UNIVERSITY OF CALIFORNIA

Santa Barbara

Structural Effects on Excited State Lifetime of Non-Canonical Nucleobases as Studied by REMPI

A Thesis submitted in partial satisfaction of

the requirements for the degree Master of

Science

in Chemistry

by

Nathan Daniel Svadlenak

Committee in charge:

Professor Mattanjah de Vries, Chair

Professor Glake Hill

Professor Martin Moskovitz

Professor Song-I Hahn

December 2020

The thesis of Nathan Svadlenak is approved.

Song-I Han

Glake Hill

Martin Moskovitz

Mattanjah de Vries, Committee Chair

Abstract

Structural Effects on Excited State Lifetime of Non-Canonical Nucleobases as Studied by REMPI

by

Nathan Daniel Svadlenak

Resonant Enhanced Multi-Photon Ionization (REMPI) mass spectroscopy is a powerful tool for the investigation of dynamic molecular states. The principles of REMPI as they relate to action mass spectroscopy are outlined. Two major investigations are presented. REMPI, in combination with quantum chemical theory, is used to study the mechanism of internal conversion in a family of related purines and pyrimidines. These same techniques are also used to study competition between hydrogen bonding and pi-stacking interactions in a set of non-standard nucleobases.

I. Introduction and Motivation

The simplicity of the genetic alphabet was as unexpected as it was striking. Before the genetic material was isolated, it was known that there were 4 nucleobases in DNA and 20 amino acids in protein. It seemed obvious that the language of life was written with amino acids instead of nucleobases. The isolation and characterization of DNA was perhaps one of the most profound scientific discoveries of the 20th century. The double helix became almost immediately iconic, acting as a universally recognized symbol of life. The simple, repeating periodic structure revealed that life was written in a language that could ultimately be understood by humans.

The parsimony of the genetic alphabet is convenient, but also mysterious. Nature only chose five chemicals to participate in the information. But guanine, cytosine, adenine, thymine and uracil all belong to a much broader set of naturally occurring chemicals, the purines and pyrimidines. The five canonical nucleobases represent just a thin slice of the possible genetic components that nature could have selected.

These so-called non-canonical nucleobases are non-trivial. Methylated versions of the nucleobases are extant in living organisms today. But even the more exotic alternative nucleobases display some intriguing characteristics. Benner has an ongoing project called AEGSIS (Artificially Expanded Genetic Information System), in which he has created a 12 letter genetic alphabet [1], with ten of those letters being entirely non-canonical. These nucleobases selectively base pair with a recognition scheme similar to that found in biological DNA. Moreover, these non-canonical nucleobases, when given the proper enzymes, are able to form a stable double helix.

The non-canonical nucleobases may not be mere academic curiosities. Projects like AEGIS may be a window through which we can peer into biochemistries that could have existed had

conditions on earth been slightly different, or perhaps even biochemistries that really do exist right now, under the light of a distant star.

At this point, it is natural to ask if there is something special about the five canonical chemicals, or perhaps instead if there is something maladaptive about the others. This is a problem that has been attacked by many different groups across many different angles. To be viable, a candidate molecule must have an abiotic synthetic pathway, it must have a long chemical half-life, it must be able to bind with ribose, it must be able to engage in selective hydrogen bonding, and it must surely cross many other hurdles as well. We desire to contribute to this ongoing narrative by investigating a criterion that we are uniquely suited to probe, that of photochemical stability.

Consider an ultraviolet photon with a 285 nm wavelength. When it gets absorbed by a molecule, it deposits 4.3 eV. This is a routine absorption event, but I want to stop for a moment and reflect on how extraordinary it really is. What temperature would a molecule have to be at in order to reach this same excited state thermally? By examining the Boltzmann ratio, one can quickly see that to even get 1% of the molecular population into this state by thermal processes, you would need a temperature of 11,000 K.

What this means is that absorbing even a single UV photon is a disaster for a molecule. Chemicals must have methods to efficiently remove optical energy or face destruction. The faster a molecule is able to shed this extra energy, the less time there for something catastrophic, such as bond dissociation or reaction, to occur. Excited state lifetime is therefore related to photo stability.

A pattern that has come out of research is that while some of the non-canonical nucleobases have “long” (nanosecond regime) excited state lifetimes, all five of the canonical nucleobases have excited state lifetimes under 10 picoseconds [2]. We may be looking at the results of an ancient

selection criterion, present when no /ozone layer existed. Solar radiation would have been very strong then, and may have played a role in determining which chemicals were available for the first abiogenetic events.

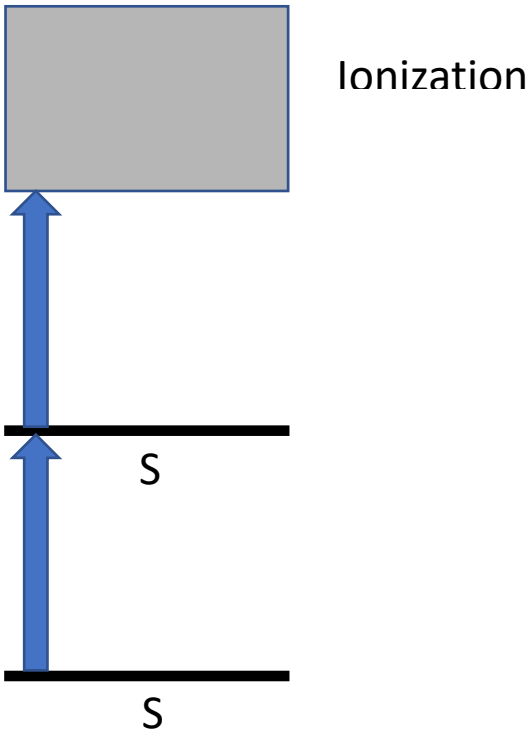
Knowing that the canonical nucleobases can ground themselves quickly is one thing, knowing how they do it is another. The internal electronic dynamics of a molecules are dictated by structure, and changes to structure can lead to changes in lifetime. “Structure” in this context can have a subtle meaning. It can refer to the chemical identity of the molecule (e.g. guanine versus adenine). It can refer to tautomers of the same chemical species (e.g. -keto -enol variants). It also refers to cluster behavior, such as guanine base pairing with cytosine, or resveratrol complexing with water molecules. Even in this last category, there are additional variants. Guanine-cytosine clusters have different dynamics depending on whether the pairing is Watson-Crick, reverse Watson-Crick, Hoogstein, or pi-stacked [3]. For the remainder of this work, “structure” refers to the complete electronic environment of the molecules being studied.

A. Laser Ionization

REMPI (Resonantly Enhanced Multi-Photon Ionization) is a spectroscopic technique for producing action mass spectroscopy spectra[4]. The feature that distinguishes REMPI for other mass spectroscopic techniques is the ionization method. A tunable laser scans in the UV-vis region. When a given wavelength corresponds to an electronic transition within the molecule, the molecule enters a high energy excited state. Additional incoming photons can then ionize the molecule.

This resonant pathway yields an ion signal that is both strong and selective. Significant signal is only produced when the laser is tuned to a specific transition. It is certainly possible to use a wavelength that can directly ionize the target with a single photon. This would save the hassle

of having to find a resonant wavelength. However, this is usually not a desirable tradeoff, because the goal of these experiments is not to generate ions, but rather to collect a UV-vis spectrum of the molecule by using the ions as a source of signal. The fact that the signal decreases when the laser is off resonance is in fact the point of the technique.



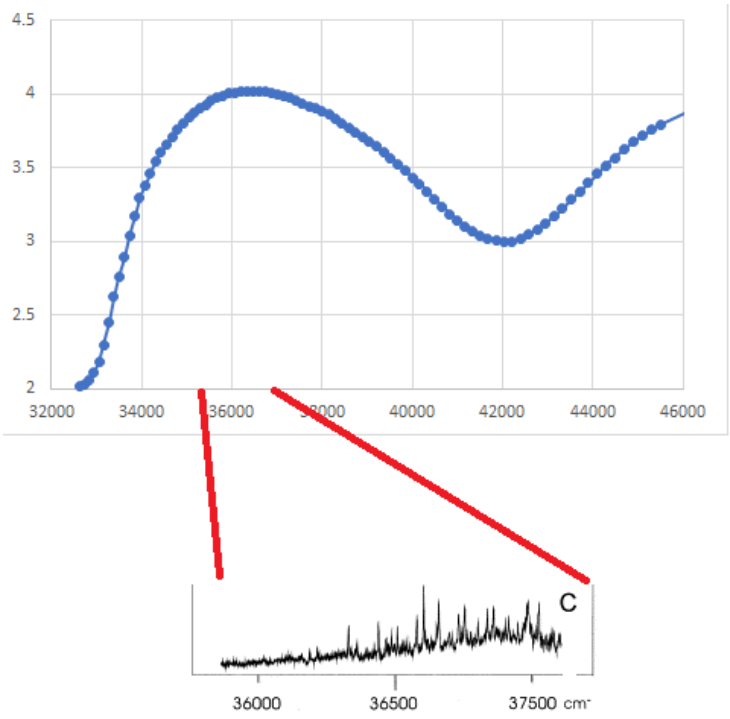
The scheme discussed here is the simplest one used in this work. There are other, more complex laser ionization techniques, but fully understanding them requires more context, and so they will be discussed later.

B. Gas Phase Spectroscopy

At first blush, studying biomolecules in the gas phase seems like a very odd decision. Biomolecules exist in complex condensed phases, not as isolated particles floating in vacuum.

Nonetheless, the decision to use gas phase is not arbitrary, and has several powerful advantages that can't be replicated by aqueous phase experiments.

It is easiest to start by showing rather than by telling.



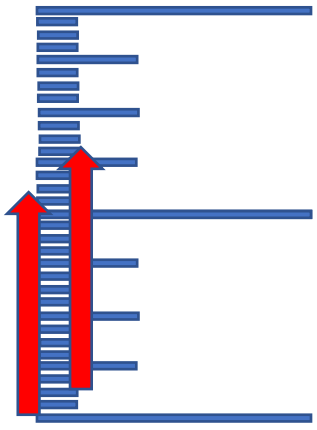
The top portion of the figure is an aqueous phase UV-vis absorption scan of cytosine, taken from the NIST Chemistry Handbook and converted to wavenumbers for convenience [5]. The inset is a gas phase REMPI scan of the same molecule [6]. Whereas the condensed phase scan is only able to show a single large peak, the REMPI resolves into the underlying fine structure. Experimenters can extract highly detailed structural and electronic information from this data.

