

UC Santa Barbara

UC Santa Barbara Electronic Theses and Dissertations

Title

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

Permalink

<https://escholarship.org/uc/item/3134h3fj>

Author

Svadlenak, Nathan

Publication Date

2021

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Santa Barbara

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

REMPI

A dissertation submitted in partial satisfaction of

the requirements for the degree Doctor of

Philosophy in Chemistry

by

Nathan Svadlenak

Committee in charge:

Professor Mattanjah de Vries, Chair

Professor Glake Hill

Professor Song-I Hahn

Profesor Steve Buratto

December 2021

The dissertation of Nathan Svadlenak is approved.

Song-I Han

Glake Hill

Steve Buratto

Mattanjah de Vries, Committee Chair

December 2021

Abstract

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

by

Nathan 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.

Contents

Abstract.....	5
1. Background and Motivation	6
1.1. Introduction	6
1.2. Methods.....	8
1.2.1. Laser Ionization	8
1.2.2. Gas Phase Spectroscopy	9
1.2.3. Action Mass Spectroscopy.....	12
1.2.4. Hole Burning.....	13
1.2.5. IR-UV Burning	18
1.3. Structure and Lifetime of Substituted Pyrimidines and Purines.....	20
1.4. Pi Stacking Interactions in Non-Standard Nucleobases.....	29
1.4.1. Methods.....	30
1.4.2. Calculations.....	31
1.4.3. Conclusions.....	41
2. Effect of substituents on the excited-state dynamics of the modified DNA bases 2,4-diaminopyrimidine and 2,6-diaminopurine	43
2.1. Introduction.....	44
2.2. Experimental and theoretical methods.....	48

2.2.1. Laser spectroscopy.....	48
2.2.2. Photoelectron spectroscopy	49
2.2.3. Quantum chemical calculations	50
2.3. Results and discussion	53
2.3.1. 2,4-diaminopyrimidine R2PI and IR-UV double resonance spectroscopy.....	53
2.3.2. Photoelectron spectroscopy	55
2.3.3. 2,6-diaminopurine R2PI and IR-UV double resonance spectroscopy	57
2.3.4. Excited state lifetimes	59
2.4. Excited-state analysis of 2,4-diaminopyrimidine	61
2.4.1. Stationary points in the ground and excited states.....	61
2.4.2. Minima on the crossing seam	63
2.4.3. Interpolation curves	66
2.4. Conclusions.....	69
Tables	72
3. Nucleobases as molecular fossils of prebiotic photochemistry: Excited state dynamics of C2 and C6 substituted purines	90
3. 1. Introduction.....	91
3. 2. Techniques	97
3. 3. Excited state dynamics.....	100
3. 3. 1. Purine	102

3. 3. 2. oxopurines.....	103
3. 3. 3. aminopurines.....	106
3. 3. 4. amino-oxopurines	110
3. 4. The role of C ₂ and C ₆ substitutions.....	114
3.5. Outlook	118
References.....	120
4. Non-standard base pairing and stacked structures in methyl xanthine clusters.....	129
4.1. INTRODUCTION	129
4.2. METHODS	130
4.2.1 Experimental.....	130
4.2.2. Theoretical	133
4.3. RESULTS AND DISCUSSION	133
4.3.1 Methyl xanthine monomers.....	133
4.3.2 7-Methylxanthine dimer.....	137
4.3.3 Theobromine dimer.....	141
4.3.4 Caffeine and theophylline dimers	143
4.4. CONCLUSION.....	144
References.....	146
CHAPTER 5	149
6. Low-temperature Formation of Carbonaceous Dust Grains from PAH.....	181

6.1. Introduction.....	181
6.2. Experiments	184
6.2.1. Sample Preparation.....	184
6.2.2. Laser Desorption Mass Spectrometry	188
6.3. Results.....	190
6.4. Analysis.....	197
6.4.1. Aromatic/Aliphatic Structure.....	197
6.5. Discussion	202
6.5.1. Evidence for PAH Growth within Grains.....	203
6.5.2. Effect of Structure on PAH Growth	205
6.5.3. Astrophysical Implications	210
6.6. Conclusions.....	211
References.....	213

1. Background and Motivation

1.1. Introduction

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 alphabet¹, 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 picoseconds². 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-stacked³. For the remainder of this work, “structure” refers to the complete electronic environment of the molecules being studied.

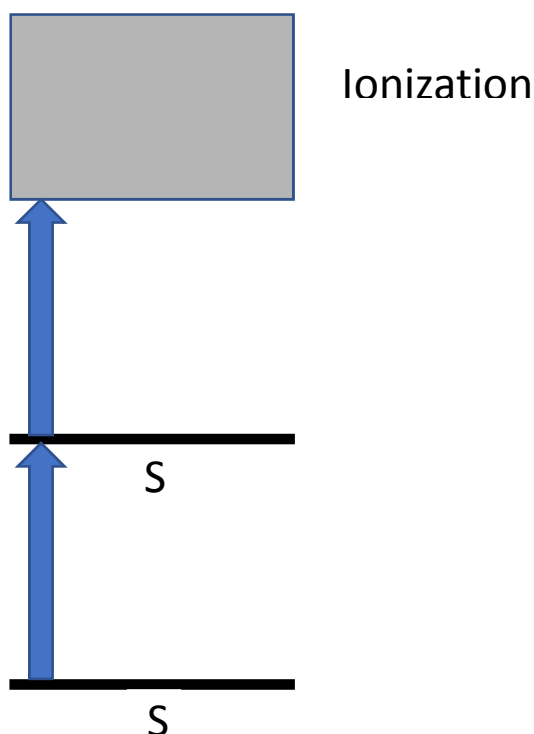
1.2. Methods

1.2.1. Laser Ionization

REMPI (Resonantly Enhanced Multi-Photon Ionization) is a spectroscopic technique for producing action mass spectroscopy spectra⁴. 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.



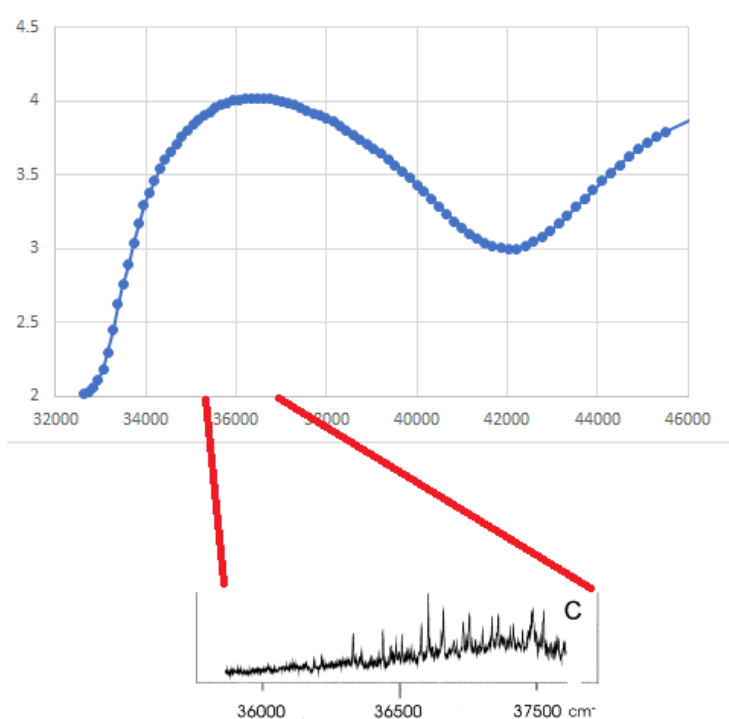
The scheme discussed here is the simplest one used in this work. There are other, more complex laser ionization techniques, but fully understanding them requires more context, and so they will be discussed later.

1.2.2. 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.

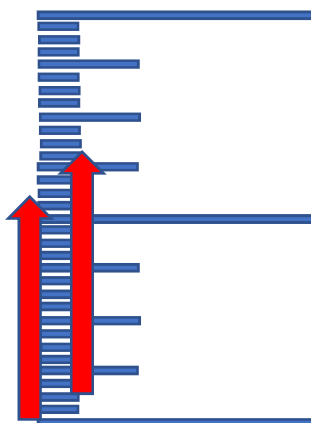


The top portion of the figure is an aqueous phase UV-vis absorption scan of cytosine, taken from the NIST Chemistry Handbook and converted to wavenumbers for convenience⁵. The inset is a gas phase REMPI scan of the same molecule⁶. Whereas the condensed phase scan is only able to show a single large peak, 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 important⁸. 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.

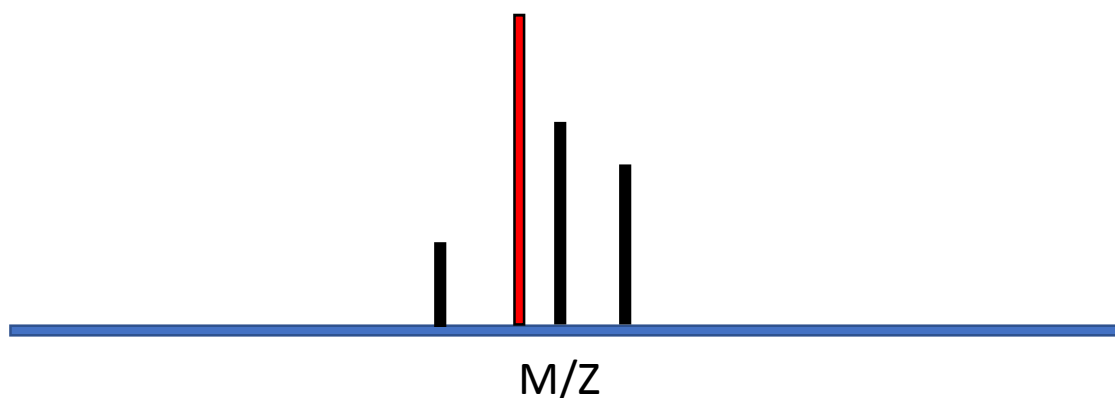
A simple diagram showing how the spectral peaks in a vibronically resolved scan correspond to vibrational levels within a single excited electronic state.

1.2.3. Action Mass Spectroscopy

This work 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.

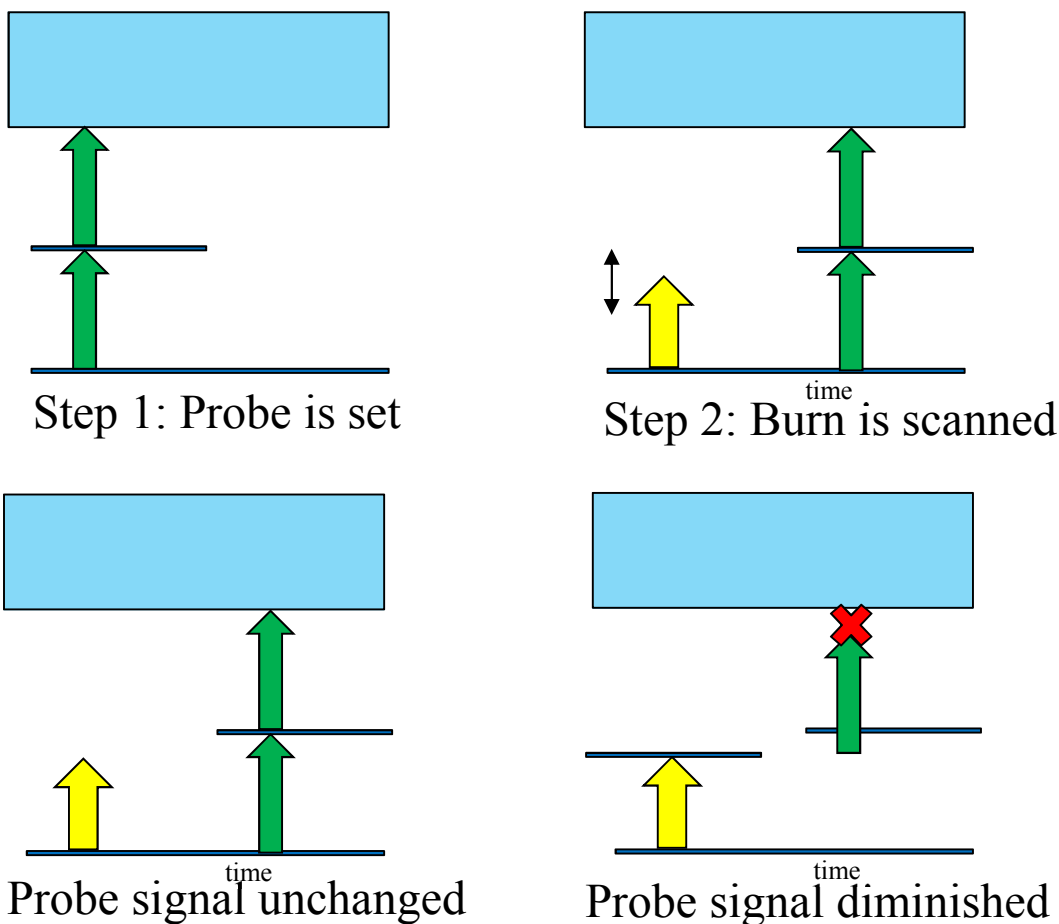
1.2.4. Hole Burning

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 through the wavelengths.

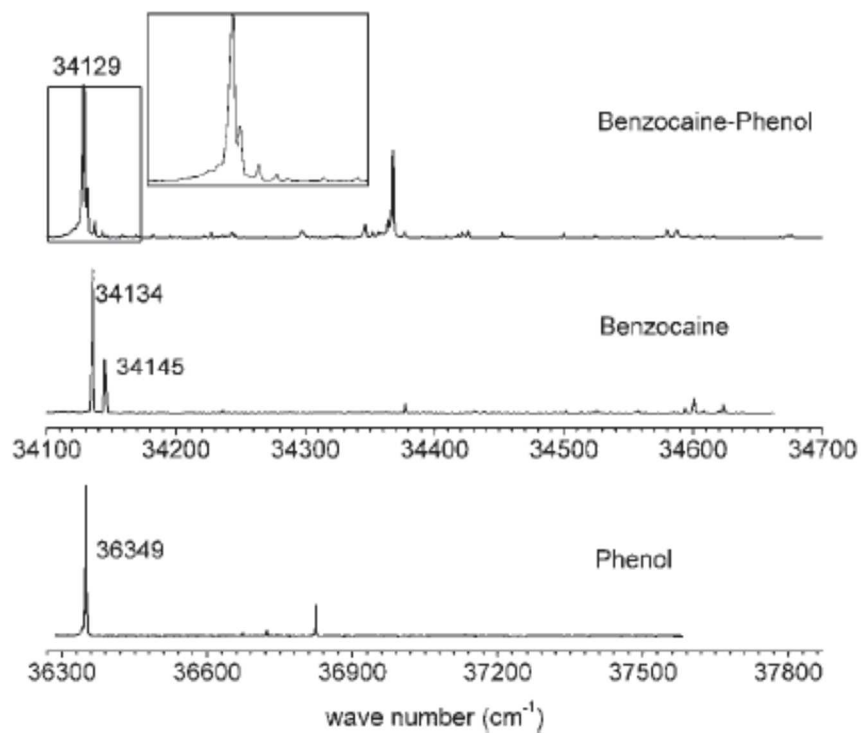
In the absence of any other effect, the probe should be generating a constant and strong REMPI signal. The burn laser alters this constant background signal by interacting with the molecule before the probe laser has a chance to reach it. If the burn laser is at a resonant wavelength, the molecule will 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.



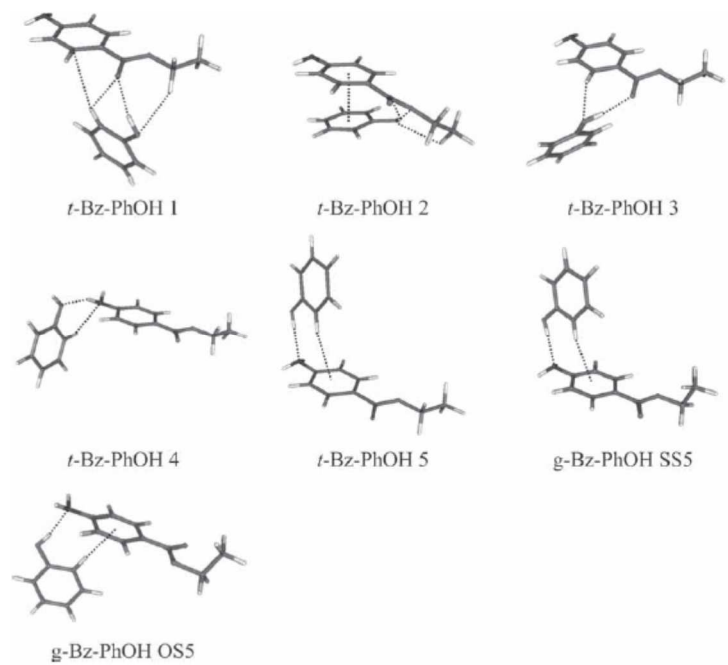
The result is an “upside down” spectrum, with an elevated baseline and sharp “holes” or “dips” burned into it. These burns correspond to the absorbance peaks of the burn laser. However, because these burns are modifications of the original fixed-wavelength probe signal, they only correspond to the structures 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

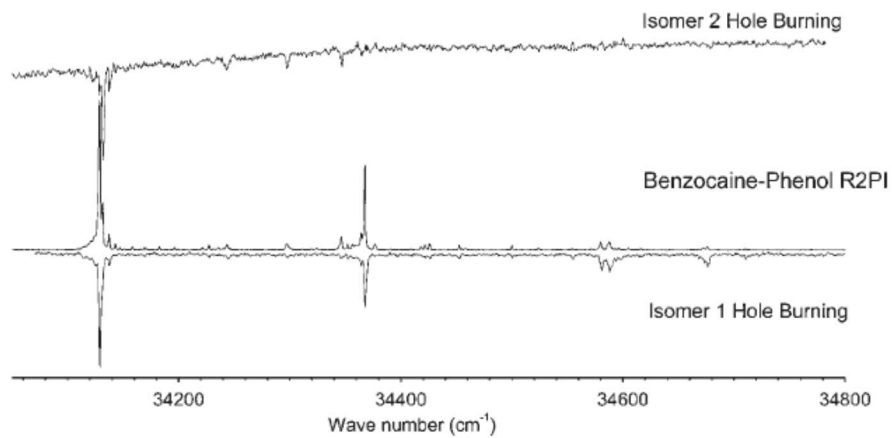
beam machine. They collected a REMPI spectrum for three different mass peaks, belonging to benzocaine, phenol, and a benzocaine+phenol cluster⁷:



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 Mons⁹ for an example of this technique in action.

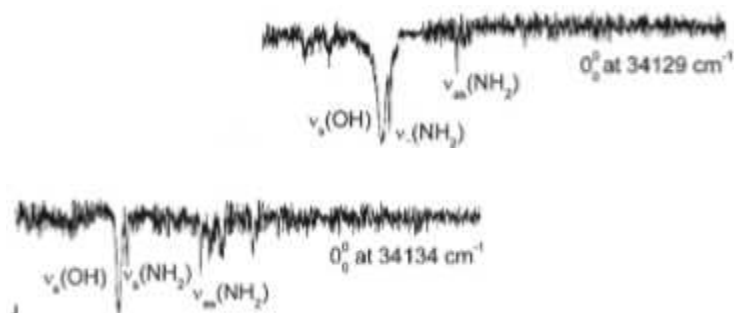
1.2.5. IR-UV Burning

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 eV¹⁰. 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 tie 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:.”



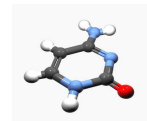
Matching these against MP2 calculations allowed them to assign these spectra to the “t Bz-PhOH 5” and “g Bz-PhOH SS5” structures, respectively.

1.3. Structure and Lifetime of Substituted Pyrimidines and Purines

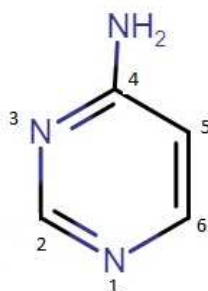
It's common to talk about excited state dynamics entirely in terms of electronic energies. It is important to remain mindful that these electronic rearrangements also cause nuclear rearrangements. An excited molecule is in a distressed and distorted state, bent out of its equilibrium geometry and flailing 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 restraint¹¹.

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 work¹² 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

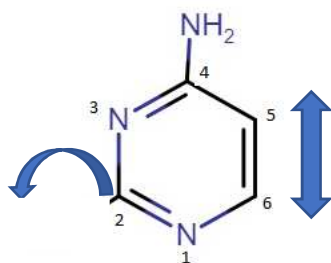


6-aminopurine

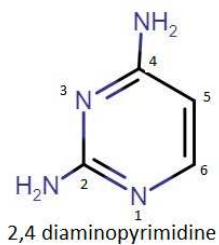
aminopurine is just 4-aminopyrimidine with the imidazole ring added. The written names might suggest that the amino group has also been moved, but inspection of the structural formulas shows that this is not the case. Unfortunately, the standard naming conventions cause the position labels to change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

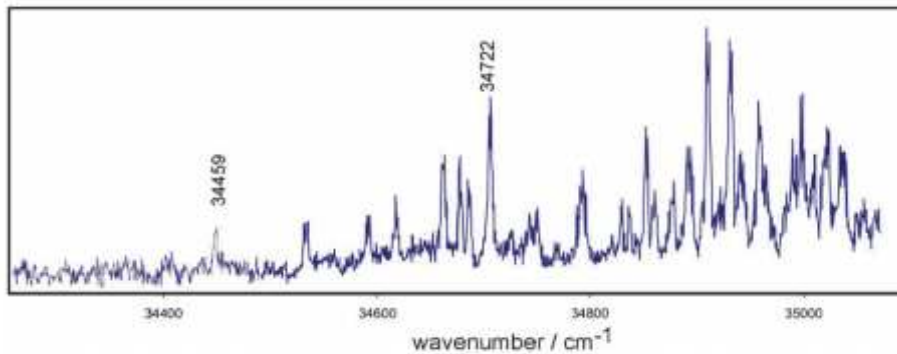
Previous theoretical work¹³ predicted that 4-aminopyrimidine reaches the ground state via a conical intersection. To reach this intersection, the molecule can either use an out-of-ring swing at the 2 position, or employ a puckering motion along the 5-6 bond. The calculated excited state lifetime was 400 fs, which is shorter than the experimental limits available to our lab at the time.



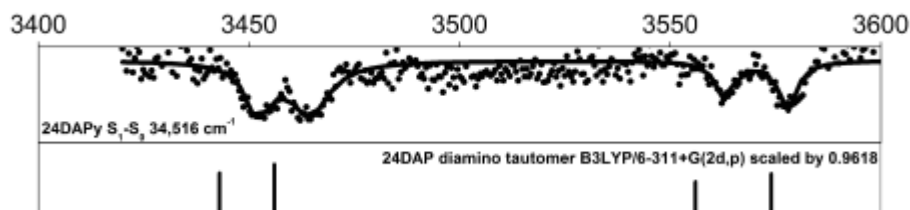
To test the C2 out-of-plane motion, we obtained a sample of 2,4-diaminopyrimidine:



A REMPI spectrum was collected:



Tautomeric determination was done with the aid of IR-UV hole burning. The peak at $34,516\text{ cm}^{-1}$ 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.

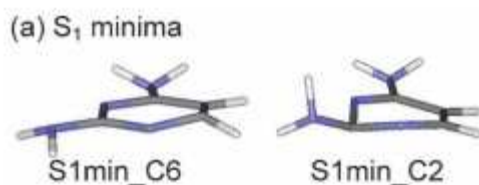
Direct measurement of the excited state lifetime was not possible, due to the short timescales involved with this molecule’s dynamics. Indeed, we were not even able to obtain a UV-UV holeburn for this molecule. However, our REMPI was of high enough resolution that the rotational envelope was visible 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 Lorentzian peak shapes, corresponding to three different linewidths:

The shortest simulated Lorentzian (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).

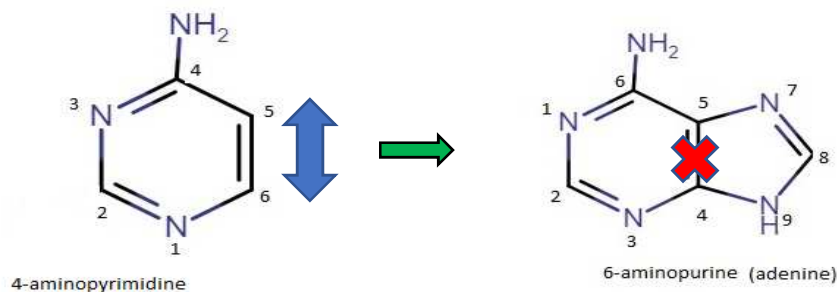
The computations found two minima at the S_1 potential energy surface. Because these are minima, they must be considered as adiabatic excited states, rather than as vertical states.



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 transition 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.

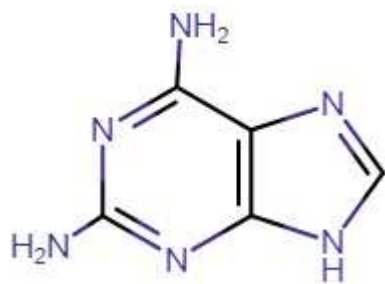
As outlined earlier, the 2 position is not the only locus implicated in the transition

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 al².

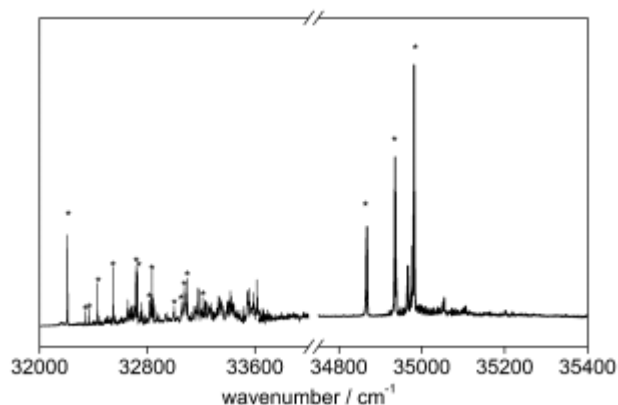


Adenine was found to have an excited state lifetime of 1 ps. This is definitely slower than 4-aminopurine, and has an unclear relationship to 2,4-diaminopyrimidine's lifetime, which was bracketed between 1 ps and 1 ns.

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.



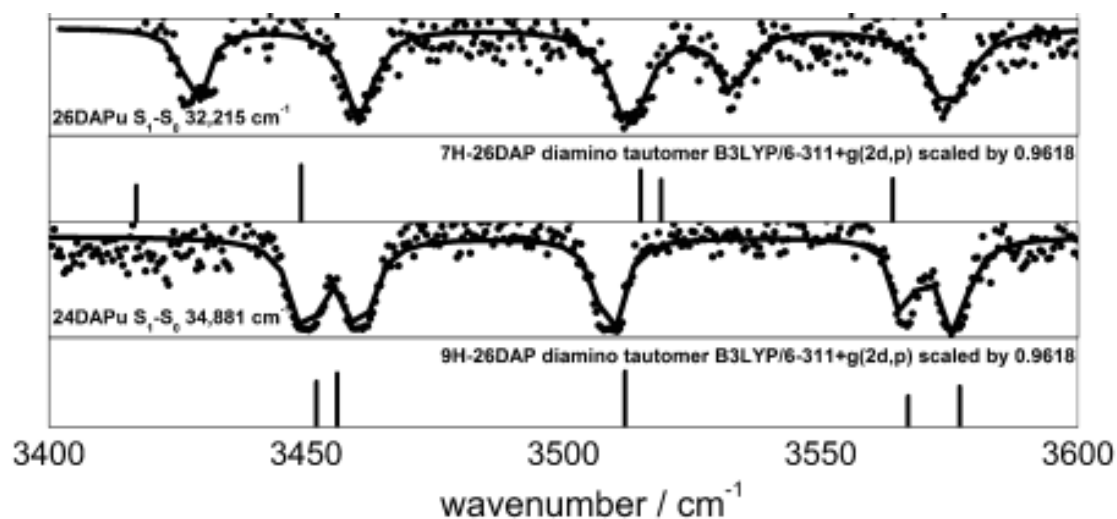
This molecule proved to be much more tractable to our techniques, yielding a clean REMPI spectrum:



A careful

reader will notice that the spectrum is split into two portions. The left portion comes from a previous study done by our group. In that study, 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 ($3,215\text{ cm}^{-1}$) and another at the beginning of the red section ($34,881\text{ cm}^{-1}$).

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 identified 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:

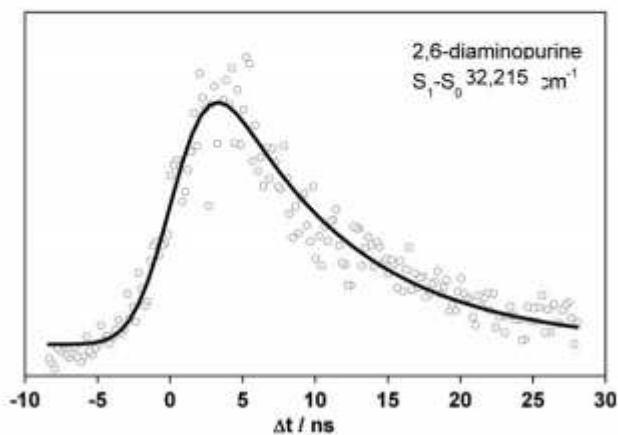
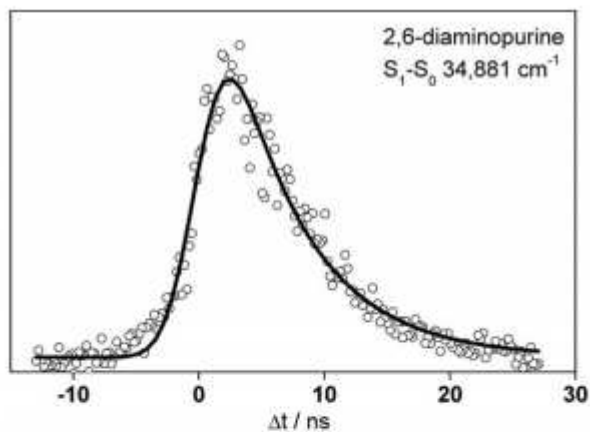
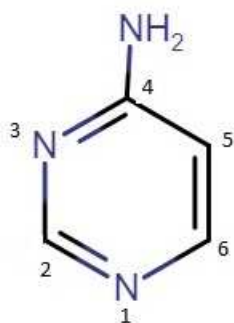


Fig. 6 Exponential decay of the ion signal of 2,6-diaminopurine in the two-color R2PI experiment. The extracted lifetime of the excited state is 8.7 ± 0.8 ns for the 7H tautomer and 6.3 ± 0.3 ns for the 9H tautomer.

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.

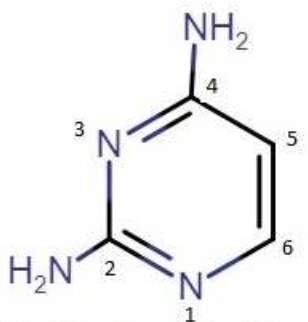
For ease of understanding, here is a graphical summary of the results:



4-aminopyridimidine
400 fs (calculated)



6-aminopurine
1 ps (experimentally measured)



2,4-diaminopyrimidine
1ps-1ns (experimentally bracketed)



2,6-diaminopurine
~6-8 ns (experimentally measured)

Hindering 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.

1.4. Pi Stacking Interactions in Non-Standard Nucleobases

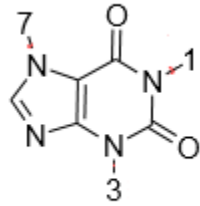
In native DNA, nucleobases pi-stack with their neighbors on the same strand and hydrogen bond with their neighbors on the opposing strand. These interactions are largely enforced

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 Hobza¹⁴ 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 derivatives¹⁵.

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	

1.4.1. 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, the dimer peak of each molecule was also burned out, and the resulting spectra were compared to theoretical calculations to see what method the molecules used to join together.

1.4.2. Calculations

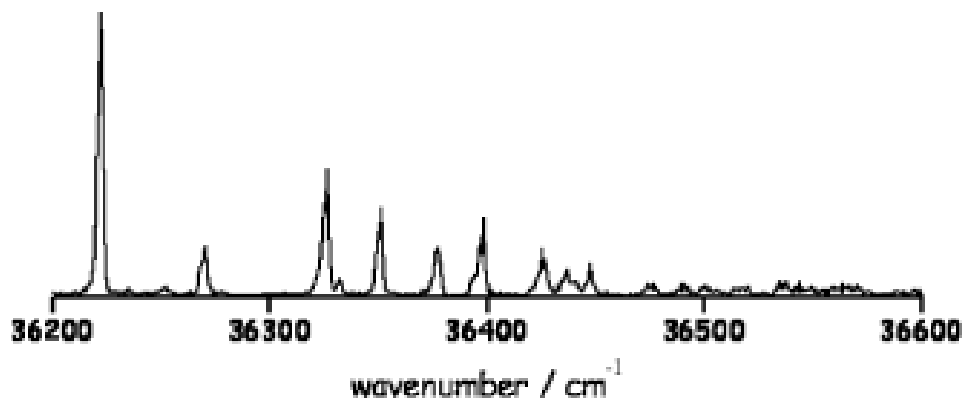
Addressing the problem of pi-stacking v H-bonding isn't even possible if we don't know what tautomers the individual monomers prefer. We used DFT, employing the popular B3LYP functional expanded across the 6-311+ G (2d,p) basis set. These calculations were used to determine the relative energies of all 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.

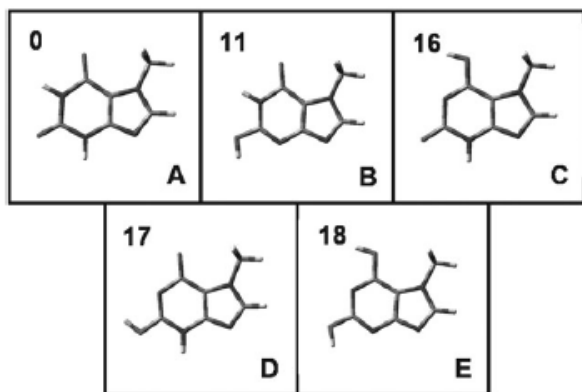
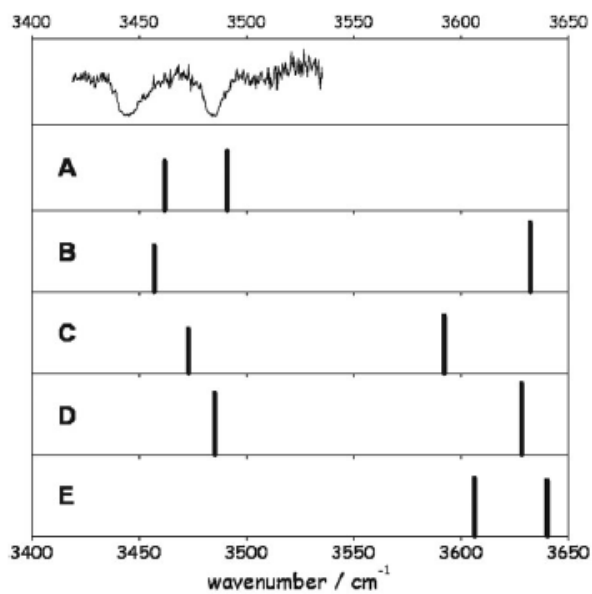
This same level of theory was then used to find the relative energies and frequencies of the dimers of 7-methylxanthine and theophylline. Theobromine and caffeine proved to be more difficult. As it turned out, their spectra were suggestive of pi-stacking. DFT commonly has trouble with structures where 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. This was the least methylated species. This turned out to be the beginning of a trend of increased methylation leading to decreased origin energy.

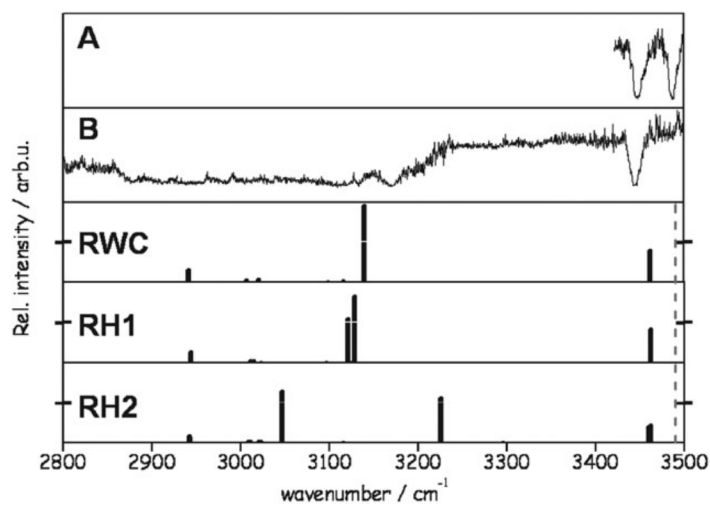
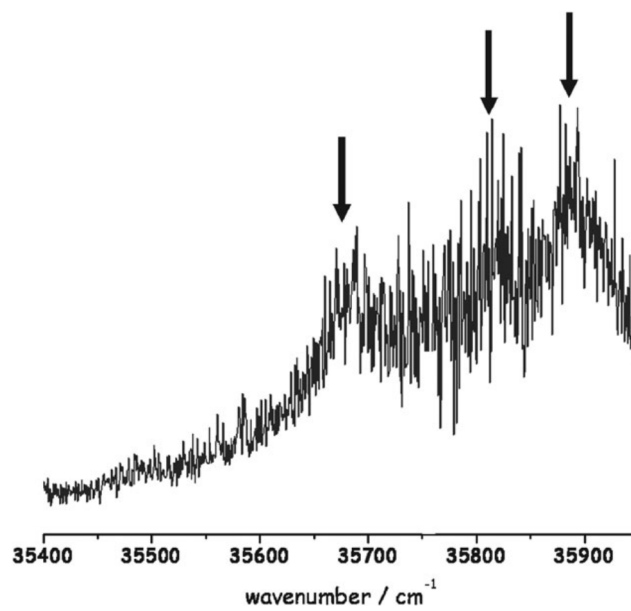




The origin peak was burned in the IR region, creating a IR-UV ion dip spectrum, which was then compared to the theoretical frequencies. Structure A, the diketo structure, not only has the lowest energy by a wide margin, but is also the only structure whose calculated frequencies are even close to the 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-methylxanthine 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.

