsuka, E. *Nucleic Acids Research* **1986**, *14*, 7727-7736.

(27)   Patil, R.; Dekkar, E. *Nucleic Acids Research* **1990**, *18*, 3080.

(28)   Palva, A.; Vidgren, G.; Paulin, L. *J. Microbiol. Methods* **1994**, *19*, 315-321.

(29)   Verma, S.; Eckstein, F. *Annu. Rev. Biochem.* **1998**, *67*, 99-134.

(30)   de Mesmaeker, A.; Haner, R.; Martin, P.; Moser, H. E. *Accounts of Chemical Research* **1995**, *28*, 366-374.

(31)   Ceballos, C.; Prata, C. A. H.; Giorgio, S.; Garzino, F.; Payet, D.; Barthelemy, P.; Grinstaff, M. W.; Camplo, M. *Bioconjugate Chem.* **2009**, *20*, 193-196.

(32)   Johnson, W. T.; Zhang, P.; Bergstrom, D. E. *Nucleic Acids Research* **1997**, *25*, 559-567.

(33)   Bergstrom, D. E.; Zhang, P.; Johnson, W. T. *Nucleic Acids Research* **1997**, *25*, 1935-1942.

(34)   Rutledge, L. R.; Wheaton, C. A.; Wetmore, S. D. *Phys. Chem. Chem. Phys.* **2007**, *9*, 497-509.

(35)   Paragi, G.; Palinko, I.; van Alsenoy, C.; Gyemant, I. K.; Penke, B.; Timar, Z. *New J. Chem.* **2002**, *26*, 1503-1506.

(36)   Hupp, T.; Sturm, C.; Janke, E. M. B.; Cabre, M. P.; Weisz, K.; Engels, B. *Journal of Physical Chemistry A* **2005**, *109*, 1703-1712.

(37)   Li, J.; Cramer, C. J.; Truhlar, D. G. *Biophysical Chemistry* **1999**, *78*, 147-155.

(38)   Shukla, M. K.; Leszcynski, J. *Theochem* **2000**, *529*, 99-112.

(39)   Shukla, M. K.; Leszcynski, J. *Journal of Physical Chemistry A* **2003**, *107*, 5538-5543.

(40)   San Roman-Zimbron, M. L.; Costas, M. E.; Acevedo-Chávez, R. *Theochem* **2004**, *711*, 83-94.

(41)   Costas, M. E.; Acevedo-Chávez, R. *Journal of Physical Chemistry A* **1997**, *101*, 8309-8318.

(42)   Kondratyuk, I. V.; Samiilenko, S. P.; Kolomiets, I. M.; Potyahaylo, A. L.; Hovorun, D. M. *Biopolimery i Kletka* **2000**, *16*, 124-137.

(43)   Kondratyuk, I. V.; Samijlenko, S. P.; Kolomiets, I. M.; Hovorun, D. M. *Journal of Molecular Structure* **2000**, *523*, 109-118.

(44)   Ramaekers, R.; Dkhissi, A.; Adamowicz, L.; Maes, G. *Journal of Physical Chemistry A* **2002**, *106*, 4502-4512.

(45)   Cubero, E.; Guimil-Garcia, R.; Luque, F. J.; Erijita, R.; Orozco, M. *Nucleic Acids Research* **2001**, *29*, 2522-2534.

(46)   Sun, X.; Lee, J. K. *Journal of Organic Chemistry* **2007**, *72*, 6548-6555.

(47)   Kurinovich, M. A.; Lee, J. K. *Journal of the American Chemical Society* **2000**, *122*, 6258-6262.

(48)   Kurinovich, M. A.; Lee, J. K. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 985-995.

(49)   Kurinovich, M. A.; Phillips, L. M.; Sharma, S.; Lee, J. K. *Chemical Communications* **2002**, 2354-2355.

(50)   Lee, J. K. *Internat. Journal of Mass Spectrom* **2005**, *240*, 261-272.