Understanding why the gas phase data are so well-resolved is best approached from understanding why the condensed phase data is not. Solvated species exist in a complex environment. Each individual molecule will have a slightly different geometric relationship to its neighbors and will therefore exist in a slightly different electric field. This causes the absorption

spectrum of the bulk chemical to broaden, since the molecules don't all share the same physical circumstances. These effects are collectively referred to as pressure broadening.

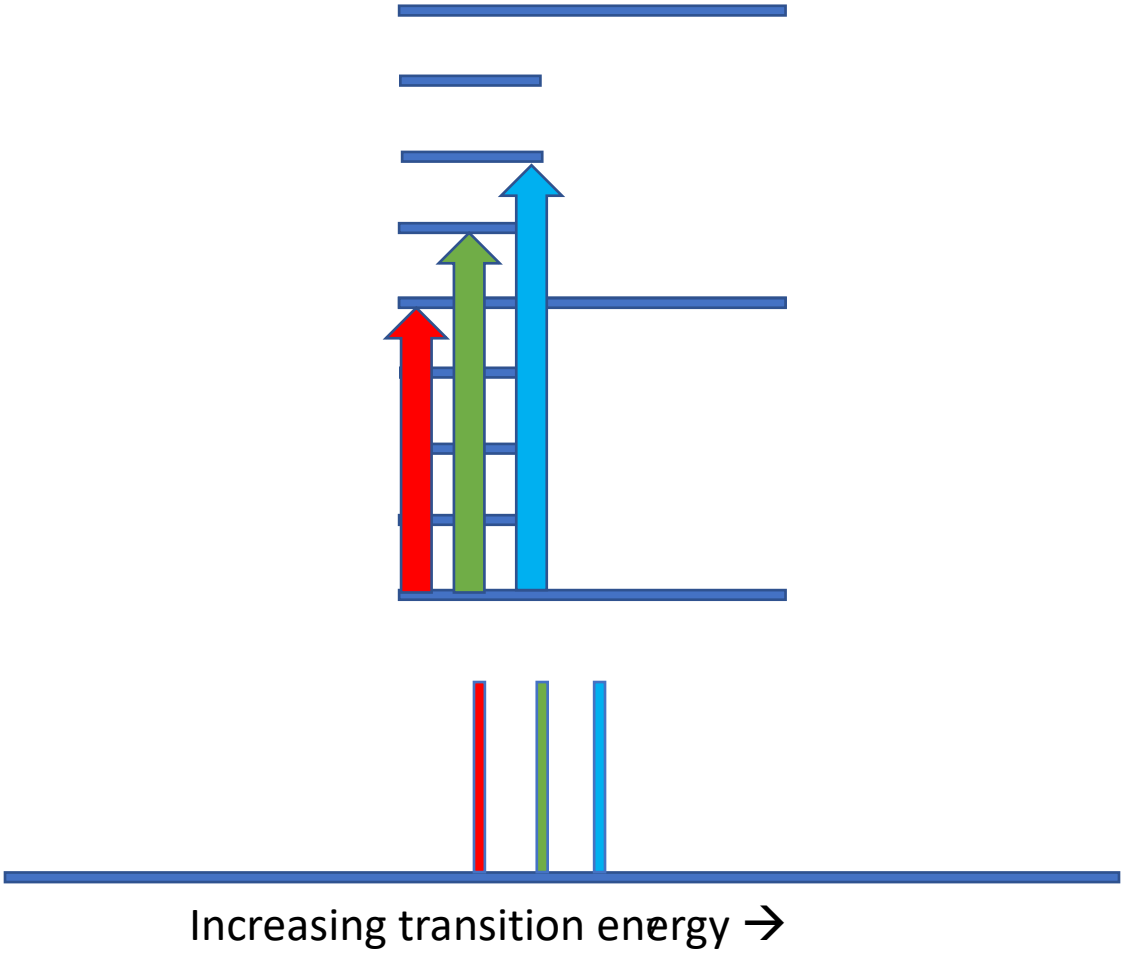
Molecular beam techniques solve this problem mostly by just being a gas phase process. At low pressures, the molecules are distant and isolated from each other, and the bulk spectrum converges towards the individual spectrum. Moreover, the actual spectroscopic readings are performed on the well-collimated portion of the beam. In this region, called the "collision-free" region, the molecules are traveling in a parallel manner and never come into contact with each other.

Pressure broadening alone is not enough to explain the contrast between condensed phase absorption spectroscopy and REMPI. The next element that needs to be considered are the vibrational and rotational levels of the molecules. The vibrational energy levels "fill in" the space between the electronic energy levels, leading to a higher density of allowable transitions. The rotational levels then fill up the space between the vibrational levels. When these effects combine with pressure broadening, it is easy to understand why condensed phase spectra are so broad and undefined.



An abstract diagram showing how the rotational and vibrational levels lead to many more possible transitions than just the simple ground-to-ground electronic transitions. This diagram does not accurately depict the spacing and density of the levels.

This problem cannot be solved simply by putting the sample into gas phase. This is where the jet cooling properties of a molecular beam become important [8]. The collisional cooling that happens in the beam shortly after desorption rapidly decreases the internal temperature of the species to around ~10 K. At these very low temperatures, nearly all the species are in the vibrational ground state. This greatly reduces the number of possible permutations when exciting from one level to another, since all molecules are now starting from the same electro-vibratory level. This simplification of the spectrum is sometimes referred to as “vibronic resolution,” because the peaks correspond to specific vibrational levels within the excited state.

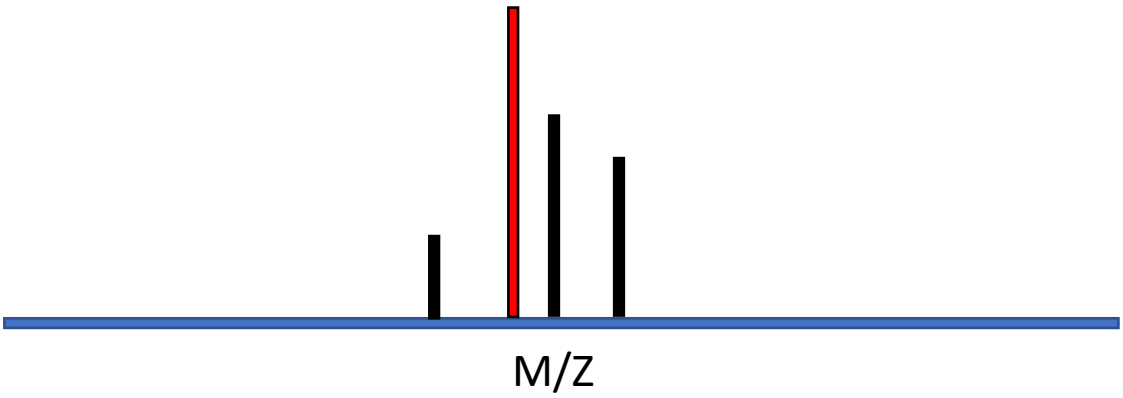


A simple diagram showing how the spectral peaks in a vibronically resolved scan correspond to vibrational levels within a single excited electronic state.

C. Action Mass Spectroscopy

This works concerns itself with action mass spectroscopy, which uses a mass spectrum as a marker for optical transitions. This allows us to collect mass-sorted high resolution UV and IR spectra while retaining the low limits of detection associated with traditional mass spec.

Consider a “dirty” sample that contains the analyte of interest and several unknown contaminants, all of which have a UV-vis spectrum that overlaps with the analyte. The mass spectrum itself would look something like this



The red line is the mass peak that corresponds to the analyte of interest. Other compounds are present, but do not overlap or interfere with the targeted mass peak in any way. These mass peaks will change in intensity as the wavelength is scanned and the laser goes in and out of resonance. By graphing our UV-vis spectra as a function of mass signal response vs wavelength,

separate and completely non-interacting data sets can be acquired for each species, even when those species share specific absorbances.