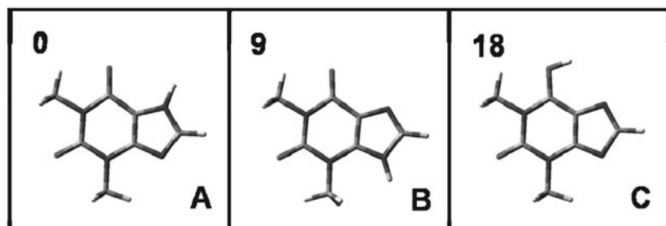
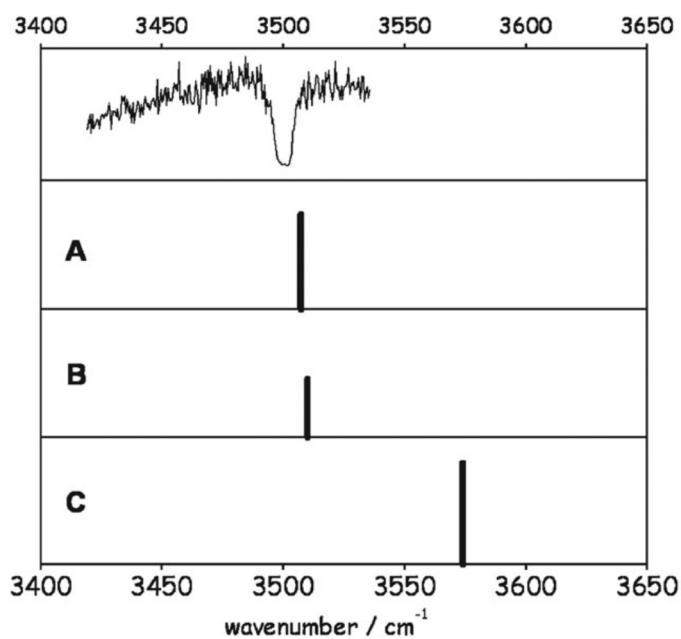


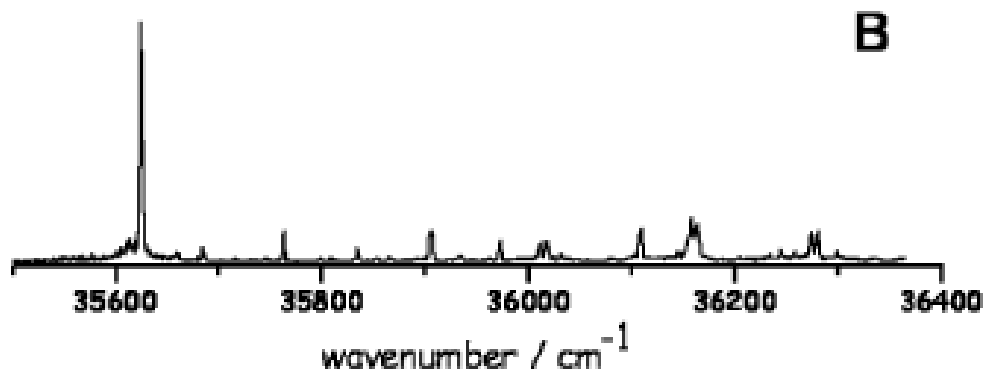
For ease of comparison, the 7-methyloxanthine monomer is shown in A, and the dimer is shown in B. The N3H stretch is missing in the dimer, suggesting that the three position is engaged in a hydrogen bond. There is also a very broad structure to the red, spanning all the way

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 Hoogsteen 1, and the reverse Hoogsteen 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):



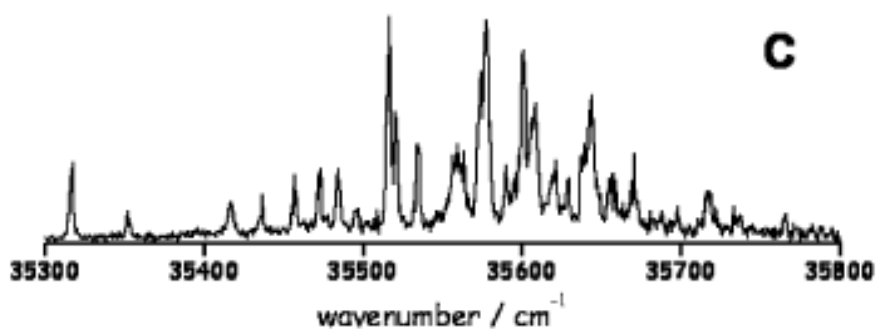


As expected, there is only one stretch. There is no need for theoretical assignment for this vibration, since the only possible stretch in The REMPI of theophylline was collected using the same one color R2PI scheme employed for 7-methylxanthine. It exhibited a strong origin peak, located significantly to the red relative to the mono-methylated 7-methylxanthine.

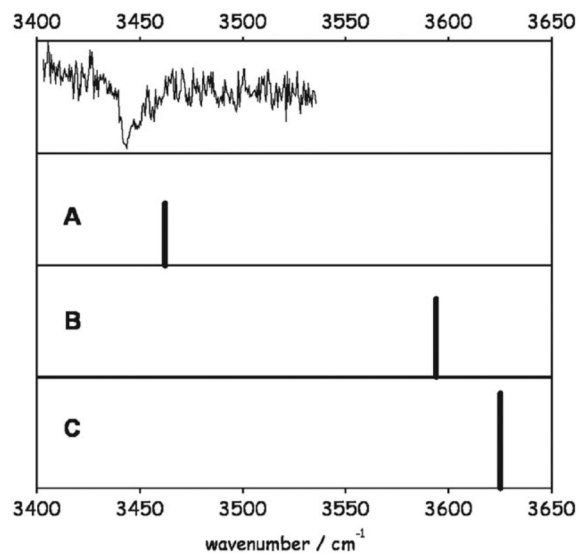
Determining the exact tautomer is another matter. Structure C (an enol tautomer) can be discounted immediately. It has massive energy relative to the other tautomers, and the frequency is much further off. The two remaining structures (7H diketo and 9H diketo) are too close to call. They both match the 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.

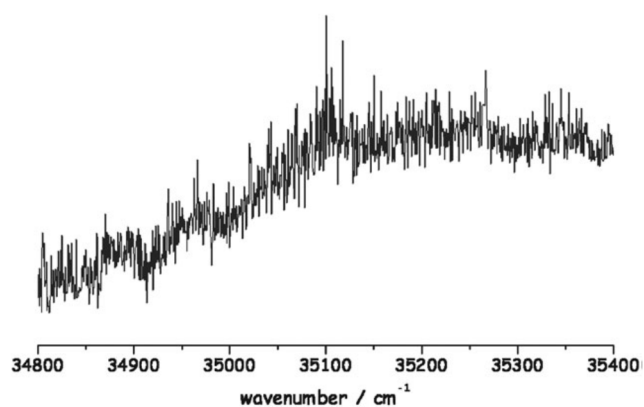
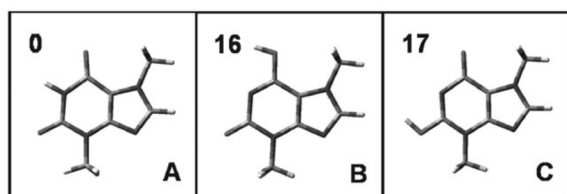
Theobromine (1,7 dimethylxanthine): Theobromine resisted initial attempts at collecting a REMPI. The one color R2PI method resulted in broad, congested spectra with no discernable features. This problem is common when there is power saturation in the ionization step. To get around this difficulty, we 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:



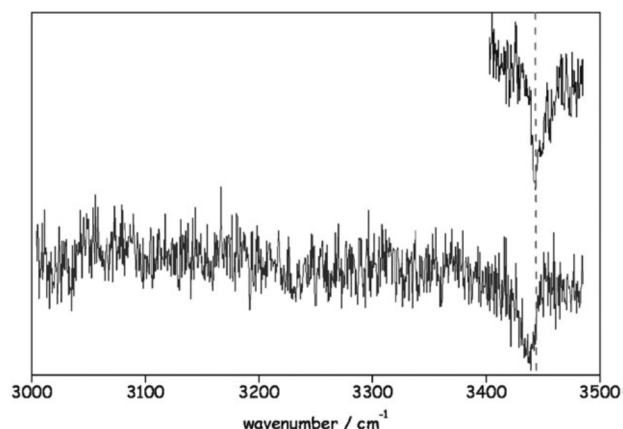
Theobromine continues the trend of methylated species being more red-shifted. The strength of the origin band is not as high, but was good enough to do hole burning on.



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.

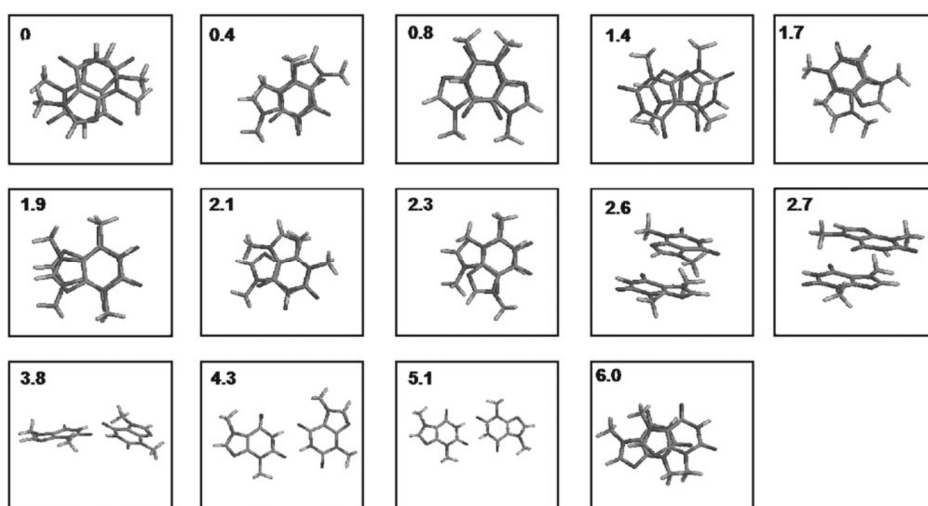


The theobromine dimer was broad and red-shifted, just like its 7-methylxanthine counterpart. We burned the spectrum at the strong structure found near 35100 wavenumbers.



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 be engaged in

a hydrogen-bonded dimer, so such dimers can be ruled out. That only leaves a pi-stacked structure as a possibility.



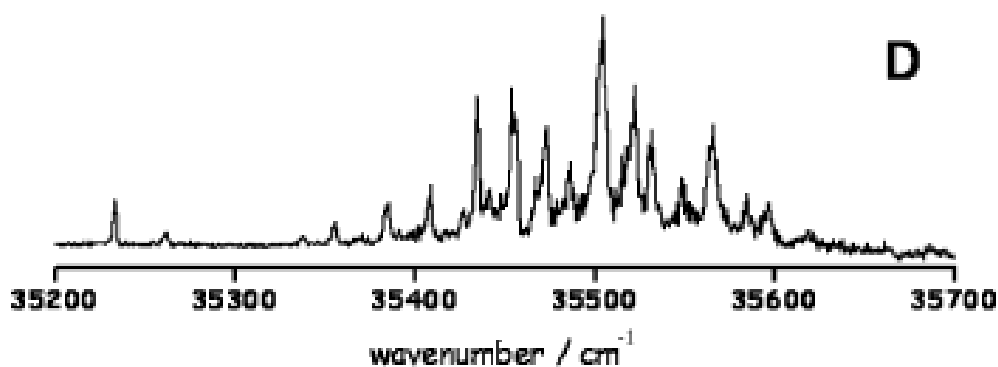
The calculated structures are mostly pi-stacked. There were three hydrogen-bonded structures (only two of which are displayed here). The hydrogen complexes all have high relative energy and do not match the experimental IR spectrum at all. Figuring out which of the pi-stacked structures is the correct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

	NH _H -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.

1.4.3. 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, E. & 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 4583-4586 The 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.; Ruckebauer, M.; Szymczak, J. J.; Aquino, 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

2. Effect of substituents on the excited-state dynamics of the modified DNA bases 2,4-diaminopyrimidine and 2,6-diaminopurine

To explore the excited state dynamics of pyrimidine derivatives we performed a combined experimental and theoretical study. We present resonant two-photon ionization (R2PI) and IR-UV double resonance spectra of 2,4-diaminopyrimidine and 2,6-diaminopurine seeded in a supersonic jet by laser desorption. For 2,4-diaminopyrimidine ($S_0 \rightarrow S_1$ 34,459 cm^{-1}) we observed only the diamino tautomer with an excited state lifetime bracketed between experimental limits of 10 ps and 1 ns. For 2,6-diaminopurine we observed two tautomers, the 9H- ($S_0 \rightarrow S_1$ 34,881 cm^{-1}) and 7H- ($S_0 \rightarrow S_1$ 32,215 cm^{-1}) diamino forms, with excited state lifetimes of 6.3 ± 0.4 ns and 8.7 ± 0.8 ns, respectively. We investigated the nature of the excited state of 2,4-diaminopyrimidine by means of multi-reference ab initio methods. The calculations of stationary points in the ground and excited states, minima on the S_0/S_1 crossing seam and connecting reaction paths show that several paths with negligible barriers exist, allowing ultrafast radiationless deactivation if excited at energies slightly higher than the band origin. The sub-nanosecond lifetime found experimentally is in good agreement with this finding.

2.1. Introduction

Gas phase laser spectroscopy provides the means to study the intrinsic properties of biologically relevant molecules in isolation. Such studies on RNA and DNA bases have revealed unique photophysical properties that are sensitive to subtle structural differences. In many cases, the biologically most relevant tautomeric form has a sub picosecond excited state lifetime, while other *tautomeric forms of the same compound* are much longer lived^{1,2}. We have even found that the Watson-Crick structure, adopted by the guanine-cytosine (GC) base pair in DNA, appears to have a much shorter excited state lifetime than other *structures of the same base pair*³. Theoretical models explain these short lifetimes by a rapid internal conversion in which the excited state (S_1) is coupled to the ground state (S_0) *via* pathways with no or a very small barrier leading to conical intersections⁴. For the longer lived structures small differences in relative energies cause the existence of barriers that lead to discrete spectra and lifetimes that can be two orders of magnitude longer. This rapid internal conversion pathway provides selected isomers with significantly enhanced photochemical stability, absent in the other longer lived structures. It is conceivable that these differences between excited state lifetimes of different bases and base-pair structures could have played a significant role in prebiotic chemistry³.

The five naturally occurring nucleic acid bases⁵⁻⁹ exhibit an ultra fast excited state relaxation in the gas phase. For DNA bases it has been shown^{1,10-24} that the energetically lowest conical intersections responsible for fast deactivation to the ground state are characterized by ring puckering modes. The most favorable ring puckering conical intersections result from the change of the HC6C5R (R = H, CH₃) dihedral angle in uracil and thymine and the HN1C2H dihedral angle in adenine. This model also explains why by contrast 2-aminopurine, in which the C2 position is

substituted by the amino group, has a long excited state lifetime and strong fluorescence²⁵⁻²⁷. Alternatively, the puckering at the C6 atom was also suggested to be responsible for adenine relaxation²⁸.

Recently, surface-hopping dynamics studies of 4-aminopyrimidine (4-APy)²⁹⁻³¹ and 9H-adenine²⁴ showed that while the former relaxes into the ground state via different conical intersections formed by puckering at the C6, N1, and C2 atoms, the latter relaxes exclusively through conical intersections formed by puckering at the C2 atom. This demonstrates the role of hindering imposed by the imidazole ring of adenine. Figure 1 shows the structures and numbering of these two compounds as well as the 2-amino substituted analogues studied in the present paper, 2,4-diaminopyrimidine (2,4-DAPy) and 2,6-diaminopurine (2,6-DAPu). Figures 1b and 1c show the lowest energy tautomers for the latter two compounds. One may expect that structural modification of the pyrimidine ring in each of these molecules affects locations and accessibility of conical intersections. In particular, amino-substitution in the C2 position should affect conical intersections associated with puckering at that site, while addition of an imidazole ring to form purine bases should affect conical intersections associated with ring deformation at the C5 and C6 sites. Thus, in 2,4-DAPy puckering at the C2 position is affected, and in 2,6-DAPu puckering at both positions is modified. Therefore, it is especially useful to compare the photochemistry of these compounds.

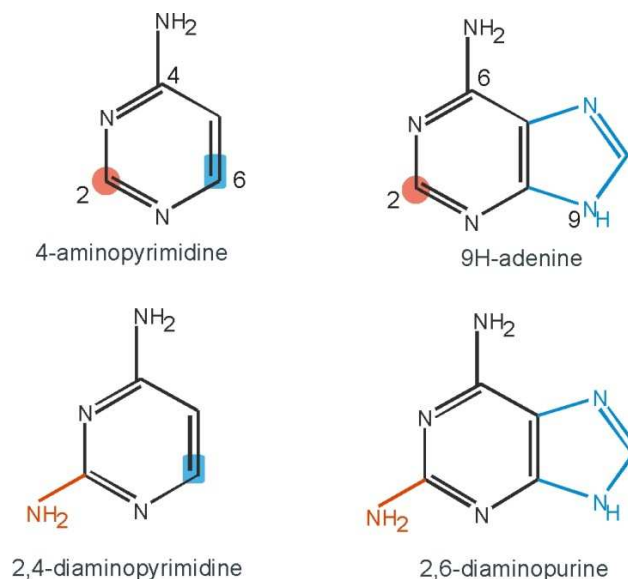


Figure 1a. Structures of the four related compounds discussed in the text. Calculations predict conical intersections for 4-Apyr due to ring deformations at positions C2 and C6, C5, and N1. Amino substitution in the C2 position affects the conical intersection associated with that position, while addition of the 5 membered ring for the purines affects the conical intersections associated with ring deformation in the other positions.

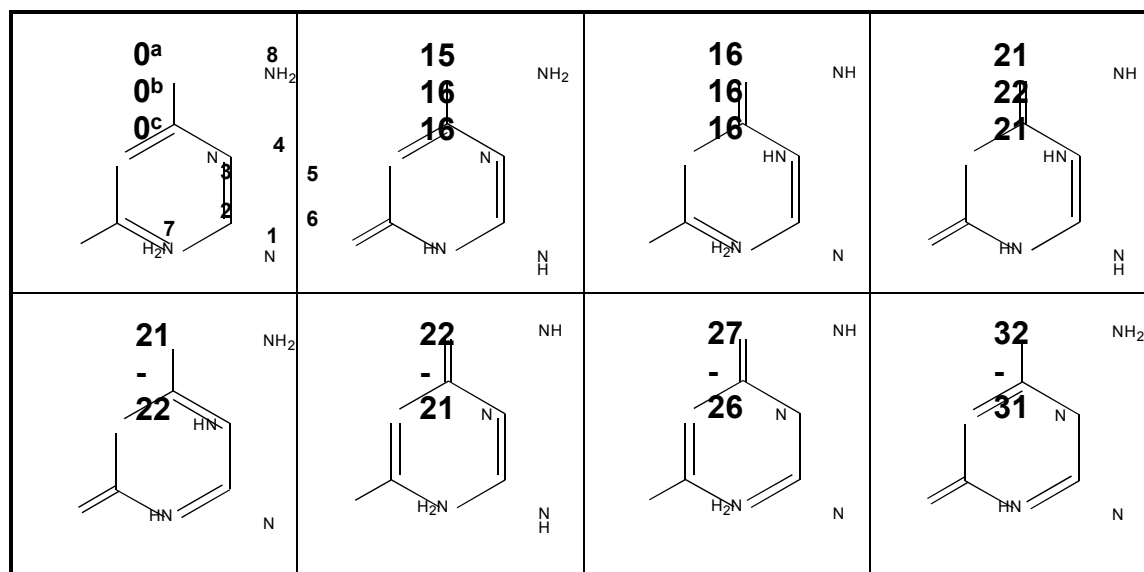


Figure 1b The tautomers and numbering scheme of 2,4-diaminopyrimidine. Relative energies obtained at the (a) B3LYP/6-311+G(2d,p) (b) MP2(FC)/6-311+G(2d,p) and (c) G3 levels are in kcal·mol⁻¹.

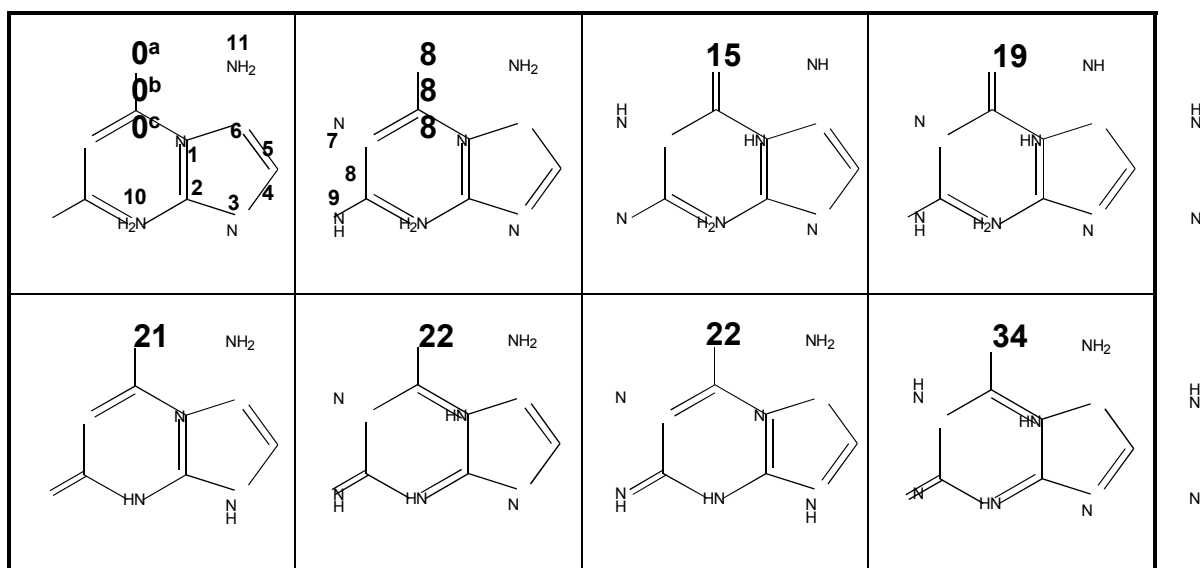


Figure 1c The tautomers and numbering scheme of 2,6-diaminopurine. Relative energies obtained at the (a) B3LYP/6-311+G(2d,p) (b) MP2(FC)/6-311+G(2d,p) and (c) G3 levels are in kcal·mol⁻¹.

In addition to these theoretical aspects there exists a further motivation for studying 2,4-DAPy and 2,6-DAPu bases. Joyce et al. have proposed that alternative nucleobases, which may form base pairs with geometries similar to Watson-Crick structures, may have played a role in constructing the first genetic code on the early Earth³². Xanthine and 2,4-DAPy can form an unnatural base pair that fits the Watson-Crick geometry and can be incorporated into RNA and DNA by polymerases³³⁻³⁶. Another base pair mimicking the Watson-Crick geometry is formed by 2,6-diaminopurine and uracil. For each of these molecules, there exists a plausible, prebiotic synthetic route³⁷⁻⁴¹.

Purines and pyrimidines can exist in a variety of different tautomeric forms, which can exhibit drastically different photophysical behavior⁴²⁻⁴⁵. Therefore, investigating the effects of modification of the hetero-aromatic ring, in particular a possible immobilization of certain parts of

this ring by its substitution at locations affecting the accessibility of conical intersections can refine our understanding of the intrinsic photostability of these alternative nucleobases^{9,29-31}.

In the series of prebiotic and alternate RNA/DNA bases, we previously studied the nucleobase xanthine⁴⁶. In this paper we report the one-color R2PI spectra of 2,4-diaminopyrimidine (2,4DAPy) and 2,6-diaminopurine (2,6DAPu) as well as their IR-UV double resonance spectra. Because pyrimidine derivatives have relatively high ionization energies unavailable by two-photon ionization⁴⁷, we measured the ionization energy of 2,4DAPy by photoelectron spectroscopy to rule out three-photon resonant ionization schemes. To explain the nature and dynamics of the excited state of 2,4DAPy, we characterized the stationary points in its ground and excited states and its S₁/S₀ conical intersections (minima on the crossing seam, MXS) as well as pathways to internal conversion by means of CASSCF, MR-CISD, and CASPT2 methods.

2.2. Experimental and theoretical methods

2,4-diaminopyrimidine and 2,6-diaminopurine were purchased from Sigma-Aldrich and used without further purification. The experimental setups have been described in detail elsewhere and only brief descriptions will be given here^{48,49}.

2.2.1. Laser spectroscopy

The desorption laser, a Nd:YAG operating at 1064 nm, is attenuated to 1 mJ cm⁻² and focused to a spot approximately 0.5 mm diameter within 2 mm in front of the nozzle orifice. We translate the sample in order to expose fresh sample to successive laser shots. The nozzle consists of a pulsed valve with a nozzle diameter of 1 mm and a backing pressure of 6 atm of argon drive

gas. The neutral molecules are skimmed and then ionized with a frequency doubled dye laser. We detect the ions in a reflectron time-of-flight mass spectrometer.

We obtain resonant two-photon ionization (R2PI) spectra by monitoring mass selected peaks while tuning the one-color, two-photon ionization wavelength. We measure UV-UV double resonance spectra with two laser pulses separated in time by 200 ns. Ionization laser intensities are approximately 3 mJ pulse⁻¹ and are strongly attenuated to avoid saturation. The first pulse serves as a “burn” pulse, which removes the ground state population and causes depletion in the ion signal of the second “probe” pulse, provided both lasers are tuned to a resonance of the same tautomer. IR-UV double resonance spectra are obtained in an analogous way with the burn laser operating in the near-IR region. IR frequencies are produced in an OPO setup (LaserVision) pumped by a Nd:YAG laser operating at its fundamental frequency. For this work, we operated within the range of 3200–3800 cm⁻¹, which encompasses NH and OH modes. Typical IR intensities in the burn region are 12 mJ pulse⁻¹ and the bandwidth is 3 cm⁻¹. Excited state lifetimes are measured in a two-color, two-photon ionization experiment. The pump and probe photons are separated by a time delay of 100 ps to 100 ns, and the ion signal is recorded as a function of the time delay. The lifetime of the excited state is given as the time constant of the exponential decay of the ion signal.

2.2.2. Photoelectron spectroscopy

He-I photoelectron spectra were recorded on a custom-built ATOMKI ESA-32 instrument⁴⁹. The spectrometer is equipped with a Leybold-Heraeus UVS 10/35 high-intensity helium discharge photon source. The sample was introduced into the ionization chamber via solid inlet probe at an elevated temperature of 260°C. The spectra were calibrated against Ar ²P_{3/2} and ²P_{1/2}

peaks. The maximum error in the ionization energies is estimated to be less than 0.04 eV, and the energy resolution of the He-I spectra was better than 40 meV (FWHM) as determined from Ar 3p ionizations. In the extraction of the vertical ionization potentials, Shirley background and PseudoVoigt peaks were fitted on the experimental spectra. The adiabatic ionization potential can be estimated as the onset of the ionization and has a somewhat higher uncertainty than the vertical ionization potentials.

2.2.3. Quantum chemical calculations

2.2.3.1. Identifying the tautomers

To assign the IR-UV double resonance spectra, we computed the structures of the possible tautomers using density functional theory (DFT) and *ab initio* methods. We combined Becke's three-parameter hybrid functional⁵⁰ with the Lee-Yang-Parr exchange correlation functional^{51,52} (B3LYP) and the 6-311+G(2d,p) basis set^{53,54}. We used the same basis set in frozen core second order Møller-Plesset perturbation (MP2) calculations⁵⁵⁻⁶⁰. We verified the equilibrium structures by the absence of imaginary vibrational frequencies. The calculated vibrational frequencies were scaled by 0.9618⁶¹ (DFT) or 0.9496⁶² (MP2) to account for electronic structure method deficiency and anharmonicity. To establish the relative energies of the tautomers, we also applied the Gaussian-3 (G3) composite method⁶³.

2.2.3.2. Ionization energies

To assign the photoelectron spectrum, we calculated ionization energies for 2,4-diaminopyrimidine, using the B3LYP functional and 6-311+G(2d,p) basis set. We calculated the adiabatic ionization potential as the energy difference between the equilibrium geometries of the ionic state and the ground state neutral molecule, estimating the first vertical ionization energy as the difference between the energies of the ionic state and the neutral molecule at the equilibrium

geometry of the latter. Then, at the equilibrium geometry of the ion, time-dependent DFT (TD-DFT) we performed calculations to compute the vertical excitation energies of the ion ⁶⁴⁻⁶⁶. Vertical ionization energies are then the sum of the adiabatic ionization energies and vertical excitation energies of the ion. We also performed outer Valence Green's Function (OVGF) calculations with the 6-311+G(2d,p) basis set at the equilibrium geometries of the neutral tautomers obtained in the DFT calculations ⁶⁷⁻⁷⁴. All calculations were carried out using the Gaussian 03 Rev. C. 02. quantum code package ⁷⁵.

2.2.3.3. Excited state investigations of 2,4-diaminopyrimidine

We performed the calculations using the complete active space self-consistent field (CASCCF) and multi-reference interaction (MR-CI) methods for 2,4-DAPy. In the MR-CI approach, single and double excitations from the CI reference space are included (MR-CISD) and generalized interacting space restrictions are adopted ⁷⁶. We constructed the orbital space for CASSCF wavefunctions using 14 electrons in 10 orbitals, i.e. composed of eight π orbitals and two lone pairs located on the ring nitrogen atoms. We used a state-averaging procedure using three states (SA-3) at the CASSCF level throughout the calculations. Based on the CAS (14,10) we constructed the MR-CISD reference space by moving orbitals with natural occupation larger than 0.9 and smaller than 0.1 to the doubly occupied and virtual spaces, respectively, resulting in a MR-CI reference space composed of six electrons in five orbitals (MR-CISD(6,5)). All single and double excitations from this reference space were allowed; all core orbitals were frozen. To reduce the computational cost of the calculations, up to a total number of 17 doubly occupied orbitals were frozen in MRCI calculations of the reaction paths. We tested this procedure against full calculations for selected typical examples, taking into account the size-consistency effects by means of Pople's correction method, indicated by +Q ⁷⁷. We used the 6-31G** basis set throughout

the calculations^{78,79}. For comparison and to verify the reliability of the applied basis set we also applied the complete active space self-consistent-field second-order perturbation theory (CASPT2) method^{80,81} with the same reference CAS space and resolution-of-identity coupled cluster to the second-order (RICC2)^{82,83} method.

We determined the minima on the ground and excited S_1 surfaces and on the seam of conical intersections (MXS) by the CASSCF(14,10) method. We constructed reaction paths between structures of the S_1 minima and the MXSs using the method of linear interpolation of internal coordinates (LIIC). The CASSCF and MR-CISD energies of relevant points of the reaction paths were plotted as a function of mass-weighted distances between each point of the path and the S_1 minimum. We also determined the structures of the transition states for selected reaction paths using the CASSCF(14,10) method. We confirmed the character of the stationary point by Hessian calculations within a selected space of internal coordinates relevant for the ring puckering modes.

We simulated the absorption spectrum of 2,4-DAPy employing the RICC2 and CASSCF methods, using the Gaussian broadening method described by Barbatti et al.⁸⁴. We used the same active space and number of states in the average procedure as specified above. We generated five hundreds points by a Wigner distribution in the ground vibrational state of the ground electronic state, computed based on a Wigner distribution for the ground vibrational and electronic state, taking each nuclear degree of freedom within the harmonic approximation. The Wigner distribution composed by 500 different geometries was projected on the excited state by multiplying it by the Einstein coefficient B computed for each point. Was used a phenomenological

broadening of 0.05 eV.

We performed the optimization of conical intersections using the analytic gradient and non-adiabatic coupling vectors⁸⁵⁻⁸⁹ available in the COLUMBUS program system⁹⁰⁻⁹². We performed the CASPT2 calculations with the MOLCAS program package⁹³⁻⁹⁵ and RICC2 computations were performed using the Turbomole program system⁹⁶. The absorption spectrum was simulated using the NEWTON-X program package^{84,97}.

2.3. Results and discussion

2.3.1. 2,4-diaminopyrimidine R2PI and IR-UV double resonance spectroscopy

Figure 2 shows the R2PI spectrum of 2,4-diaminopyrimidine in the wavelength range of 34,350-35,100 cm⁻¹. In a one-color experiment, we might not be able to observe the S₁←S₀ origin transition if the ionization potential of the molecule is too high. Therefore, we determined the ionization potential by photoelectron spectroscopy, as detailed below. For 2,4-diaminopyrimidine we found a relatively low vertical ionization potential of 8.30 ± 0.05 eV and an adiabatic ionization potential of 7.86 ± 0.05 eV, which corresponds to 33,471 cm⁻¹ and 31,698 cm⁻¹ photon wave numbers, respectively, in terms of the one-color experiments. We, therefore, tentatively assign the red-most peak at 34,459 cm⁻¹ as the S₁←S₀ origin transition. If this assignment is correct, the intensity distribution in the spectrum suggests significant geometric deformation of the excited state relative to the ground state.

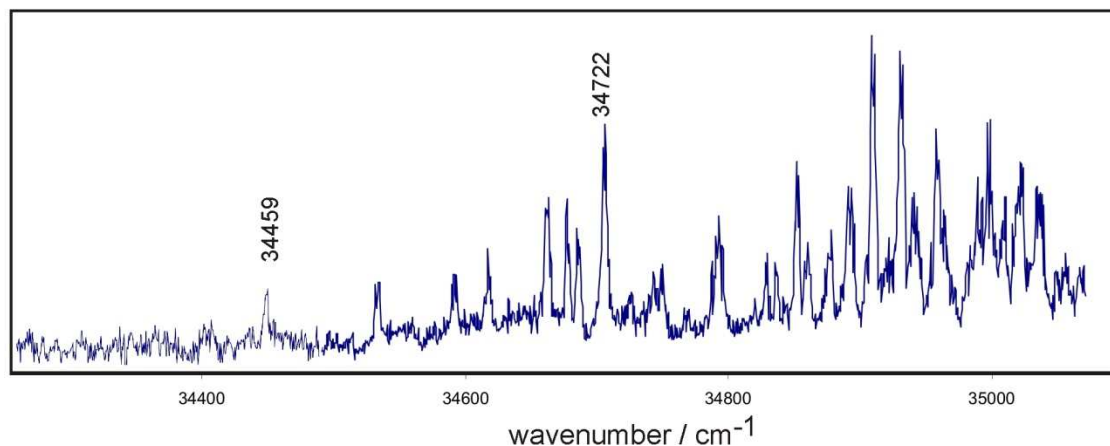


Figure 2. R2PI spectrum of 2,4-diaminopyrimidine. The $S_1 \leftarrow S_0(\pi\pi^*)$ origin is at $34,459 \text{ cm}^{-1}$. The bottom trace shows the depletion of ion through the entire range of UV wavelength when the IR burn pulse was set to 3578 cm^{-1} , indicating that we observed only one tautomer in the jet. Data were not corrected for variations in laser intensities.