(51)   Liu, M.; Xu, M.; Lee, J. K. *Journal of Organic Chemistry* **2008**, *73*, 5907-5914.

(52)   Pan, S.; Sun, X.; Lee, J. K. *Journal of the American Society for Mass Spectrometry* **2006**, *17*, 1383-1395.

(53)   Sharma, S.; Lee, J. K. *Journal of Organic Chemistry* **2002**, *67*, 8360-8365.

(54)   Sharma, S.; Lee, J. K. *Journal of Organic Chemistry* **2004**, *69*, 7018-7025.

(55)   Liu, M.; Li, T.; Amegayibor, F. S.; Cardoso, D. S.; Fu, Y.; Lee, J. K. *Journal of Organic Chemistry* **2008**, *73*, 9283-9291.

(56)   LeNovere, N. *Bioinformatics* **2001**, *17*, 1226-1227.

(57)   SantaLucia, J.; Allawi, H. T.; Seneviratne, P. A. *Biochemistry* **1996**, *35*, 3555-3562.

(58)   SantaLucia, J. J. *Proceedings of the National Academy of the Sciences, USA* **1998**, *95*, 1460-1465.

(59)   SantaLucia, J. J.; Hicks, D. *Annu. Rev. Biophys. Biomol. Struct.* **2004**, *33*, 415-440.

(60)   Lopez, L. L.; Tiller, P. R.; Senko, M. W.; Schwartz, J. C. *Rapid Communications in Mass Spectrometry* **1999**, *13*, 663-668.

(61)   Gabelica, V.; De Pauw, E. *Internat. Journal of Mass Spectrom* **2002**, *219*, 151-159.

(62)   Gabelica, V.; De Pauw, E. *Journal of Mass Spectrometry* **2001**, *36*, 397-402.

(63)   Gabelica, V.; De Pauw, E. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 91-98.

(64)   Gabelica, V.; Rosu, F.; Houssier, C.; De Pauw, E. *Rapid Communications in Mass Spectrometry* **2000**, *14*, 464-467.

(65)   Pan, S.; Sun, X.; Lee, J. K. *Internat. Journal of Mass Spectrom* **2006**, *253*, 238-248.

(66)   Schnier, P. D.; Klassen, J. S.; Strittmatter, E. F.; Williams, E. R. *Journal of the American Chemical Society* **1998**, *120*, 9605-9613.

(67)   Armentrout, P. B. In *Advances in Gas Phase Ion Chemistry*; Adams, N. G., Babcock, L. M., Eds.; JAI Press Inc.: Greenwich, CT, 1992; Vol. 1, p 83-119.

(68)   Wenthold, P. G.; Squires, R. R. *Journal of the American Chemical Society* **1994**, *116*, 6401-6412.

(69)   Wenthold, P. G.; Squires, R. R. *Journal of Mass Spectrometry* **1995**, *30*, 17-24.

(70)   Wan, K. X.; Gross, M. L.; Shibue, T. *Journal of the American Society for Mass Spectrometry* **2000**, *11*, 450-457.

(71)   Wan, K. X.; Shibue, T.; Gross, M. L. *Journal of the American Chemical Society* **2000**, *122*, 300-307.

(72)   Armentrout, P. B. *Journal of the American Society for Mass Spectrometry* **2002**, *13*, 419-434.

(73)   Deschenes, L. A. V. B., D. A. *Journal of the American Chemical Society* **2000**, *122*, 9567-9568.

(74)   Pan, S.; Verhoeven, K.; Lee, J. K. *Journal of the American Society for Mass Spectrometry* **2005**, *16*, 1853-1865.

(75)   Gaffney, B. L.; Jones, R. A. *Biochemistry* **1989**, *28*, 5881-5889.

(76)   Klassen, J. S.; Schnier, P. D.; Williams, E. R. *Journal of the American Society for Mass Spectrometry* **1998**, *9*, 1117-1124.