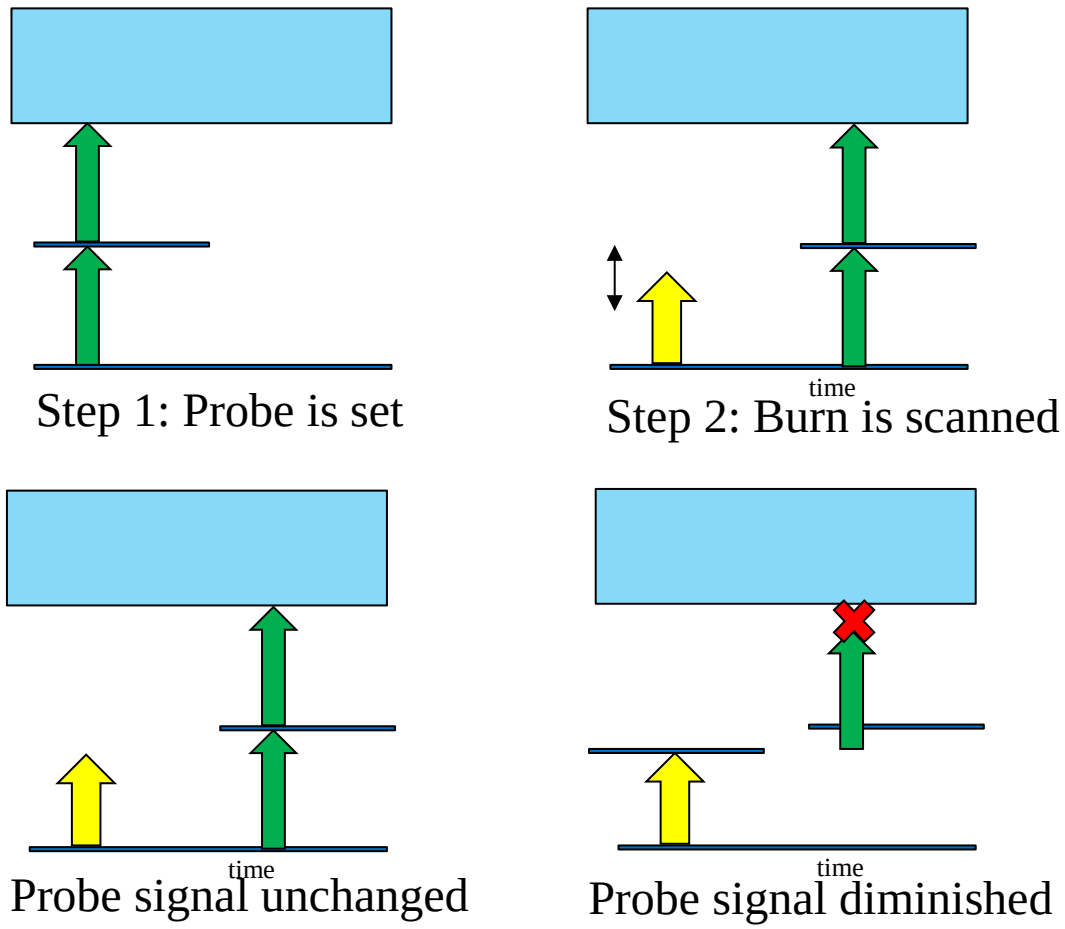
D. Hole Burning

For many investigations, separating species out by their mass is not enough. Sometimes, the experimenter would like to further separate the analyte by structure. If, for example, a given molecule has both cis and trans isomers, it may be valuable to know which isomers are responsible for which spectral features. Or, as a subtler example, if two molecules form a dimer, there may be multiple ways that the two molecules can join together. Disentangling these isomeric structures is usually accomplished by a technique known as spectral hole burning, which is also sometimes called ion dip spectroscopy.

Hole burning techniques require at least two separate lasers to interact with the molecular beam. One laser, called the probe, is set to a wavelength corresponding to a peak in the REMPI spectrum of the structure of interest. This laser is then fixed at that wavelength. An additional laser, called the burn laser, is then introduced into the experiment. The burn laser fires before the probe (the exact time delay between the two lasers is typically on the order of nanoseconds, but this can change radically depending on the exact experimental parameters. There is no one "correct" time interval for hole burning). Unlike the probe, which is fixed at a known resonant wavelength, the burn laser scans though the wavelengths.

In the absence of any other effect, the probe should be generating a constant and strong REMPI signal. The burn laser alters this constant background signal by interacting with the molecule before the probe laser has a chance to reach it. If the burn laser is at a resonant wavelength, the molecule will be in an excited state when the probe laser reaches it. The probe laser, which was set to be resonant for the ground state, will not be resonant with this new excited

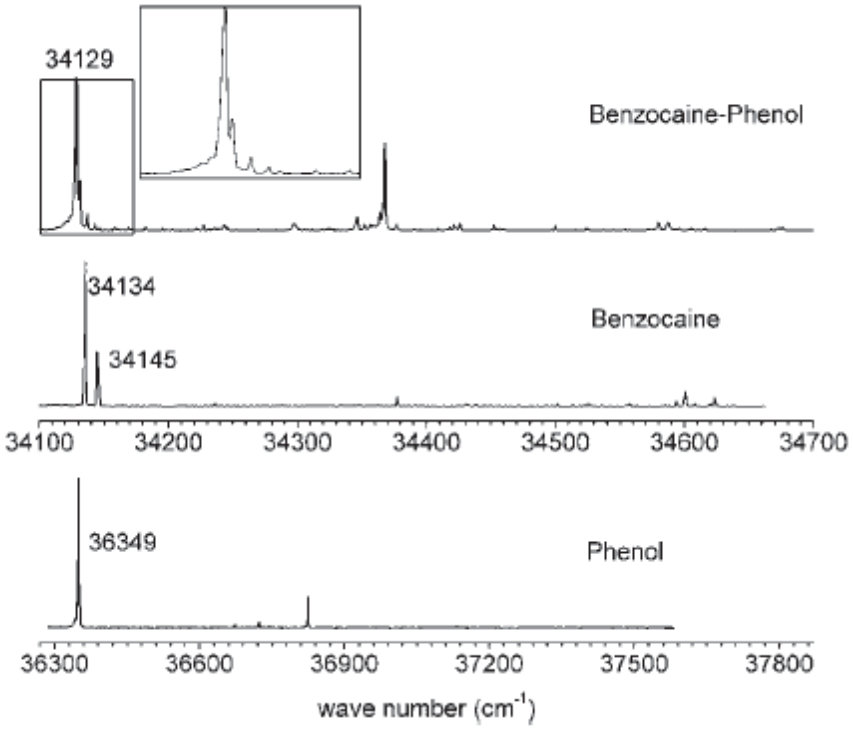
state, and so the REMPI signal will go down. As the burn laser scans past the peak, the signal will go back up, since the burn laser is no longer interacting with the molecule and not affecting the probe's interaction with it anymore.



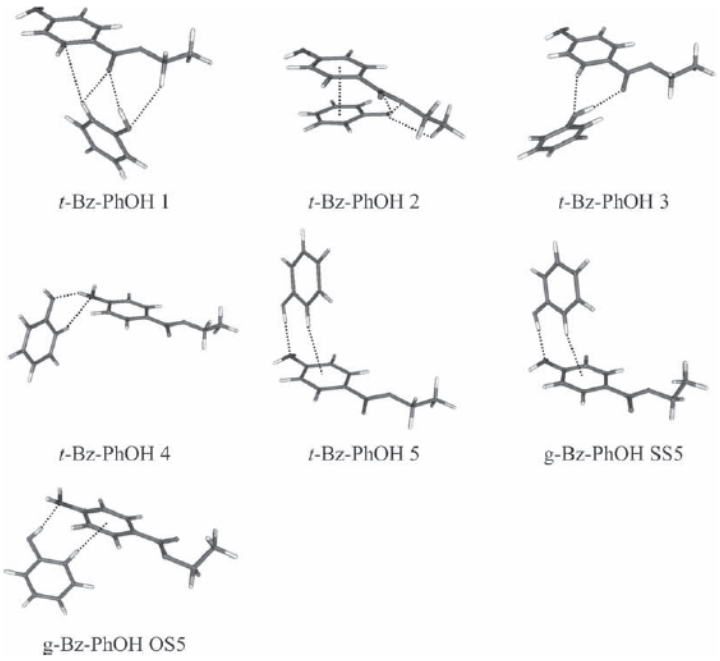
The result is an "upside down" spectrum, with an elevated baseline and sharp "holes" or "dips" burned into it. These burns correspond to the absorbance peaks of the burn laser. However, because these burns are modifications of the original fixed-wavelength probe signal, they only correspond to the structures that were generating that specific probe peak. Peaks that belong to one structure will share the same burn spectrum, and will have different burn spectra from peaks that

belong to other structures. In this manner, the number of different structures present can be determined.

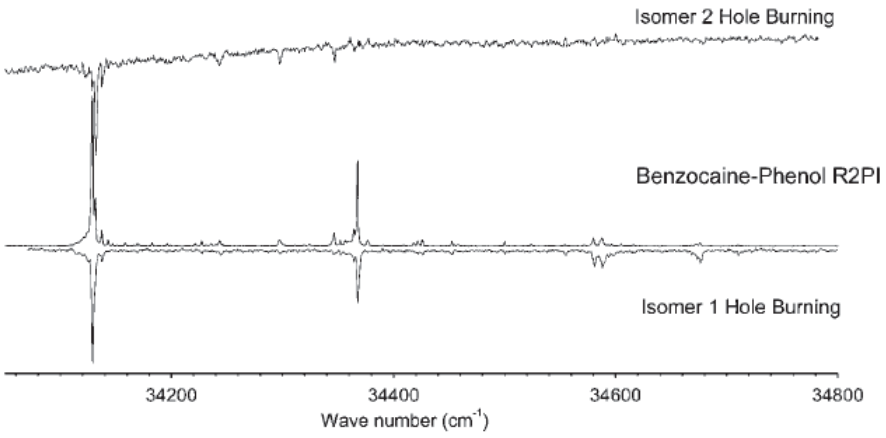
UV-UV holeburning can be difficult to understand, so here is an example from the literature. Aguado *et al* were studying a mixed sample of benzocaine and phenol in a molecular beam machine. They collected a REMPI spectrum for three different mass peaks, belonging to benzocaine, phenol, and a benzocaine+phenol cluster [7]:



The investigators were specifically interested in the phenol + benzocaine interaction, and wanted to know how exactly the two molecules were connecting with each other. From theory, they knew that there were several possible candidates:



The B+P spectrum at the top of the figure could be due one, some, or all of these structures. UV-UV holeburning was used to separate the REMPI into its component burn spectra. They burned two REMPI peaks, one at 34129 cm^{-1} and the other at 34134 cm^{-1}. Here are the two burn spectra, with the original REMPI spectrum shown in the middle for comparison:



Note that the burn spectra are “upside down,” because the signal in a burn spectrum is actually a dip into detector response, rather than a peak. The original REMPI spectrum, of course, is “right side up.”

Inspection of the graph shows that most of the features in the REMPI belong to what they call “Isomer 1.” The remaining features show up in “Isomer 2.” This set of spectra is evidence that there are at least two separate structures responsible for the original REMPI spectrum.

The assignment of these spectra to specific chemical structures is usually done via IR-UV hole burning, a technique that we will cover in the next section.

A slightly different application of UV-UV hole burning is eliminate a structure from a spectrum. If the experimenter is faced with a very congested spectrum, it may make sense to burn out a peak associated with one of the “noisier” molecules. The resulting spectrum will contain features from all of the structures, minus the ones from the structure that was burned out. This can reveal hidden or otherwise difficult to resolve features that may have been buried under the burned structure. See Mons [9] for an example of this technique in action.