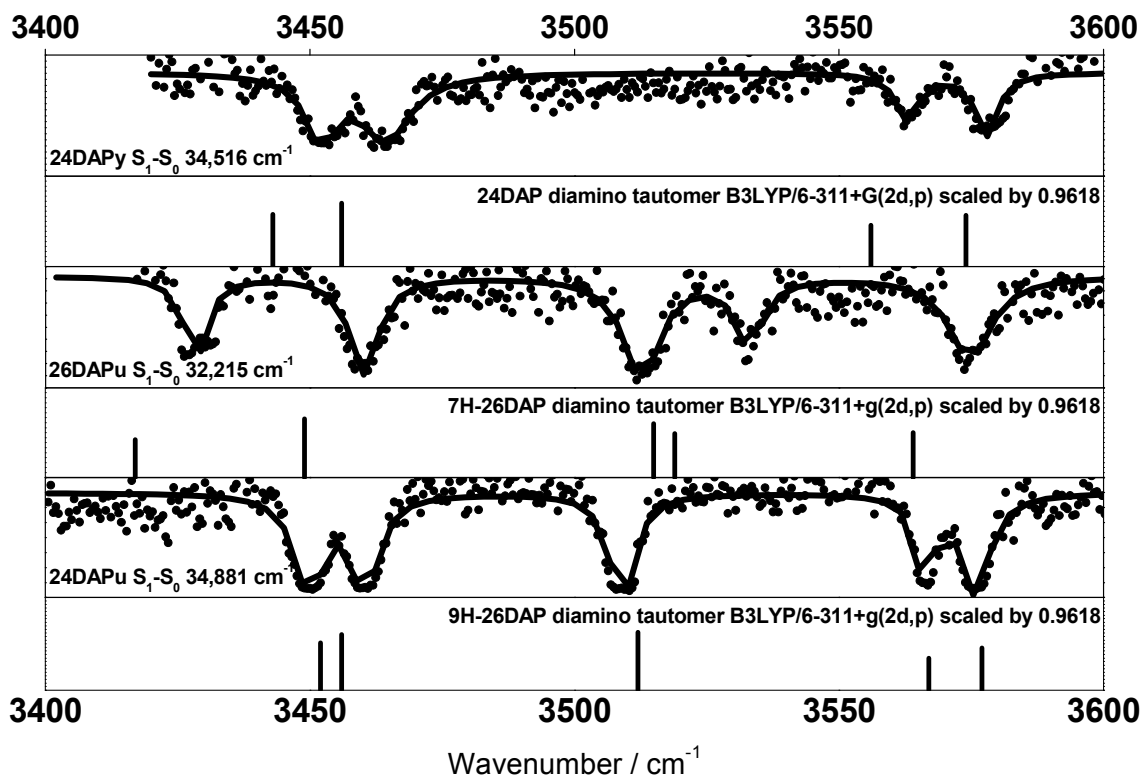


Figure 3 The IR-UV double resonance spectra of 2,4-diaminopyrimidine and 2,6-diaminopurine in the N-H stretch IR wavelength range. The stick spectra represent the best matching calculated vibrational frequencies at the B3LYP/6-311+G(2d,p) level. A scaling factor of 0.9618 was applied.

The R2PI spectrum consists of a single tautomer, based on IR-UV double resonance experiments at different UV probe wavelengths, as detailed below. Figure 3 shows the IR spectrum, obtained at a probe frequency of $34,516\text{ cm}^{-1}$. The spectrum shows four distinct peaks in the frequency range of $3,400\text{-}3,600\text{ cm}^{-1}$. By fitting the experimental data with a sum of Lorentzian curves we extracted vibrational frequencies of 3452 , 3464 , 3563 , and 3578 cm^{-1} . This indicates the presence of four free N-H stretching modes in the molecule and suggests that we observe the diamino form (see Fig. 1b). We computed the relative energies of all tautomers at the B3LYP/6-311+G(2d,p), MP2(FC)/6-311+G(2d,p) and G3 levels. At every level the di-amino form was the lowest energy tautomer with the second most stable tautomer at a relative energy 15 kcal/mol higher than the diamino form. We calculated the vibrational frequencies for all tautomers at the DFT and MP2 levels. The di-amino tautomer provides an excellent match between the calculated and the experimental frequencies. Table 1 summarizes the vibrational frequencies.

To determine that *only* the diamino tautomer is present in the R2PI spectrum, we set the IR frequency to 3578 cm^{-1} and scanned the R2PI spectrum again. The IR band at 3578 cm^{-1} is the most diagnostic band for the diamino tautomer compared with the theoretically calculated IR bands for the other tautomers. The R2PI spectrum obtained while hole burning at this IR frequency showed a complete depletion in the ion signal for the entire spectral range when the IR was on. This result implies that all peaks in the R2PI spectrum are due to the diamino tautomer.

2.3.2. Photoelectron spectroscopy

Figure 4 shows the photoelectron spectrum of 2,4-diaminopyrimidine. The adiabatic ionization potential, derived from the onset, is $7.81 \pm 0.05\text{ eV}$. We derived the vertical ionization potentials by fitting the sum of PseudoVoigt curves on the observed bands. Based on the intensities and the widths of the bands, we find vertical ionization potentials corresponding to six ionic states.

An additional experimentally unresolved broad band starts at an onset of 13 eV. It is implausible that precise ionization energies can be extracted from the spectrum above 13 eV, and we only note that the first maximum of this range is at 13.51 eV.

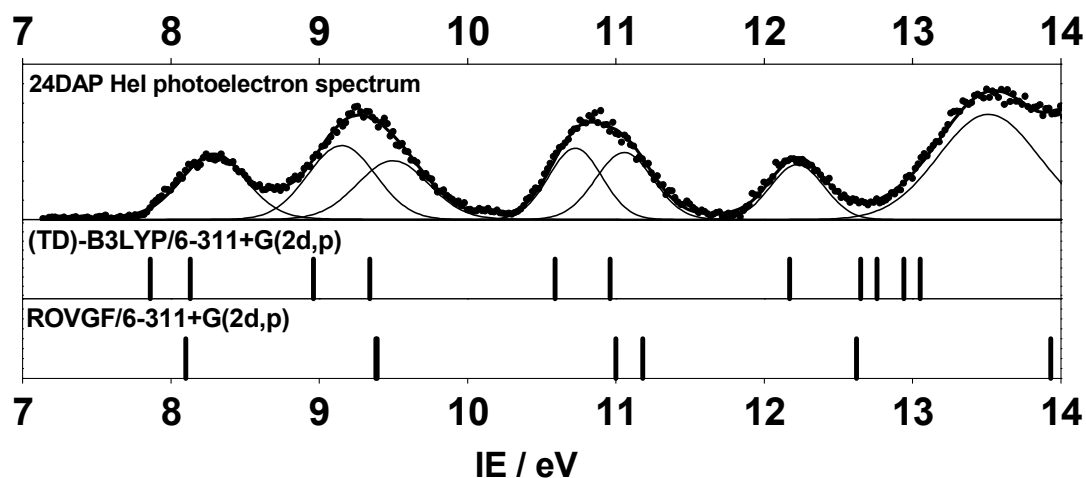


Figure 4 He-I photoelectron spectrum of 2,4-diaminopyrimidine. PseudoVoigt shaped peaks were fitted on the observed bands to extract the vertical ionization potentials. The adiabatic ionization potential was taken as the ionization onset at 7.86 eV. The adiabatic and vertical ionization energies computed were also computed at the (TD)-B3LYP/6-311+G(2d,p) and ROVGF/6-311+G(2d,p) levels and represented by the stick spectra. See text for details.

Considering the relative energies of the tautomers of 2,4-diaminopyrimidine, it is very unlikely that we observe more than one tautomer in the gas phase, even at the applied inlet temperature of 260°C. However, we performed restricted OVGf calculations at the 6-311+G(2d,p) level for all the possible isomers, employing single point calculations on the DFT equilibrium geometries. The outer valence Green's function method is a useful tool in the assignment of photoelectron spectra of organic molecules. The vertical ionization potentials are usually predicted within 0.3 eV⁹⁸. We obtained the best match with the experimental values for the most stable tautomer, the diamino form. Figure 4 lists the calculated ionization potentials. Photoelectron

spectroscopy can also be regarded as a tool for probing the ionic states. To obtain adiabatic and vertical ionization energies for the di-amino form, we carried out TD-B3LYP/6-311+G(2d,p) calculations at the equilibrium geometry of the ion. According to both methods, the first peak of the spectrum can be assigned to the ionization of the HOMO, that is, the removal of an electron from the π system.

Table 2 lists the adiabatic and the lowest vertical ionization energies. The predicted vertical ionization energies of all the possible tautomers are compared to the experimental spectrum in the Supporting Information (Fig. S1 and Table S1.)

2.3.3. 2,6-diaminopurine R2PI and IR-UV double resonance spectroscopy

Previously, we reported a resonant two-photon ionization spectrum of 2,6-diaminopurine between 32,000 cm^{-1} and 34,000 cm^{-1} without identifying the observed tautomer. In the present study, we extended the R2PI spectrum up to 35,400 cm^{-1} , as shown in Figure 5. Two groups of peaks can be distinguished, based on hole burning, with possible $S_1 \leftarrow S_0$ origins at 32,215 cm^{-1} and 34,881 cm^{-1} , respectively. We carried out IR-UV double resonance experiments on these two possible origins. The resulting IR spectra, shown in Figure 3, are not identical, indicating the presence of at least two different tautomers in the gas phase. Both spectra exhibit five IR bands. With the UV probe set at 34,881 cm^{-1} IR frequencies appear at 3450, 3460, 3509, 3566, and 3576 cm^{-1} . These frequencies match very well with the calculated IR frequencies of 2,6-diaminopurine. We assign the peaks at 3566 and 3576 cm^{-1} to the antisymmetric combinations of the N-H stretches in the amino groups, the peaks at 3450 and 3460 cm^{-1} to the symmetric combinations of the same stretches, and the peak at 3509 cm^{-1} to either the N9-H or the N7-H stretching mode. A hydrogen atom on the N9 position seems to be more plausible, because it might

have a smaller effect on the N-H stretching modes of the amino groups, as discussed in the next paragraph.

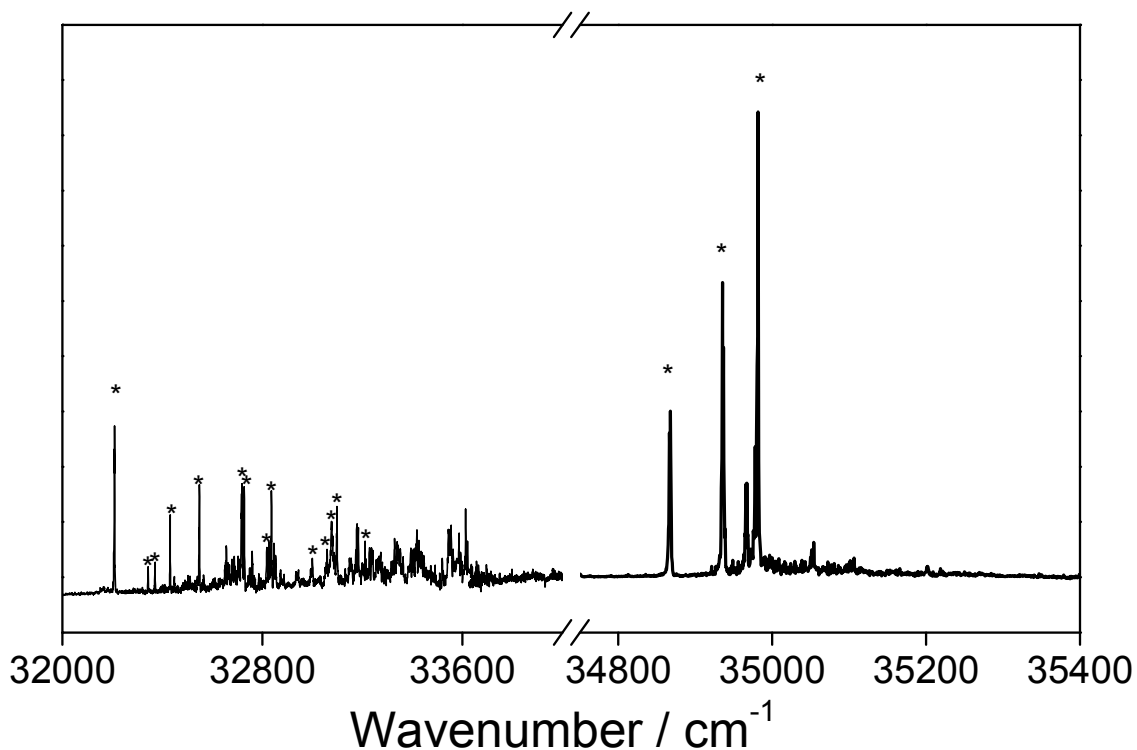


Figure 5 The R2PI spectra of the 7H and 9H tautomers of 2,6-diaminopurine. The $S_1 \leftarrow S_0(\pi\pi^*)$ origins are at 32,215 cm^{-1} and 34,881 cm^{-1} , respectively.

With the UV probe at 32,215 cm^{-1} , the five IR bands appear at 3428, 3460, 3513, 3533, and 3575 cm^{-1} . When compared to the former IR spectrum, significant red shifts of two peaks are apparent. The peak at 3450 cm^{-1} is shifted to 3428 cm^{-1} , and the peak at 3566 cm^{-1} is shifted to 3533 cm^{-1} . If the assignment of the first IR spectrum is correct, this shift can be explained by placing the hydrogen atom for this tautomer on the N7 position instead of the N9 position. That would perturb the N-H stretching modes in the N11 amino group. DFT (B3LYP/6-311+G(2d,p)) and *ab initio* (MP2(FC)/6-311+G(2d,p)) calculations of the vibrational frequencies of the various

tautomers support this qualitative assignment. Table 3 shows the comparison of the calculated vibrational frequencies with the experimental values. The two di-amino forms provide the best match. Furthermore, when calculating the relative energies of the tautomers the N9H di-amino tautomer is the most stable form, followed by the N7 di-amino tautomer, at every level of calculation.

2.3.4. Excited state lifetimes

To obtain a measure of excited state lifetimes we performed two color pump-probe experiments, in which we resonantly excited the molecule to the S_1 state, followed by ionization out of the excited state with a 266 nm photon from a second laser, with a variable delay between the two pulses. Figure 6 shows the results for 2,6DAPu. Fitting the decay curves with a single exponential decay, we obtained excited state lifetimes of 6.3 ± 0.4 ns for the N9H tautomer and 8.7 ± 0.8 ns for the N7H tautomer. Using laser pulses of about 5 ns pulse width imposes a lower limit of the order of a few nanoseconds on these lifetime measurements.

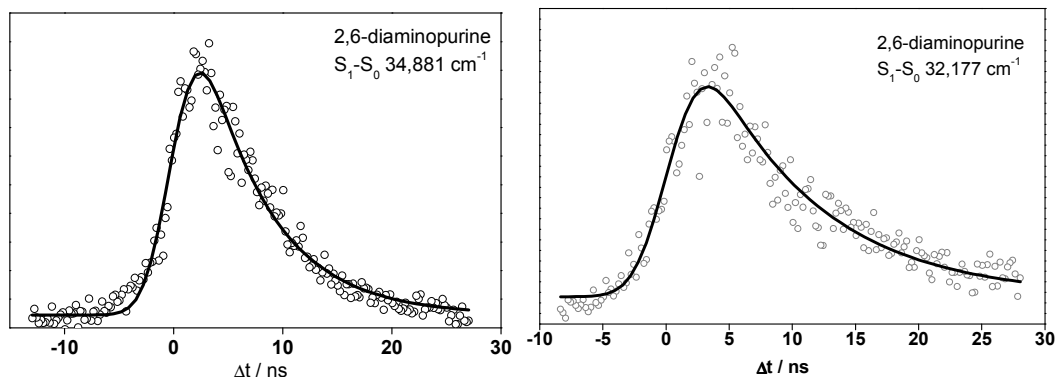


Figure 6 Exponential decay of the ion signal of 2,6-diaminopurine in the two-color R2PI experiment. The extracted lifetime of the excited state is 8.7 ± 0.8 ns for the 7H tautomer and 6.3 ± 0.3 ns for the 9H tautomer.

In the case of 2,4DAPy the lifetime is shorter than this experimental limit so we were

unable to perform two-color experiments and our attempts to carry out UV-UV double resonance experiments also failed. In this case we can use the peak width in the R2PI spectrum to obtain a rough estimate of the lower limit of the excited state lifetime. Figure 7 shows the peak at $34,720\text{ cm}^{-1}$ together with simulations of the rotational envelope with different Lorentzian linewidths. The simulated lineshapes assume a dominantly b-type transition and a rotational temperature of 20 °K . These parameters are consistent with our simulations of the peaks in the 2,6DAPur spectrum, for which the width is constrained by the experimental resolution only and not by lifetime broadening. The results, as shown in

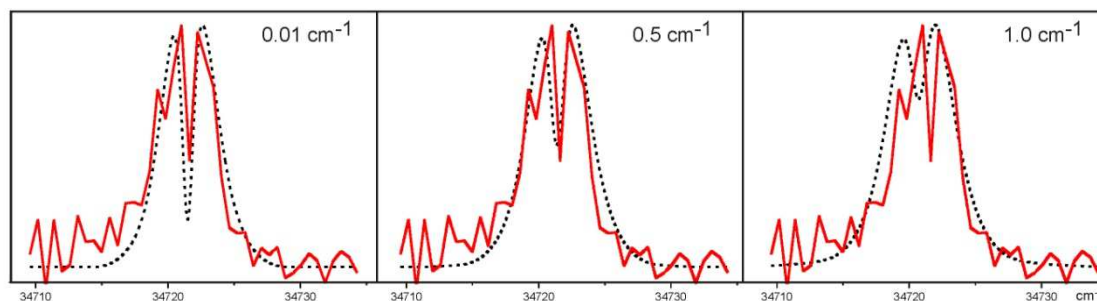


Figure 7. Detail of the 34720 cm^{-1} peak in the 2,4 DAPy REMPI spectrum, compared with simulations of the rotational envelope at three different Lorentzian linewidths.

Figure 7, suggest an upper limit to the linewidth of 0.5 cm^{-1} corresponding to a lower limit of the excited state lifetime of 10 ps . Therefore, we can bracket the experimental window of our lifetime measurements for 2,4DAPy roughly between 10^{-11} and 10^{-9} seconds. This is clearly shorter than that of 2,6DAPur, but it is longer than the lifetimes of 1.8 or 2.4 ps and 5.2 or 6.4 ps , reported for uracil and thymine, respectively, with excitation at 267 nm .⁹⁹ We also note that the UV spectra of uracil and thymine are broad while that of 2,4DAPy is sharp⁵.

This lifetime is clearly shorter than that of 2,6-DAPu, but consistent with that of other

pyrimidine bases excited at the band origin ^{100,101}. Not surprisingly, this 10 ps–1 ns time range spans values much longer than the typical 1–6 ps found for pyrimidine bases excited close to the first band maximum, well above any barriers towards conical intersections ^{9,99}. The sub-nanosecond lifetime of band-origin excited 2,4-DAPy is a strong indication that this molecule deactivates by means of internal conversion just like 4-APy does. In order to examine how the internal conversion takes place, we carried out extensive analysis of the excited-state surfaces of this molecule, which will be discussed in the next section.

2.4. Excited-state analysis of 2,4-diaminopyrimidine

2.4.1. Stationary points in the ground and excited states

Table 4 shows the vertical excitation energies calculated at the CASSCF, MR-CISD, and CASPT2 levels. At the CASSCF level, the first excited state is of $\pi\pi^*$ character while the second excited state is due to excitation from the lone pairs located on the nitrogen atoms of the pyrimidine ring ($n\pi^*$ transition). The calculated energy gap between these states is 0.34 eV. We obtained the same ordering of states at the CASPT2 and RICC2 levels, with a smaller energy gap amounting to 0.2 and 0.1 eV, respectively. Increasing the flexibility of the basis set does not change the ordering of the states and the energy gap between them (see Table 4 and Supplementary Material for CASPT2 and RICC2 results, respectively). At the MR-CISD+Q level the ordering is reversed with an energy gap of 0.3 eV.

The experimentally measured excitation energy at the band origin is 4.27 eV. To estimate the energy of the band maxima we have performed a simulation of the absorption spectra (see Fig. 8 for the absorption spectrum simulated with RICC2 method). From the figure we calculated the

energy difference between the band maximum and band origin as 0.63 eV (the same result was obtained employing the CASSCF method). Adding this shift to the experimental value of the band origin results in a band maximum at 4.91 eV, in a good agreement with the calculated vertical excitation energies.

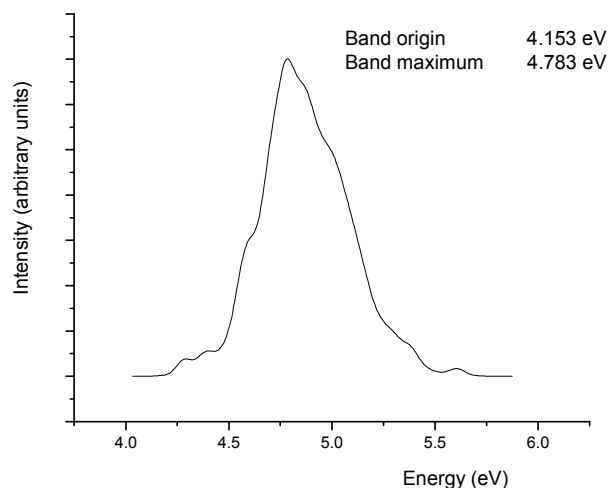


Figure 8 Absorption spectrum of 2,4-DAPy calculated at the RICC2 method using the SVP basis set.

We found two minima on the S_1 surface (Fig. 9a) using CASSCF optimization, with energies of 4.75 and 5.08 eV above the ground state minimum (Table 4). Both structures show a distortion from the planarity of the ring which is more pronounced in the latter case. For both minima, the CASSCF wavefunctions show strong coupling between the $n\pi^*$ and $\pi\pi^*$ configurations with almost equal weight. Unlike the case of 4-aminopyrimidine^{29,30}, we did not succeed in locating a third minimum which is planar and of $\pi\pi^*$ character. Any attempt to find this minimum ended up in conformation S1min_C6.

The energy ordering of S1min_C6 and S1min_C2 remains the same at the MR-CISD(8)+Q

level and at the CASPT2 level with an energy gap of about 0.2 and 0.3 eV, respectively. The energy ordering is reversed when 17 orbitals are frozen in the CI procedure, with the energetic changes being, however, quite small. The latter calculations place the S1min_C2 below the S1min_C6 by about 0.1 eV. The difference between the results with 8 and 17 orbitals frozen is however not significant for the discussion of reaction paths calculated by means of linear interpolation curves and thus the less accurate but more computationally feasible approach with 17 orbitals frozen can be used.

There is a substantial energetic relaxation of the first excited states when moving from the Franck-Condon region to the S1min_C6 and S1min_C2 optimized structures. This result is consistent with the experimentally observed low intensity of the 0-0 transition. Although the amount of relaxation energy from 2^1A into the S1min_C2 structure, calculated with different methods, is not as uniform as in the case of relaxation into the S1min_C6 structure (see Table 4), results for all methods used show that both minima located on the S_1 surface are energetically accessible from the Franck-Condon region.

2.4.2. Minima on the crossing seam

Table 4 lists the CASSCF and MR-CISD energies and the character of five different minima located on the S_0/S_1 crossing seam optimized at the CASSCF level; Fig. 9b shows the corresponding structures. All structures display strong out-of-plane deformations with the twist around either the CN or CC bonds. We adopted the Cremer-Pople classification scheme in performing the conformational analysis of these structures^{102,103}. Using this approach, we identified four types of structure deformations, in particular screw-boat (S), boat (B), half-chair

(H) and envelope (E) conformations (for Cartesian coordinates see Supplementary Material. The 3H_4 , 1S_2 , 1S_6 and B_{14} MXSs are energetically grouped closely together at the CASSCF level within 0.3 eV; all MXSs shown are energetically accessible starting from the Franck-Condon region. The structures and the energetics of all MXS points are very similar to those found by Barbatti et al. for 4-aminopyrimidine although the results of conformational analysis are not the same (see Table 4). This situation indicates that there is only a small (if any) energetic effect of the additional amino group on the puckering of the pyrimidine ring and subsequent formation of conical intersection.

In all cases there are two major configurations which contribute to the CASSCF wavefunction ($\pi\pi^*$ and $n\pi^*$) (see Table 4). Although it is difficult to distinguish lone-pair and π orbitals in such distorted structures, the main character of the singly occupied orbitals is indicated in the Table 4. The second configuration contributing to the wavefunction in all cases is the closed-shell configuration (CS). In agreement with the results found for 4-aminopyrimidine, the character of the singly occupied orbital was identified as π orbital.

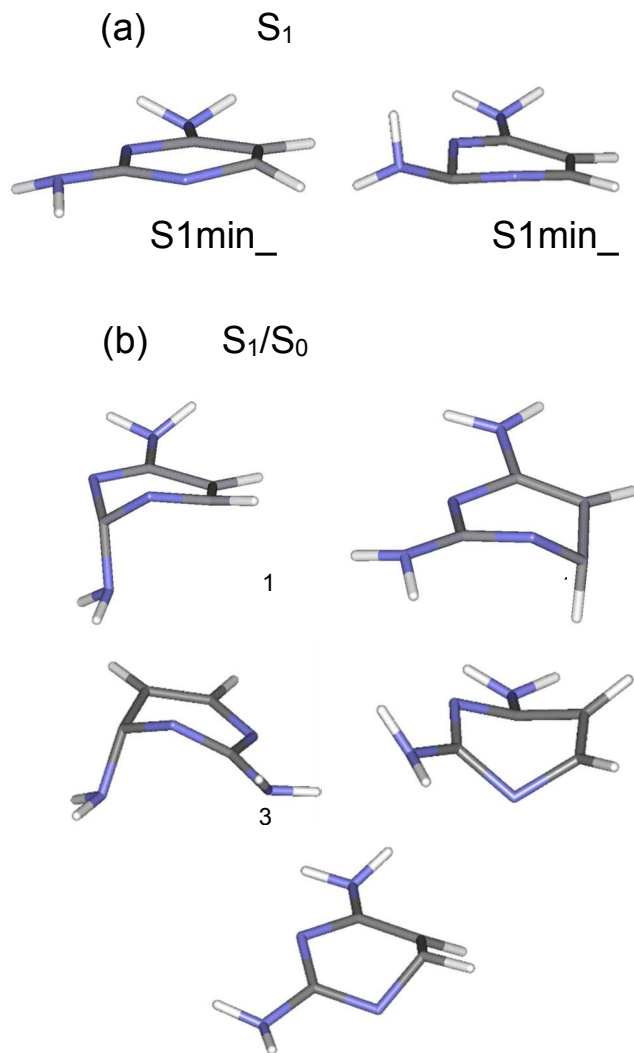


Figure 9a Structures of the two minima found on the S_1 potential energy surface optimized at the CASSCF(14,10)/6-31G** level.

Figure 9b Structures of the five different minima located on the S_0/S_1 crossing seam optimized at the CASSCF(14,10)/6-31G** level

We performed MR-CISD calculations with 8 and 17 orbitals frozen in the CI procedure. The MR-CI and CASPT2 procedures resulted in relatively large splittings, usually around 0.8 eV,

of the S_0 and S_1 surfaces at the MXS points determined at the CASSCF level. We estimated the energies and MXS structures at the MR-CISD level from an extrapolation of the LIIC curves (these results are obtained from the MR-CISD(17)+Q calculations, see below). Table 4 lists the CASSCF and these estimated MR-CISD(17)+Q results. When the dynamic correlation is included, there is a change in the ordering of various MXS structures and the 1S_2 structure becomes the most stable one. However, all structures are placed in a relatively small energy range, with the difference between the most and least stable structure 0.8 eV, in good agreement with the results of 4-aminopyrimidine. The CASPT2 energies, also estimated from the LIIC curves, agree reasonably well with the MR-CI results. Thus, the inclusion of the dynamic correlation effects does not change the parameters for prediction of the dynamics of the molecular systems studied.

2.4.3. Interpolation curves

We calculated LIIC curves for the reaction paths between the S1min_C2 and S1min_C6 minima and the four lowest MXS structures at the CASSCF and MR-CISD levels. The LIIC curve for the reaction path S1min_C6 to 1S_6 was calculated using the CASPT2 method as well. (See Fig. 10 for the reaction path between S1min_C6 and 1S_6 . Other reaction paths are reported in Supplementary Material). To make the studies of reaction paths computationally more feasible we performed MR-CISD calculations with 17 frozen orbitals. A comparison of the character of the interpolation curves calculated at the MR-CISD(8)+Q and MR-CISD(17)+Q levels for the reaction path between S1min_C6 and 1S_6 (see Fig. 10 and Supplementary Material) shows that the characters of the interpolation curves are very similar justifying the use of the latter approach. The barriers for the reaction paths estimated from the highest point of the interpolation curves are presented in Table 5. These values are similar to those found by Zechmann and Barbatti for 4-APy

¹⁰⁴. Since this estimation provides only an upper bound for the reaction barrier, we have optimized the transition states along the selected reaction paths, particularly the paths from S1min_C6 towards ¹S₆ and ³H₄ MXS structures. The energies of those structures are indicated in the graphs (see Fig. 10 and Supplementary Material) and in Table 5.

Relatively small barriers exist for the reaction paths between the S1min_C6 and the ¹S₆ and B_{1,4} and ³H₄ MXSs. The only sizeable barrier we found was for the reaction paths toward the ¹S₂ structure. This finding reflects some immobilization of the ring by substitution of the hydrogen atom by an amino group at C₂ atom. The barriers connecting the S1min_C2 with MXSs are generally larger with values of approximately 0.8 eV. Starting from the highest points of the interpolation curves we determined the true saddle points for the reaction paths towards the ¹S₆ and ³H₄ structures. At the CASSCF and MR-CISD(17)+Q levels these calculations reduced the reaction barriers to 0.17 and 0.12 eV at for ³H₄, and to 0.04 and 0.02 eV for ¹S₆, respectively, indicating that both conical intersections should be easily accessible leading to a subsequent ultrafast deactivation to the ground state.

Comparing the results for 2,4-DAPy and those previously reported for 4-APy shows that the additional amino-group on the pyrimidine ring does not significantly influence the energetics of the stationary points and conical intersections, as well as the characters of interpolation curves. The only exception is the ¹S₂ structure and the interpolation curve towards it, as a result of an immobilization of the pyrimidine ring at the C2 position.

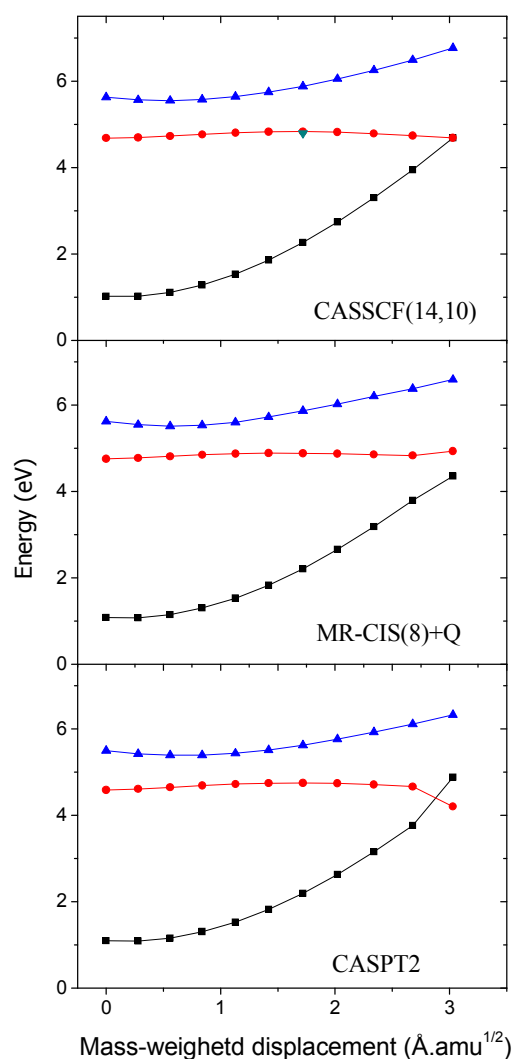


Figure 10 Reaction path between S1min_C6 and 1S_6 calculated using CASSCF(14,10), MR-CISD(8)+Q, and CASPT2 methods. Other reaction paths are reported in the Supplementary Material.

Since 4-APy usually does not access the pathway with C2 atom deformation, one would not expect significant differences in the dynamics of these two systems, provided comparable initial conditions are used. Therefore, in analogy to the previously described surface hopping

dynamics for 4-APy, the present reaction pathways for 2,4-DAPy indicate that ultrafast internal conversion to the ground state should be expected provided that sufficient initial kinetic energy is available to surpass the small energy barriers to the 1S_6 and 3H_4 conical intersections.

In the current experiments, the pump energy is in the range of 33,000 to 34,000 cm^{-1} which corresponds to 4.1 – 4.2 eV. Neglecting any vibrational corrections, this energy range is close to or even below the calculated energies of the saddle points (using the CASPT2 energies of MXSs and energy barriers of the relevant saddle points collected in Table 4). Therefore, the pump energy used in the experiment is not sufficient to overcome the barrier to reach the conical intersections, producing the much longer lifetime experimentally observed for 2,4-DAPy.

When the relevant coordinates involve ring deformations, lifetimes can be significantly different for specific structures. In particular, if deformation involving C5 and C6 leads to internal conversion in 2,4-DAPy then that pathway would be significantly altered for 2,6-DAPu, in which the puckering at C6 is strongly hindered. This observation explains why the lifetime for 9H-adenine is longer than that calculated for 4-APy and also explains the very long lifetime (6-8 ns) for 2,6-DAPu, in which both the C2 position and the C5=C6 positions are modified. In this latter case the subtle difference between the N7H and N9H tautomer lifetimes requires further analysis. This effect of immobilization of the C5=C6 twist is analogous to the findings of Zgierski et al., which showed that in solution 5,6-trimethylenecytosine and 5,6-trimethylenouracil do not exhibit subpicosecond excited-state lifetimes characteristic of the naturally occurring pyrimidine bases ²¹.

2.4. Conclusions

We recorded resonant two-photon ionization spectra of 2,4-diaminopyrimidine and 2,6-diaminopurine in the frequency range of 32,000 to 36,000 cm^{-1} . IR-UV double resonance experiments supported by DFT and *ab initio* calculations suggest that we observed the most stable tautomers of both compounds. We also recorded and assigned the photoelectron spectrum of 2,4-diaminopyrimidine, placing the adiabatic ionization potential at 7.86 ± 0.05 eV, and the first vertical ionization potential at 8.30 ± 0.04 eV. The $S_1 \leftarrow S_0$ ($\pi\pi^*$) origin of 2,4-diaminopyrimidine, as measured with one-color, two-photon resonant ionization, is $34,459 \text{ cm}^{-1}$. The excited state lifetime measured at this pump wavelength is shorter than the time resolution (approx. 1 ns) of the instrument and longer than 10 ps, based on the sharp and resolved peaks in the R2PI spectrum.

According to the IR-UV double resonance experiments, both the N7H and N9H tautomers of 2,6-diaminopurine are easily observable in the cold jet. The $S_1 \leftarrow S_0$ ($\pi\pi^*$) origin is at $32,215 \text{ cm}^{-1}$ for the N7H tautomer and $34,881 \text{ cm}^{-1}$ for the N9H tautomer. The measured excited state lifetime is 8.7 ± 0.8 ns for the former and 6.3 ± 0.4 ns for the latter.

This long lifetime is consistent with reduction of ring deformations in both the C2 position and the C5=C6 twist, affecting all conical intersections that lead to fast excited state dynamics in 4-APy.

To understand the photodynamics of 2,4-diaminopyrimidine, we carried out quantum chemical modeling. Based on the results of the relative energies of stationary points on the S_1 surface and its crossings with the ground state at MXS points, together with the nature of the interpolation curves we can expect an ultrafast dynamics in the excited state of 2,4-diaminopyrimidine which is predicted to be comparable to that of 4-aminopyrimidine. The

additional NH₂ group in the C2 position blocks the path to only one of the conical intersections predicted for the analogous case of 4-aminopyrimidine, namely the one associated with ring deformation at C2. However, there are still other alternatives for almost barrierless access to conical intersections, associated with deformations at the C5, C6, and N1 positions. The actually observed lifetimes will depend on the chosen excitation energy which needs to be large enough to surmount the existing small energy barriers if ultrafast processes are to occur.

The present experiments show a significant reduction of the lifetime of 2,4-DAPy as compared to the nanosecond lifetimes determined for 2,6-DAPu, indicating a larger internal conversion rate for the former molecule. However, the current experiments involve a very narrow excitation window very close to the band origin, and therefore exhibit the longest possible lifetime. The pump energies in the range of 34,000 to 35,000 cm⁻¹ are close to or possibly just below the computed energy barriers, which explains the still rather large lifetime of 2,4-DAPy compared to that of other pyrimidine bases excited at the center of the absorption band.

Tables

Table 1. Calculated vibrational frequencies for tautomers of 2,4-diaminopyrimidine in the N-H region. All frequencies have been scaled and given in cm^{-1} . MP2 scaling factor: 0.9496; B3LYP scaling factor: 0.9618. For numbering scheme, see Fig 1. For comparison, the experimental vibrational frequencies are 3452, 3464, 3563, and 3578 cm^{-1} .

Tautomer	MP2	B3LYP	vib.
	6-	6-	Mode
	311+G(2d,p) ω/cm^{-1}	311+G(2d,p) ω/cm^{-1}	
24DAP diamino tautomer			N7H
	3532	3574	as
	3518	3556	as
	3410	3456	N7H s
	3400	3443	N8H s
1H, 7 imino	3527	3564	N8H as
	3440	3483	N1H
	3403	3446	N8H s
	3351	3386	N7H
3H, 8 imino	3496	3534	N7H as
	3403	3445	N3H

	3387	3431	N7H
			s
	3346	3380	N8H
1H,3H	3467	3507	N1H
7,8 imino	3422	3469	N3H
	3359	3392	N10H
	3344	3378	N11H
3H, 7			N8H
imino		3535	as
		3454	N3H
			N8H
		3435	s
		3387	N7H
1H, 8			N7H
imino		3508	as
		3492	N1H
		3411	N7H
			s
		3328	N8H
5H, 8			N7H
imino		3593	as
			N7H
		3465	s
		3316	N8H
		2986	CH

5H, 7		N8H
	3559	
imino		as
		N8H
	3436	
		s
	3345	N7H
	2972	CH

Table 2. Experimental excitation and ionization energies (in eV) of 2,4-diaminopyrimidine and 2,6-diaminopurine.