(77)   Ganem, B.; Li, Y.-T.; Henion, J. D. *Tetrahedron Letters* **1993**, *34*, 1445-1448.

(78)   Gidden, J.; Ferzoco, A.; Baker, E. S.; Bowers, M. T. *Journal of the American Chemical Society* **2004**, *126*, 15132-15140.

(79)   Daniel, J. M.; Friess, S. D.; Rajagopalan, S.; Wendt, S.; Zenobi, R. *Internat. Journal of Mass Spectrom* **2002**, *216*, 1-27.

(80)   Cheng, X.; Chen, R.; Bruce, J. E.; Schwartz, B. L.; Anderson, G. A.; Hofstadler, S. A.; Gale, D. C.; Smith, R. D.; Gao, J.; Sigal, G. B.; Mammen, M.; Whitesides, G. M. *Journal of the American Chemical Society* **1995**, *117*, 8859-8860.

(81)   Rueda, M.; Kalko, S. G.; Luque, F. J.; Orozco, M. *Journal of the American Chemical Society* **2003**, *125*, 8007-8014.

(82)   Daneshfar, R.; Klassen, J. S. *Journal of the American Society for Mass Spectrometry* **2004**, *15*, 55-64.

(83)   McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. *Journal of the American Society for Mass Spectrometry* **1992**, *3*, 60-70.

(84)   McLuckey, S. A.; Habibigoudarzi, S. *Journal of the American Chemical Society* **1993**, *115*, 12085-12095.

(85)   McLuckey, S. A.; Vaidyanathan, G.; Habibi-Goudarzi, S. *Journal of Mass Spectrometry* **1995**, *30*, 1222-1229.

(86)   McLuckey, S. A.; Vaidyanathan, G. *Internat. Journal of Mass Spectrom* **1997**, *162*, 1-16.

(87)   Hang, B.; Singer, B.; Margison, G. P.; Elder, R. H. *Proceedings of the National Academy of the Sciences, USA* **1997**, *94*, 12869-12874.

(88)   Saparbaev, M.; Laval, J. *Proceedings of the National Academy of the Sciences, USA* **1994**, *91*, 5873-5877.

(89)   Stivers, J. T.; Jiang, Y. L. *Chemical Reviews* **2003**, *103*, 2729-2759.

(90)   Berti, P. J.; McCann, J. A. B. *Chemical Reviews* **2006**, *106*, 506-555 and references therein.

(91)   O'Brien, P. J.; Ellenberger, T. *Journal of Biological Chemistry* **2004**, *279*, 26876-26884.

(92)   O'Brien, P. J.; Ellenberger, T. *Biochemistry* **2003**, *42*, 12418-12429.

(93)   Singer, B.; Hang, B. *Chem. Res. Toxicol.* **1997**, *10*, 713-732.

(94)   Wyatt, M. D.; Allan, J. M.; Lau, A. Y.; Ellenberger, T. E.; Samson, L. D. *BioEssays* **1999**, *21*, 668-676.

(95)   *Repair of Alkylation Damage to DNA*; Seeberg, E.; Berdal, K. G., Eds.; Landes Bioscience: New York, 1997.

(96)   Lau, A. Y.; Wyatt, M. D.; Glassner, B. J.; Samson, L. D.; Ellenberger, T. *Proceedings of the National Academy of the Sciences, USA* **2000**, *97*, 13573-13578.

(97)   In E. coli, 3-methyladenine DNA glycosylase II (AlkA) excises hypoxanthine. For reviews, see: references JKL LOOK FOR STIVERS AND BERTI'S CHEM REV ARTICLES!

(98)   Lau, A. Y.; Schärer, O. D.; Samson, L.; Verdine, G. L.; Ellenberger, T. *Cell* **1998**, *95*, 249-258.