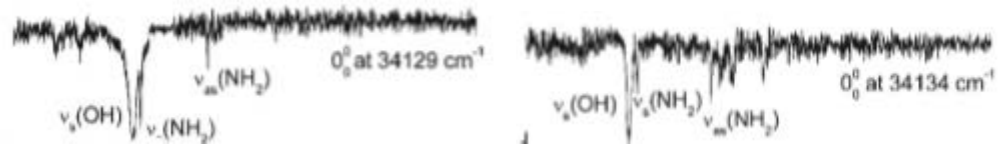
E. IR-UV Burning

Confident structural assignment of a REMPI peak to a specific structural isomer requires a vibrational spectrum. A naïve proposal would be to use an infrared laser as the ionization source. The problem with this approach is that the ionization potential of most molecules is very large relative to the energy of the individual photons. For example, an O-H stretch at 3200 cm-1 has an energy of 0.399 eV. The ionization potential of purine is 8.87 eV [10]. Detecting such a vibration

through an ionizing IR laser would require a 21st order absorption event, which is effectively impossible. A more subtle technique is needed.

IR-UV hole burning is just such a technique. The principal is the same as UV-UV hole burning, except that the burn laser is now an IR laser instead of UV. When the IR laser matches a vibrational resonance, the molecule stops being resonant with the UV probe, and signal yield is decreased. This creates an ion-dip spectrum of the vibrational frequencies of the species in question.

Continuing with the literature example used in the previous section, the experimenters had used REMPI to generate three mass-resolved UV spectra. They then used UV-UV hole burning to deduce that there were two structures present in their beam, and were able to the REMPI peaks to their corresponding chemical structures. Having sorted the isomers, they were then able to identify them by using IR-UV hole burning on peaks known to be unique to given isomers. This created IR ion-dip spectra for “isomer 1” and “Isomer:.”



Matching these against MP2 calculations allowed them to assign these spectra to the “t Bz-PhOH 5” and “g Bz-PhOH SS5” structures, respectively.

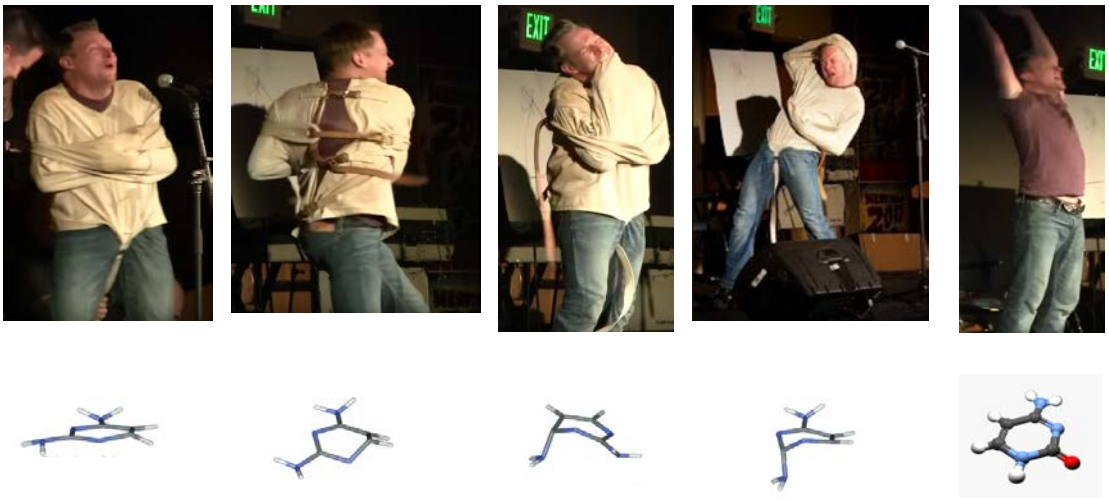
II. Structure and Lifetime of Substituted Pyrimidines and Purines

It's common to talk about excited state dynamics entirely in terms of electronic energies. It is important to remain mindful that these electronic rearrangements also cause nuclear rearrangements. An excited molecule is in a distressed and distorted state, bent out of its equilibrium geometry and flailing about as the internal energy quanta splash back and forth throughout the various bonds. When it finds a conical intersection and returns to the ground state, it is because it has bent itself into a specific geometry that is shared by both the ground state and the excited one. The process is very physical, and lends itself well to direct visualization.

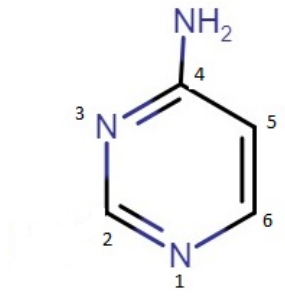
This geometry-focused interpretation of excited state dynamics is rich ground for studying specific questions about the mechanism of electronic relaxation. Molecules can reach a lower state when they reach a conical intersection. But the conical intersections are gated by a barrier, else the excited state would have no lifetime at all. Somewhere along this barrier is one or more saddle points, a sort of "valley" that pierces through the barrier and offers a somewhat easier pathway to the intersection. These saddle points correspond to the transition state geometry that bridges the excited state to the ground state. Through its random undulations, the molecule must snap itself through this shape before it can escape to the underlying levels, just as an escape artist might have to pass through a strained and difficult posture to wriggle out of a restraint [11].

This also means that we can experimentally test different saddle point candidates by substituting the molecule in a manner that would either help or hinder the necessary motions, and then seeing how the substituted molecules' excited state lifetimes changes relative to that of the original molecule. As a simple example, if it is proposed that a certain ring atom must pass through a high

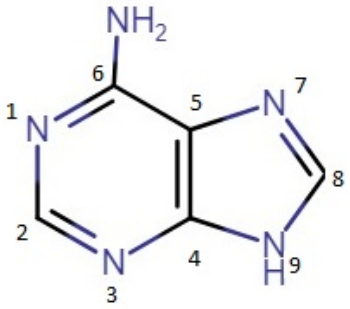
energy out-of-plane deformation to reach the saddle point, then attaching a bulky alkaloid chain to the atom should hinder this motion and increase the lifetime



This work [12] uses 4-aminopyrimidine as a starting point, and then alters this molecule to study the effects on excite dynamics. We will be looking at both pyrimidines and purines, and it is helpful to quickly review the structure and nomenclature of these two classes of chemicals:



4-aminopyrimidine

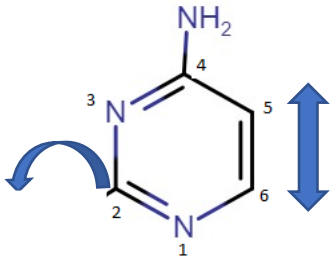


6-aminopurine

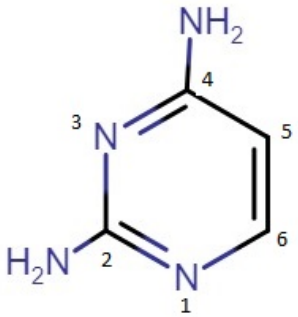
6-aminopurine is just 4-aminopurine with the imidazole ring added. The written names might suggest that the amino group has also been moved, but inspection of the structural formulas shows that this is not the case. Unfortunately, the standard naming conventions cause the position

labels to change from the pyrimidine to the purine, which can cause confusion. Please refer to these figures if the numbering system ever feels unclear. Also note that, by coincidence, positions 2 and 5 are the same between the two molecules. Both of these positions will be important later.

Previous theoretical work [13] predicted that 4-aminopyrimidine reaches the ground state via a conical intersection. To reach this intersection, the molecule can either use an out-of-ring swing at the 2 position, or employ a puckering motion along the 5-6 bond. The calculated excited state lifetime was 400 fs, which is shorter than the experimental limits available to our lab at the time.

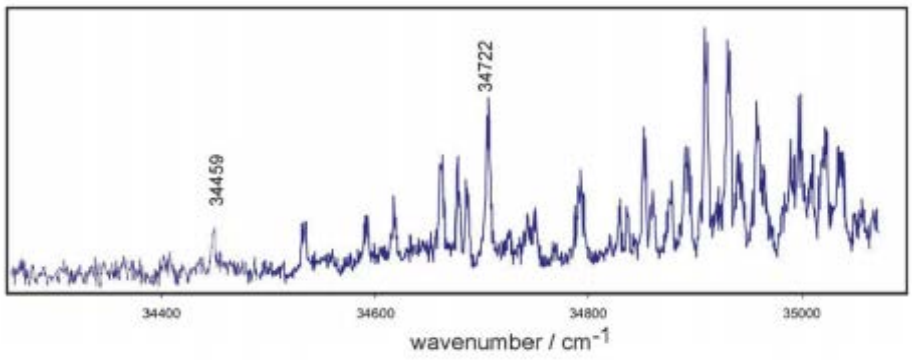


To test the C2 out-of-plane motion, we obtained a sample of 2,4 diaminopyrimidine:

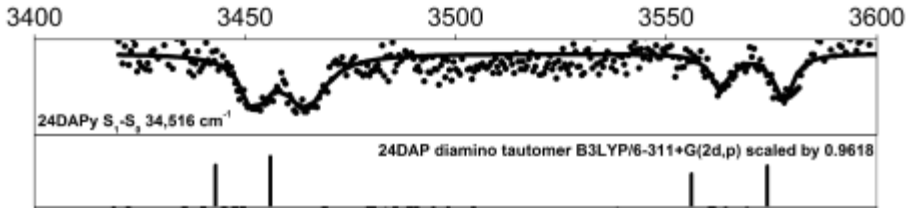


2,4 diaminopyrimidine

A REMPI spectrum was collected:



Tautomeric determination was done with the aid of IR-UV hole burning. The peak at 34,516 cm^{-1} was probed with a scanning IR laser, resulting in the following ion dip spectrum:



The best match to theory is shown below the spectrum. This corresponds to the diamino tautomer. The match is quite good, but to eliminate the possibility of additional tautomers, we did a "reverse" hole burn, using the IR frequency at 3578 cm^{-1} as the fixed frequency, and scanning the UV laser with a delay of a few nanoseconds. We found total depletion of the spectrum across the entire scan region, indicating that the IR bands we found belong to the diamino tautomer, and to that tautomer alone.