2,4-diaminopyrimidine		
Transition		$h\nu / \text{eV}$.
S_1		4.27
$\leftarrow S_0$		
D_0	I	7.81
$\leftarrow S_0$	E_{ad}	
	I	8.30
	E_{vert}	
7H-2,6-diaminopurine		
S_1		3.99
$\leftarrow S_0$		
9H-2,6-diaminopurine		
S_1		4.32
$\leftarrow S_0$		

Table 3. Calculated vibrational frequencies for tautomers of 2,6-diaminopurine in the N-H region. All frequencies have been scaled and given in cm^{-1} . MP2 scaling factor: 0.9496; B3LYP scaling factor: 0.9618. For numbering scheme see Fig 1. For comparison, the experimental vibrational frequencies are 3450, 3460, 3509, 3566, and 3576 cm^{-1} (UV probe 34,881 cm^{-1}), and 3428, 3460, 3513, 3533, and 3575 cm^{-1} (UV probe 32,215 cm^{-1}).

Tautomer	MP2	B3LYP	Vib. Mode
	6-	6-	
	311+G(2d,p) ω/cm^{-1}	311+G(2d,p) ω/cm^{-1}	
<hr/>			
9H-			
26DAP	3505	3577	N11H as
diamino tautomer	3498	3567	N10H as
	3438	3512	N9H
	3380	3456	N11H s
	3379	3452	N10H s
<hr/>			
7H-			
26DAP	3493	3564	N10H as
diamino tautomer	3454	3519	N11H as+N7H
		3515	N7H+N11H
	3356		s
	3375	3449	N10H s
	3346	3417	N11H s
<hr/>			
1,9H, 11			
imino		3525	N11H as
		3507	N9H
		3453	N1H
		3426	N10H s

	3377	N11H
<hr/>		
1,7H, 11		
imino	3508	N7H
	3500	N10H as
	3465	N1H
	3406	N10H s
	3351	N11H
<hr/>		
3,9H, 10		
imino	3586	N11H as
	3512	N9H
	3486	N3H
	3460	N11H s
	3391	N10H
<hr/>		
1,9H, 10		
imino	3543	N11H as
	3511	N9H
	3460	N1H
	3437	N11H s
	3392	N10H
<hr/>		
3,7H, 10		N11H
imino	3522	as+N7H
		N7H+N11H
	3520	as
	3484	N3H
	3417	N11H s
	3390	N10H
<hr/>		

1,7H, 10	3516	N11H as
imino	3485	N7H
	3460	N1H
	3422	N11H s
	3392	N10H

Table 4. Energies (in eV) of stationary points and conical intersections for 2,4-di-aminopyrimidine relative to the ground state minimum structure calculated at the CASSCF(14,10), MR-CISD(6,5)+Q and CASPT2 levels and comparison with 4-aminopyrimidine. The 6-31G** basis set was used. The experimental excitation energy (band maximum) was estimated to be 4.91 eV.^a

		4A				
		CA	MR-	CAS	P ^d	
		SSCF	CISD(17)+Q ^b	PT2 ^c	MR	
		-QDPT2				
2 ¹ A	Δ	5.1	5.420	5.02	4.7	
($\pi\pi^*$)	E	89	(5.406)	4 (4.938)	9 (5.34)	
	<i>f</i>	0.0		0.06		
	<i>e</i>	65		3		
3 ¹ A	Δ	5.5	5.044	5.20	4.7	
($n\pi^*$)	E	28	(5.142)	4 (5.110)	1 (5.88)	
	<i>f</i>	0.0		0.01		
	<i>e</i>	13		2		
S1m	π	4.7	4.743	4.57	4.2	
in_C6	$\pi^*+n\pi^*$	54	(4.708)	3 (4.528)	2 (4.93)	
S1m	π	5.0	4.632	4.83	4.3	
in_C2	$\pi^*+n\pi^*$	75	(4.917)	1 (4.822)	3 (5.02)	

	MX	C	4.5	4.2 ^f	4.3	4.6
S(³ H ₄)	S+ππ*	94			(4.4) ^f	2 (4.60) ^g
	MX	C	4.6	3.9 ^f	4.2	4.3
S(¹ S ₂)	S+ππ*	45			(4.2) ^f	6 (4.49) ^h
	MX	C	4.7	4.5 ^f	4.5	4.7
S(¹ S ₆)	S+ππ*	59			(4.5) ^f	1 (4.79)
	MX	C	4.9	4.1 ^f	4.3	(4.9
S(B _{1,4})	S+ππ*	49			(4.3) ^f	7)
	MX	C	5.3			(5.4
S(⁶ E)	S+ππ*	28		4.7 ^f		2)

^aFor estimation of the experimental band maximum see text.

^bThe results with eight frozen orbitals are given in parentheses.

^cThe results obtained with 6-311G(2d,p) basis set.

^dMR-QDPT2 energies (CASSCF energies are given in parentheses) calculated for the relevant structures of 4-aminopyrimidine.

^eOscillator strength

^fEstimated from an extrapolation of the LIIC curves

^gAssigned as ³S₄ for 4-APy

^hAssigned as E₂ for 4-APy.

Table 5. Barriers (in eV) located on the interpolation curves.

	CASSCF	CISD(17)+Q
S1min_C6	0.145	0.082
to 1S_6	Saddle point: 0.040	Saddle point: 0.024
S1min_C6	0.419	0.153
to $B_{1,4}$		
S1min_C6	0.407	0.251
to 3H_4	Saddle point: 0.174	Saddle point: 0.118
S1min_C6	1.041	1.042
to 1S_2		
S1min_C2	0.614	0.976
to 1S_2		
S1min_C2	0.686	0.974
to $B_{1,4}$		
S1min_C2	0.788	0.830
to 1S_6		
S1min_C2	0.870	0.857
to 3H_4		

References

- (1) Marian, C. M. *J. Phys. Chem. A (USA)* **2007**, *111*, 1545-1553.
- (2) Seefeld, K.; Brause, R.; Haber, T.; Kleinermanns, K. *J. Phys. Chem. A (USA)* **2007**, *111*, 6217-6221.
- (3) Abo-Riziq, A.; Grace, L.; Nir, E.; Kabelac, M.; Hobza, P.; de Vries, M. S. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 20-23.
- (4) Sobolewski, A. L.; Domcke, W.; Hättig, C. *PNAS* **2005** *102*, 17903-17906.
- (5) Brady, B. B.; Peteanu, L. A.; Levy, D. H. *Chem. Phys. Lett.* **1988**, *147*, 538-543.
- (6) Luhrs, D. C.; Viallon, J.; Fischer, I. *Physical Chemistry Chemical Physics* **2001**, *3*, 1827-1831.
- (7) Kang, H.; Jung, B.; Kim, S. K. *J. Chem. Phys.* **2003**, *118*, 6717-6719.
- (8) Bisgaard, C. Z.; Satzger, H.; Ullrich, S.; Stolow, A. *Chemphyschem* **2009**, *10*, 101-110.
- (9) Canuel, C.; Mons, M.; Piuizzi, F.; Tardivel, B.; Dimicoli, I.; Elhanine, M. *J. Chem. Phys.* **2005**, *122*, 7.
- (10) Hudock, H. R.; Levine, B. G.; Thompson, A. L.; Satzger, H.; Townsend, D.; Gador, N.; Ullrich, S.; Stolow, A.; Martinez, T. J. *J. Phys. Chem. A (USA)* **2007**, *111*, 8500-8508.
- (11) Climent, T.; Gonzalez-Luque, R.; Merchan, M.; Serrano-Andres, L. *Chem. Phys. Lett.* **2007**, *441*, 327-331.
- (12) Serrano-Andres, L.; Merchan, M.; Borin, A. C. *Proc. Natl. Acad. Sci. U. S.*

A. **2006**, *103*, 8691-8696.

(13) Perun, S.; Sobolewski, A. L.; Domcke, W. *J. Am. Chem. Soc.* **2005**, *127*, 6257-6265.

(14) Chen, H.; Li, S. H. *J. Phys. Chem. A (USA)* **2005**, *109*, 8443-8446.

(15) Marian, C. M. *J. Chem. Phys.* **2005**, *122*, 10.

(16) Ismail, N.; Blancafort, L.; Olivucci, M.; Kohler, B.; Robb, M. A. *J. Am. Chem. Soc.* **2002**, *124*, 6818-6819.

(17) Blancafort, L.; Cohen, B.; Hare, P. M.; Kohler, B.; Robb, M. A. *J. Phys. Chem. A (USA)* **2005**, *109*, 4431-4436.

(18) Blancafort, L.; Robb, M. A. *J. Phys. Chem. A (USA)* **2004**, *108*, 10609-10614.

(19) Blancafort, L. *J. Am. Chem. Soc.* **2006**, *128*, 210-219.

(20) Matsika, S. *J. Phys. Chem. A (USA)* **2004**, *108*, 7584-7590.

(21) Zgierski, M. Z.; Patchkovskii, S.; Fujiwara, T.; Lim, E. C. *J. Phys. Chem. A (USA)* **2005**, *109*, 9384-9387.

(22) Kistler, K. A.; Matsika, S. *J. Chem. Phys.* **2008**, *128*, 215102.

(23) Kistler, K. A.; Matsika, S. *J. Phys. Chem. A (USA)* **2007**, *111*, 2650-2661.

(24) Barbatti, M.; Lischka, H. *J. Am. Chem. Soc.* **2008**, *130*, 6831-6839.

(25) Broo, A. *J. Phys. Chem.* **1998**, *A 102*, 526-531.

(26) Nir, E.; Kleiner, K.; Grace, L.; de Vries, M. S. *J. Phys. Chem. A (USA)* **2001**, *105*, 5106-5110.

(27) Perun, S.; Sobolewski, A. L.; Domcke, W. *Mol Phys* **2006**, *104*, 1113-1121.

(28) Fabiano, E.; Thiel, W. *J. Phys. Chem. A (USA)* **2008**, *112*, 6859-6863.

- (29) Barbatti, M.; Lischka, H. *J. Phys. Chem. A (USA)* **2007**, *111*, 2852-2858.
- (30) Barbatti, M.; Ruckebauer, M.; Szymczak, J. J.; Aquino, A. J. A.; Lischka, H. *Physical Chemistry Chemical Physics* **2008**, *10*, 482-494.
- (31) Barbatti, M.; Sellner, B.; Aquino, A. J. A.; Lischka, A. *Nonadiabatic excited-state dynamics of aromatic heterocycles: toward the time-resolved simulation of nucleobasis*; Springer Science, Business Media B.V., 2008.
- (32) Joyce, G. F.; Schwartz, A. W.; Miller, S. L.; Orgel, L. E. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 4398-4402.
- (33) Lutz, M. J.; Horlacher, J.; Benner, S. A. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1149-1152.
- (34) Lutz, M. J.; Held, H. A.; Hottiger, M.; Hubscher, U.; Benner, S. A. *Nucleic Acids Res.* **1996**, *24*, 1308-1313.
- (35) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature (UK)* **1990**, *343*, 33-37.
- (36) Yamazaki, S.; Sobolewski, A. L.; Domcke, W. *Physical Chemistry Chemical Physics* **2009**, *11*, 10165-10174.
- (37) Cleaves, H. J.; Nelson, K. E.; Miller, S. L. *Naturwissenschaften* **2006**, *93*, 228-231.
- (38) Ferris, J. P.; Zamek, O. S.; Altbuch, A. M.; Freiman, H. *J. Mol. Evol.* **1974**, *3*, 301-309.
- (39) Miyakawa, S.; Cleaves, H. J.; Miller, S. L. *Origins of Life and Evolution of the Biosphere* **2002**, *32*, 209-218.
- (40) Robertson, M. P.; Levy, M.; Miller, S. L. *J. Mol. Evol.* **1996**, *43*, 543.

- (41) Saladino, R.; Crestini, C.; Neri, V.; Brucato, J. R.; Colangeli, L.; Ciciriello, F.; Di Mauro, E.; Costanzo, G. *Chembiochem* **2005**, *6*, 1368-1374.
- (42) Nir, E.; Muller, M.; Grace, L. I.; de Vries, M. S. *Chem. Phys. Lett.* **2002**, *355*, 59-64.
- (43) Nir, E.; Janzen, C.; Imhof, P.; Kleinermanns, K.; de Vries, M. S. *J. Chem. Phys.* **2001**, *115*, 4604-4611.
- (44) Nir, E.; Hunig, I.; Kleinermanns, K.; de Vries, M. S. *Physical Chemistry Chemical Physics* **2003**, *5*, 4780-4785.
- (45) Seefeld, K. A.; Plutzer, C.; Lowenich, D.; Haber, T.; Linder, R.; Kleinermanns, K.; Tatchen, J.; Marian, C. M. *Physical Chemistry Chemical Physics* **2005**, *7*, 3021-3026.
- (46) Callahan, M. P.; Crews, B.; Abo-Riziq, A.; Grace, L.; de Vries, M. S.; Gengeliczki, Z.; Holmes, T. M.; Hill, G. A. *Physical Chemistry Chemical Physics* **2007**, *9*, 4587-4591.
- (47) Tembreull, R.; Sin, C. H.; Pang, H. M.; Lubman, D. M. *Analytical Chemistry* **1985**, *57*, 2911-2917.
- (48) Meijer, G.; de Vries, M. S.; Hunziker, H. E.; Wendt, H. R. *Applied Physics B-Photophysics and Laser Chemistry* **1990**, *51*, 395-403.
- (49) Csakvari, B.; Nagy, A.; Zanaly, L.; Szepes, L. *Magy. Kem. Foly.* **1992**, *98*, 415-419.
- (50) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
- (51) Lee, C. T.; Yang, W. T.; Parr, R. G. *Phys Rev B* **1988**, *37*, 785-789.
- (52) Miehlisch, B.; Savin, A.; Stoll, H.; Preuss, H. *Chem. Phys. Lett.* **1989**, *157*,

200-206.

(53) Mclean, A. D.; Chandler, G. S. *J. Chem. Phys.* **1980**, *72*, 5639-5648.

(54) Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. *J. Chem. Phys.* **1980**, *72*, 650-654.

(55) Moller, C.; Plesset, M. S. *physical Review* **1934**, *46*, 618-622.

(56) Pople, J. A.; Headgordon, M.; Raghavachari, K. *J. Chem. Phys.* **1989**, *90*, 4635-4636.

(57) Frisch, M. J.; Headgordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 275-280.

(58) Frisch, M. J.; Headgordon, M.; Pople, J. A. *Chem. Phys. Lett.* **1990**, *166*, 281-289.

(59) Headgordon, M.; Headgordon, T. *Chem. Phys. Lett.* **1994**, *220*, 122-128.

(60) Saebo, S.; Almlöf, J. *Chem. Phys. Lett.* **1989**, *154*, 83-89.

(61) Andersson, M. P.; Uvdal, P. *J. Phys. Chem. A (USA)* **2005**, *109*, 2937-2941.

(62) Scott, A. P.; Radom, L. *J. Phys. Chem.* **1996**, *100*, 16502-16513.

(63) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764-7776.

(64) Stratmann, R. E.; Scuseria, G. E.; Frisch, M. J. *J. Chem. Phys.* **1998**, *109*, 8218-8224.

(65) Bauernschmitt, R.; Ahlrichs, R. *Chem. Phys. Lett.* **1996**, *256*, 454-464.

(66) Casida, M. E.; Jamorski, C.; Casida, K. C.; Salahub, D. R. *J. Chem. Phys.* **1998**, *108*, 4439-4449.

(67) Ortiz, J. V. *J. Chem. Phys.* **1988**, *89*, 6353-6356.

- (68) Cederbaum, L. S. *Journal of Physics B-Atomic Molecular and Optical Physics* **1975**, *8*, 290-303.
- (69) von Niessen, W.; Schirmer, J.; Cederbaum, L. S. *Comp. Phys. Rep.* **1984**, *1*, 57.
- (70) Zakrzewski, V. G.; Vonniessen, W. *J. Comput. Chem.* **1993**, *14*, 13-18.
- (71) Zakrzewski, V. G.; Ortiz, J. V. *Int. J. Quantum Chem.* **1995**, *53*, 583-590.
- (72) Ortiz, J. V. *Int. J. Quant. Chem. Symp.* **1988**, *22*, 431.
- (73) ortiz, J. V. *Int. J. Quant. Chem. Symp.* **1989**, *23*, 321.
- (74) Ortiz, J. V.; zakrzewski, V. G.; Dolgounircheva, O. In *Conceptual Perspectives in Quantum Chemistry*; Calais, J. L., Kryachko, E. S., Eds.; Kluwer Academic: 1997, p 465.
- (75) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J., J. A.; 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, O.; 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.; and

Pople, J. A.; Gaussian, Inc.: Wallingford CT, 2004.

(76) Bunge, A. *J. Chem. Phys.* **1970**, *53*, 20-&.

(77) Pople, J. A.; Seeger, R.; Krishnan, R. *Int. J. Quantum Chem.* **1977**, 149-163.

(78) Hehre, W. J.; Ditchfie.R; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257-&.

(79) Binkley, J. S.; Pople, J. A.; Hehre, W. J. *J. Am. Chem. Soc.* **1980**, *102*, 939-947.

(80) Andersson, K.; Malmqvist, P. A.; Roos, B. O. *J. Chem. Phys.* **1992**, *96*, 1218-1226.

(81) Andersson, K.; Malmqvist, P. A.; Roos, B. O.; Sadlej, A. J.; Wolinski, K. *J. Phys. Chem.* **1990**, *94*, 5483-5488.

(82) Kohn, A.; Hattig, C. *J. Chem. Phys.* **2003**, *119*, 5021-5036.

(83) Hattig, C. *J. Chem. Phys.* **2003**, *118*, 7751-7761.

(84) Barbatti, M.; Granucci, G.; Persico, M.; Ruckebauer, M.; Vazdar, M.; Eckert-Maksic, M.; Lischka, H. *Journal of Photochemistry and Photobiology a-Chemistry* **2007**, *190*, 228-240.

(85) Dallos, M.; Lischka, H.; Shepard, R.; Yarkony, D. R.; Szalay, P. G. *J. Chem. Phys.* **2004**, *120*, 7330-7339.

(86) Lischka, H.; Dallos, M.; Szalay, P. G.; Yarkony, D. R.; Shepard, R. *J. Chem. Phys.* **2004**, *120*, 7322-7329.

(87) Lischka, H.; Dallos, M.; Shepard, R. *Mol Phys* **2002**, *100*, 1647-1658.

(88) Shepard, R.; Lischka, H.; Szalay, P. G.; Kovar, T.; Ernzerhof, M. *J. Chem. Phys.* **1992**, *96*, 2085-2098.

- (89) Shepard, R. In *Moder Electronic Structure Theory*; Yarkony, D. R., Ed.; World Scientific: Singapore, 1995; Vol. 1, p 345.
- (90) Lischka, H.; Shepard, R.; Pitzer, R. M.; Shavitt, I.; Dallos, M.; Muller, T.; Szalay, P. G.; Seth, M.; Kedziora, G. S.; Yabushita, S.; Zhang, Z. Y. *Physical Chemistry Chemical Physics* **2001**, *3*, 664-673.
- (91) Lischka, H.; Shepard, R.; Brown, F. B.; Shavitt, I. *Int. J. Quantum Chem.* **1981**, 91-100.
- (92) Lischka, H.; Shepard, R.; Shavitt, I.; Pitzer, R.; Dallos, M.; Muller, T.; Szalay, P. G.; Brown, F. B.; Ahlrichs, R.; Boehm, H. J.; Chang, A.; Comeau, D. C.; Gdanitz, R.; Dachsel, H.; Ehrhardt, C.; Ernzechhof, J. H.; P. Irlle, S.; Kedziora, G. S.; Kovar, T.; Parasuk, V.; Pepper, M. J.; Sharf, P.; Shiffer, H.; Schindler, M.; M., S. S., M.; Stahlberg, E. A.; Zhao, J.-G.; Yabushita, S.; Zhang, Z.; Barbatti, M.; Matsika, S.; Schuurmann, M.; Yarkony, D. R.; Brozell, S. R.; Beck, E. V.; Blaudeau, J.-P. 2006.
- (93) Karlstrom, G.; Lindh, R.; Malmqvist, P. A.; Roos, B. O.; Ryde, U.; Veryazov, V.; Widmark, P. O.; Cossi, M.; Schimmelpfennig, B.; Neogrady, P.; Seijo, L. *Computational Materials Science* **2003**, *28*, 222-239.
- (94) Malmqvist, P. A.; Rendell, A.; Roos, B. O. *J. Phys. Chem.* **1990**, *94*, 5477-5482.
- (95) Roos, B. O.; Taylor, P. R. *Chem. Phys.* **1980**, *48*, 157-173.
- (96) Ahlrichs, R.; Bar, M.; Haser, M.; Horn, H.; Kolmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165-169.
- (97) M. Barbatti, M.; Granucci, G.; Lischka, H.; Ruckebauer, M.; Persico, M. 2007.

- (98) Csonka, I. P.; Szepes, U.; Modelli, A. *J. Mass Spectrom.* **2004**, *39*, 1456-1466.
- (99) Kang, H.; Lee, K. T.; Jung, B.; Ko, Y. J.; Kim, S. K. *J. Am. Chem. Soc.* **2002**, *124*, 12958-12959.
- (100) He, Y. G.; Wu, C. Y.; Kong, W. *J. Phys. Chem. A (USA)* **2003**, *107*, 5145-5148.
- (101) He, Y. G.; Wu, C. Y.; Kong, W. *J. Phys. Chem. A (USA)* **2004**, *108*, 943-949.
- (102) Cremer, D. *Acta Crystallographica Section B-Structural Science* **1984**, *40*, 498-500.
- (103) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354-1358.
- (104) Zechmann, G.; Barbatti, M. *Int. J. Quantum Chem.* **2008**, *108*, 1266-1276.

3. Nucleobases as molecular fossils of prebiotic photochemistry: Excited state dynamics of C2 and C6 substituted purines

Many variations of the canonical nucleobases, such as different derivatives and analogues, would likely have been present in a primordial soup. Alternative combinations of molecular building blocks would conceivably have been possible to form self-replicating RNA like structures. The nucleobases that are involved in replication selectively exhibit short excited state lifetimes which provide high intrinsic stability against otherwise harmful UV photo-damage. The stark difference in response to UV irradiation between these structures and many of the alternative bases suggests the possibility of a photochemical selection of the molecular building blocks of life long before the advent of biological selection. It is thus conceivable that the molecular photo-properties of nucleobases are molecular fossils of the prebiotic chemistry that occurred 4 billion years ago. This chapter considers the excited state dynamics of the purine bases by comparing a full set of canonical and alternative nucleobases formed by oxo- and amino- substitutions of purine.

3. 1. Introduction

The key components of the reproductive machinery are unlikely to have been selected by a biological evolution that requires that very machinery to begin with. Instead it is conceivable that choosing the molecular building blocks of life was mediated by a chemical selection that preceded biology. This hypothesis also implies that molecular properties of nucleobases, preserved unchanged from prebiotic times, can possibly serve as molecular fossils of that prebiotic chemistry.

The nucleobases that are involved in replication exhibit short excited state lifetimes which provide high intrinsic stability against otherwise harmful UV photo-damage.¹⁻³ UV protection comes about when electronic excitation is converted to heat by internal conversion at rates too fast for other more harmful reactive pathways to occur while subsequently safely dissipating the energy to the environment. The canonical nucleobases generally decay in less than 1 ps, orders of magnitude faster than in most other heterocyclic compounds. This property would have been highly advantageous for the first self-replicating molecules in prebiotic times before modern enzymatic repair or the formation of the ozone layer that would later attenuate the high levels of UV radiation penetrating the early atmosphere. The safe elimination of excess electronic energy in the canonical bases is exquisitely sensitive to molecular structure and much slower relaxation is observed in many closely related structures.

These many variations of the canonical nucleobases, such as different derivatives and analogues, would likely have been present in a primordial soup. Alternative combinations of molecular building blocks would conceivably have been possible to form self-replicating RNA like structures. However, while the structural differences appear relatively small, many of the alternative nucleobases have much longer excited state lifetimes, sometimes by orders of magnitude, which would allow a variety of possibly harmful photochemical processes and thus

render them much less photostable. The stark difference in response to UV irradiation between these structures suggests the possibility of a photochemical selection of the molecular building blocks of life long before the advent of biological selection. It is thus conceivable that the molecular photo-properties of nucleobases, which we study now, are relics from part of the prebiotic chemistry that occurred 4 billion years ago.

The rapid decay mechanism that provides photostability has primarily been studied for the canonical bases themselves, providing five data points, and in the case of the purine bases, just two. Purine itself has a long-lived excited state and the substitutions in C₂ and C₆, to form guanine and adenine, dramatically shorten the excited state lifetime. Other variations in those substitutions, however, lead to vastly different excited state dynamics. Therefore, to further understand these processes and the role of the C₂ and C₆ substitutions, we here examine a larger set of nine structures, formed by all combinations of oxo- and amino- derivatives in the C₂ and C₆ position (see Figure 1).

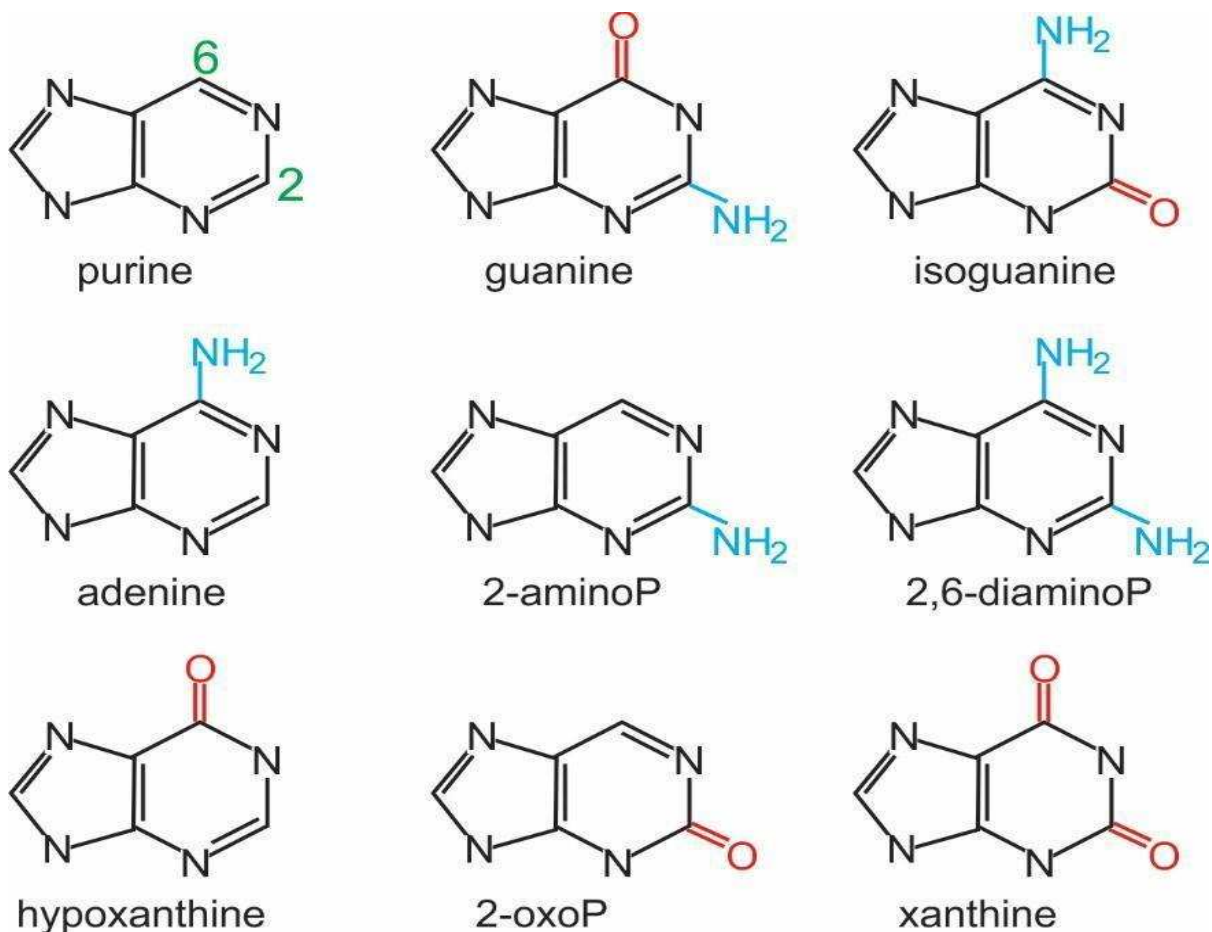


Figure 1: The 9 purine structures formed by all combinations of oxo- and amino- substitutions in the C₂ and C₆ positions. Only one tautomer is shown for each structure.

Each of these purine derivatives could conceivably form alternative base pairs with suitable complementary nucleobases, such as isoguanine with isocytosine, xanthine with 2,4-diaminopyrimidine, or 2,6-diaminopurine with uracil, as shown in Figure 2. Such alternative base pairs have been proposed as part of an alternative genetic lexicon and in various prebiotic scenarios.⁴⁻⁷ Therefore, the fundamental photochemistry needs to be understood both for the canonical bases that make up life as we know it today and for possible alternative bases that could have competed under plausible prebiotic conditions. UV sunlight is an attractive energy source for creating the organic molecules needed to start life on the early earth and not all competing

processes are harmful, particularly when they trigger reactivity for synthetic steps.⁸ Generally, the molecular architecture for replication seems to require rigid structures, which tend to be aromatic and therefore inevitably prone to UV absorption. Since UV excitation is unavoidable for such compounds, it is crucial to understand the possible competing pathways available to process the electronic excitation. At this point, the relaxation dynamics of the canonical nucleobases are well studied and focus has shifted to their derivatives and analogues. The goal is to determine if there is a mechanistic reason for the unique photostability of the canonical bases.

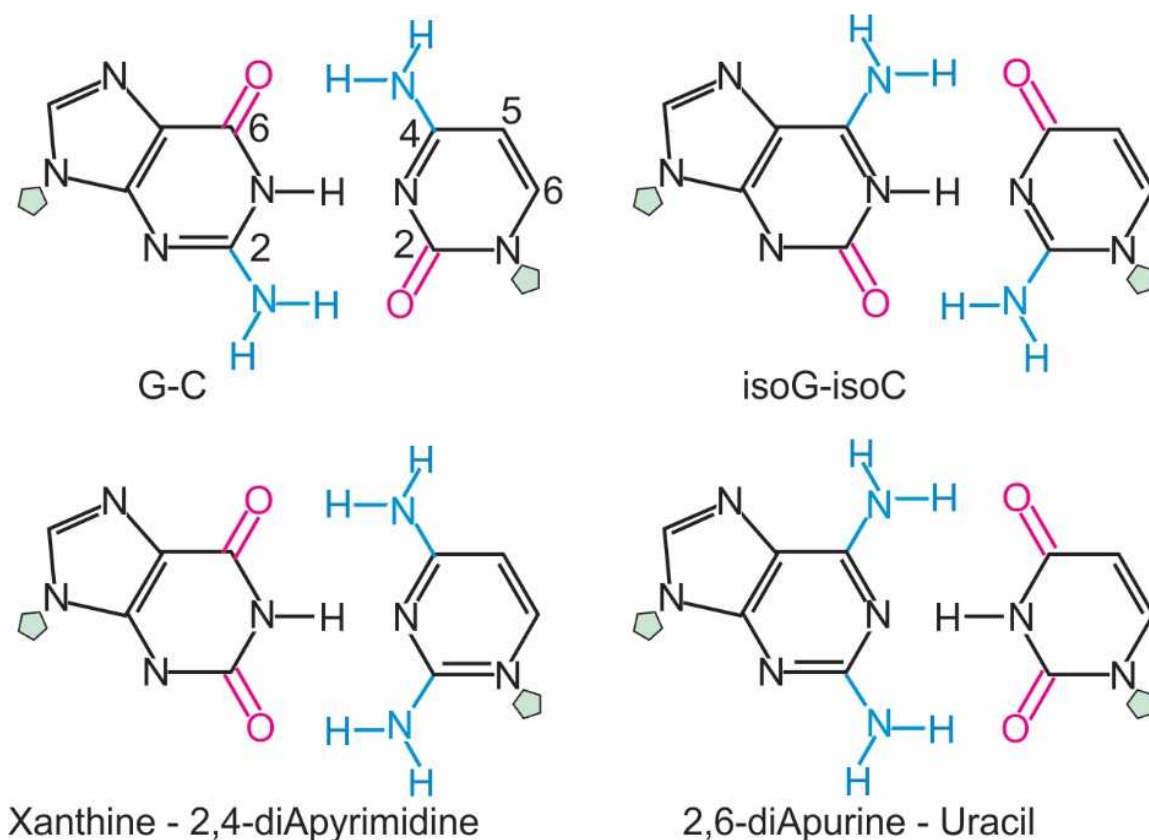


Figure 2: Examples of alternative base pairs formed with purine derivatives, compared with the canonical G-C pair. G=guanine, C=cytosine. Green pentagons indicate the position of the sugar moiety in corresponding nucleosides. The different numbering for purines and pyrimidines is shown in the G-C pair.

Figure 3 shows a schematic Jablonski diagram to illustrate some major competing excited

state decay processes following UV absorption (purple). The only radiative process shown here is fluorescence (blue), which typically involves a timescale of nanoseconds. All the other processes shown are non-radiative. Internal conversion processes (green) can be fast, on the order of ps or less, if trajectories are available to conical intersections (CIs). Lifetimes can become significantly longer if the system is trapped in a minimum below a barrier in the potential energy surface (barriers are schematically indicated in brown). For example, trapping in a bright $^1\pi\pi^*$ minimum may result in fluorescence. Intersystem crossing (red) can be fast, on the order of ps, when El-Sayed allowed (I), or very slow, on the order of ns to μ s, when not el-Sayed allowed (II). Dashed grey arrows indicate intramolecular vibrational redistribution (IVR). If the $n\pi^*$ state is higher in energy than the $\pi\pi^*$ state, the only available pathways would be process i and fluorescence and to some extent ISC to the $^3\pi\pi^*$ state, which would be very slow. For the molecule to be most photostable, it is advantageous for the pathways that return it to the ground state to have the highest rates, competing favourably with processes that lead to ISC or to long times in the excited state.

Such a Jablonsky diagram is an oversimplification since potentials are a function of all internuclear coordinates and multiple trajectories are possible over complex multidimensional potential energy surfaces (PESs). Not only are the relative energies of S_1 , S_2 , T_1 , and T_2 important, but also the detailed shapes of the corresponding PESs, including location of minima, barriers, and conical intersections and seams. Especially the coordinates around C_2 and C_6 may affect the role of substituents in those positions. Conical intersections, required for fast internal conversion, involve distortions of the molecular frame relative to equilibrium structures. Several motifs have been discussed in the literature for the structure of CIs in purines, especially out of plane distortion at C_2 and sometimes C_6 .⁹⁻¹¹ Alternatively, substitution at C_6 has also been discussed as possibly affecting barriers on trajectories toward CIs.^{12, 13} In the case of pyrimidines, distortion of the C_5 -

C₆ bond often leads to a CI but for purines the additional 5 membered ring “flattens” the structure so this motif does not play a role. Furthermore, the potential energy landscape is also affected by tautomeric form^{14, 15} and solvent environment, further complicating relaxation pathways and photostability.^{16, 17}

In the following, we will assess the role of the oxo- and amino- groups at C₂ and C₆ by systematically examining all 9 derivative structures. To elucidate the intrinsic photochemical properties, we focus on gas phase data, obtained by us and others, and their comparison with high level computational theory. Gas phase data don't describe the entire dynamics because the solvent also plays a role, but they provide insight in the most fundamental intrinsic properties of the chromophore, provide the best comparison to the highest level computations, and allow for tautomer selection. In some cases, we will refer to solution data to discuss the additional role of the solvent or when gas phase data are scarce. As some of these systems have multiple tautomers present, both in solution and in the gas phase, we will primarily focus on the biologically relevant tautomer and briefly discuss other tautomers where the data are available, to highlight how tautomerization can affect the excited state dynamics.