(99)   Vallur, A. C.; Feller, J. A.; Abner, C. W.; Tran, R. K.; Bloom, L. B. *Journal of Biological Chemistry* **2002**, *277*, 31673-31678.

(100)   Xia, L.; Zheng, L.; Lee, H.-W.; Bates, S. E.; Federico, L.; Shen, B.; O'Connor, T. R. *Journal of Molecular Biology* **2005**, *346*, 1259-1274.

(101)   Guliaev, A. B.; Hang, B.; Singer, B. *Nucleic Acids Research* **2002**, *30*, 3778-3787.

(102)　AAG does not target hypoxanthine in RNA, so in vivo applications of hypoxanthine as a universal base in RNA would not be problematic from this perspective.

(103)　Abner, C. W.; Lau, A. Y.; Ellenberger, T.; Bloom, L. B. *Journal of Biological Chemistry* **2001**, *276*, 13379-13387.

(104)　Vallur, A. C.; Maher, V. M.; Bloom, L. B. *DNA Repair* **2005**, *4*, 1088-1098.

CURRICULUM VITAE

Xuejun Sun

Education:

1996 - 2000	B.S., Chemistry Nankai University, P. R. China
2000 - 2003	M.S., Polymer Chemistry Nankai University, P. R. China
2003 - 2010	Ph.D., Bioorganic Chemistry Rutgers, The State University of New Jersey

Publications:

1. Anna Zhachkina, Min Liu, **Xuejun Sun**, F. Sedinam Amegayibor and Jeehiun K. Lee; "Gas-Phase Thermochemical Properties of the Damaged Base O^6-Methylguanine versus Adenine and Guanine"; *Journal of Organic Chemistry*; **2009**; 74(19) pp 7429-7440.
2. **Xuejun Sun**, and Jeehiun K. Lee; "Acidity and Proton Affinity of Hypoxanthine in the Gas Phase versus in Solution: Intrinsic Reactivity and Biological Implications"; *Journal of Organic Chemistry*; **2007**; 72(17) pp 6548-6555.
3. Su Pan, **Xuejun Sun**, and Jeehiun K. Lee; "DNA Stability in the Gas Versus Solution Phases: A Systematic Study of Thirty-One Duplexes with Varying Length, Sequence, and Charge Level"; *Journal of the American Society for Mass Spectrometry*; **2006**; 17(10) pp 1383-1395.
4. Su Pan, **Xuejun Sun**, and Jeehiun K. Lee; "Stability of Complementary and Mismatched DNA Duplexes: Comparison and Contrast in Gas Versus Solution Phase"; *International Journal of Mass Spectrometry*; **2006**; 253(3) pp 238-248.
5. **Xuejun Sun**, Suxia Chen, Shunzi Li, Husheng Yan, Yunge Fan, Huaifeng Mi; "Deletion of two C-terminal Gln Residues of 12-26-Residue Fragment of Melittin Improves Its Antimicrobial Activity"; *Peptides*; **2005**; 26(3) pp 369-375.
6. Suxia Chen, **Xuejun Sun**, Shunzi Li, Husheng Yan; "Antibacterial and Hemolytic Activities of the Reverse-sequence Analogs of a Melittin′s C-terminal Segment"; *Chemical Journal of Chinese Universities*; **2005**; 26(12) pp 2233-2236.
7. Shunzi Li, **Xuejun Sun**, Husheng Yan, Binglin He; "Antimicrobial and Hemolytic Activities of Melittin and Its Analogues and Their Interactions with Phospholipid Membranes"; *Chemical Journal of Chinese Universities*; **2005**; 26(1) pp 73-77.
8. Husheng Yan, Shunzi Li, **Xuejun Sun**, Huaifeng Mi, Binglin He; "Individual Substitution Analogs of Mel(12-26), Melittin's C-terminal 15-Residue Peptide: Their Antimicrobial and Hemolytic Actions"; *FEBS Letters*; **2003**; 554(1) pp 100-104.