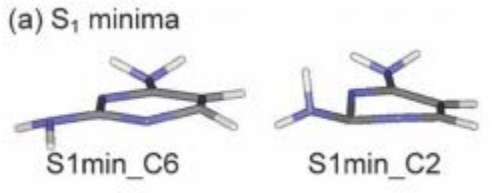
Direct measurement of the excited state lifetime was not possible, due to the short timescales involved with this molecule's dynamics. Indeed, we were not even able to obtain a UV-UV holeburn for this molecule. However, our REMPI was of high enough resolution that the

rotational envelope was visible for several of the peaks. By Heisenberg uncertainty, the width of these envelopes places a floor on the lifetime. We estimated the width of these peaks by simulating three different Lorenztian peak shapes, corresponding to three different linewidths:

The shortest simulated Lorzentian (0.01 cm^{-1}) is very slightly too narrow, and the widest one (1.0 cm^{-1}) is very slightly too wide. This allows us to bracket the linewidth in between these two numbers. Applying the Heisenberg limits, this places the minimum possible lifetime somewhere between 10 ps and 1 nanosecond. That's quite a wide range, but is still unambiguously longer than 2-aminopyrimidine's calculated lifetime of 400 fs.

Calculations were performed using Complete Active Space Self-Consistent Field (CASSCF) and the 6-31** basis set. Results from this were cross-checked with complete active space self-consistent-field second-order perturbation theory (CASPT2) and Resolution of Identity Coupled Cluster to the second order (RICC2).

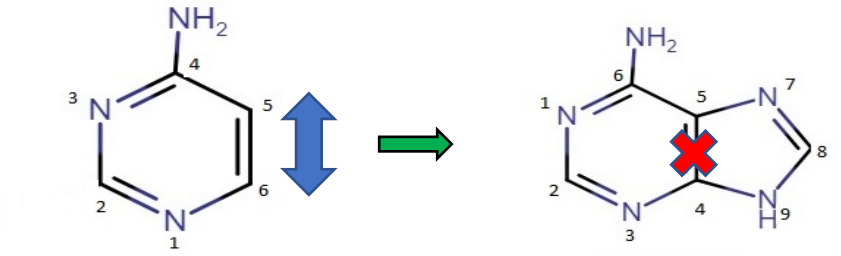
The computations found two minima at the S_1 potential energy surface. Because these are minima, they must be considered as adiabatic excited states, rather than as vertical states.



Now that we know where we start on the excited state energy surface, we need to know where we are going. Our calculations detected the presence of a conical intersection with five separate minima along the seam. These were grouped into four families: Boat (B), Screw-boat (S),

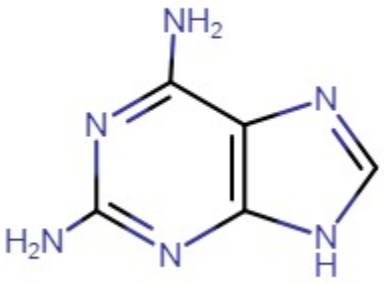
Half-chair (H), and Envelope (E). The B,S, and H families were energetically very similar, with E being much higher than the rest. In a previous study, Barbatti et al found similar transistion state structures for the singly-substituted 4-aminopyrimidine. This suggests that the addition of the amino group on the 2 position has only a marginal effect of the transition state, which is consistent with the difference in lifetimes between the two.

As outlined earlier, the two position is not the only locus implicated in the transition state. The conical intersection is gated by both an N2 motion and a C5-C6 stretching motion. The next logical step is to revert the 2 position to its original unhindered state, and start blocking the C5-C6 motion instead. There is no need to do any new experiments for this, because natural adenine already meets these criteria and has been measured by Kang et al [2].

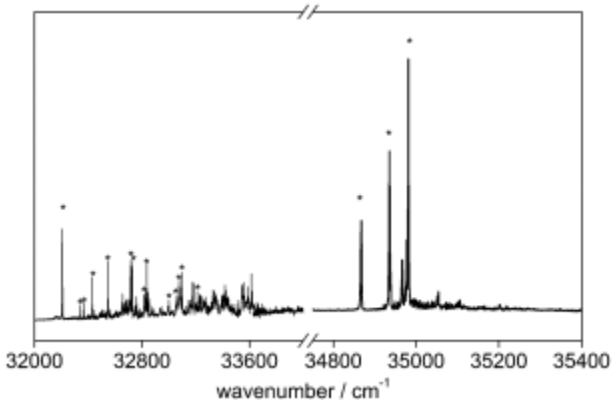


Adenine was found to have an excited state lifetime of 1 ps. This is definitely slower than 4-aminopurine, and has an unclear relationship to 2,4-diaminopyrimidine's lifetime, which was bracketed between 1 ps and 1 ns.

The next logical thing to do is combine our previous modifications and block both the N2 site and the C5-C6 site. We purchased 2,6 diaminopurine.

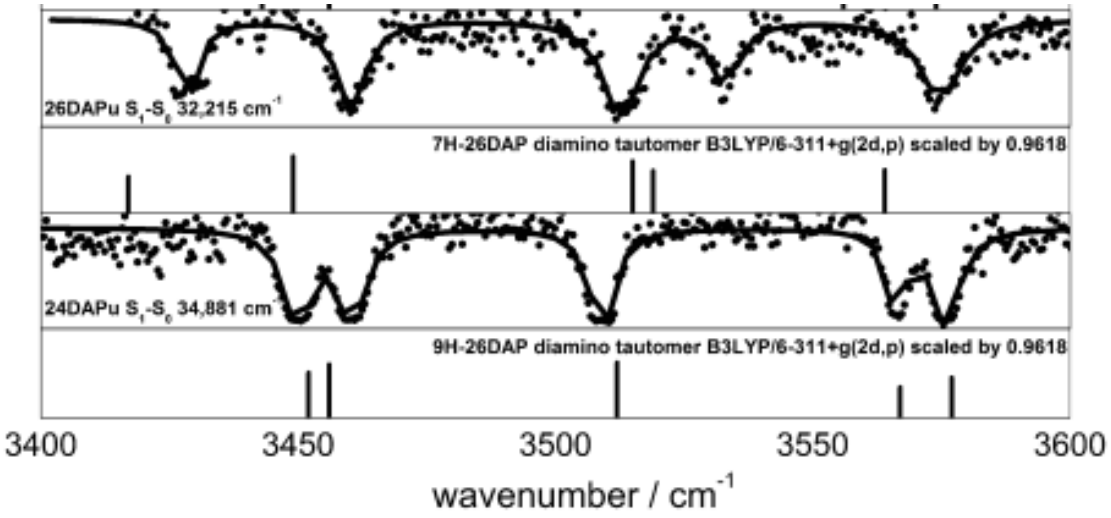


This molecule proved to be much more tractable to our techniques, yielding a clean REMPI spectrum:



A careful reader will notice that the spectrum is split into two portions. The red portion comes from a previous study done by our group. We were able to collect a REMPI, but not able to identify the tautomers. In this study, we extended the existing spectrum out to the blue, discovering new peaks that may help us with making a tautomeric determination. The two collections of peaks are quite distant from each other, hinting at the possibility of two tautomers, one possibly having its origin at the beginning of the blue section (3,215 cm^{-1}) and another at the beginning of the red section (34,881 cm^{-1}).

This is purely speculative, of course, but this guess provides a framework from which we design experiments. We carried out IR-UV hole burning on the two aforementioned peaks:



Best theory matches are shown as stick spectra.

The two peaks clearly belong to at least two different structures. DFT methods indentified these as the 7H and 9H diamino tautomers. Additionally, energy calculations indicated that these two structures were the lowest energy of all possible tautomers, lending further credence to our structural assignments.

We then directly measured the lifetimes of both tautomers. This was done by resonantly exciting the molecule at the two origin peaks and then ionizing with a 266 nm pulse with a variable delay between the two lasers. This yielded two transient absorption spectra:

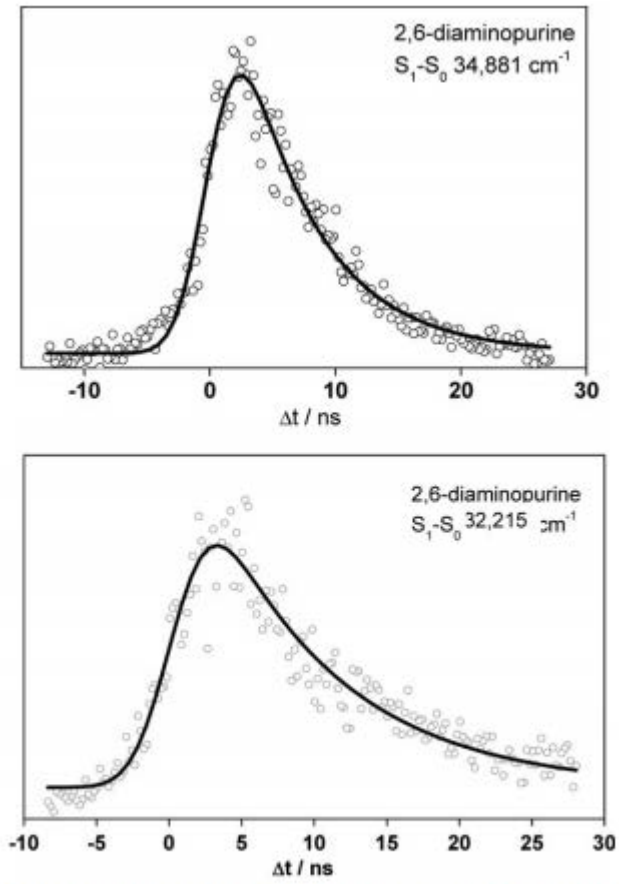
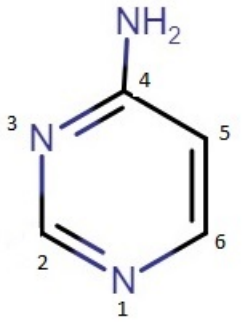


Fig. 6 Exponential decay of the ion signal of 2,6-diaminopurine in the two-color R2PI experiment. The extracted lifetime of the excited state is 8.7 ± 0.8 ns for the 7H tautomer and 6.3 ± 0.3 ns for the 9H tautomer.

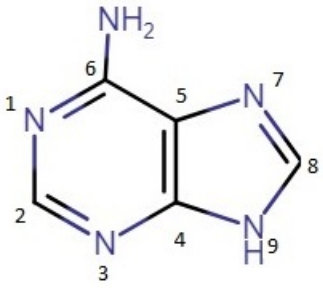
Both structures are long-lived, with the 7H being around 8.7 ns and the 9H being around 6.3 ns. These are much, much longer lifetimes than the other three molecules studied, and supports our model of the excited state transition being mediated by the N2 and C5-C6 deformations.