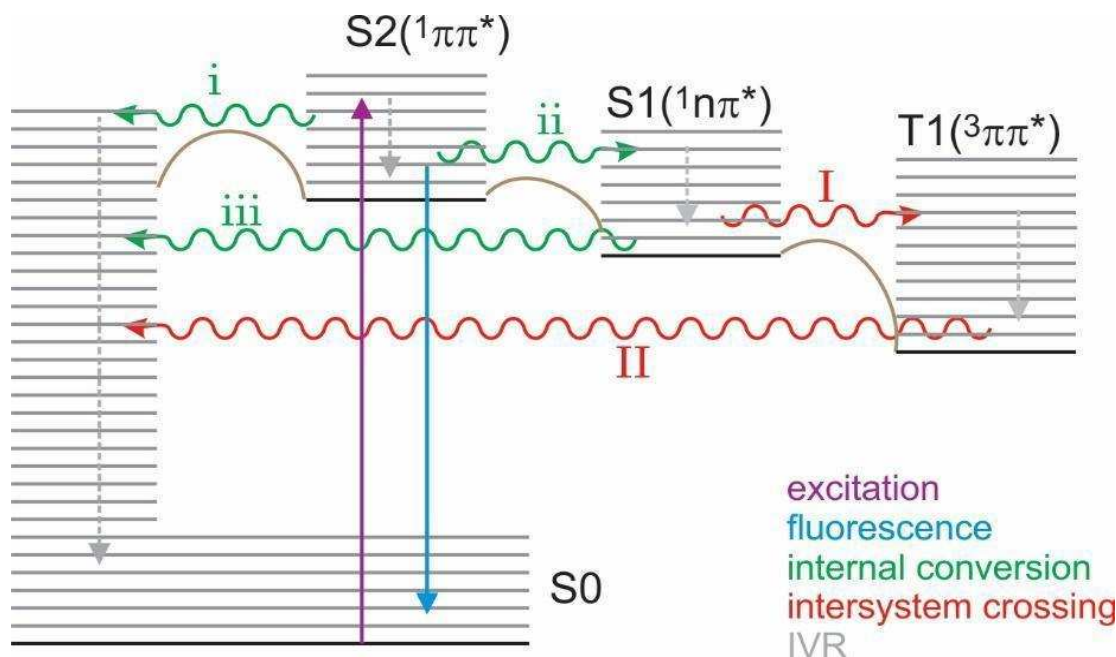


Figure 3: Schematic Jablonski diagram, indicating major competing excited state decay processes. All internal conversion processes (green) can be fast if trajectories are available to conical intersections. Lifetimes can become significantly longer if the system is trapped in a minimum below a potential barrier (barriers are indicated in brown). Intersystem crossing (red) can be fast when El Sayed allowed (I) or very slow when not El-Sayed allowed (II). If the $n\pi^*$ is higher in energy than the $\pi\pi^*$ state, only process i and fluorescence are possible while ISC to the $^3\pi\pi^*$ state would be very slow.

3. 2. Techniques

The photodynamics of nucleobases is typically studied either in the gas phase or in solution. Although the role of the solvent is important, there are four reasons why gas phase studies are useful. (i) The absence of intermolecular interactions reveals intrinsic properties of the chromophore, (ii) comparison with high level computations is facilitated, (iii) isomer selective spectroscopy is possible, and (iv) in combination with jet-cooling high resolution can be obtained. A major gas phase spectroscopy technique is resonance enhanced multiphoton ionization (REMPI), while photoelectron spectroscopy,¹⁸⁻²⁰ IR,^{21, 22} and microwave absorption²³⁻²⁶ have also been employed.

In REMPI, commonly combined with laser desorption and jet-cooling, a first photon excites the molecule resonantly to an excited state and a second photon ionizes the excited molecule. The ions are detected in a mass spectrometer. Varying the time between the two photons – the pump and the probe pulse - allows the measurement of excited state lifetimes. Various double resonant techniques, such as IR-UV hole burning, allow identification and selection of individual isomers. This makes it possible to obtain excited state lifetimes of selected tautomers, cluster structures, or conformers.^{3, 27-29} We note that REMPI is a form of action spectroscopy in which the excited state population is observed through its photoionization. Therefore, small ionization efficiency leads to small or absent signal. This is especially important if the excited state lifetime is significantly shorter than the laser pulse width, preventing ionization. As a result, some states may only be observed with short laser pulses and the technique can be blind to very short-lived excited states.

Two major techniques in solution are excited state absorption, both steady state and transient, and fluorescence upconversion. Both techniques employ a pump pulse, usually from a fs laser, to excite the system, followed by monitoring the wavepacket evolution after absorption through the excited state. Fluorescence upconversion is mostly limited to monitoring the $^1\pi\pi^*$ states, as emission is predominately from these states. One benefit of solution is the, often employed, femtosecond resolution since narrow line width is not as important.

High-level theoretical calculations significantly aid in the analysis of these experimental results. There are a number of reviews and resources that present an in-depth review of theoretical background formulation and methods, specifically detailing strengths and weaknesses.³⁰⁻³⁴ The foundation of the ab initio theoretical approach is the solution to the time independent Schrodinger Equation as a complete description of the nuclei and the electrons. Any practical solution requires

approximations, especially to treat electron correlation and excited states.

Full Configuration Interaction (CI) best describes electron correlation because of its expansive description of multiple permutations of the electrons within virtual orbitals. This method can accurately describe the electron correlation, however, it is prohibitively expensive for systems of any significant size. Attempts to extend the application to reasonable systems include truncating the determinants by limiting the electrons to valence electrons and only including certain transitions, many body perturbation theory (MPn methods),³⁵ and using a cluster operator (Couple Cluster (CC) methods).³⁶ Density Functional Theory (DFT)³⁷⁻⁴⁰ attempts to determine the electron correlation, not by adding more excited state determinants, but by using the density of the electrons to enhance the wavefunction. This method is then corrected with coefficients parameterized with experimental values to produce hybrid functionals, which can give relatively good descriptions inexpensively.

However, these methods are challenged in describing molecules within the excited state. CIS (Configuration Interaction with single excitation) gives a basic description approximating the excitation via a singly excited determinant without including the ground state determinant.⁴¹ But, in the excited state, it does not include electron correlation and is often not accurate. To correct this, other methods solve the time-dependent Schrodinger equation instead of the time-independent Schrodinger equation. This produces TD-DFT approaches, which gives a reasonable approximation of excited state phenomena.^{42, 43} However, TDDFT does not properly describe reactions within the excited state potential energy surface of a molecule.^{44, 45} For these circumstances, Complete Active Space SCF (CASSCF) becomes a more appropriate choice. CASSCF, uses the SCF determinants for the core electrons, but uses the most accurate Full CI for small groups of specific electrons.⁴⁶ The more electrons and orbitals are included, however, the

more expensive the calculations. Therefore, great care must be taken to determine which electrons and which orbitals are included. The benefit of CASSCF is that the method better describes the potential energy surface of the molecule and is better equipped to study conical intersections and other calculations that require following PES reaction pathways in the excited state. most widely used and simplest way to study excited states.

Often, configuration interaction techniques are used for single point calculations at important geometries (Frank-Condon, minima, CIs). Then, complete active space (CAS) is used to calculate electronic energies and pathways between these points to determine relaxation pathways.

A note about fitting decay traces: Lifetimes can only be fit for dynamics that are about half the cross-correlation of the pulse widths of the pump/probe lasers or longer. The experiment is essentially blind to faster dynamics because they occur on a time resolution faster than the pulse width of the measurement. Furthermore, derived lifetimes can vary between studies, even when for the same system with the same technique. These differences can be due to many factors, from sample preparation to the fitting algorithm used. A single trace can sometimes be fit with different lifetimes and different numbers of exponentials depending on the pre-exponential factors.⁴⁷ Thus, lifetimes need to be interpreted with some caution to simply get a sense of the overall orders of magnitude and photostability of the system.

3. 3. Excited state dynamics

Here we briefly summarize what is known about the excited state dynamics for each of the nine compounds. Of the set of compounds, we have studied guanine, adenine, isoguanine, xanthine, and 2,6-diaminopurine. Together with data from many other groups, a fairly extensive

picture emerges. Table 1 lists the excited state lifetimes as measured at the origin of the lowest $^1\pi\pi^*$ transition.

compound	2	6	tautomer	lifetime (ps)	1	references
purine	H	H		long	π^*	48
guanine	N	O	N1 keto	<1	π	49, 50
	H ₂		enol	13,000/ 40,000	π^*	51
isoguanine	O	N	N3,7 keto	>900	π^*	52
	H ₂		enol	<30	π^*	
adenine	H	N		9	π^*	53
	H ₂					
2-AP	N	H		156	π^*	16
	H ₂					
2,6-dAP	N	N		655/6,3	π^*	this work
	H ₂	H ₂		00		
hypoxanthine	H	O		<1	π	54
X					π^*	

2- oxoP	O	H			π π^*	¹³
xanth ine	O	O		30 ps/2 μ s	n π^*	this work

Table 1: Excited state lifetimes as measured at the origin of lowest $\pi\pi^*$ transition.

3. 3. 1. Purine

Purine is the base structure and chromophore for all nine compounds. However, without substituents purine has a long-lived excited state, so the purine base structure itself is not responsible for the ultrafast relaxation dynamics of the canonical purines.

In the gas phase, purine appears exclusively as the 9H tautomer.⁵⁵ Schneider et al. have reported an R2PI spectrum and determined the lowest electronic state to be of $^1n\pi^*$ character.⁴⁸ Individual bands in the spectrum appear up to +4000 cm^{-1} , past the $^1n\pi^*$ origin, though they become diffuse after +2000 cm^{-1} . By comparing with deuterated purine, the authors attribute the low vibronic transitions to skeletal motions, as opposed to hydrogen atom vibrations. The narrow linewidths over a long range in the R2PI spectrum suggest a long-lived excited state. The excited state lifetime of purine has not been measured in the gas phase, but was found to be 360 ps and 645 ps in acetonitrile and aqueous solution, respectively. Solution phase experiments are complicated by the near equal populations of the 7H and 9H tautomers.

Crespo-Hernandez et al. studied purine and 9-methylpurine with transient absorption spectroscopy, coupled with quantum calculations.¹² They compared the dynamics to that of 9-methylpurine because that can help determine differences in the dynamics of purine's two

tautomers.^{56, 57} Their calculations assign the lowest electronic states to a dark $S_1(^1n\pi^*)$ and a bright $S_2(^1\pi\pi^*)$ for both tautomers. After UV absorption to $S_2(^1\pi\pi^*)$, both systems internally convert to $S_1(^1n\pi^*)$ in ~ 100 fs.^{9, 58} Vibrational cooling then occurs to the relaxed $^1n\pi^*$ state in 6-15 ps. Once in S_1 , the system bifurcates to S_1/S_0 internal conversion and $S_1(^1n\pi^*)/T_2(^3\pi\pi^*)$ intersystem crossing. The triplet yield is expected to be near unity. This predominance of the triplet state is supported by other solution phase experiments.⁵⁹ The authors argue that amino- or oxo-substitution at C_6 prevents $^1n\pi^*$ access, promoting $^1\pi\pi^*/S_0$ internal conversion via a C_2 distortion conical intersection instead.¹²

3. 3. 2. *oxopurines*

2-oxopurine was seen to absorb in solution through two transitions at 321 nm and 239 nm respectively.⁶⁰ Computationally, Mburu and Mitsika characterized both transitions to states of $^1\pi\pi^*$ character⁹, while Martinez-Fernandez et al. found two dark $^1n\pi^*$ states in between.¹³ The bright $S_1(^1\pi\pi^*)$ state with a minimum 3.4 eV above the ground state minimum has a 0.6 eV barrier separating it from a $S_1(^1\pi\pi^*)/S_0$ CI. The CI is characterized by stretching of the C_2 -O and N_3 - C_4 bonds, puckering of the C_6 atom, and tilting of the H atom at C_6 . Calculations of spin orbit coupling show that S_1 also intersects with two triplet states of $^3\pi\pi^*$ character with coupling large enough for relaxation to prefer intersystem crossing over internal conversion to the ground state.¹⁷ These dynamics are proposed in two theoretical studies, but have yet to be confirmed by gas phase experimental data.

6-oxopurine or hypoxanthine is adenine with an oxo- group instead of an amino- group. Hypoxanthine can function as a rare RNA base in transfer RNA and exhibits extremely short excited state lifetime in solution.⁶¹ Lin et al. characterized the electronic states by photoelectron spectroscopy as the $S_1(^1\pi\pi^*)$ and higher energy $S_2(^1n\pi^*)$ states.⁶² No other gas phase experiments

have been reported, but several solution phase data and computational models are available. Rottger et al.,⁶¹ using fluorescence up-conversion and transient absorption spectroscopy, reported a biexponential decay $\ll 1$ ps. Because the hypoxanthine spectrum is a combination of the 7H and 9H tautomers, they studied inosine as well. Based on the similarity of the overall hypoxanthine spectrum to that of inosine, both tautomers must have similar ultrafast decays. Villabona-Monsalve et al. and Chen and Kohler also found similar ultrafast decays with a lifetime of ~ 200 fs for 7H and 9H hypoxanthine, though these investigators described them by monoexponentials.^{54, 63}

Theoretical calculations by Guo and co-workers support the lifetime assignment of ~ 200 fs.⁶⁴ They found two different conical intersections for each tautomer due to out-of-plane distortions of the C₂ atom. Both conical intersections reported were lower in energy than the Frank-Condon region with no barriers, allowing hypoxanthine to decay directly from its bright S₁(¹ $\pi\pi^*$) to the ground state. Solvation slightly affects the dynamics of the two tautomers,⁶⁵ slightly decreasing the lifetime of 7H and increasing the lifetime of 9H, though their decays are still predicted to be < 1 ps. They also confirmed the ultrafast dynamics observed experimentally in solution for inosine by modelling the dynamics of 9-methylhypoxanthine in solution.⁶⁶

2,6-dioxopurine or xanthine is a purine derivative with oxo- moieties on both C₂ and C₆. It is the combination of 2-oxopurine and hypoxanthine. Xanthine is found in the human body, as a degradation product of adenine and guanine. There is strong evidence that xanthine existed on an early Earth, as it was found on meteorites⁶⁷ and synthesized using prebiotic chemistry.^{68, 69}

Xanthine was first studied spectroscopically in the gas phase by Callahan et al.⁷⁰ They collected an R2PI spectrum and determined that only the lowest energy 7H tautomer was present. A combined theoretical and experimental study by Gate et al. further studied the photodynamics of 7H and 9-methylxanthine in the gas phase, with 9-methylxanthine serving as a substitute for the 9H tautomer.⁷¹ 7H and 9-methylxanthine had similar dynamics with a < 30 ps and an ultraslow decay component observed at all vibronic transitions (Fig. X4). These relaxation mechanisms were attributed to prompt internal conversion from $S_2(^1\pi\pi^*)/S_1(^1n\pi^*)$ which then overcomes a slight barrier to decay to the ground state for the fast decay or ISC from the $^1n\pi^*$ state to a long-lived triplet state, explaining the ultraslow decay.

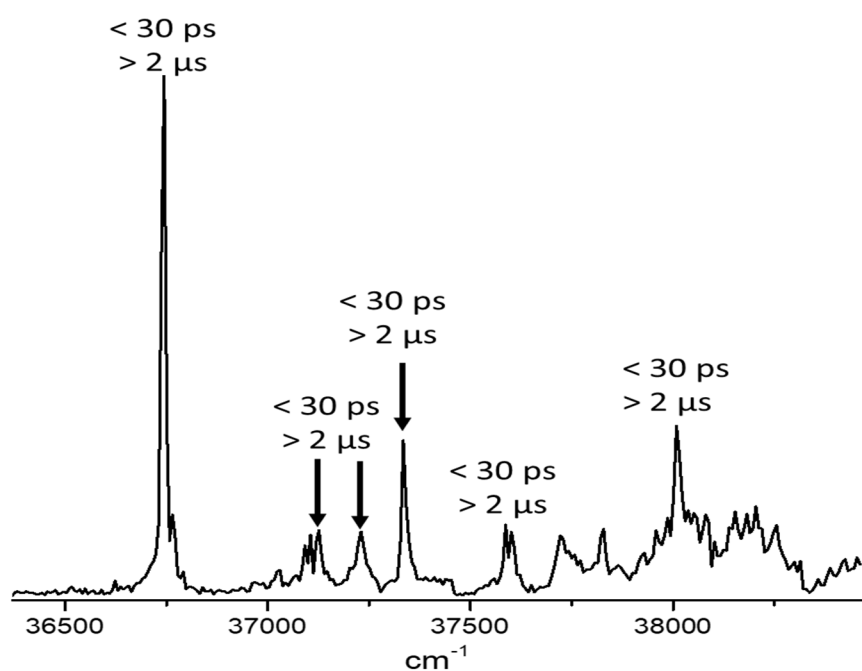


Figure 4. R2PI spectrum of 7H xanthine. Lifetimes for each peak are displayed above the respective peak. Similar decays were found for 9-methylxanthine.

Yamazaki et al. explain the lack of observed 9H tautomer in the gas phase by its poorer Frank-Condon factors and worse stability relative to 7H.⁷² For 7H they found a near barrierless path from the Frank-Condon region of the bright $^1\pi\pi^*$ state to a conical intersection defined by deformation of the 5-membered ring. The 9H exhibited a much larger barrier to reach the conical

intersection. This deformation of the 5-membered ring to reach the conical intersection stands in contrast to all other purine relaxation mechanisms that involve deformation of the 6-membered ring. The authors attribute this to the occurrence of only one double bond, across C₄C₅, in the 6-membered ring. Also, the ¹nπ* state was close in energy to the ¹ππ*, possibly allowing for significant vibronic coupling into the ¹nπ* state.

Xanthine does not dissolve in solution, so solution phase studies have used various forms of xanthosines and methylated xanthines. Chen and Kohler studied the ultrafast nonradiative decay of hypoxanthine and several methylxanthines and found all molecules to have monoexponential decays with lifetimes < 600 fs.⁵⁴ Gustavsson and co-workers found the lifetimes of various methylated xanthines to all be ~1 ps in both H₂O and methanol.⁷³ Studies on the ribonucleotides XMP and XMP⁻ agree with these results.⁷⁴ The ultrafast nonradiative decays of these various xanthines have been attributed to a bright ¹ππ* state relaxing directly to the ground state, in line with theoretical calculations.^{54, 73, 74}

3. 3. 3. aminopurines

2-aminopurine (2AP), the isomer of adenine with the amino group in the C₂ instead of the C₆ position, has been used for decades as an analogue of adenine in DNA. Its most extensive use has been as an environmental probe due to its high fluorescence in solution, with diminished yields when incorporated in DNA.⁷⁵

2AP was first observed in the gas phase in 2001⁷⁶ and is mainly found as the lowest energy and biologically relevant 9H tautomer.⁷⁷ Significant contributions to the understanding of the excited state dynamics of 2AP come from studies by Leutwyler and coworkers.^{14, 78, 79} They observed 7H-2AP in the molecular beam, ~1600 cm⁻¹ to the red and 1000x weaker than 9H-2AP.¹⁴

They measured lifetimes of 156 ps and $< 5 \mu\text{s}$ at the origin of isolated 2AP.^{16, 78} The ultraslow decay was attributed to $S_1(^1n\pi^*)/T_1(^3\pi\pi^*)$ intersystem crossing. By studying clusters with water they found that microhydration by up to two water molecules increased the lifetime 4-100 fold, up to 14.5 ns, depending on the site of hydration. They correlated this lifetime increase with a greater $^1n\pi^* - ^1\pi\pi^*$ energy gap which led to less vibronic coupling. They observed two isomers of 9H-2AP·H₂O, differing only by their hydrogen bonding network.⁸⁰ This study implied the cluster without the amino group hydrogen-bonding had greater $^1(\pi\pi^*)/^1n\pi^*$ vibronic coupling than the other. Hydration of the amino moiety destabilizes the $n\pi^*$ state, reducing the vibronic coupling and increasing the fluorescence yield and lifetime. Using the ribonucleoside 2APdr, Crespo-Hernandez et al. observed increasing triplet yield of 10 – 40% with decreased solvent polarity.^{17, 75, 77, 81-87} Further, fluorescence decreased with decreasing solvent polarity. This triplet state exists for hundreds of ns, and is most likely populated via $^1\pi\pi^* \rightarrow ^1n\pi^* \rightarrow ^3\pi\pi^*$, confirming the work by Leutwyler and co-workers.

6-aminopurine (6AP) or adenine is found almost exclusively in the biologically relevant 9H tautomer in the gas phase^{53, 88} with a single, weaker peak for the 7H tautomer.⁸⁹ In solution, about 80% is in the 9H configuration and the rest in 7H.^{15, 90} The lowest electronic states are $^1n\pi^*$, $^1L_b(\pi\pi^*)$, $^1L_a(\pi\pi^*)$, and $^1\pi\sigma^*$. Kim et al. first observed adenine in the gas phase, including the dark $^1n\pi^*$ state at 35503 cm^{-1} and the higher energy $^1L_b(\pi\pi^*)$ state at 36105 cm^{-1} . This was observed in helium droplets as well.^{89, 91, 92} Broad absorption above $\sim 36800 \text{ cm}^{-1}$ was assigned to the $^1L_a(\pi\pi^*)$ state.^{92, 93}

Adenine is noted for having an ultrafast decay in the gas phase. Fischer and co-workers in 2001 reported a lifetime of 9 ps at the $^1L_b(\pi\pi^*)$ origin.⁵³ Various fs pump-probe transient spectroscopy and photoelectron spectroscopy studies have concluded that the initially bright states

(1L_b and 1L_a) decay to the dark $^1n\pi^*$ in < 100 fs, and this state decays to the ground state in < 1 ps.^{18, 20, 93-99} Direct $^1\pi\pi^*/S_0$ internal conversion may contribute to this < 100 fs component and there is also experimental evidence for a small long-lived triplet state component.^{18, 94, 95}

Once in the $^1n\pi^*$, adenine undergoes significant ring puckering and out-of-plane deformation at the C_2-N_3 and C_6 to reach a conical intersection and relax back to the ground state.^{81, 100-104} Higher energy excitations to $^1L_a(\pi\pi^*)$ lead to branching of $^1L_a(\pi\pi^*)$ to $^1n\pi^*$ and $^1\pi\sigma^*$. The $^1\pi\sigma^*$ state involves the N_9-H coordinate.^{98, 100, 105, 106} Repulsion of the N_9-H bond in the excited state is stabilized by its bond lengthening, leading to possible H atom ejection. This state is estimated to only be accessible with an excitation 0.5 – 1.0 eV above the vertical excitation of the $^1L_b(\pi\pi^*)$.¹⁰⁰⁻¹⁰² Serrano-Andres et al. and Corral and co-workers proposed an alternative relaxation model.¹² In this model, adenine relaxes via similar structural deformations to reach $^1\pi\pi^*/S_0$ conical intersections, excluding the $^1n\pi^*$.

The 7H tautomer has been noted for a barrier to access the $^1n\pi^*$ conical intersection, which explains its experimentally observed increased lifetime and fluorescence.⁸²

In aqueous solution, adenine is noted for a biexponential decay with lifetimes of ~ 0.2 and 8 ps.^{15, 90, 107, 108} The 8 ps lifetime was shown to be due to the significantly more fluorescent 7H tautomer.¹⁰⁸ Adenosine was found to have a similarly fast lifetime of 0.3 fs.¹⁰⁷⁻¹¹⁰ Some studies find biexponential decays for adenosine, but these lifetimes are still consistently < 0.5 ps. Because adenosine does not have a 7H tautomer, and therefore no long 8 ps decay, adenosine was also used to study the dynamics of intermolecular vibrational cooling in the ground state.¹⁰⁹⁻¹¹¹ Others have studied the dynamics of adenosine in the context of making comparisons to adenine homooligonucleotides,¹¹¹⁻¹¹⁴ finding similar lifetimes for adenosine. Experimentally, it is unclear whether in solution adenine decays using the $^1n\pi^*$ state. Some studies point to its contribution,^{15,}

¹⁰⁸ while others do not.^{20, 112}

2,6-diaminopurine (2,6-dAP) is the combination of 2AP and 6AP, with amino- moieties at both C₂ and C₆. 2,6-dAP likely existed on an early Earth, as it was found in meteorite samples⁶⁷ and could be synthesized under the same conditions and starting materials as the canonical nucleobases.^{115, 116}

2,6-dAP was first observed in the gas phase in 2001.⁷⁶ Gengeliczki et al. determined the presence of both the 7H- and 9H-diamino tautomers in the molecular beam with the 7H origin red-shifted $\sim 2600\text{ cm}^{-1}$ relative to the 9H origin.¹⁰ Using ns pump probe, they measured excited state lifetimes of 8.7 ns and 6.3 ns for the 7H and 9H tautomers, respectively. Recently, Gate and co-workers performed a gas-phase excited state dynamics study using both REMPI and quantum calculations. For the 7H tautomer they observed a sizable barrier to internal conversion from the bright S₁(¹ $\pi\pi^*$) state leading to a ~ 25 ns lifetime, when probed with 213 nm, and significant triplet yield with a $> 2\ \mu\text{s}$ lifetime (Fig. 5). The 9H tautomer has a small barrier to the S₁(¹ $\pi\pi^*$)/S₀ CI, with a small fluorescent quantum yield for the first few vibronic transitions, yielding the 6.3 ns component. A 600 ps decay component was also observed for the first few vibronic bands that reduces to 9 ps at $+800\text{ cm}^{-1}$, due to internal conversion through a conical intersection.

These experimental results are supported by theoretical calculations. The 7H tautomer has a barrier from its Frank-Condon excitation to the S₁/S₀ conical intersection, requiring significant nuclear motion, explaining the 25 ns decay. The intersystem crossing geometry is very similar to the Frank-Condon geometry. This similar geometry, coupled with the amount of time 7H stays in the S₁(¹ $\pi\pi^*$), allows for significant triplet yield and competitive intersystem crossing at all probed wavelengths. The 9H tautomer has a small barrier between the Frank-Condon excitation and the ground state, allowing for a small fluorescent yield and consistent with the faster 6.3 ns decay. The

majority of the wavepacket decays nonradiatively, attributed to the ps decay. With a small amount of excess energy, the barrier is overcome and the fluorescent yield disappears, outcompeted by the much more efficient nonradiative decay. The significant drop in lifetime of the ps decay and broadening of the R2PI spectrum are evidence to this.

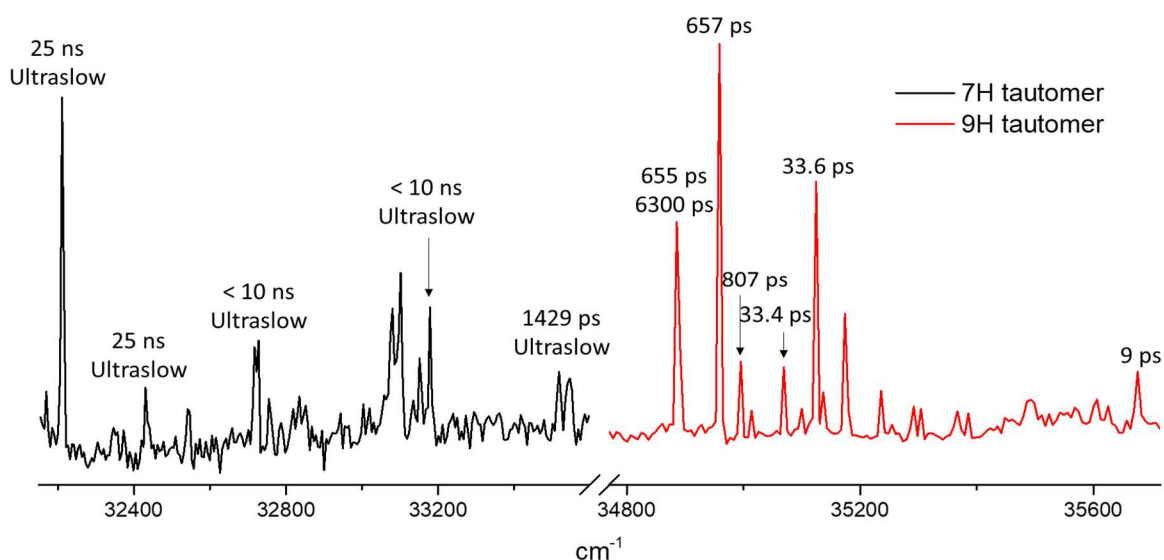


Figure 5. R2PI spectra of the 7H (black) and 9H (red) tautomers of 2,6-dAP. The origins are at 32216 cm^{-1} and 34885 cm^{-1} for the 7H and 9H, respectively. The lifetimes derived for each peak are listed above the peaks. Ultraslow has a decay of $> 2\mu\text{s}$.

Santhosh and Mishra also studied 2,6-dAP in aqueous solution using absorption, fluorescence and ns pump-probe to determine a 4.0 ns fluorescent decay.⁸⁵ Virta et al. also used fluorescent spectroscopy but found a 1.7 ns lifetime with a small quantum yield of 0.02.¹¹⁷ As far as we can tell, no one has determined the tautomer distribution of 2,6-dAP in aqueous solution; based on the trend of 2AP⁸⁶ and 6AP we expect the majority of the dissolved population to be in the 9H tautomer.^{15, 90}

3. 3. 4. amino-oxapurines

Isoguanine is the isomer of guanine with the oxo- and amino- moieties switched. Interest in isoguanine grew after Rich in 1962 suggested isoguanine and isocytosine could form a highly stable Watson-Crick base pair that could biologically be incorporated into DNA.¹¹⁸ Studies have shown that isoguanine exists in multiple different tautomeric forms in DNA^{119, 120} and can basepair with guanine,¹²¹ cytosine,^{119, 122, 123} thymine,¹¹⁹ and isocytosine.^{118, 124} Isoguanine can also form parallel strand helices¹²² and tetraplexes¹²³ with cytosine. Because of the many tautomers of isoguanine, two of the more influential studies by Sepiol et al.¹²⁵ and Seela et al.¹²⁶ focused on assigning the predominant tautomeric forms to the keto-amino and enol-amino forms. Other studies have found evidence for other tautomers, showing that the tautomerization of isoguanine is complex.^{119, 121}

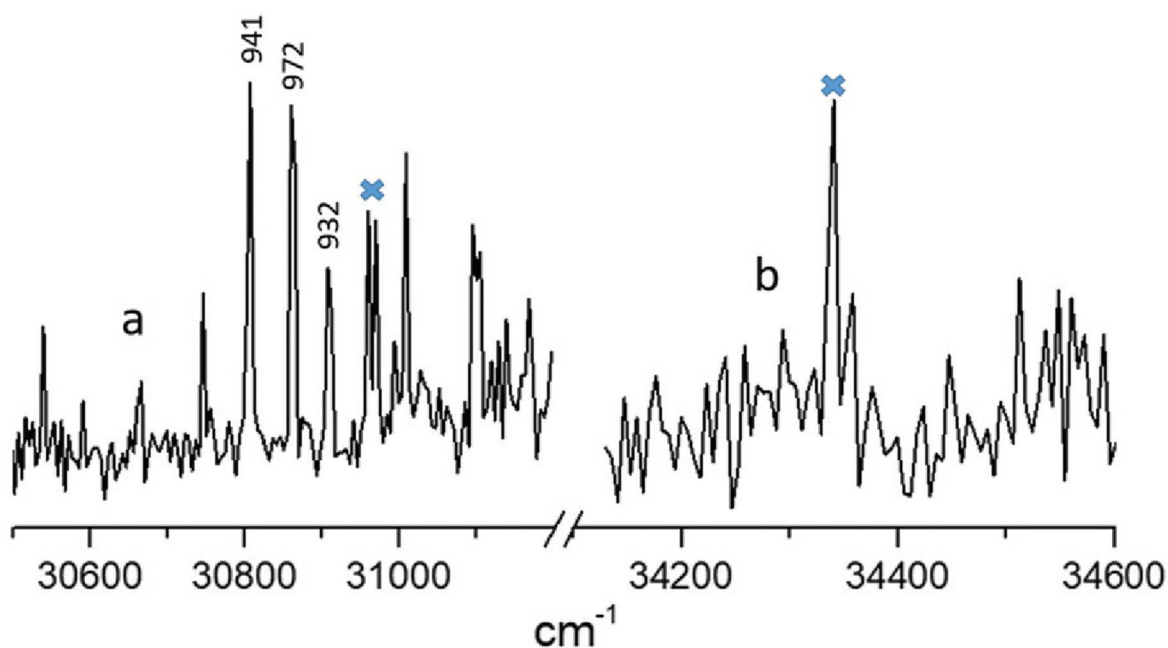


Figure 6. R2PI spectra of keto-N3,7 (a) and enol-N7 (b) isoguanine. Lifetimes, in ps, are displayed above each peak. Peaks where pump probe was unsuccessfully attempted are marked with crosses.

Gate et al. in 2019 observed the keto-N3,7 and enol-N7 forms in the gas phase (Fig. 6).⁵² The keto form was found to have a lifetime around 950 ps at all transitions. The lifetime for the

enol-N9 form could not be measured. Calculations (NEVPT2 for keto tautomers and ADC(2) for enol tautomers) show that the keto-N3,7 $S_1(^1n\pi^*)$ minimum is 0.98 eV below the conical intersection; such a sizable barrier explains the 950 ps lifetime observed. (Fig. 7) Two conical intersections with smaller barriers were computed for the enol-N7 tautomer, explaining the small amount of signal and inability to derive a lifetime. Calculations were also done on the biologically relevant keto-N1,9 and enol-N9, keto-N1,9 had a large barrier to a CI, while enol-N9 had no barrier to reach the $S_1(^1n\pi^*)/S_0$ CI.

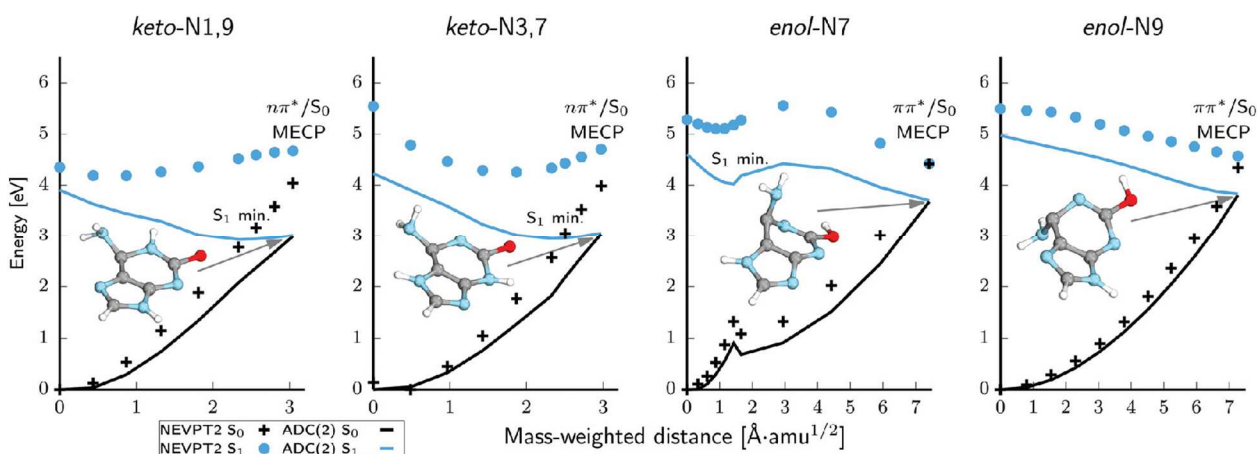


Figure 7. Potential energy profiles for the predicted dominant photodeactivation pathways of keto and enol isoguanine. The inserted structure corresponds to the S_1/S_0 minimum-energy crossing point (MECP) geometry (conical intersection). For a description of the applicability of NEVPT2 and ADC(2), refer to reference ⁵².

Guanine A high resolution gas phase spectrum of guanine was first reported by Nir et al. in 1999,¹²⁷ followed by the vibronic spectrum of guanosine.¹²⁸ IR-UV double resonance experiments, coupled with calculated structure stabilities were used to assign tautomers in the gas phase. The assignment of the four tautomers was initially wrongly assigned to the lowest energy tautomers: keto-9H, keto-7H, enol-9H, and enol-7H^{27, 129, 130} but it was later confirmed that tautomers 3-7 kcal/mol higher in energy are the ones actually observed in R2PI experiments,

namely the enol-9H, enol-7H, and two imino-keto 7H forms.¹³¹⁻¹³³ The keto-9H, keto-7H, enol-9H (trans), and enol-9H(cis) were observed in helium nanodroplets and IR laser spectroscopy, thus also present in the gas phase.^{23, 134} Since R2PI is a form of action spectroscopy, these results suggest that the failure to observe the lowest energy tautomers in R2PI experiments is not indicative of their absence in the molecular beam but of failure to ionize them due to excited state lifetimes that are significantly shorter than the 6 ns laser pulse widths employed in the R2PI experiments. Femtosecond gas phase experiments found < 1 ps lifetimes, but these did not have tautomer selection.^{94, 99} The other tautomers have lifetimes on the order of tens of nanoseconds.^{51,}

130

Theoretical calculations explain that biological guanine (keto-9H) is not observed in molecular beam experiments due to poor Frank-Condon factors.¹³³ It is also predicted to have ultrafast decay due to a barrierless $S_1(1\pi\pi^*)/S_0$ CI from puckering at C_2 . Another barrierless conical intersection opens up via hydrogen-abstraction from a $1\pi\sigma^*$ state due to stabilization of guanine's large static dipole moment.^{49, 50, 133, 135} The $1n\pi^*$ state is negligibly involved in the excited state dynamics due to its high energy. Thiel and co-workers conducted a nonadiabatic dynamics study in simulated aqueous solution and found these dynamics to be faster than in vacuo.¹³⁶ This difference is due to a smaller energy gap between $1L_a(\pi\pi^*)$ and $1L_b(\pi\pi^*)$ and a new conical intersection by out-of-plane motion of the carbonyl group.