For ease of understanding, here is a graphical summary of the results:



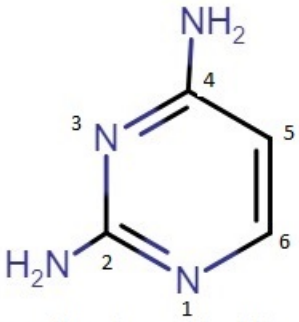
4-aminopyridmidine

400 fs (calculated)



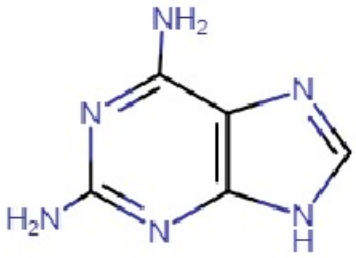
6-aminopurine

1 ps (experimentally measured)



2,4 diaminopyrimidine

1ps-1ns (experimentally bracketed)



2,6-diaminopurine

~6-8 ns (experimentally measured)

Blocking either the 2 or the 5-6 areas have small but detectable effects on the lifetime of the molecule. Blocking both has a dramatic order of magnitude effect. We believe that we are systematically “cornering” the molecule’s excited state, successively cutting off its access to the conical intersection with substitution.

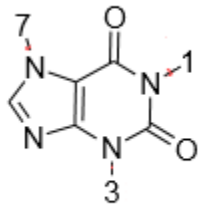
III. Pi Stacking Interactions in Non-Standard Nucleobases

In native DNA, nucleobases pi-stack with their neighbors on the same strand and hydrogen bond with their neighbors on the opposing strand. These interactions are largely enforced by the structural constraints imposed by the phospho-ribose backbone. The question of what these nucleobases do when not constrained by the backbone has relevance for prebiotic conditions, as well as for non-canonical nucleobases that exist in nature as free floating chemicals.

There is, of course, no one single answer to this question. The behaviors observed depend on the specific molecules being studied, as well as the extant conditions. In particular, microhydration and methylation have been shown in calculations to push DNA dimers away from hydrogen bonding and towards pi-stacking. Kabelac and Hobza [14] did simulations of all the possible pairings of DNA nucleobases (including non-canonical pairs, such as AA or GT) and found a general, if non-uniform, trend towards pi-stacking as the number of water molecules included in the simulation increased. Additionally, methylated versions of these nucleobases were found to favor pi-stacking more strongly, usually without even needing water at all.

We are well-suited to study the effects of methylation on the competition between hydrogen-bonding and pi-stacking. To investigate this trend, we prepared a sequence of increasingly methylated xanthine derivatives [15].

Positions 1, 3, and 7 are all hydrogen in unsubstituted xanthine. Methyl substitutions were made at these three sites and are summarized in this table:



	1	3	7
7-methylxanthine	H	H	CH_3
Theobromine	CH_3	CH_3	H
Theophylline	CH_3	H	CH_3
Caffeine	CH_3	CH_3	CH_3

A. Methods

REMPI spectra were collected for all four species using the doubled output of an organic dye laser. In two cases, a second doubled dye laser was necessary to give us access to two wavelengths at once. Once the major peaks were identified in the REMPI spectrum, we burned them out with a Nd:YAG pumped OPO/OPA IR laser system, within a range of 2800-3550 wavenumbers. The IR ion-dip spectra were compared to theoretical calculations to assign tautomeric structures. Then, dimer peak of each molecule was also burned out, and the result spectra were also compared to theoretical calculations to see what method the molecules were used to join together.

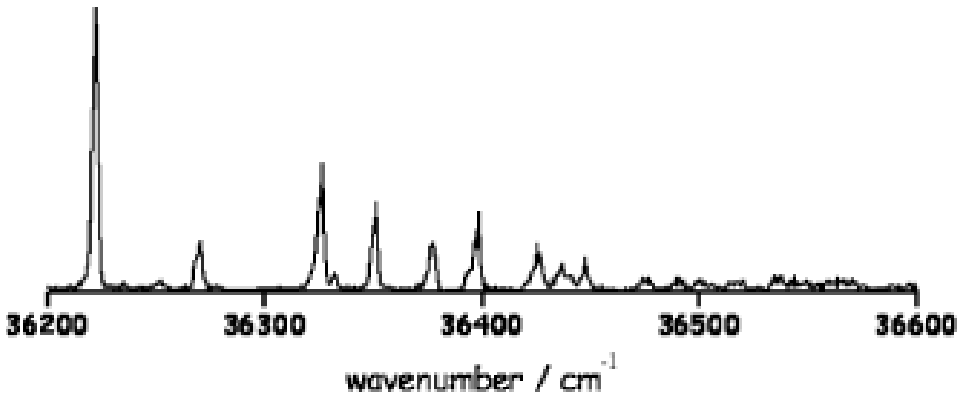
B. Calculations

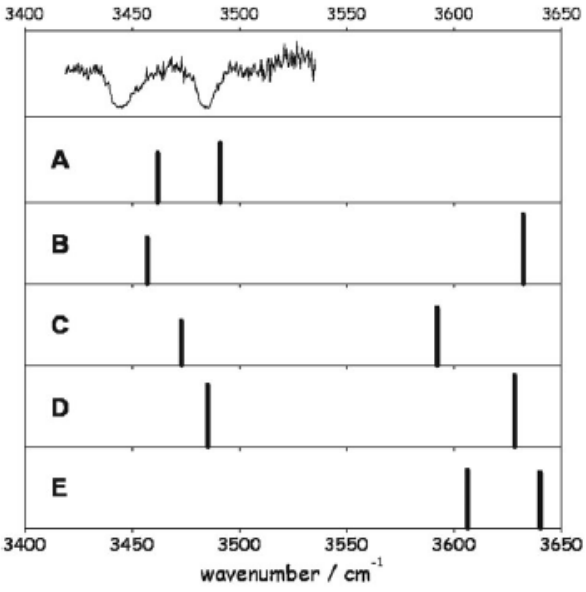
Addressing the problem of pi-stacking v H-bonding isn't even possible if we don't know what tautomers the individual monomers prefer. We used DFT, employing the popular B3LYP functional expanded across the 6-311+ G (2d,p) basis set. These calculations were used to determine the relative energies of all of the possible tautomers. A second derivative Hessian analysis was performed to ensure that the critical points found in the first step were indeed minima. The same second derivative method was then also used to predict the vibrational frequencies of the tautomers for future comparison against experimental data.

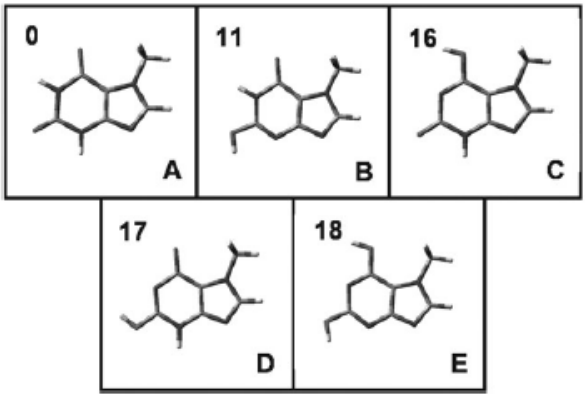
This same level of theory was then used to find the relative energies and frequencies of the dimers of 7-methylxanthine and theophylline. Theobromine and caffeine proved to be more difficult. As it turned out, their spectra were suggestive of pi-stacking. DFT commonly has trouble with structures where dispersion forces are dominant, so the MD/Q method was used instead, at the SCC-DF-TB-D level of theory. The lowest energy conformers (they were clustered at +9 kj/mol or lower) were then reoptimized using a Resolution of Identity Density Functional Theory (RI-DFT-D) using the TPSS functional with a dispersion-related augmentation (TPSS-D) and expanded across the 6-311++ G (3df,3dp) basis set.

For the sake of readability, discussion of the experimental work and theoretical matching will be done on a molecule-by-molecule basis.

7-methylxanthine: The REMPI on this molecule was collected with a relatively simple 1 color R2PI scheme. The spectrum had a strong origin just above 36200 wavenumbers. This was the highest energy origin of the four monomers studied. Since this is the least methylated species, the observed trend is that increased methylation leads to decreased origin energy.

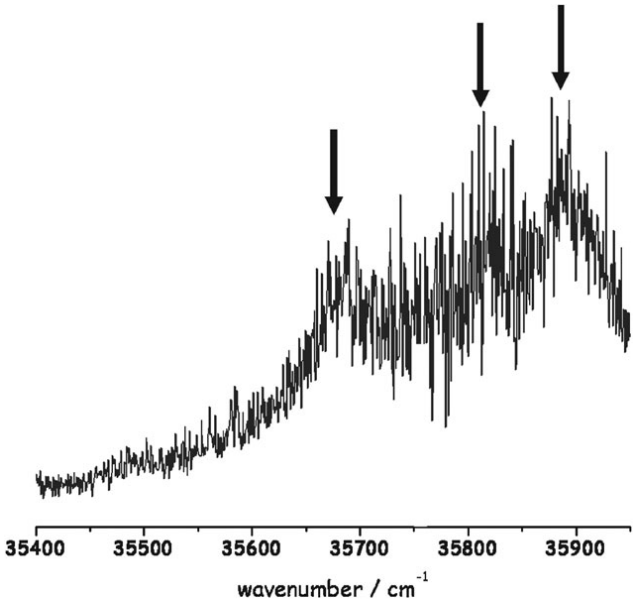


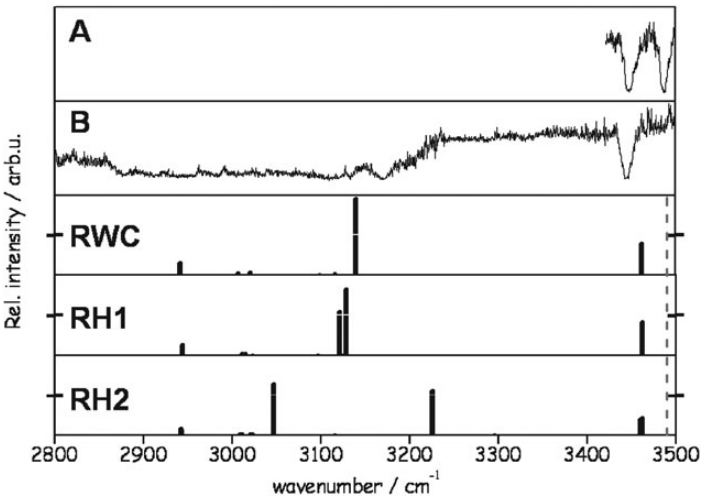




The origin peak was burned in the IR region, creating a IR-UV ion dip spectrum, which was then compared to the theoretical frequencies. Structure A, the diketo structure, not only has the lowest energy by a wide margin, but is also the only structure whose calculated frequencies are even close to the experimental data. The slight blue-shifting of the theory relative to the experiment is expected, since theory uses a harmonic oscillator approximation, which will overestimate the frequencies. 7-methyxanthine can be confidently assigned to the diketo tautomer. The two stretches seen in the spectra are from the N1H and N3H vibrations.

The mass spectrum for 7-methylxanthine contained a strong dimer peak, which was scanned with the same REMPI technique as used for the dimer. The result was a very broad, very red-shifted spectrum. The three structures marked with arrows were scanned using IR-UV holeburning. All three scans returned essentially identical spectra, suggesting strongly that only one conformer of the 7-methylxanthine dimer exists in our molecular beam.

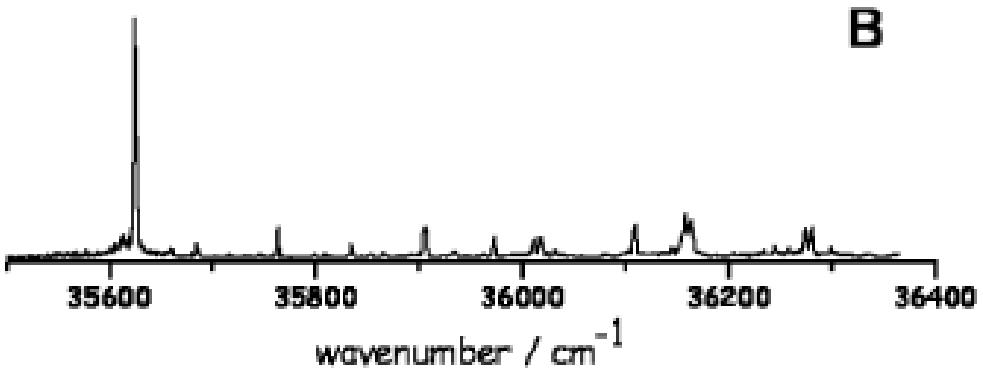




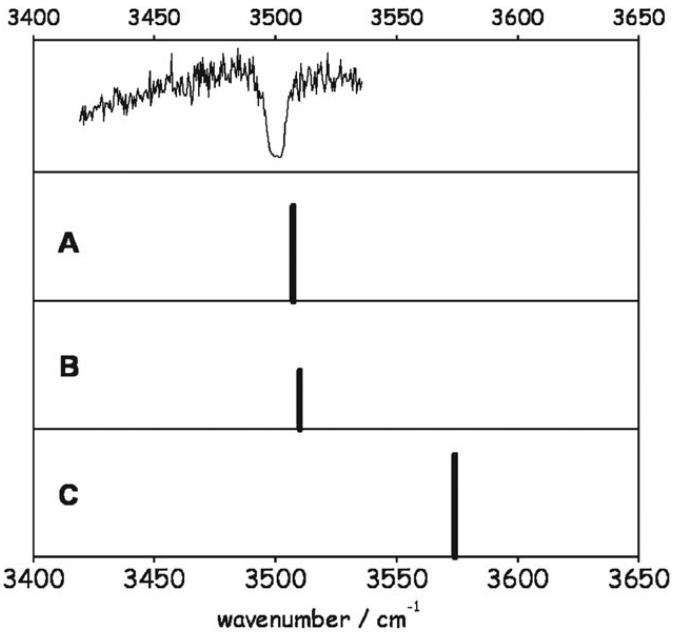
For ease of comparison, the 7-methyxanthine monomer is shown in A, and the dimer is shown in B. The N3H stretch is missing in the dimer, suggesting that the three position is engaged in a hydrogen bond. There is also a very broad structure to the red, spanning all the way from 2800 to just beyond 3200 wavenumbers. This is not a depressed baseline effect. It was spot checked in several locations by blocking the IR laser. Each time the IR was blocked, the signal jumped back up to baseline levels, demonstrating that the holeburning effect is real and that this is just a very wide structure. The three theoretical structures listed are the reverse Watson-Crick, the reverse Hoogstein 1, and the reverse Hoogstein 2. Not only were these three the lowest energy conformers found, but they are also the only ones that have the N3H position blocked by hydrogen

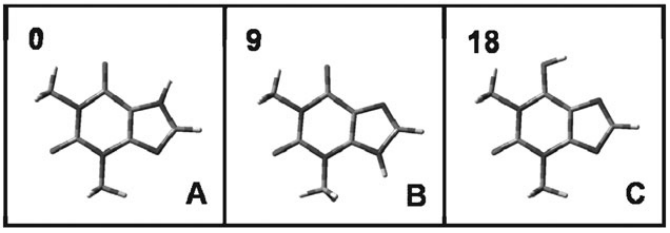
bonding. This is strong evidence that the observed species are exclusive to this set. The only significant differences in the theoretical spectra lie in the red portion, but the experimental data there is too broad to make comparisons. We can't make a definitive call here, but it is likely that the extreme broadening is due to there being multiple structures overlapping. The most likely conclusion is that we are seeing two or three of the set of RWC, RH1 and RH2.

Theophylline (1,3-dimethylxanthine):



The REMPI of theophylline was collected using the same one color R2PI scheme employed for 7-meythlxanthine. It exhibited a strong origin peak, located significantly to the red relative to the mono-methylated 7-methylxanthine.



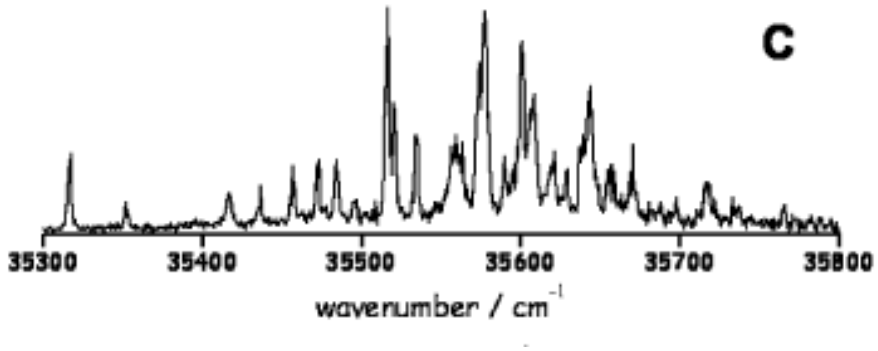


As expected, there is only one stretch. There is no need for theoretical assignment for this vibration, since the only possible stretch in the scanned energies is the unblocked N3H position.

Determining the exact tautomer is another matter. Structure C (an enol tautomer) can be discounted immediately. It has massive energy relative to the other tautomers, and the frequency is much further off. The two remaining structures (7H diketo and 9H diketo) are too close to call. They both match the observed frequency very well, and are differ by only 9 kcal/mol. This energy difference is significant, but not decisive. Therefore, we can only tentatively assign the structure found in our molecular beam to the 7H diketo, with the 9H diketo being a distant but real possibility.

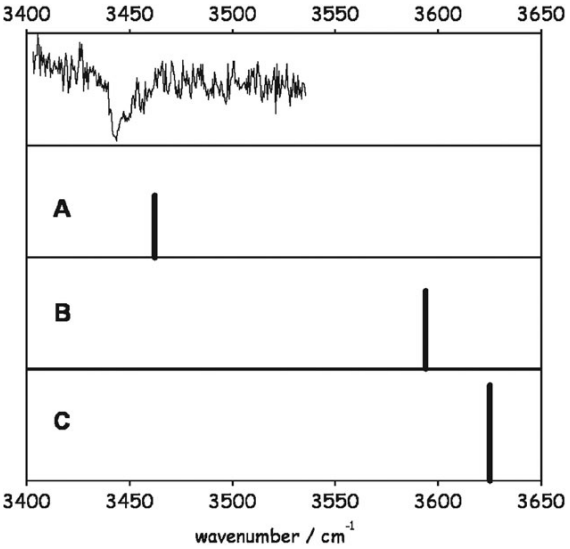
No dimers of theophylline were found in our mass spectrum. This is curious, because as we will see, the very similar theobromine *did* form observable dimers under the same conditions. This should not be taken as evidence that such dimers do not exist at all in the gas phase. There may be photophysical mechanisms specific to theophylline that inhibit detection. Poor oscillator strength, strong spectral shifting, or very short excited-state lifetimes are all possibilities.

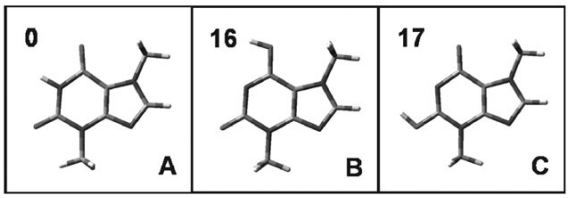
Theobromine (1,7 dimethylxanthine): Theobromine resisted initial attempts at collecting a REMPI. The one color R2PI method resulted in broad, congested spectra with no discernable features. This problem is common when there is power saturation in the ionization step. To get around this difficulty, we employed a more complicated two-color technique. By decoupling the excitation and ionization steps, we can use very high power in the ionizing wavelength without getting broadening effects due to the saturation of the excited state. This two-color scheme yield a sharply-resolved spectrum:

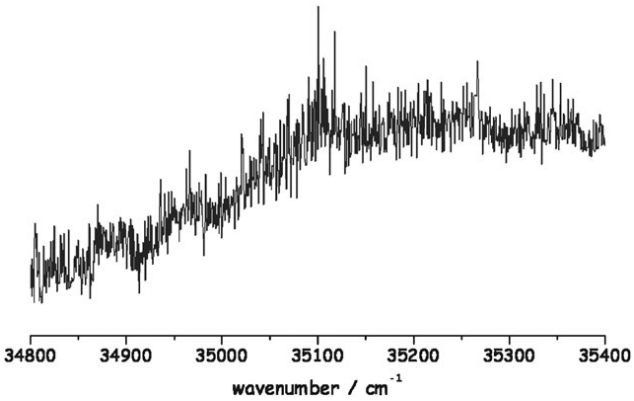


Theobromine continues the trend of methylated species being more red-shifted. The strength of the origin band is not as high, but was good enough to do hole burning on.