Guanine has been scantily studied in solution due to its poor solubility. Instead, solution phase studies have focused on guanosine, GMP, and dGMP. The biologically relevant amino-keto tautomer is found in aqueous solution for guanosine and GMP.¹³⁷ Like adenine, these systems have been noted for two low-energy bright states, the $1L_a(\pi\pi^*)$ and $1L_b(\pi\pi^*)$. For guanine, the state

ordering is reversed though, with ${}^1L_a(\pi\pi^*)$ lower in energy than ${}^1L_b(\pi\pi^*)$. The fluorescence spectrum of GMP is much broader and red-shifted relative to other nucleotides with a tail out to 700 nm.^{110, 138} This red tail is due to the relatively flat ${}^1L_a(\pi\pi^*)$ PES leading to the S_1/S_0 conical intersection. Excitation of the higher energy ${}^1L_b(\pi\pi^*)$ leads to < 100 fs internal conversion to ${}^1L_a(\pi\pi^*)$.^{109, 110, 138-141} The ${}^1L_a(\pi\pi^*)$ then decays along a barrierless, though flat, PES to the S_1/S_0 conical intersection, leading to a < 1 ps lifetime,

3. 4. The role of C₂ and C₆ substitutions

Photostability follows from the ability to return to the electronic ground state at time scales short enough to outcompete all other possible processes. This scenario requires a potential energy landscape that allows barrierless trajectories to efficient conical intersections. The dependence on structure comes about because even minor changes in the PES can affect the positions and levels of minima, CIs, barriers, or ordering of states which can have large effects on the excited state dynamics. The structure dependence is subtle and even tautomers of the same compound can have excited state lifetimes that differ by orders of magnitude. As a result, it is hard to formulate general trends. A number of models have been proposed to generalize the structural effects and each of those describe important elements. However, the complete description still appears to require the inclusion of additional details.

While comparing the effects of different C₂ and C₆ substitutions, it should be noted that the excited state lifetimes are not a single number but are, rather, a function of absorption wavelength. The focus here is on absorption close to the origin, because the excited state dynamics at threshold energies are the most sensitive measure of the topography of the PE landscapes. Moreover, at high excitation energy, above all barriers, deactivation rates tend to increase and thus may not be an

accurate reflection of photostability. For the molecule to rapidly and safely diffuse electronic excitation, no significant part of the spectrum should be trapped in the excited state, including the part absorbed near threshold. In some of the non-canonical bases, such as xanthine and 2AP, ultrafast dynamics occur, but at the same time ISC occurs as well with a significant quantum yield, thus threatening the long-term stability upon radiation. Furthermore, under prebiotic conditions, absorption further to the red in the spectrum may have implied more exposure to damaging radiation.

The base structure of these substitutions is purine. Purine itself is long-lived via the mechanism of a $S_2(^1\pi\pi^*)/S_1(^1n\pi^*)$ CI in 100 fs followed by a near 100% QY for ISC to a long-lived triplet state. So the chromophore by itself is not responsible for the photostability of the purine nucleobases. Substitutions of oxo- or amino- groups at the C_2 and C_6 position modify the dynamics to produce different outcomes for each of the other eight structures, including producing photostability for selected ones. This is not unique to the purine nucleobases as Arpa et al. have pointed out that “The pyrimidine chromophore is not responsible for the photostability of the nucleobases”.¹⁴² In this case C_2 and C_4 functionalization with amino- and/or oxo- substituents modifies the PES landscape, creating ultrafast pathways through CIs to the ground state. For purine based structures, four major motifs have emerged in describing how the excited state dynamics is affected by the potential energy landscapes of purine and its oxo- and amino- derivatives.

(i) The first motif is the order of $\pi\pi^*$ and $n\pi^*$ states which the substituents modify through stabilization/destabilization effects of the added groups.⁹ Generally, the $\pi\pi^*$ state is the bright state. If the $n\pi^*$ is lower in energy in the FC region or its minimum is below that of the $\pi\pi^*$ state, then there may be conical intersections between the two that allow the $n\pi^*$ to be populated and to serve as a doorway state or a dark state.^{102, 143} This is the reported arrangement in all cases, except

guanine, hypoxanthine, 2-oxopurine and enol isoguanine. However, while an accessible $n\pi^*$ state creates a more complex PE landscape, it does not suffice to characterize excited state dynamics. Even when the $n\pi^*$ state is not accessible, there may or may not be a barrierless path to the ground state, as in keto vs. enol guanine. Conversely, when the $^1n\pi^*$ state can be reached, it may subsequently lead to the ground state, as in adenine, or lead to for example intersystem crossing to a $^3\pi\pi^*$ state as in purine and 2-oxopurine, and possibly xanthine, 2,6-diaminopurine, and 2-aminopurine.

(ii) A major motif, in many cases, is C_2 puckering in the CI geometry.^{13, 81} It has been argued that this geometry is responsible for the short lifetime of two structures without a C_2 substituent, adenine and hypoxanthine. The most often cited example is the lifetime difference between adenine and 2-aminopurine. For 2-aminopurine, it has been argued that the out of-plane rotation of the C_2 -amino group takes longer than the C_2 -H bond rotation in adenine, resulting in the longer observed lifetime. However, this trend does not hold for purine which has no C_2 substituent but still has a long lifetime. It is also contradicted by the fact that ketoG and enol isoG with an amino- and oxo- substituent in C_2 , respectively, have fast excited state lifetimes.

(iii) Puckering at C_6 can also define conical intersection geometries to a lesser extent.⁸² For example, it provides an additional deactivation channel in adenine, which becomes significant only at higher excitation energies.¹⁰⁴ C_6 substitution has also been proposed to play a role in reducing the barrier in the $\pi\pi^*$ state. The effect is illustrated with the observation that adenine and hypoxanthine decay fast relative to purine. However, the observations that 2,6-dAP is slower than 2AP and that enol-G and keto-isoG are slow complicate this explanation.

(iv) When considering the fate of photoexcited nucleobases, the possibility of intersystem

crossing is of particular concern. Only rapid return to the ground state, either directly or through a doorway state, effectively transforms electronic excitation to heat but some nonradiative transitions lead to a longer-lived dark state. Purine, 2,6-dAP, and xanthine all have significant ISC with the major pathway involving a $^1n\pi^*/^3\pi\pi^*$ CI, which is El-Sayed allowed. Large spin-orbit coupling, as in the case of 2-oxopurine and 2-aminopurine can also contribute to this strength of this ISC pathway.

While this review is based on insights from gas phase data, other aspects, such as the solvent, of course play a role. A prime example is 2AP, which is known for its long excited state lifetime in solution, but in the gas phase has a 156 ps lifetime at the origin. Lobsiger et al. have shown that it takes only three water molecules for that lifetime to lengthen to 14.5 ns. Computations predict a reduction in the barrier for tautomerisation upon solvation for all canonical nucleobases. For example, a reduction from 190 to 112 kJ/mol and from 230 to 126 kJ/mol was predicted for thymine and cytosine respectively by Valadbeigi et al.¹⁴⁴ Although the barriers are not easy to compute and their values strongly depend on the computation level, all the computational works report that solvation by water reduces the barriers by about ~ 120 kJ/mol.¹⁴⁵ Intrastrand hydrogen bonding can also affect the excited state dynamics. For example, Zeleny et al. found computationally that in the GC base pair the out of plane motions of the guanine NH_2 are restricted, affecting the associated CI.¹⁴⁶ On the other hand, the base pair forms a charge transfer state along the $\text{N}_1\text{-H}$ coordinate, which provides another fast decay trajectory.^{28, 147}

It should also be noted that a longer stay in an excited state may not always need to lead to photochemical damage. An example is the formation of excitons in stacked bases, which return to the ground state by charge recombination at a time scale of 100 ps.¹⁴⁸

One might speculate that the oxo- and amino- substitutions did not only provide the purine and pyrimidine skeletons with the necessary hydrogen bonding structures – for which many combinations were possible – but at the same time also with UV hardness – for which only a subset was preferred. To understand which structures could provide the fastest deactivation dynamics back to the ground state requires careful computations to map out all potential energy surfaces in comparison with detailed experimental data. In putting these observations and trends together, we observe that the dynamics is governed by distortions of the 6 ring and the resulting potentials do depend on the structure around the ring. But structural trends combine in very nuanced ways; the effects of individual substitutions don't simply add up and the end result cannot be intuitively predicted.

3.5. Outlook

While from all the data, described here, an extensive picture emerges of the entire landscape of 2,6 oxo-/amino- substituted purines, a few pieces of the puzzle are still missing. Notably, no gas phase data are yet available for 2-oxopurine, hypoxanthine, and keto guanine. For the latter two, presumably the sub-picosecond excited state lifetime is too short for REMPI detection. Alternative approaches, such as population of the hot ground state following internal conversion may provide additional data.

Possible next steps may include the following considerations. (i) When considering prebiotic scenarios, one may consider recent proposals that the difficulty of plausible glycosylation pathways for the nucleobases suggest the possibility that nucleosides were synthesized directly to form the first building blocks.¹⁴⁹ Following this work on nucleobases, it will therefore be important to study the nucleoside photochemistry, some of which has already been reported for the canonical

nucleosides.^{128, 150} (ii) Some of the effects of hydrogen-bonding in base pairs and in water have been mentioned above, but much more work is called for, especially on clusters with water. (iii) Another possible structural motif that leads to intermolecular interactions is pi-stacking. Upon excitation, stacking can lead to exciplexes, which provides additional photochemical pathways. Some of those dynamics has been studied in solution^{111, 113} but not in the gas phase because nucleobases tend only to form hydrogen bonded structures in the gas phase.^{112, 151-154} (iv) In addition to substitutions, it will be of interest to the prebiotic community to consider the photostability of analogues, such as triazines.⁷

This field has been progressing, and continues to do so, thanks to parallel developments in experiments and in theory, allowing the study of molecules of increasing size in the gas phase and with ever higher levels of computational theory. Finally, in addition to the prebiotic importance of photostability of the molecular building blocks of life, complete understanding of the structure dependence of photostability might one day help in the design of compounds with desired photochemical properties.

References

1. A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, *Proc. Natl. Acad. Sci. U.S.A.*, 2005, **102**, 20-23.
2. R. Weinkauf, J. P. Schermann, M. S. de Vries and K. Kleinermanns, *European Physical Journal D*, 2002, **20**, 309-316.
3. E. Nir, K. Kleinermanns and M. S. de Vries, *Nature*, 2000, **408**, 949-951.
4. S. Boldissar and M. S. de Vries, *Phys Chem Chem Phys*, 2018, **20**, 9701-9716.
5. D. Hutter and S. A. Benner, *J. Org. Chem.*, 2003, **68**, 9839-9842.
6. S. Hoshika, N. A. Leal, M. J. Kim, M. S. Kim, N. B. Karalkar, H. J. Kim, A. M. Bates, N. E. Watkins, H. A. SantaLucia, A. J. Meyer, S. DasGupta, J. A. Piccirilli, A. D. Ellington, J. SantaLucia, M. M. Georgiadis and S. A. Benner, *Science*, 2019, **363**, 884-+.
7. N. V. Hud, B. J. Cafferty, R. Krishnamurthy and L. D. Williams, *Chem Biol*, 2013, **20**, 466-474.
8. M. W. Powner, B. Gerland and J. D. Sutherland, *Nature*, 2009, **459**, 239-242.
9. E. Mburu and S. Matsika, *J. Phys. Chem. A*, 2008, **112**, 12485-12491.
10. Z. Gengeliczki, M. P. Callahan, N. Svadlenak, C. I. Pongor, B. Sztaray, L. Meerts, D. Nachtigallova, P. Hobza, M. Barbatti, H. Lischka and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2010, **12**, 5375-5388.
11. D. Nachtigallova, M. Barbatti, J. J. Szymczak, P. Hobza and H. Lischka, *Chem. Phys. Lett.*, 2010, **497**, 129-134.
12. C. E. Crespo-Hernandez, L. Martinez-Fernandez, C. Rauer, C. Reichardt, S. Mai, M. Pollum, P. Marquetand, L. Gonzalez and I. Corral, *J Am Chem Soc*, 2015, **137**, 4368-4381.
13. L. Martinez-Fernandez, S. Arslançan, D. Ivashchenko, C. E. Crespo-Hernandez and I. Corral, *Phys Chem Chem Phys*, 2019, **21**, 13467-13473.
14. S. Lobsiger, R. K. Sinha, M. Trachsel and S. Leutwyler, *J. Chem. Phys.*, 2011, **134**.

15. B. Cohen, P. M. Hare and B. Kohler, *J. Am. Chem. Soc.*, 2003, **125**, 13594-13601.
16. S. Lobsiger, S. Blaser, R. K. Sinha, H. M. Frey and S. Leutwyler, *Nat Chem*, 2014, **6**, 989-993.
17. C. Reichardt, C. W. Wen, R. A. Vogt and C. E. Crespo-Hernandez, *Photoch. Photobio. Sci.*, 2013, **12**, 2203-2203.
18. S. Ullrich, T. Schultz, M. Z. Zgierski and A. Stolow, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2796-2801.
19. A. Mohamadzade, S. Bai, M. Barbatti and S. Ullrich, *Chem. Phys.*, 2018, DOI: 10.1016/j.chemphys.2018.08.011.
20. V. R. Smith, E. Samoylova, H. H. Ritze, W. Radloff and T. Schultz, *Phys. Chem. Chem. Phys.*, 2010, **12**, 9632-9636.
21. R. Casaes, R. Provencal, J. Paul and R. J. Saykally, *J. Chem. Phys.*, 2002, **116**, 6640-6647.
22. M. R. Viant, R. S. Fellers, R. P. Mclaughlin and R. J. Saykally, *J. Chem. Phys.*, 1995, **103**, 9502-9505.
23. J. L. Alonso, I. Pena, J. C. Lopez and V. Vaquero, *Angew Chem Int Edit*, 2009, **48**, 6141-6143.
24. M. E. Sanz, S. Blanco, J. C. Lopez and J. L. Alonso, *Angew Chem Int Edit*, 2008, **47**, 6216-6220.
25. V. Vaquero, M. E. Sanz, J. C. Lopez and J. L. Alonso, *J. Phys. Chem. A*, 2007, **111**, 3443-3445.
26. J. C. Lopez, M. I. Pena, M. E. Sanz and J. L. Alonso, *J. Chem. Phys.*, 2007, **126**.
27. E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *J. Chem. Phys.*, 2001, **115**, 4604-4611.
28. E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2002, **4**, 732-739.
29. E. Nir, I. Hünig, K. Kleinermanns and M. S. de Vries, *ChemPhysChem*, 2004, **5**, 131-137.
30. C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models* Wiley, 2nd Edition edn.
31. F. Jensen, *Introduction to Computational Chemistry* Wiley, 3rd Edition edn.

32. J. B. Foresman and Æ. Frisch, *Exploring Chemistry with Electronic Structure Methods*, (Gaussian, Inc., , Wallingford, CT, 3rd edition edn., 2015.
33. C. Adamo and D. Jacquemin, *Chem. Soc. Rev.*, 2013, **42**, 845-856.
34. H. B. Schlegel and J. J. McDouall, in *Computational Advances in Organic Chemistry*, eds. C. Ögretir and I. G. Csizmadia, Kluwer Academic, the Netherlands, 1991, DOI: 10.1007/978-94-011-3262-6, pp. 137-185.
35. C. Moller and M. S. Plesset, *Phys Rev*, 1934, **46**, 618-622.
36. R. J. Bartlett and G. D. Purvis, *Int. J. Quantum Chem.*, 1978, **14**, 561-581.
37. W. Kohn, A. D. Becke and R. G. Parr, *J. Phys. Chem.*, 1996, **100**, 12974-12980.
38. T. Ziegler, *Chem. Rev.*, 1991, **91**, 651-667.
39. P. Hohenberg and W. Kohn, *Phys Rev B*, 1964, **136**, B864-+.
40. W. Kohn and L. J. Sham, *Phys Rev*, 1965, **140**, 1133-&.
41. M. Headgordon, R. J. Rico, M. Oumi and T. J. Lee, *Chem. Phys. Lett.*, 1994, **219**, 21-29.
42. R. Bauernschmitt and R. Ahlrichs, *Chem. Phys. Lett.*, 1996, **256**, 454-464.
43. A. Dreuw and M. Head-Gordon, *Chem. Rev.*, 2005, **105**, 4009-4037.
44. D. J. Tozer and N. C. Handy, *J. Chem. Phys.*, 1998, **109**, 10180-10189.
45. Y. H. Shao, M. Head-Gordon and A. I. Krylov, *J. Chem. Phys.*, 2003, **118**, 4807-4818.
46. F. Bernardi, A. Bottoni, J. J. W. Mcdouall, M. A. Robb and H. B. Schlegel, *Faraday Symp Chem S*, 1984, 137-147.
47. M. R. Haggmark, G. Gate, S. Boldissar, J. Berenbeim, A. L. Sobolewski and M. S. de Vries, *Chemical Physics*, 2018, **515**, 535-542.
48. M. Schneider, T. Hain and I. Fischer, *ChemPhysChem*, 2009, **10**, 634-636.
49. M. Barbatti, J. J. Szymczak, A. J. A. Aquino, D. Nachtigallova and H. Lischka, *J. Chem. Phys.*, 2011, **134**, 014304.

50. S. Yamazaki, W. Domcke and A. L. Sobolewski, *J. Phys. Chem. A*, 2008, **112**, 11965-11968.
51. F. M. Siouri, S. Boldissar, J. A. Berenbeim and M. S. de Vries, *J Phys Chem A*, 2017, **121**, 5257-5266.
52. G. Gate, R. Szabla, M. R. Haggmark, J. Sponer, A. L. Sobolewski and M. S. de Vries, *Phys. Chem. Chem. Phys.*, 2019, **21**, 13474-13485.
53. D. C. Luhrs, J. Viallon and I. Fischer, *Phys. Chem. Chem. Phys.*, 2001, **3**, 1827-1831.
54. J. Q. Chen and B. Kohler, *Phys. Chem. Chem. Phys.*, 2012, **14**, 10677-10682.
55. W. Caminati, G. Maccaferri, P. G. Favero and L. B. Favero, *Chem. Phys. Lett.*, 1996, **251**, 189-192.
56. N. C. Gonnella and J. D. Roberts, *J. Am. Chem. Soc.*, 1982, **104**, 3162-3164.
57. M. Majoube, P. Millie, L. Chinsky, P. Y. Turpin and G. Vergoten, *J. Mol. Struct.*, 1995, **355**, 147-158.
58. A. C. Borin, L. Serrano-Andres, M. P. Fulscher and B. O. Roos, *J. Phys. Chem. A*, 1999, **103**, 1838-1845.
59. E. Quinones and R. Arce, *J. Am. Chem. Soc.*, 1989, **111**, 8218-8223.
60. S. F. Mason, *J. Chem. Soc.*, 1954, 2071-2081.
61. K. Rottger, R. Siewertsen and F. Temps, *Chem. Phys. Lett.*, 2012, **536**, 140-146.
62. J. Lin, C. Yu, S. Peng, I. Akiyama, K. Li, L. K. Lee and P. R. Lebreton, *J. Phys. Chem.*, 1980, **84**, 1006-1012.
63. J. P. Villabona-Monsalve, R. Noria, S. Matsika and J. Peon, *J. Am. Chem. Soc.*, 2012, **134**, 7820-7829.
64. 3. G. Guo, Z. G. Lan and Z. 3. Cao, *Phys. Chem. Chem. Phys.*, 2013, **15**, 10777-10782.
65. 3. G. Guo, Y. Zhao and Z. 3. Cao, *Phys. Chem. Chem. Phys.*, 2014, **16**, 15381-15388.
66. 3. G. Guo, H. J. Yuan, B. B. An, Q. L. Zhu and J. L. Zhang, *J. Chem. Phys.*, 2016, **144**.

67. M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House and J. P. Dworkin, *Proc. Natl. Acad. Sci. U.S.A.*, 2011, **108**, 13995-13998.
68. J. A. Piccirilli, T. Krauch, S. E. Moroney and S. A. Benner, *Nature*, 1990, **343**, 33-37.
69. R. Saladino, C. Crestini, V. Neri, J. R. Brucato, L. Colangeli, F. Ciciriello, E. Di Mauro and G. Costanzo, *Chembiochem*, 2005, **6**, 1368-1374.
70. M. P. Callahan, B. Crews, A. Abo-Riziq, L. Grace, M. S. de Vries, Z. Gengeliczki, T. M. Holmes and G. A. Hill, *Phys. Chem. Chem. Phys.*, 2007, **9**, 4587-4591.
71. G. Gate, M. Haggmark and M. S. de Vries.
72. S. Yamazaki, A. L. Sobolewski and W. Domcke, *Phys. Chem. Chem. Phys.*, 2009, **11**, 10165-10174.
73. P. Changenet-Barret, L. Kovacs, D. Markovitsi and T. Gustavsson, *Molecules*, 2016, **21**.
74. K. Rottger, R. Stellmacher, M. C. Stuhldreier and F. Temps, *Molecules*, 2017, **22**.
75. M. Barbatti and H. Lischka, *Phys. Chem. Chem. Phys.*, 2015, **17**, 15452-15459.
76. E. Nir, K. Kleinermanns, L. Grace and M. S. de Vries, *J. Phys. Chem. A*, 2001, **105**, 5106-5110.
77. K. A. Seefeld, C. Plützer, D. Löwenich, T. Häber, R. Linder, K. Kleinermanns, J. Tatchen and C. M. Marian, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3021-3026.
78. S. Blaser, H. M. Frey, C. G. Heid and S. Leutwyler, *Chimia (Aarau)*, 2014, **68**, 260-263.
79. R. K. Sinha, S. Lobsiger, M. Trachsel and S. Leutwyler, *J. Phys. Chem. A*, 2011, **115**, 6208-6217.
80. S. Lobsiger, H. M. Frey, S. Leutwyler, P. Morgan and D. Pratt, *J. Phys. Chem. A*, 2011, **115**, 13281-13290.
81. A. Broo, *J. Phys. Chem.*, 1998, **A 102**, 526-531.
82. L. Serrano-Andres, M. Merchan and A. C. Borin, *Proc. Natl. Acad. Sci. U.S.A.*, 2006, **103**, 8691-8696.
83. S. Perun, A. L. Sobolewski and W. Domcke, *Mol. Phys.*, 2006, **104**, 1113-1121.
84. E. L. Rachofsky, J. B. A. Ross, M. Krauss and R. Osman, *J. Phys. Chem. A*, 2001, **105**, 190-197.

85. C. Santhosh and P. C. Mishra, *Spectrochimica Acta A*, 1991, **47**, 1685-1693.
86. R. K. Neely, S. W. Magennis, D. T. F. Dryden and A. C. Jones, *J. Phys. Chem. B*, 2004, **108**, 17606-17610.
87. A. Holmen, B. Norden and B. Albinsson, *J. Am. Chem. Soc.*, 1997, **119**, 3114.
88. C. Plützer, E. Nir, M. S. de Vries and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2001, **3**, 5466-5469.
89. C. Plützer and K. Kleinermanns, *Phys. Chem. Chem. Phys.*, 2002, **4**, 4877-4882.
90. F. Buchner, H. H. Ritze, J. Lahl and A. Lubcke, *Phys. Chem. Chem. Phys.*, 2013, **15**, 11402-11408.
91. N. J. Kim, G. Jeong, Y. S. Kim, J. Sung, S. K. Kim and Y. D. Park, *J. Chem. Phys.*, 2000, **113**, 10051-10055.
92. S. Smolarek, A. M. Rijs, W. J. Buma and M. Drabbels, *Phys. Chem. Chem. Phys.*, 2010, **12**, 15600-15606.
93. H. Kang, J. Chang, S. H. Lee, T. K. Ahn, N. J. Kim and S. K. Kim, *J. Chem. Phys.*, 2010, **133**.
94. H. Kang, K. T. Lee, B. Jung, Y. J. Ko and S. K. Kim, *J. Am. Chem. Soc.*, 2002, **124**, 12958-12959.
95. S. Ullrich, T. Schultz, M. Z. Zgierski and A. Stolow, *J. Am. Chem. Soc.*, 2004, **126**, 2262-2263.
96. N. L. Evans and S. Ullrich, *J. Phys. Chem. A*, 2010, **114**, 11225-11230.
97. E. Samoylova, T. Schultz, I. V. Hertel and W. Radloff, *Chem. Phys.*, 2008, **347**, 376-382.
98. H. Kang, B. Jung and S. K. Kim, *J. Chem. Phys.*, 2003, **118**, 6717-6719.
99. C. Canuel, M. Mons, F. Piuze, B. Tardivel, I. Dimicoli and M. Elhanine, *J. Chem. Phys.*, 2005, **122**, 074316-074317.
100. C. M. Marian, *J. Chem. Phys.*, 2005, **122**, 10.
101. S. Perun, A. L. Sobolewski and W. Domcke, *Chem. Phys.*, 2005, **313**, 107-112.
102. S. Perun, A. L. Sobolewski and W. Domcke, *J. Am. Chem. Soc.*, 2005, **127**, 6257-6265.
103. H. Chen and S. H. Li, *J. Phys. Chem. A*, 2005, **109**, 8443-8446.

104. F. Plasser, R. Crespo-Otero, M. Pederzoli, J. Pittner, H. Lischka and M. Barbatti, *J Chem Theory Comput*, 2014, **10**, 1395-1405.
105. K. L. Wells, G. M. Roberts and V. G. Stavros, *Chem. Phys. Lett.*, 2007, **446**, 20-24.
106. I. Hünig, C. Plützer, K. A. Seefeld, D. Löwenich, M. Nispel and K. Kleinermanns, *ChemPhysChem*, 2004, **5**, 1427-1431.
107. T. Pancur, N. K. Schwalb, F. Renth and F. Temps, *Chem. Phys.*, 2005, **313**, 199-212.
108. T. Gustavsson, A. Sharonov, D. Onidas and D. Markovitsi, *Chem. Phys. Lett.*, 2002, **356**, 49-54.
109. J. M. L. Pecourt, J. Peon and B. Kohler, *J. Am. Chem. Soc.*, 2001, **123**, 10370-10378.
110. D. Onidas, D. Markovitsi, S. Marguet, A. Sharonov and T. Gustavsson, *J. Phys. Chem. B*, 2002, **106**, 11367-11374.
111. C. Su, C. T. Middleton and B. Kohler, *J. Phys. Chem. B*, 2012, **116**, 10266-10274.
112. W. M. Kwok, C. S. Ma and D. L. Phillips, *J. Am. Chem. Soc.*, 2006, **128**, 11894-11905.
113. J. Q. Chen and B. Kohler, *J. Am. Chem. Soc.*, 2014, **136**, 6362-6372.
114. C. E. Crespo-Hernandez and B. Kohler, *J. Phys. Chem. B*, 2004, **108**, 11182-11188.
115. S. Miyakawa, H. J. Cleaves and S. L. Miller, *Origins Life Evol. Biosphere*, 2002, **32**, 209-218.
116. S. Miyakawa, H. Yamanashi, K. Kobayashi, H. J. Cleaves and S. L. Miller, *Proc. Natl. Acad. Sci. U.S.A.*, 2002, **99**, 14628-14631.
117. P. Virta, A. Koch, M. U. Roslund, P. Mattjus, E. Kleinpeter, L. Kronberg, R. Sjöholm and K. D. Klika, *Org Biomol Chem*, 2005, **3**, 2924-2929.
118. A. Rich, in *Horizons in Biochemistry*, eds. M. Kasha and B. Pullman, Academic Press, New York, 1962, pp. 103-126.
119. J. R. Blas, F. J. Luque and M. Orozco, *J. Am. Chem. Soc.*, 2004, **126**, 154-164.
120. H. Robinson, Y. G. Gao, C. Bauer, C. Roberts, C. Switzer and A. H. J. Wang, *Biochemistry*, 1998, **37**, 10897-10905.

121. A. Eschenmoser, *Pure Appl. Chem.*, 1993, **65**, 1179-1188.
122. C. Roberts, J. C. Chaput and C. Switzer, *Chem Biol*, 1997, **4**, 899-908.
123. H. Sugiyama, S. Ikeda and I. Saito, *J. Am. Chem. Soc.*, 1996, **118**, 9994-9995.
124. C. Roberts, R. Bandaru and C. Switzer, *J. Am. Chem. Soc.*, 1997, **119**, 4640-4649.
125. J. Sepiol, Z. Kazimierczuk and D. Shugar, *Z Naturforsch C*, 1976, **31**, 361-370.
126. F. Seela, C. F. Wei and Z. Kazimierczuk, *Helv. Chim. Acta*, 1995, **78**, 1843-1854.
127. E. Nir, L. Grace, B. Brauer and M. S. de Vries, *J. Am. Chem. Soc.*, 1999, **121**, 4896-4897.
128. E. Nir, P. Imhof, K. Kleineremanns and M. S. de Vries, *J. Am. Chem. Soc.*, 2000, **122**, 8091-8092.
129. F. PiuZZi, M. Mons, I. Dimicoli, B. Tardivel and Q. Zhao, *Chem. Phys.*, 2001, **270**, 205-214.
130. W. Chin, M. Mons, I. Dimicoli, F. PiuZZi, B. Tardivel and M. Elhanine, *European Physical Journal D*, 2002, **20**, 347-355.
131. M. Mons, F. PiuZZi, I. Dimicoli, L. Gorb and J. Leszczynski, *J. Phys. Chem. A*, 2006, **110**, 10921-10924.
132. K. Seefeld, R. Brause, T. Häber and K. Kleineremanns, *J. Phys. Chem. A*, 2007, **111**, 6217-6221.
133. C. M. Marian, *J. Phys. Chem. A*, 2007, **111**, 1545-1553.
134. M. Y. Choi and R. E. Miller, *J. Am. Chem. Soc.*, 2006, **128**, 7320-7328.
135. H. Chen and S. H. Li, *J. Chem. Phys.*, 2006, **124**.
136. B. Heggen, Z. G. Lan and W. Thiel, *Phys. Chem. Chem. Phys.*, 2012, **14**, 8137-8146.
137. H. T. Miles, J. Frazier and F. B. Howard, *Science*, 1963, **142**, 1458-&.
138. F. A. Miannay, T. Gustavsson, A. Banyasz and D. Markovitsi, *J. Phys. Chem. A*, 2010, **114**, 3256-3263.
139. V. Karunakaran, K. Kleineremanns, R. Improta and S. A. Kovalenko, *J. Am. Chem. Soc.*, 2009, **131**, 5839-5850.
140. J. Peon and A. H. Zewail, *Chem. Phys. Lett.*, 2001, **348**, 255-262.

141. J. M. L. Pecourt, J. Peon and B. Kohler, *J. Am. Chem. Soc.*, 2000, **122**, 9348-9349.
142. E. M. Arpa, M. M. Brister, S. J. Hoehn, C. E. Crespo-Hernandez and I. Corral, *J Phys Chem Lett*, 2020, DOI: 10.1021/acs.jpcllett.0c00935, 5156-5161.
143. D. Picconi, F. J. A. Ferrer, R. Improta, A. Lami and F. Santoro, *Faraday Discuss.*, 2013, **163**, 223-242.
144. Y. Valadbeigi, M. Soleiman-Beigi and R. Sahraei, *Chem. Phys. Lett.*, 2015, **629**, 1-7.
145. L. Mejía-Mazariegos and J. Hernández-Trujillo, *Chem. Phys. Lett.*, 2009, **482**, 24-29.
146. T. Zeleny, M. Ruckebauer, A. J. A. Aquino, T. Muller, F. Lankas, T. Drsata, W. L. Hase, D. Nachtigallova and H. Lischka, *J. Am. Chem. Soc.*, 2012, **134**, 13662-13669.
147. A. L. Sobolewski and W. Domcke, *Phys. Chem. Chem. Phys.*, 2004, **6**, 2763-2771.
148. J. Q. Chen, Y. Y. Zhang and B. Kohler, *Top Curr Chem*, 2015, **356**, 39-87.
149. J. Xu, V. Chmela, N. J. Green, D. A. Russell, M. J. Janicki, R. W. Gora, R. Szabla, A. D. Bond and J. D. Sutherland, *Nature*, 2020, **582**, 60-66.
150. A. Abo-Riziq, B. O. Crews, I. Compagnon, J. Oomens, G. Meijer, G. Von Helden, M. Kabelac, P. Hobza and M. S. de Vries, *J. Phys. Chem. A*, 2007, **111**, 7529-7536.
151. C. Su, C. T. Middleton and B. Kohler, *J. Phys. Chem. B*, 2012, **116**, 10266-10274.
152. C. E. Crespo-Hernandez, B. Cohen and B. Kohler, *Nature*, 2005, **436**, 1141-1144.
153. C. E. Crespo-Hernandez, K. de La Harpe and B. Kohler, *J. Am. Chem. Soc.*, 2008, **130**, 10844-+.
154. I. Buchvarov, Q. Wang, M. Raytchev, A. Trifonov and T. Fiebig, *Proc. Natl. Acad. Sci. U.S.A.*, 2007, **104**, 4794-4797.

4. Non-standard base pairing and stacked structures in methyl xanthine clusters

We present resonant two-photon ionization and IR-UV double resonance spectra of methylated xanthine derivatives including 7-methylxanthine dimer and theobromine dimer seeded in a supersonic jet by laser desorption. For 7-methylxanthine, theophylline, and theobromine monomer, the lowest energy tautomer was assigned based on comparison with IR-UV double resonance spectra and calculated IR frequencies. For the 7-methylxanthine dimer, we observe hydrogen bonding on the N3H position suggesting 3 possible combinations, one that is reverse Watson-Crick and two that are reverse Hoogsteen. For the theobromine dimer, we observe a stacked structure.

4.1. INTRODUCTION

Both hydrogen bonding and π - π interactions (or stacking interactions) play an important role in biological structures. Hydrogen bonding is responsible for base pair recognition in DNA and stabilizes α -helices and β -sheets in protein structures. Stacking interactions help stabilize the structure of DNA duplexes.¹ Gas phase laser spectroscopy provides a means to study the intrinsic properties of biologically relevant molecules in a solvent-free environment, such as the non-covalent forces that govern their interactions. Pairings of various nucleobases by hydrogen bonding have been extensively studied by these techniques²⁻⁸, however, only a few studies exist of stacked structures.

Theoretical calculations show that in the gas phase hydrogen bonded structures

between nucleobases dominate.⁹ Population of stacked structures can be enhanced in one of two ways. First, water molecules can stabilize stacked structures by bridging.⁹⁻¹¹ Kabeláč *et al.* predict that two to six water molecules are required for most nucleobases pair combinations to cause stacked structures to be preferred over H-bonded ones.⁹ Second, methylation can lead to stacking by reducing the number of H-bonding sites. Kabeláč *et al.* have shown that 7-methyladenine-adenine adopts a nearly planar hydrogen bonded structure and 9-methyladenine-adenine adopts a stacked structure.¹² Here we further explore this aspect of competition between stacking and H-bonding by studying clusters of various methylated xanthines. Figure 1 shows the methyl derivatives used in this work.

Methyl xanthines are important molecules commonly used as mild stimulants and bronchodilators.¹³ In the gas phase, these purines can exist in a variety of different tautomeric forms, which can exhibit drastically different photophysical behavior. We investigated the structure of 7-methylxanthine dimer and theobromine dimer by resonant two-photon ionization (R2PI) and IR-UV double resonance spectroscopy. By comparing the observed IR spectrum with calculated frequencies of optimized structures, we conclude that 7-methylxanthine dimer is most likely hydrogen bonded in the three-position and theobromine dimer is in a stacked configuration.

4.2. METHODS

4.2.1 Experimental

7-Methylxanthine, theophylline (1,3-dimethylxanthine), theobromine (3,7-dimethylxanthine), and caffeine (1,3,7-trimethylxanthine) were obtained from Sigma-Aldrich and used without further purification. Their structures are shown in figure 1.

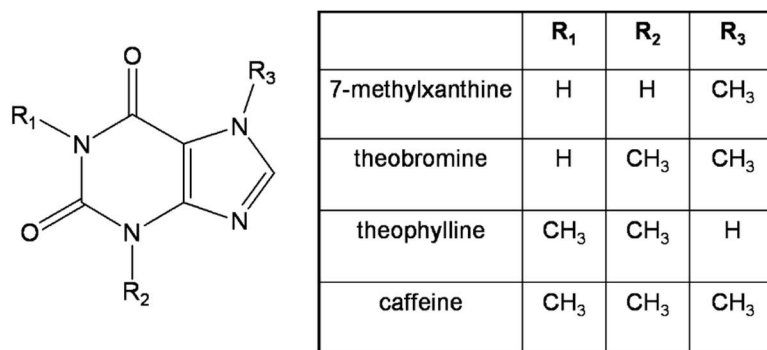


Figure 1: The structures of methyl xanthine derivatives examined in this report.

The experimental setup has been described in detail elsewhere.¹⁴ We laser desorb a thin layer of sample from a graphite substrate in front of a pulsed nozzle. The desorption laser, a Nd:YAG operating at 1064 nm, is attenuated to 1 mJ cm⁻² and focused to a spot approximately 0.5 mm diameter within 2 mm in front of the nozzle orifice. We translate the sample in order to expose fresh sample to successive laser shots. The nozzle consists of a pulsed valve with a nozzle diameter of 1 mm and a backing pressure of 6 atm of argon drive gas. The neutral molecules are skimmed and then ionized with a frequency-doubled dye laser. We detect the ions in a reflectron time-of-flight mass spectrometer. Typical mass resolution ($m/\Delta m$) measured at the molecular ion peak is around 700.

We obtain resonant two-photon ionization (R2PI) spectra by monitoring mass selected peaks while tuning the one-color, two-photon ionization (1C-R2PI) wavelength. For theobromine and caffeine, we also ionized these molecules using a two-color two-photon process (2C-R2PI). For the second photon we used 266 nm provided by the fourth harmonic of another Nd:YAG laser. We measure IR-UV double resonance spectra with two laser pulses separated in time by 200 ns. The first IR pulse serves as a “burn” pulse, which removes the ground state population and causes depletion in the ion signal of the second UV “probe” pulse, provided both lasers are tuned to a

resonance of the same tautomer. IR frequencies are produced in an OPO/OPA setup (LaserVision) pumped by a Nd:YAG laser operating at its fundamental frequency. For this work, we operated within the range of 2800–3550 cm^{-1} , which encompasses NH and OH modes. Typical IR intensities in the burn region are 12 mJ pulse^{-1} and the bandwidth is 3 cm^{-1} .

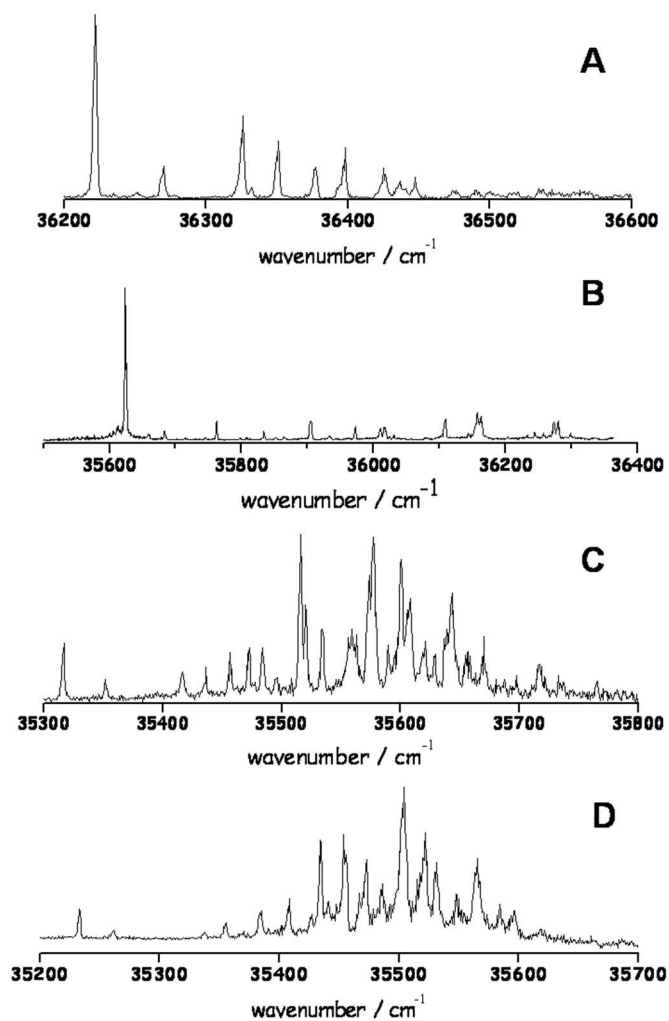


Figure 2: The R2PI spectra of (A) 7-methylxanthine, (B) theophylline, (C) theobromine, and (D) caffeine. The theobromine and caffeine spectra are two-color R2PI with 266 nm.

4.2.2. Theoretical

The tautomeric equilibria of methylated xanthines were assessed using the Gaussian 03 quantum code package.¹⁵ Application of Density Functional Theory's B3LYP hybrid functional (a parameterized combination of Becke's exchange functional, the Lee, Yang, and Parr correlation functional and the exact exchange functional)¹⁶⁻¹⁸ with a 6-311+G(2d,p) basis set^{19, 20} yielded equilibrium geometries of the tautomers for each methylated xanthine. We performed second derivative calculations for purposes of vibrational frequency analysis and to verify that the geometries for all species corresponded to local minima. Vibrational frequencies were computed on an ultrafine grid and a corrective scaling factor of 0.9618 was applied to account for anharmonicity.²¹

4.3. RESULTS AND DISCUSSION

4.3.1 Methyl xanthine monomers

We measured one-color R2PI spectra for 7-methylxanthine and theophylline, and two-color R2PI spectra for theobromine and caffeine, shown in figure 2. One-color R2PI was possible for both theobromine and caffeine, but peaks were broadened due to power saturation, which led to increased spectral congestion. By performing two-color R2PI, separating excitation and ionization, we could reduce the intensity of the excitation laser. This approach reduced power broadening and produced better spectra. The R2PI spectra for all of these methylated xanthines are sharp and vibronically resolved. Increasing the number of methyl substitutions on xanthine shifts the R2PI spectra to lower energies. To identify specific tautomers in this R2PI region, we measured IR-UV double resonance spectra on the origin peak for 7-methylxanthine, theophylline, and theobromine. For all three monomers, we observe the lowest energy tautomer under our

experimental conditions.

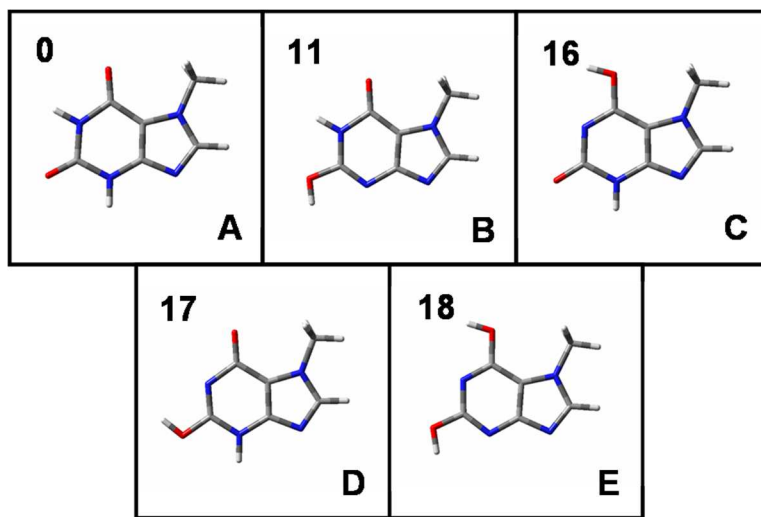
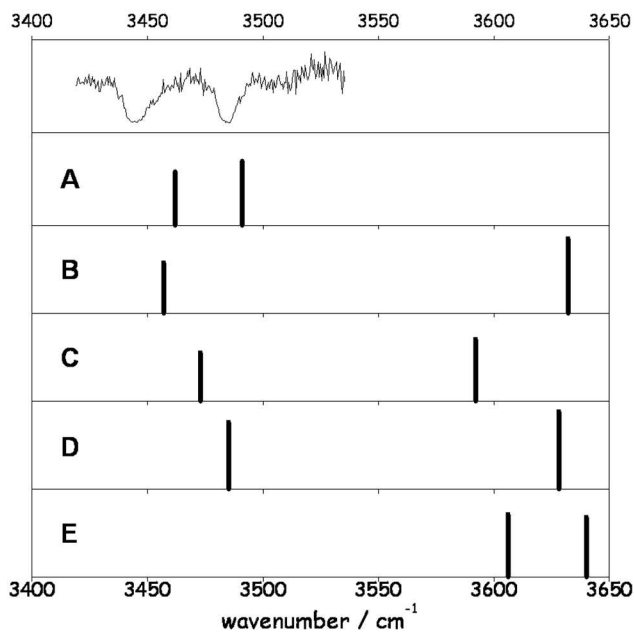


Figure 3: IR-UV double resonance spectrum of 7-methylxanthine. Stick spectra are calculated frequencies at the B3LYP/6-311+G(2d,p) level for tautomers of 7-methylxanthine (lettered A-E). Their relative energies (in kcal/mol) are shown in the upper left hand corner.

Figure 3 shows the IR-UV double resonance spectrum of 7-methylxanthine along with the calculated frequencies for five tautomers optimized at the B3LYP/6-311+G(2d,p) level. By comparing the IR spectrum to the calculated frequencies and intensities, we conclude that we

observe the diketo 7-methylxanthine, which we calculated to be the lowest energy tautomer. The two strong bands at 3445 and 3485 cm^{-1} in the IR spectrum can be assigned as the N1H and N3H stretching vibrations. The other four tautomers are higher in energy by at least 11 kcal/mol and their calculated IR frequencies do not match the experimental result.

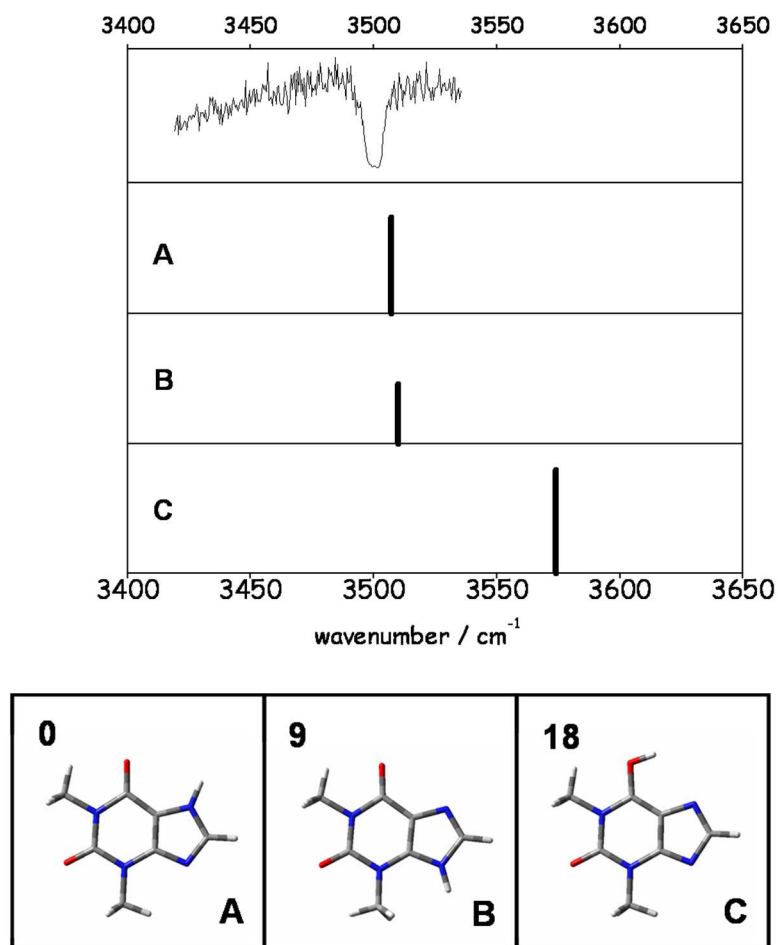


Figure 4: IR-UV double resonance spectrum of theophylline. Stick spectra are calculated frequencies at the B3LYP/6-311+G(2d,p) level for tautomers of theophylline (lettered A-C). Their relative energies (in kcal/mol) are shown in the upper left hand corner.

Figure 4 shows the IR-UV double resonance spectrum of theophylline along with the calculated frequencies for three tautomers optimized at the B3LYP/6-311+G(2d,p) level. The

lowest energy tautomer is the N7H form, while the N9H tautomer is 9 kcal/mol higher in energy. However, the calculated N-H stretching vibrations for these two tautomers are almost indistinguishable from each other. In another report, we measured the IR-UV double resonance spectrum of xanthine and assigned it to the N7H form.²²

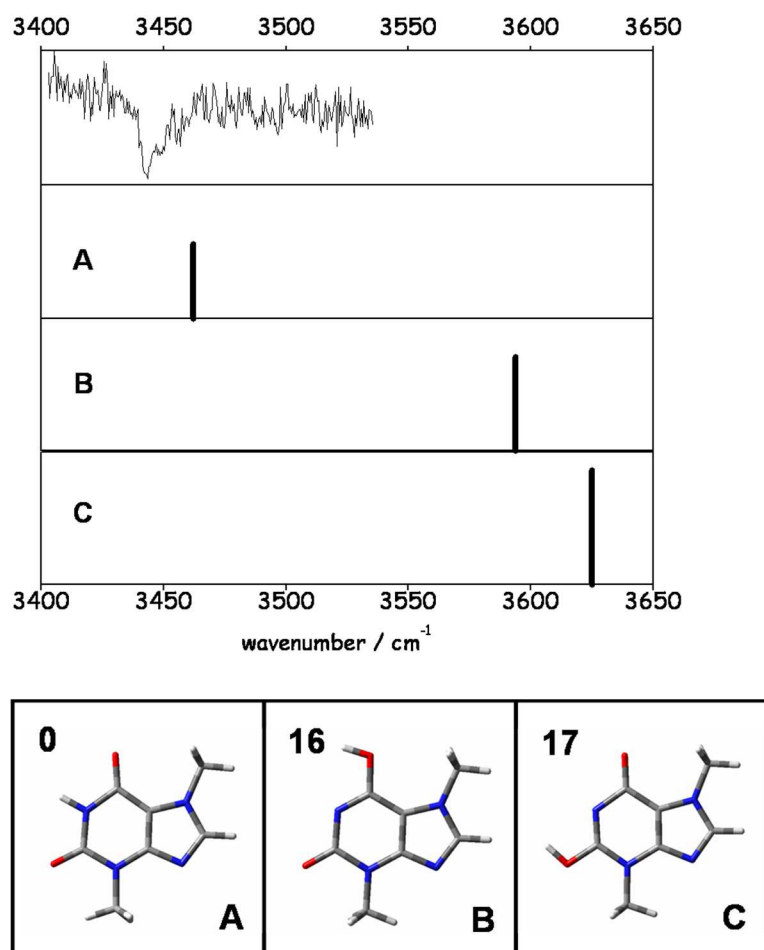


Figure 5: IR-UV double resonance spectrum of theobromine. Stick spectra are calculated frequencies at the B3LYP/6-311+G(2d,p) level for tautomers of theobromine (lettered A-C). Their relative energies (in kcal/mol) are shown in the upper left hand corner.

In xanthine, the differences in the calculated frequencies between the N7H and N9H forms were larger despite similar energy differences. By comparison of IR spectra, it would seem that theophylline is also in the N7H form (see supporting material). Therefore, we tentatively

assign the one band at 3500 cm^{-1} in the IR spectrum as the N7H stretching vibration. The third tautomer is an unlikely assignment since it is much higher in energy (18 kcal/mol).

Figure 5 shows the IR-UV double resonance spectrum of theobromine along with the calculated frequencies for three tautomers optimized at the B3LYP/6-311+G(2d,p) level. By comparing the IR spectrum to the calculated frequencies and intensities, we conclude that we observe the diketo form in the supersonic jet. This is the most confident assignment since the other two tautomers are much higher in energy (greater than 16 kcal/mol). In addition, the calculated OH band is more than 100 cm^{-1} away from the one band in the experimental IR spectrum at 3446 cm^{-1} , assigned as the N1H stretching vibration. For caffeine, we did not measure the IR-UV double resonance spectrum since all three N-H sites are replaced by methyl groups.

4.3.2 7-Methylxanthine dimer

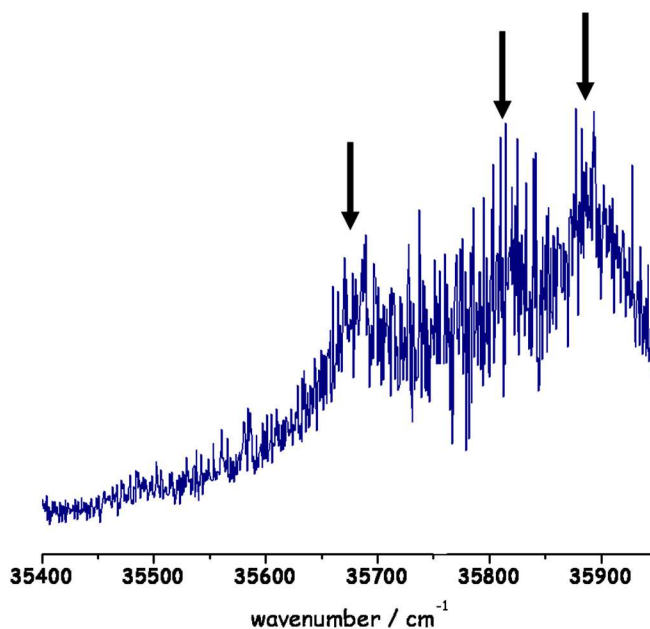


Figure 6: The broad, one-color R2PI spectrum of 7-methylxanthine dimer. The arrows at 35677 , 35822 , and 35892 cm^{-1} indicate where IR spectra were measured by double resonant techniques.

Figure 6 shows the one-color R2PI spectrum for 7-methylxanthine dimer. Its appearance

differs significantly from that of the 7-methylxanthine monomer. The R2PI spectrum is red-shifted compared to the monomer and broad. Figure 7 shows a mass spectrum recorded at 35677 cm^{-1} exhibiting a strong peak for the 7-methylxanthine dimer. A very small peak for the 7-methylxanthine trimer was the only other peak observed in the mass spectrum. Despite optimization on the dimer peak, we cannot exclude the possibility that the dimer spectrum contains contributions from higher clusters fragmenting. However, we are confident that the R2PI spectrum for 7-methylxanthine monomer does not contain contributions from the dimer because they absorb in different spectral regions.

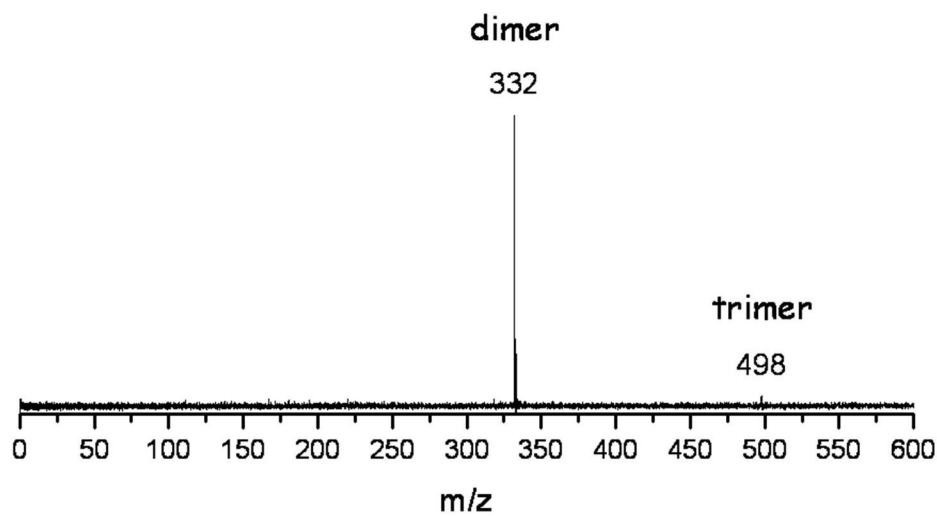


Figure 7: Mass spectrum recorded at 35677 cm^{-1} showing a strong 7-methylxanthine dimer peak.

We calculated energies and IR frequencies for the 7-methylxanthine dimer at the B3LYP/6-311+G(2d,p) level. We did this for 10 different combinations where both molecules are in the lowest energy diketo form. These base pairs are all close in energy, where the lowest and highest energy combinations are only separated by 3 kcal/mol, as summarized in Figure 8. Other combinations were not considered since enol and dienol tautomers are much higher in energy and

their presence is unlikely in the supersonic expansion. The IR-UV double resonance spectrum for the 7-methylxanthine monomer shows two strong bands at 3445 and 3485 cm^{-1} , which correspond to the computed N1H and N3H stretch frequencies.

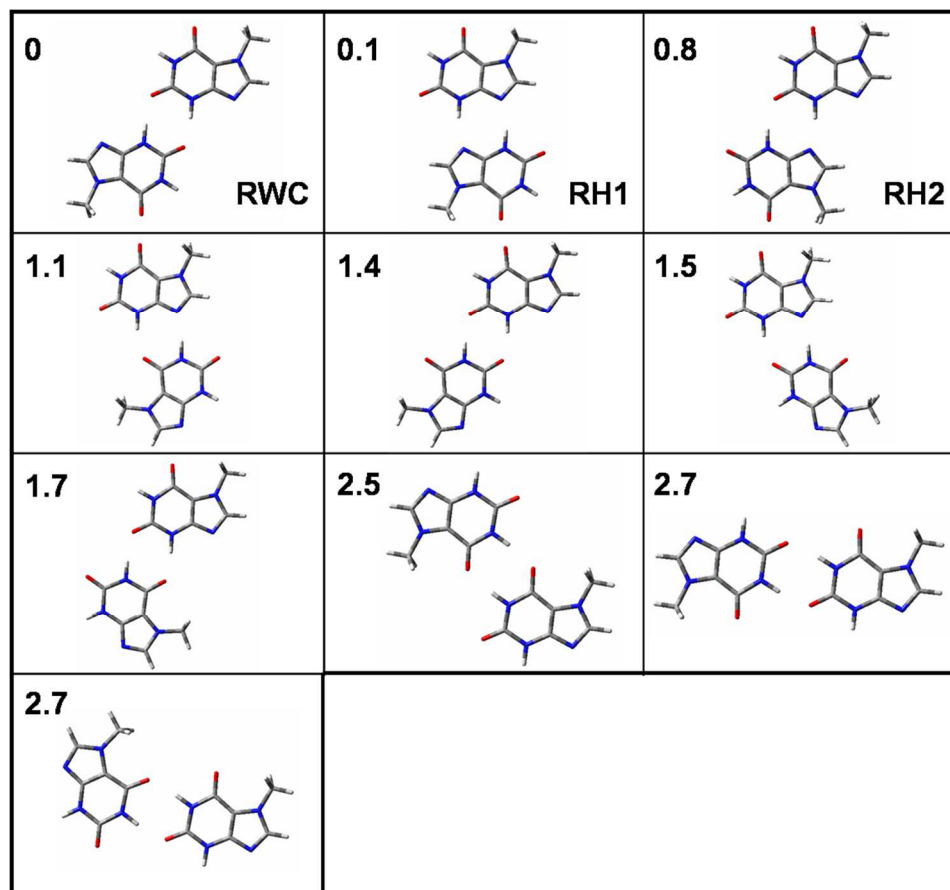


Figure 8: The relative energies (in kcal/mol) of the 7-methylxanthine dimer calculated at the B3LYP/6-311+G(2d,p) level. The three lowest energy base pairs all have their N3H involved in hydrogen bonding. One base pair is reverse Watson-Crick (RWC) and the other two are reverse Hoogsteen (RH1 and RH2).

We measured the IR-UV double resonance spectrum for the 7-methylxanthine dimer while probing at three different UV frequencies: 35677, 35822, and 35892 cm^{-1} . The IR spectra probed at each UV frequency were essentially identical. In these spectra, shown in Figure 9, the N3H band is absent, which indicates that this position is now involved in a hydrogen bond. In addition, two strong broad bands are observed below 3200 cm^{-1} , with one especially broad band

spanning from 3130 to 2875 cm^{-1} . Attenuating the IR output in this broad region did not change the appearance of the IR spectrum.

Of the 10 calculated dimer structures, the three lowest energy base pairs are the only ones where the N3H in both molecules are involved in hydrogen bonds. The lowest energy pair is in a reverse Watson-Crick (RWC) type configuration and the other two pairs are in reverse Hoogsteen (RH1 and RH2) type configurations. The calculated IR frequencies for the N1H stretching vibration in these configurations are in the same position and fit the experimental band well.

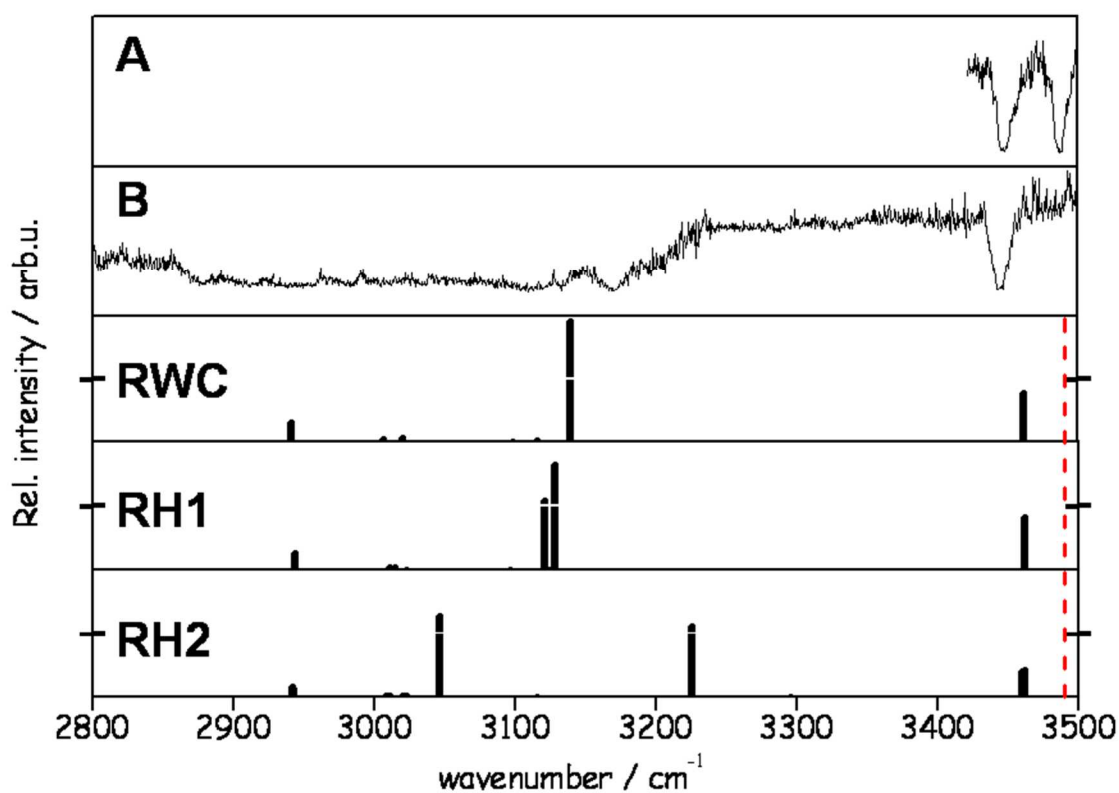


Figure 9: IR-UV double resonance spectrum of (A) 7-methylxanthine monomer and (B) 7-methylxanthine dimer. Stick spectra are calculated frequencies at the B3LYP/6-311+G(2d,p) level for the three lowest energy dimer structures. In all three dimer structures, the N1H is free. The dotted red line is the calculated N3H stretch, which matches the IR band in the monomer spectrum very well.

However, establishing the exact base pair configuration observed in our experiment is difficult due to the broad nature of the IR bands below 3200 cm^{-1} . For the reverse Watson-Crick

(RWC) base pair we calculated one strong IR band at 3140 cm^{-1} , which corresponds to N3Hs hydrogen bonded to oxygen atoms at the two-position. For the lower energy reverse Hoogsteen base pair (RH1), two strong IR bands are calculated around 3121 and 3128 cm^{-1} , which correspond to N3Hs hydrogen bonded to nitrogen atoms at the nine-position. For the higher energy reverse Hoogsteen base pair (RH2), two strong IR bands are calculated around 3047 and 3226 cm^{-1} , which correspond to the N3H hydrogen bonded to the nitrogen atom at the nine-position and the other N3H hydrogen bonded to oxygen atoms at the two-position, respectively. The latter calculated band is a good fit to the experimental band at 3170 cm^{-1} . The extremely broad bands in the experimental IR spectrum could be explained by the overlap of frequencies from all three dimer structures. In addition, the broad and congested R2PI spectrum can be the result of overlapping spectra from multiple similar structures. Furthermore, photophysics similar to that found in various nucleobases might also contribute to spectral broadening, however, this usually pertains to the canonical nucleobases²³⁻²⁶ and to the specific Watson-Crick configuration.²

There is also a possibility for different tautomeric structures in which three hydrogen bonds are formed either by two enol tautomers or one diketo and one dienol tautomer. This could occur if the stabilization of the three hydrogen bonds would be greater than the reorganization energy of the tautomers. However all of these combinations have a free N3H mode and therefore do not fit the experimental IR data. In addition, Dietrich *et al.* have calculated that these base pairs containing three hydrogen bonds are much less stable compared to those with two hydrogen bonds involving the N3H position.²⁷

4.3.3 Theobromine dimer

Figure 10 shows the one-color R2PI spectrum for theobromine dimer. Again, the R2PI

spectrum of the dimer is broad compared to the sharp and resolved R2PI spectrum of the monomer. The onset of the R2PI spectrum for theobromine dimer is red-shifted compared to the monomer.

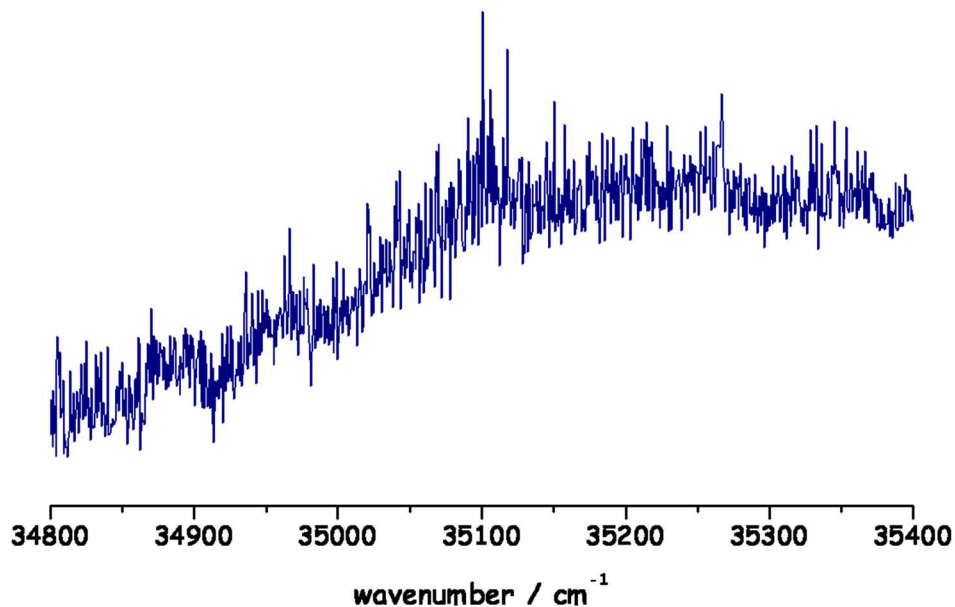


Figure 10: The broad, one-color R2PI spectrum of theobromine dimer.

Figure 11 shows the IR-UV double resonance spectra for the theobromine monomer and dimer. Both the monomer and dimer feature a strong band around 3446 and 3438 cm^{-1} , respectively. Despite a slight shift of the IR band observed in the dimer, they clearly fit the N1H stretch frequency. The experimental IR spectrum indicates that the dimer structure is in a stacked configuration rather than hydrogen bonded one. A stacked structure for theobromine dimer may partly explain the broad appearance of the R2PI spectrum, since mixing of electronic states is more likely to occur. Broad excitation spectra are commonly observed for aromatic dimers due to stacking interactions and excimer formation.^{28,29}

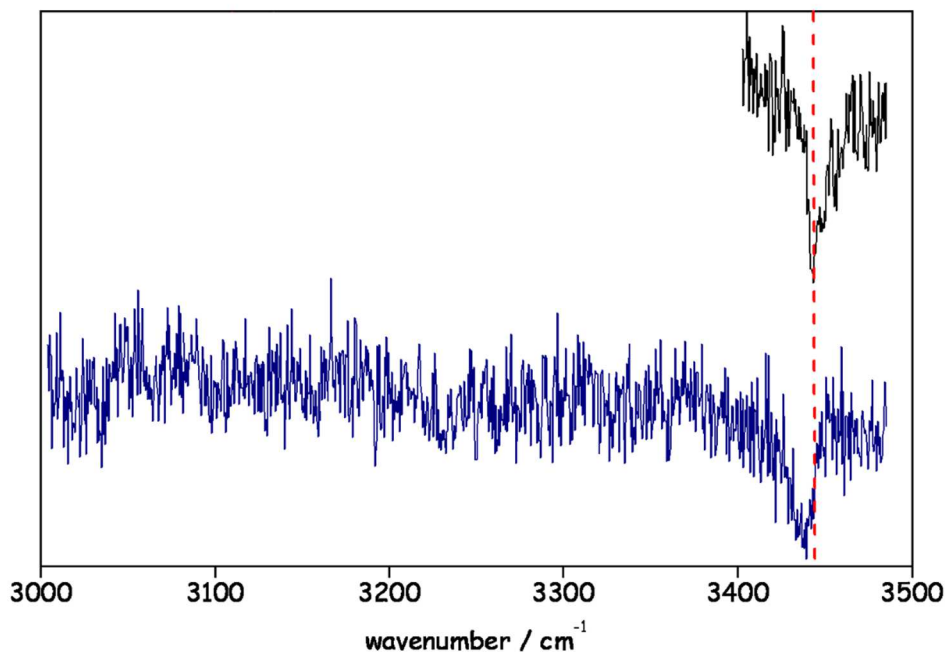


Figure 11: IR-UV double resonance spectra of theobromine monomer (top trace) and theobromine dimer (bottom trace). The IR band for the dimer is slightly red-shifted with respect to the IR band for the monomer.

4.3.4 Caffeine and theophylline dimers

For caffeine, we observe a very strong dimer signal in the mass spectrum. Since all hydrogen bonding sites are blocked we did not measure an IR spectrum and these must be van der Waals dimers based on π -stacking. Prediction of the exact structure awaits further computation. Curiously, we do not observe any theophylline dimer despite the fact we can observe caffeine dimers, which would have more steric resistance. We note that cluster formation (in principle) occurs readily in our experiment because by adjusting the sample position relative to the pulsed nozzle, we can observe theophylline-argon clusters in the mass spectrum. It is possible for species to be present in the molecular beam without being observed by R2PI detection. Possible reasons include unfavorable Frank-Condon factors, spectral shifts, low oscillator strengths, or short excited state lifetimes compared to the nanosecond timescale laser pulses. These effects could also play a

role in the failure to observe hydrogen bonded theobromine dimers and all but the three lowest energy structures of 7-methylxanthine dimer. Since for the latter the calculated energy differences are very small this would suggest structure dependent photochemistry.

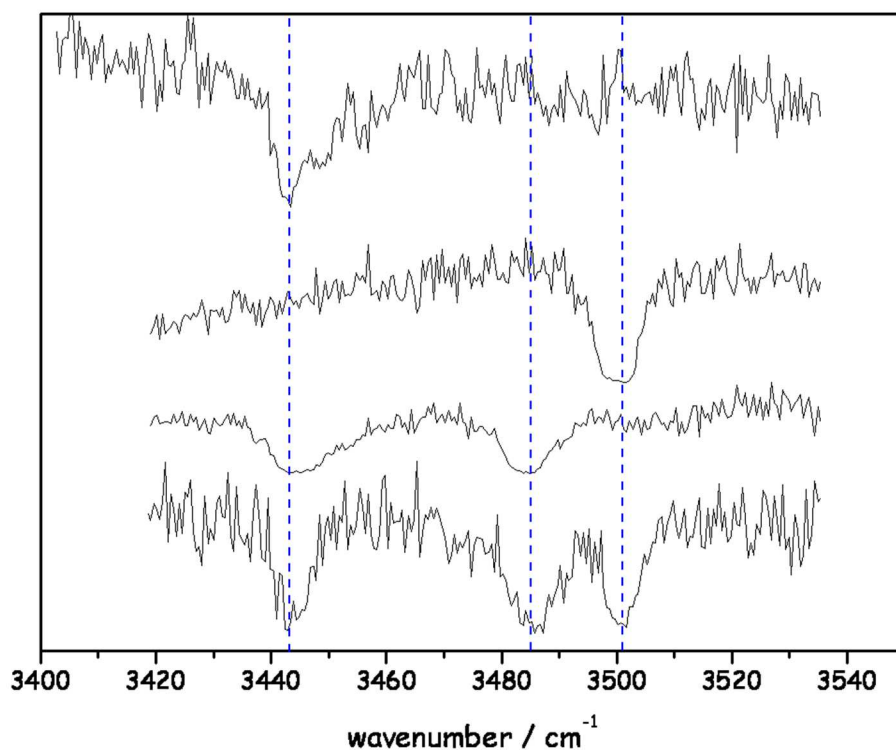


Figure S1: IR-UV double resonance spectra theobromine (top trace), theophylline (second trace), 7-methylxanthine (third trace), and xanthine (bottom trace).

4.4. CONCLUSION

We investigated the gas phase structure of various methylated xanthine derivatives including 7-methylxanthine dimer and theobromine dimer by resonant two-photon ionization and IR-UV double resonance spectroscopy. The R2PI spectra for 7-methylxanthine, theophylline, theobromine, and caffeine were sharp and vibronically resolved. Based on calculated frequencies,

we assigned the IR spectra for these monomers to their lowest energy tautomer. The R2PI spectra for 7-methylxanthine dimer and theobromine dimer are redshifted and broad compared to their respective monomers. By comparing the observed IR spectrum with calculated frequencies for optimized base pairs, we conclude that the 7-methylxanthine dimer is hydrogen bonded on the N3H position suggesting 3 possible combinations, one that is reverse Watson-Crick and two that are reverse Hoogsteen. We assign a stacked structure to theobromine dimer since its IR spectra is similar to that of theobromine monomer. By increasing the number of methylated sites, a stacking structure is preferred over a hydrogen bonded base pair. This observation indicates that stabilization by stacking can compete with hydrogen bonding, especially since hydrogen bonding sites still remain in theobromine dimer.