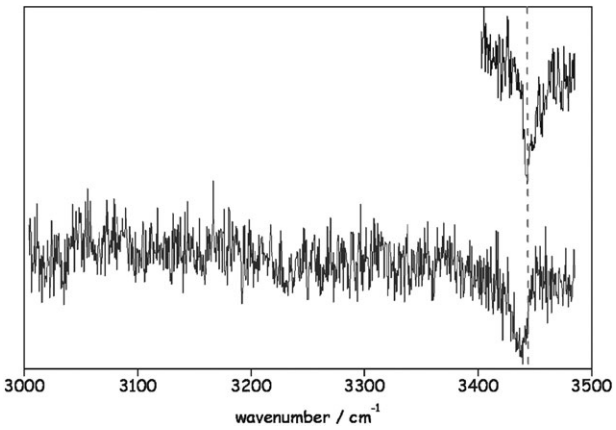
The IR spectrum can be confidently assigned to the diketo tautomer. The relative energies of the next two tautomers are very high, and the frequencies are very off. The band is assigned to the N1H vibration.



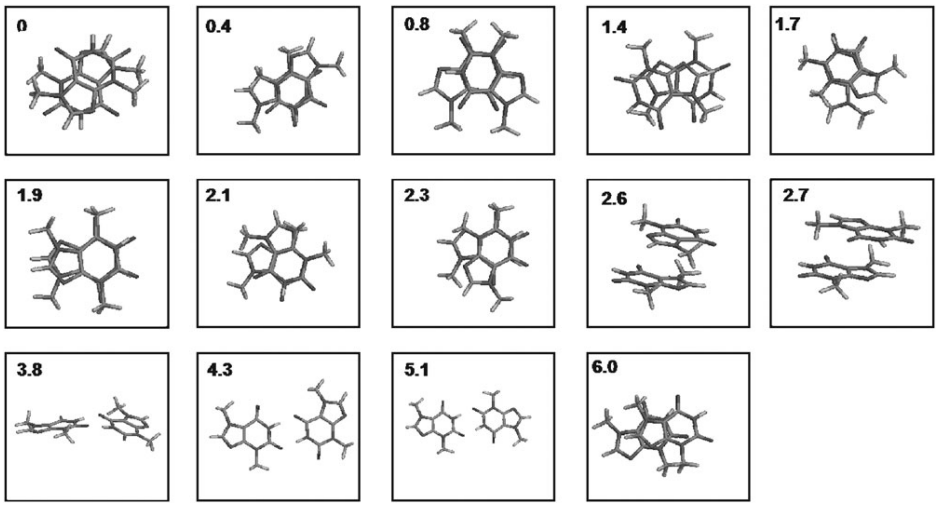




The theobromine dimer was broad and red-shifted, just like its 7-methylxanthine counterpart. We burned the spectrum at the strong structure found near 35100 wavenumbers.



The results of the burn. The monomer burn signal is superimposed above. The two spectra are functionally identical. Since the N1H stretch is the only feature in the monomer, this suggests that the N1H position is free in the dimer. This position would normally engaged in a hydrogen-bonded dimer, so such dimers can be ruled out. That only leaves a pi-stacked structure as a possibility.



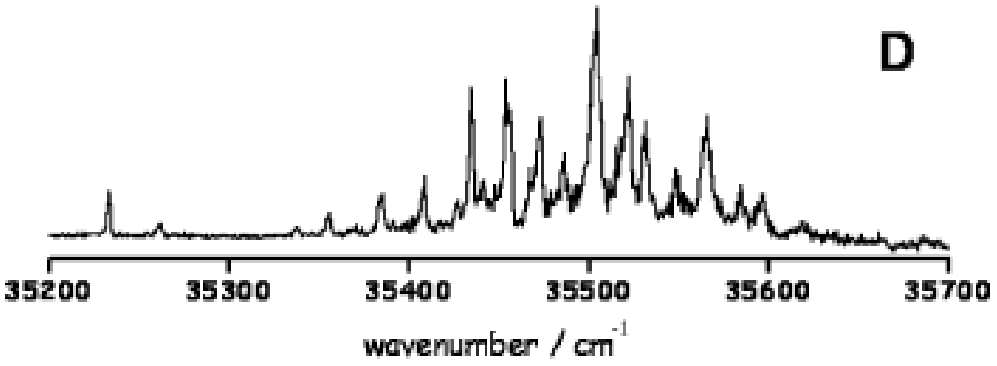
The calculated structures are mostly pi-stacked. There were three hydrogen-bonded structures (only two of which are displayed here). The hydrogen complexes all have high relative energy and do not match the experimental IR spectrum at all. Figuring out which of the pi-stacked structures is the correct one requires a closer look at the theoretical results.

	$NH_{H\text{-bonded}}$	NH_{free}	NH_{free}
str_01	—	3500	3499
str_02	—	3489[a]	3489[a]
str_03	—	3499	3498
str_04	—	3499	3498
str_05	—	3483[a]	3481[a]
str_06	—	3485	3484
str_07	—	3488	3482
str_08	—	3484[a]	3483[a]
str_09	—	3499	3495
str_10	—	3500	3496
str_11	3092	—	—
str_12	3116	—	—
str_13	3055	—	—
str_14	—	3498	3480

[a] Neighboring vibration modes are coupled.

The theoretical frequencies are shown here in tabular form. Structures 11-13 (the h-bonded ones) can be ruled out, but all of the others are close enough to be possible. Moreover, if there were several of these structures present all at once, they would still only show up as a single dip in our spectrum, due to how close they are together and how broad our ion dip is.

We can safely say that theobromine forms pi-stacks in gas phase, but the exact structure could be one, some, or all of the calculated pi-stacks.



Caffeine (1,3,7 trimethylxanthine) Caffeine posed the same difficulties as theobromine. It was necessary to use a two color R2PI method to get clearly resolved spectra. Caffeine exhibited a weak origin band. As the mostly heavily methylated molecule yet, it is no surprise that it also has the reddest origin by a large margin.

There was no need to do IR-UV holeburning on caffeine. All three hydrogen sites have been blocked by methyl groups. Since the NH stretches are the only available vibrations visible in the range of our OPO/OPA, no holeburning is expected to occur at all.

The mass spectrum of caffeine contained a very strong dimer peak. Since there are no bonding hydrogens on caffeine at all, a hydrogen bonded structure is categorically impossible. The dimer must then be the result of pi-stacking.

This was a computationally difficult structure. We run molecular dynamics at various temperatures, and found many possible structures, all of which were pi-stacked. Much higher level calculations would be needed to draw any further conclusions.

Conclusions

Our results support the emerging narrative that methylation drives xanthine nucleobase away from hydrogen bonding and towards pi-stacking. The mono-substituted 7-methylxanthine was assigned three possible hydrogen bonded complexes (possibly all three at once). The double substituted theobromine was assigned to a pi-stacked structure. The triply substituted caffeine showed strong dimer formation, all of which must be due to pi-stacking, on account of all of the

H-bonding sites being blocked. Curiously, the double substituted theophylline did not form visible dimers at all, in spite of being very closely related to theobromine.

References

[1] Merritt, K. K.; Bradley, K. M.; Hutter, D.; Matsuura, M. F.; Rowold, D. J.; Benner, S. A. Beilstein J. Org. Chem. 2014 (10): 2348–2360.

[2] Kang, H,; Lee, k. Y.; Jung, B.; Y. J.;Kim, S. K. J. Am. Chem. Soc., 2002, 124 (44): 12958–12959.

[3] Abo-Riziq A, Grace L,; Nir E,; Kabelac M,; Hobza P,; de Vries MS. Proc Natl Acad Sci U S A. 2004 102 (1):20-3.

[4] de Vries, MS; Hunziker, H. E.; Meijer, G; Wendt, h. R.; Proc. SPIE 1437, Applied Spectroscopy in Material Science, (1 April 1991)

[5] https://webbook.nist.gov/cgi/cbook.cgi?ID=C71307&Mask=400#UV-Vis-Spec accessed April 2019

[6] E. Nir, M. Müller, L.I. Grace, M.S. de Vries, Chemical Physics Letters, 2002, 355 (1–2): 59-64.

[7] Aguado, Edurne & león, Iker & Cocinero, Emilio & Lesarri, Alberto & Fernández, José & Castaño, Fernando. PCCP., (2009): 11608-16.

[8] G. Meijer; M. S. de Vries; H. E. Hunziker; H. R. Wendt Applied Physics B. 1990 51 (6):395–403.

[9] Mons, M.; Dimicoli, I.; Piuzzi, F.; Tardivel, B.; Elhanine, M. J. Phys. Chem. A., 2002, 106 (20), 5088–5094.

[10] Tetrahedron Letters Volume 8, Issue 46, 1967, Pages 4583-4586The ionization potentials of biological purines and pyrimidines ChavaLifschitzErnst D.BergmannBernardPullman

[11] Images are from a show by Bryan Brushwood, professional magician. Used with written permission http://www.shwood.com/

[12] Gengeliczki, Z.; Callahan, M. P.; Svadlenak, N.; Pongor, C. I.; Sztáray, B.; Meerts, L.; Nachtigallová, D.; Hobza, P.; Barbatti, M.; Lischka, H.;de Vries, M. S. PCCP 12(20):5375-88

[13] Barbatti, M,; Ruckenbauer, M,; Szymczak, J. J.; Aquinoa, A. J. A.; Lischka PCCP. 2008, 10, 482-494

[14] Kabeláč, M.; Hobza, P.; Chemistry A European Journal Volume 2001, 7(10) 2067-2074

[15] Callahan, M. P.; Gengeliczki, Z.; Svadlenak, N.; Valdés, H.; Hobza, P.; de Vries, M. S. PCCP 2008 10(19):2819-26