References

- 1 D. Voet and J. G. Voet, *Biochemistry*, John Wiley and Sons, Inc. , New York, 2004.
- 2 A. Abo-Riziq, L. Grace, E. Nir, M. Kabelac, P. Hobza and M. S. de Vries, *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**, 20-23.
- 3 E. Nir, I. Hunig, K. Kleinermanns and M. S. de Vries, *Physical Chemistry Chemical Physics*, 2003, **5**, 4780-4785.
- 4 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Physical Chemistry Chemical Physics*, 2002, **4**, 732-739.
- 5 E. Nir, C. Janzen, P. Imhof, K. Kleinermanns and M. S. de Vries, *Physical Chemistry Chemical Physics*, 2002, **4**, 740-750.
- 6 E. Nir, K. Kleinermanns and M. S. de Vries, *Nature*, 2000, **408**, 949-951.
- 7 C. Plutzer, I. Hunig and K. Kleinermanns, *Physical Chemistry Chemical Physics*, 2003, **5**, 1158-1163.
- 8 C. Plutzer, I. Hunig, K. Kleinermanns, E. Nir and M. S. de Vries, *Chemphyschem*, 2003, **4**, 838-842.
- 9 M. Kabelac and P. Hobza, *Chemistry-a European Journal*, 2001, **7**, 2067-2074.
- 10 M. Kabelac, F. Ryjacek and P. Hobza, *Physical Chemistry Chemical Physics*, 2000, **2**, 4906-4909.
- 11 D. Sivanesan, I. Sumathi and W. J. Welsh, *Chemical Physics Letters*, 2003, **367**, 351-360.
- 12 M. Kabelac, C. Plutzer, K. Kleinermanns and P. Hobza, *Physical Chemistry Chemical Physics*, 2004, **6**, 2781-2785.

13 D. T. Hurst, *An Introduction to the Chemistry and Biochemistry of Pyrimidines, Purines and Pteridines*, Wiley, New York, 1980.

14 G. Meijer, M. S. Devries, H. E. Hunziker and H. R. Wendt, *Applied Physics B-Photophysics and Laser Chemistry*, 1990, **51**, 395-403.

15 M. J. T. Frisch, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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, O.; 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.; and Pople, J. A., Gaussian 03, Revision C.02, in *Gaussian, Inc.*, Wallingford CT, 2004.

16 A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648.

17 C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

18 B. Miehlich, A. Savin, H. Stoll and H. Preuss, *Chem. Phys. Lett.*, 1989, **157**, 200.

19 R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *Journal of Chemical Physics*, 1980, **72**, 650-654.

20 A. D. McLean and G. S. Chandler, *Journal of Chemical Physics*, 1980, **72**, 5639-

5648.

21 M. P. Andersson and P. Uvdal, *Journal of Physical Chemistry A*, 2005, **109**, 2937-2941.

22 M. P. Callahan, B. Crews, A. Abo-Riziq, L. Grace, M. S. de Vries, Z. Gengeliczki, T. M. Holmes and G. A. Hill, *Physical Chemistry Chemical Physics*, 2007, **9**, 4587-4591.

23 C. Canuel, M. Mons, F. PiuZZi, B. Tardivel, I. Dimicoli and M. Elhanine, *Journal of Chemical Physics*, 2005, **122**.

24 E. Samoylova, H. Lippert, S. Ullrich, I. V. Hertel, W. Radloff and T. Schultz, *Journal of the American Chemical Society*, 2005, **127**, 1782-1786.

25 S. Ullrich, T. Schultz, M. Z. Zgierski and A. Stolow, *Journal of the American Chemical Society*, 2004, **126**, 2262-2263.

26 S. Ullrich, T. Schultz, M. Z. Zgierski and A. Stolow, *Physical Chemistry Chemical Physics*, 2004, **6**, 2796-2801.

27 B. Dietrich, T. A. Hupp and B. Engels, NIC Symposium, Julich, Germany, 2001.

28 F. PiuZZi, I. Dimicoli, M. Mons, P. Millie, V. Brenner, Q. Zhao, B. Soep and A. Tramer, *Chemical Physics*, 2002, **275**, 123-147.

29 H. Saigusa and E. C. Lim, *Accounts of Chemical Research*, 1996, **29**, 171-178.

5. Conformer selective photodissociation in FGG tripeptide ions

In a molecular beam environment, neutral Phenylalanyl-glycyl-glycine tripeptides (FGG) exist as 4 different rotational conformers, one folded and three extended structures. We find that photo-ionization of these conformers, followed by photodissociation of the ions, produces ion photo-fragmentation ratios that depend on the original neutral conformation. The folded and extended structures of FGG fragment with fragment-to-parent ratios of 1.4 and 0.3, respectively. In addition to FGG, we also studied GGF. This species also exhibited a folded and extended structure. Similar to FGG, the folded GGF tautomer displayed high fragmentation (3:1 vs 1:2 for folded versus unfolded). This dependence of photodissociation of the cation on the structure of the original neutral molecule indicates a strong “memory” on the part of the cation. We propose that photoisomerization is insignificant on the timescale of our experiments. Two proposed methods by which neutral structure could affect cationic fragmentation are charge transfer and charge delocalization.

5.1 INTRODUCTION

The connection between conformation and reaction rates in large molecules, such as proteins, motivates investigation of the effect of secondary structure on excited state dynamics and on pathways of energy relaxation in peptides. Employing laser desorption techniques coupled with mass spectrometry and spectroscopy we can examine energy redistribution and dissociation of peptide cations as a function of their initial secondary structure. Our approach is to measure conformer specific ion dissociation spectra. Statistical theories, such as Rice-Ramsperger-Kassel-

Marcus (RRKM), predict rates of dissociation that scale with the number of [normal modes, which has been demonstrated experimentally for gas phase peptides by Hu et al.¹. For large molecules, such as peptides, these rates are on the order of 10^2 - 10^5 s⁻¹, and do not explain fast dissociative processes that occur in the presence of relatively small amounts of excess energy. This leads to the idea that for some molecules there exist fragmentation pathways fast enough to circumvent statistical relaxation. Using multi photon ionization and subsequent excitation schemes it is possible to control the amount of excess energy deposited into the ion, providing insight into energy migration pathways and dissociation mechanisms. Using 10^7 Watt/cm² UV laser pulses to induce multiphoton ionization and fragmentation in peptides, Lubman and coworkers determined amino acid sequences of peptides^{2,3}. Antoine et al. compared the fragmentation pathways in protonated pentapeptides due to laser excitation, collisions, and electron capture and found indications that the secondary structure plays a role⁴. By combining resonant UV ionization with visible (VIS) excitation, Weinkauff et al. examined dissociation and charge transfer in peptide cations as a function of amino acid sequence^{5,6}. They concluded that specific amino acids create barriers to energy transfer, affecting fragmentation products and rates of dissociation. Lee et al. reported conformationally dependent fragmentation in Phenyl alanine cations excited by one VIS photon, and explained their observation as resulting from variation in charge distribution among ionic conformers^{7,8}. We now present evidence of conformationally selective dissociation behavior in the excited tri-peptide cations Phenylalanyl-glycyl-glycine (FGG) and Glycyl-glycyl-phenylalanine (GGF) and determine that starting neutral secondary structure modulates fragmentation patterns.

5.2 EXPERIMENTAL METHODS

5.2.1. Experimental.

All measurements were performed with an apparatus described in detail elsewhere⁸. In brief, a thin layer of sample is laser-desorbed from a graphite bar in front of a pulsed nozzle. The desorption laser, a Nd:YAG operating at its fundamental frequency, is attenuated to less than 1 mJ/pulse and focused in front of the nozzle opening. We use a pulsed valve with a nozzle diameter of 1 mm, and a backing pressure of 6 atm of Argon. The neutral molecules from the desorption step are carried downstream in a pulsed nozzle expansion which passes through a skimmer (3mm by 1mm) before entering the extraction region of a reflectron time-of-flight mass spectrometer. UV wavelengths are produced by the frequency doubled output of a dye laser. The output of the dye laser has a bandwidth of 0.03 cm^{-1} . After frequency doubling, pulse energies are 2-3 mJ, and we attenuate to avoid saturation. We obtain resonant two-photon ionization (1+1 R2PI) spectra by monitoring mass selected peaks while tuning over a specific wavelength region. We measure UV-UV double resonance spectra with two laser pulses separated in time by 400 ns. The first pulse serves as a "burn" pulse, which reduces the ground state population and causes depletion of the ion signal of the second "probe" pulse, provided both lasers are tuned to a resonance of the same conformer. We obtain IR-UV double resonance spectra in an analogous way with the burn laser operating in the near IR region. We produce IR frequencies ranging from 2800 cm^{-1} to 3800 cm^{-1} (3cm^{-1} bandwidth) in an OPO/OPA setup (LaserVision), pumped by the fundamental output of a Nd:YAG. Typical IR intensities in the burn region are 8-10 mJ/pulse.

Ion dissociation spectra were obtained by ionizing the peptides with the 1+1 R2PI technique mentioned above, and then introducing an additional visible wavelength dye laser (600 nm) 10 ns after the ionizing pulse. This additional energy input causes changes in the

fragmentation behavior of the peptide, and these changes are structure-dependent.

5.2.2. Computational.

In previous studies^{9,10} we detected four distinct conformers of neutral FGG and two conformers of GGF in the molecular beam and we assigned the structure of each conformation through a comparison of experimental IR hole burning spectra to theoretical frequency calculations at the RI-MP2/cc-pVDZ level. In the current study, each one of these previously-assigned neutral structures were used as a starting point for optimizing the corresponding cation structure. That is, one electron was removed from the neutral and the structure was re-optimized, in order to parallel the process of vertical ionization and relaxation to a stable cation structure. The 6-311++G(d,p) basis set was determined to be a good compromise between accuracy and computational cost for the systems studied.

These re-optimizations were performed with four different DFT functionals: B3LYP (Becke 3-parameter-Lee-Yang-Parr)¹¹, X3LYP (extended hybrid functional combined with Lee-Yang-Parr correlation functional)¹², M06-2X (Minnesota '06 density functional with twice the amount of HF exchange)¹³, and B97D (Grimme's functional including dispersion)¹⁴. Harmonic frequency calculations were run at the same level of theory as the optimizations in order to ensure that the optimized structures were true minima rather than transition states. The vertical ionization energies and adiabatic ionization energies were also calculated with each functional for each conformer, by taking the difference of the energy of the cation without geometry optimization and the neutral energy, and the energy of the optimized cation and the neutral, respectively. These

calculations were performed with the Gaussian 03¹⁵ and Gaussian 09¹⁶ suite of programs.

5.3. RESULTS

5.3.1. FGG

We ionized jet-cooled FGG using 1+1 R2PI. FGG adopts four conformations in the molecular beam with origin transitions at (A) 37,370 cm⁻¹, (B) 37,549 cm⁻¹, (C) 37,556 cm⁻¹ and (D) 37,568 cm⁻¹. We previously assigned the structure of each conformation through a comparison of the experimental IR hole burning spectra to theoretical frequency calculations at the RI-MP2/cc-pVDZ level⁹. The results show that three of the conformers, whose origin transitions are close in energy, namely structures B-D in Figure 2, adopt similar gamma turn structures with minimal interactions involving the phenyl ring. The fourth and relatively red-shifted conformer, structure (A) with its origin at 37,370 cm⁻¹, adopts a unique folded structure with a strong dispersion interaction between the phenyl ring and the peptide carboxylic tail. By tuning the UV excitation laser to a specific resonant pi-pi* transition of the phenyl ring we can thus selectively produce FGG cations originating from specific neutral conformers.

According to photoelectron spectra¹⁷, the vertical ionization potentials of the phenyl ring and amino terminus lone pair orbitals on phenylalanine are 8.9 (adiabatic = 8.5) and 9.3 (adiabatic = 8.8) eV, respectively. At the highest frequency excitation in our experiments, two UV photons equal 9.3 eV so the excess energy imparted on the ion is in the range of 0 to 0.4 eV minus the kinetic energy carried off by the photoelectron.

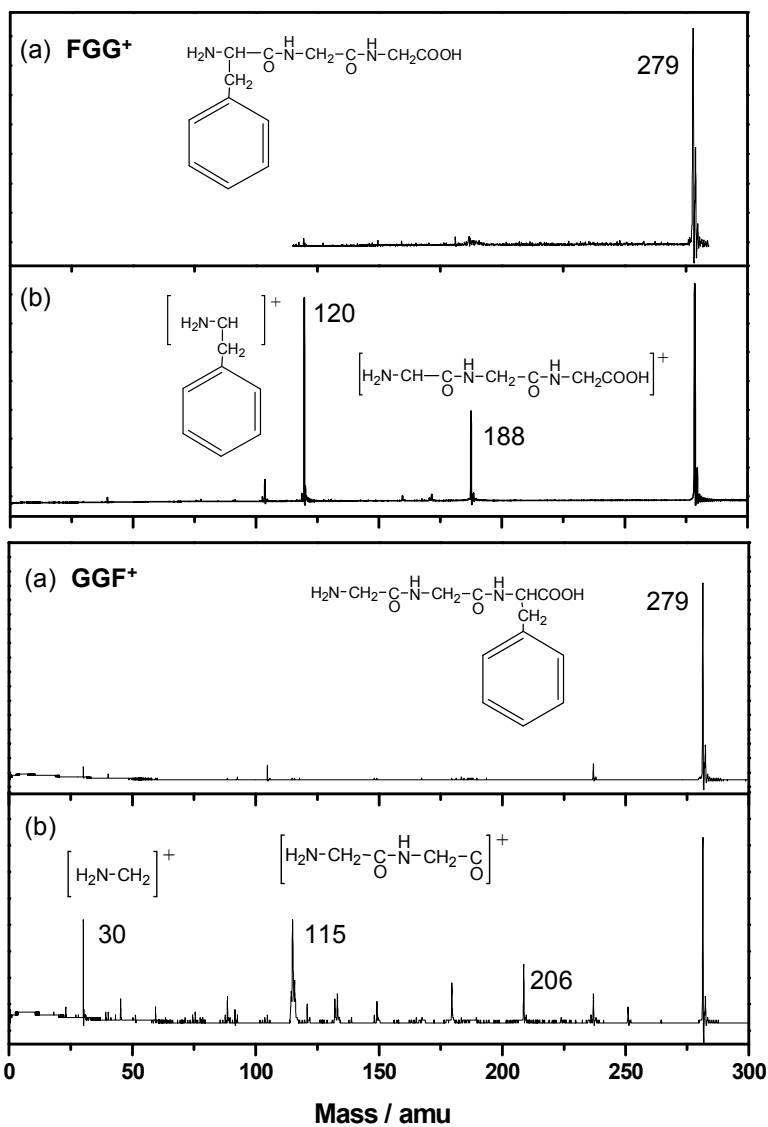


Figure 1. Time-of-flight mass spectra of **FGG⁺** and **GGF⁺** and fragment ions produced by (a) R2PI and (b) R2PI plus subsequent excitation at 600nm.

The flight time of ions in our reflectron mass spectrometer is on the order of tens of microseconds. Any statistical fragmentation caused by the excess energy from the 1+1 R2PI ionization process would have to occur within this time frame to be detectable. Specifically, any

fragmentation events that occur during the extraction phase would result in asymmetrical mass peaks. Any fragmentations that happen during the first leg of the reflectron phase would instead show up as new peaks with shorter arrival times. Neither of these phenomena occur in our 1+1 R2PI mass spectrum of FGG (Figure 1(a)). We conclude that our ionization step does not cause any fragmentation of FGG in the time scale of the experiment.

Figure 1(b) shows the ion dissociation spectrum of FGG, composed of ions generated by 1 + 1 R2PI and subsequently excited with a visible laser pulse at 600 nm. The ion dissociation spectrum features fragment masses appearing at 120 and 188 Dalton. Since neutral FGG is transparent to visible light, these fragment masses result from the excitation of the FGG cation by the visible laser pulse. The shift in the absorption spectra of aromatic molecules to the visible upon photo-ionization, and the states involved, are well known from spectroscopic studies of the benzene cation, substituted benzene cations, naphthalene cations and other charged aromatic species¹⁸. The open shell nature of phenylalanine cations makes it possible to promote a lower level electron into the half-filled HOMO by absorption of a visible photon. However, if charge redistribution or transfer occurs shortly after ionization, this channel might not remain available.

Only if the positive charge remains localized on the aromatic moiety in FGG can the peptide cation absorb a visible photon (2.0 - 2.3eV). According to previously proposed mechanisms based on ion dissociation studies, the cation may even absorb additional visible photons if it undergoes de-excitation or internal conversion to the ionic ground state, from where the visible electronic transition can re-occur^{19,20}. We note that the timescale for such sequential two photon excitation in our experiments is limited by the laser pulse width of about 5 ns.

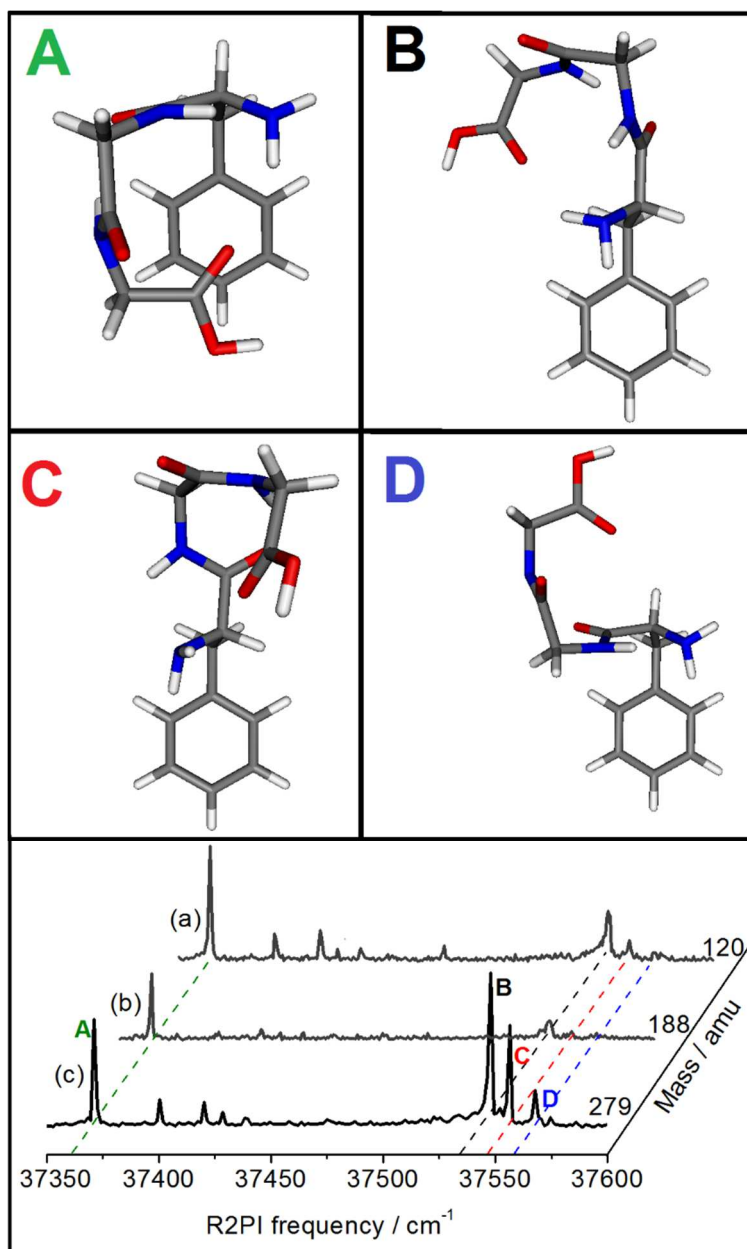


Figure 2. Theoretical neutral structures of **FGG** calculated at the RI-MP2/cc-pVDZ level and named according to respective origin transitions A-D. Ion dissociation spectra of fragment ions (a) 120 amu and (b) 188 amu, produced by R2PI plus subsequent VIS excitation. (c) R2PI of FGG, with labeled origin transitions A-D (VIS excitation off).

Figure 2 depicts the 1 + 1 R2PI spectrum of FGG (Mass = 279) and ion dissociation action spectra of the two-fragment species (mass 188, and mass 120). The origin transitions of each of the four conformers of FGG are labeled A-D in the 1+1 R2PI spectrum (trace c) and the structures corresponding to each transition are shown and also labeled A-D. Structure A (folded FGG

conformer) is very compact due to a strong dispersion interaction involving the pi system of the phenylalanine ring. Structures B-D (“extended” FGG conformers) do not contain interactions involving the phenylalanine ring, which is oriented outward to the side of the peptide chain. All four peptide structures contain a gamma tight turn. We calculated these structures at the RI-MP2/cc-pVDZ level and previously assigned them by comparing theoretical frequencies to the experimental IR spectra measured in our lab⁹.

Traces (a) and (b) in Figure 2 (mass 120 and mass 188) are the R2PI wavelength dependent ion dissociation spectra of the fragment masses which appear in Figure 1. The degree of fragmentation caused by the visible laser pulse depends on which initial conformer we resonantly excite and ionize in the R2PI step. The most red shifted FGG conformer (A) fragments with a fragment-to-parent ratio of 1.4 (estimated from ion intensities) whereas the three blue shifted FGG conformers (B-D) have fragment-to-parent ratios near 0.3. We observe no dependence of the ion dissociation spectra on visible wavelength in the region 532-680 nm.

5.3.2. GGF

There are three main conformations of GGF in the molecular beam. Strong origin transitions appear at $37,489\text{ cm}^{-1}$, $37,514\text{ cm}^{-1}$, and $37,540\text{ cm}^{-1}$. A very weak transition also appears at $37,420\text{ cm}^{-1}$. We assign the structure of each conformation by comparing the experimental IR hole burning spectra to theoretical frequency calculations at the RI-MP2/cc-pVDZ level. These results show that the origin transitions at $37,514\text{ cm}^{-1}$ and $37,540\text{ cm}^{-1}$ arise from distinct extended and folded conformers, respectively, shown in Figure 3 as structures A and B. The signal from the other two conformers was too weak to allow analysis of their

fragmentation. In experiments analogous to those on FGG, we selectively produce GGF cations originating from a unique neutral conformer and subsequently photodissociate the selected ions with a VIS pulse.

The ion dissociation mass spectrum of GGF^+ , generated by 1+1 R2PI plus subsequent VIS excitation appears in Figure 1. VIS photodissociation of GGF^+ produces two main fragments at 30 and 115 amu. The fragment at 115 amu exhibits an asymmetric TOF peak indicating metastable decay. Fragmentation peaks appear at a variety of mass values, however since mass 115 is metastable we cannot accurately assign the mass values of the smaller fragmentation peaks as they may originate from the same channel as mass 115. A less intense fragment appears at 206 amu, produced by cleavage of the N-C α bond in glycine2 and resulting in a positively charged phenyl containing fragment. Due to low signal strength we were unable to measure the wavelength dependence of this fragment.

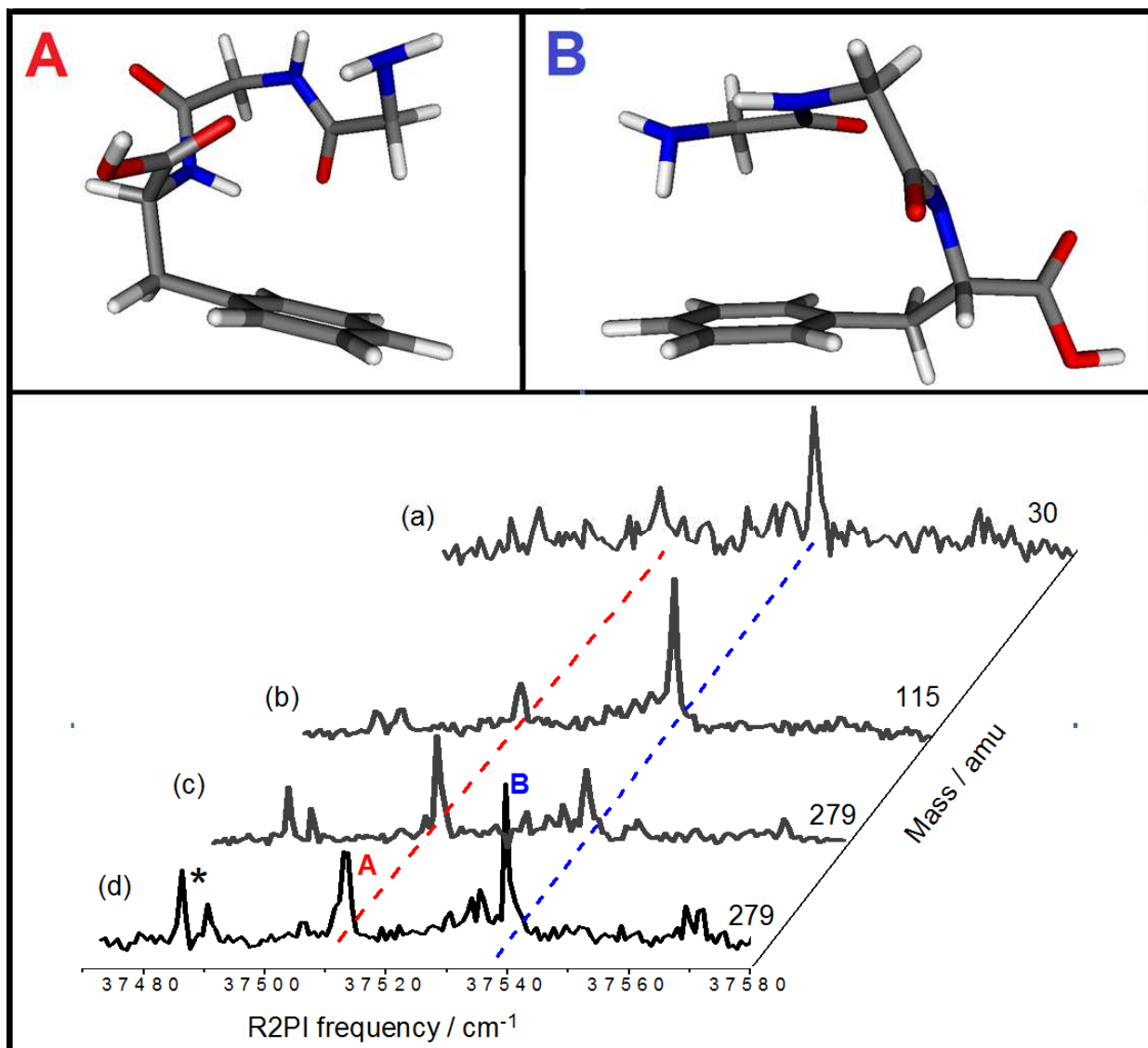


Figure 3. Theoretical neutral structures of **GGF** calculated at the RI-MP2/cc-pVDZ level and named according to respective origin transitions A,B. Ion dissociation spectra of fragment ions (a) 30 amu, (b) 115 amu and (c) depleted GGF parent ion produced by R2PI plus subsequent VIS excitation. (d) R2PI of GGF, with labeled origin transitions A, B and *. (VIS excitation off)

Figure 3 depicts the 1+1 R2PI spectrum of GGF parent (mass 279) and ion dissociation action spectra of the two fragment species (mass 115 and mass 30). The origin transitions of the two most intense conformers of GGF are labeled A and B in the 1+1 R2PI spectrum and the structures corresponding to these origin transitions are shown and also labeled A and B. Structure A (extended GGF conformer) is a gamma turn structure with little interaction between the peptide

chain and the pi system. GGF conformer B (folded GGF conformer) contains a gamma turn and an interaction between the phenyl ring and the peptide amino tail forming a folded structure.

Traces (a) and (b) in Figure 3 (mass 30 and mass 115) are the wavelength dependent ion dissociation spectra of the fragment masses. Trace (c) is the wavelength dependent ion depletion spectrum of the parent mass of GGF, which reflects changes in the ion intensity of the GGF^+ signal caused by VIS photodissociation. Since mass 115 breaks up metastably upon VIS photodissociation it is not possible to get a true measure of the degree of fragmentation from the intensity of the metastable ion peak alone. For this reason we include the ion depletion spectrum of the GGF parent mass. Trace (d) shows the R2PI spectrum of GGF in the absence of VIS excitation. The degree of ion fragmentation caused by the visible laser pulse depends on which conformer we resonantly excite and ionize. Ions formed from the extended GGF conformer (A) photodissociate with a fragment-to-parent ratio of less than 0.5 (estimated from ion intensities and the parent ion depletion R2PI spectrum) whereas ions formed from the folded GGF conformer (B) photodissociate with a fragment-to-parent ratio of more than 3. Similar to the FGG case, we observe no dependence of the ion dissociation spectra on visible wavelength in the region 532-680 nm. It is worth noting that FGG also displayed much higher rates of fragmentation when folded relative to when it was extended.

5.4. DISCUSSION

5.4.1 *Ion conformations*

UV resonant multi-photon ionization allows preparation of FGG and GGF ions with initial conformer selection. Thus it is possible to study subsequent VIS excitation and photodissociation as a function of the initial neutral tripeptide conformation. To the extent that a memory of the neutral conformation is preserved in the ensuing ion, we can then investigate the

effects of this initial conformation on the fragmentation dynamics in the ion. Therefore we first need to discuss the conformational landscape of the ions produced by the photoionization step. Our results suggest that whatever the final conformations are of the ions following R2PI, they are different depending on the starting structure. At the 400 ns timescale, apparently sufficient memory of the original structure of the neutral peptide is preserved in the cation to lead to significantly different photodissociation dynamics. While photoisomerization might occur, at the very least these results show that conformations are not completely randomized. Figure 4 shows the calculated ionic shapes that result from removal of the pi electron from the phenyl ring and re-optimization. This calculation simulates photoionization without significant internal energy in the ion. Such an approach can be rationalized as follows:

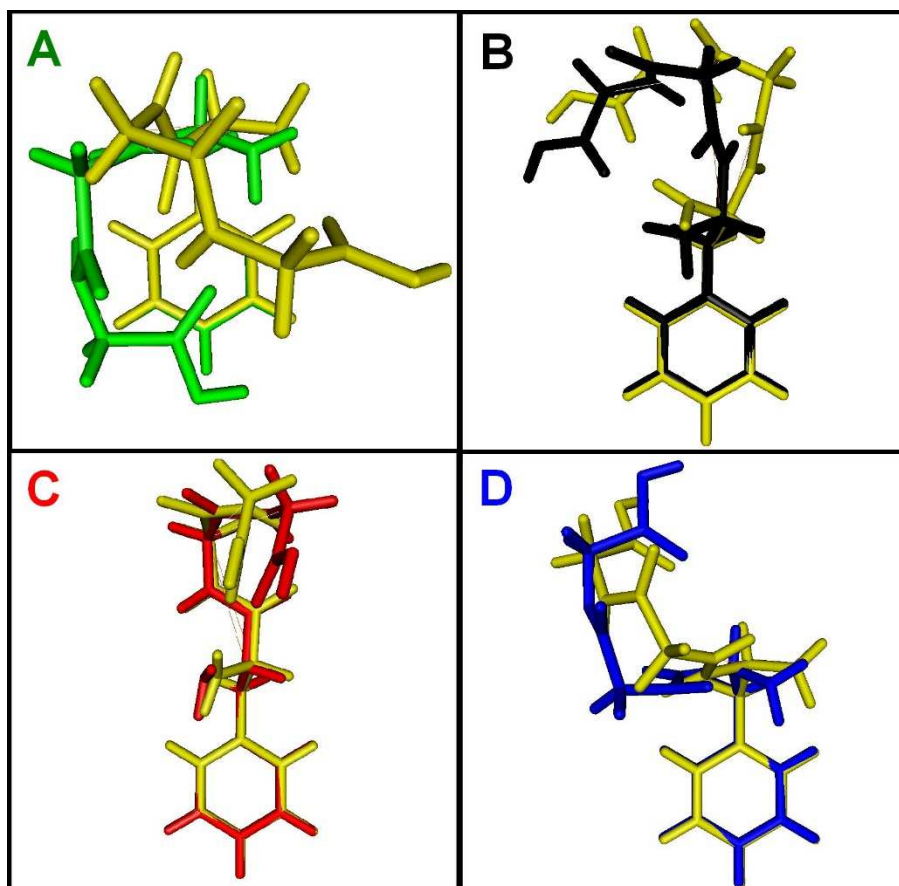


Figure 4. Neutral conformers of FG (same as those at the top in **Figure 2**) color-coded and labeled according to their respective origin transitions and overlaid with their corresponding cation structures (in yellow) calculated at the B3LYP/6-311++G(d,p) level.

Excess energy after photoionization is at most 0.7 eV, assuming the adiabatic IP of phenylalanine, a significant part of which is carried off by the photoelectron as kinetic energy. If the ion is not formed at its minimum geometry, it will undergo IVR and several tenths of an eV of internal energy may be available but that is still likely to lead to long timescales for isomerization. Figure 4 shows the optimized cation geometries for FGG, as calculated at the B3LYP/6-311++G(d,p) level, in yellow. Upon ionization, the folded conformer A of FGG unfolds and the termini attract one another, forming a hydrogen bond. As the dispersive force disappears upon ionization, the hydrogen-bonding propensities take precedence. In the extended conformers B and C the repulsion between the positive charges of the phenyl ring and the amino terminus is modulated by the formation of a hydrogen bond between the termini. The ring twists away to avoid the charge interaction, while at the same time both termini twist in order to draw closer to one another. Conformer D is the most extended neutral conformer, with very little interaction between the ring and the amino terminus, and not surprisingly there is scarcely any geometry change from neutral to cation. The geometry optimization with the X3LYP functional converges to the same cation structures as those obtained with B3LYP for all FGG conformers except for the one originating from Conformer C, which is shown in Figure 5. The X3LYP structure predicts the carboxylic tail twisting further away from the phenyl ring in comparison to the B3LYP structure.

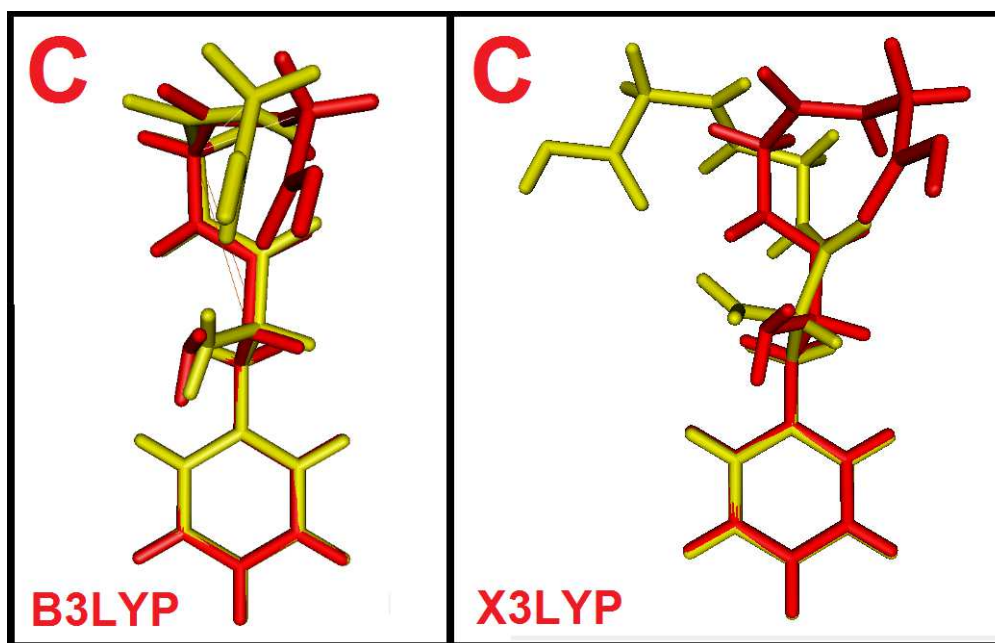


Figure 5. At right, the neutral conformer C of FG G overlaid with its corresponding cation structure (in yellow) calculated at the X3LYP/6-311++G(d,p) level. At left is FG G Conformer C overlaid with the B3LYP-optimized structure as seen in **Figure 4**, for comparison. All other FG G neutral conformers optimized with X3LYP converged to the same cation structures as those calculated with B3LYP (see **Figure 4**)

5.4.2. Fragmentation dynamics

Figure 6 shows the geometries of the GGF cation calculated with B3LYP. In these conformers, the transition from neutral to cation is accompanied by the predicted geometry change—the now-positively charged aromatic ring repels the amino terminus, causing the ends to twist away from one another in the cation structure. Multiple instances of hydrogen bonding also appear, in contrast to the cation structures of FG G, which may explain the more folded forms of both the “folded” (dispersion-stabilized conformer B) and “extended” (conformer A with little dispersion interaction) structure. **Figure 7 shows** the X3LYP-optimized cation structures of GGF. The main difference seen is that in the X3LYP structures, the tail pulls vertically away from the phenyl ring, instead of twisting away as the tails in the B3LYP structures.

The results intimate conformational dependence either of the photo-excitation of the cation or of the following fragmentation dynamics. Considering the latter first we must note the time

scales in our experimental arrangement. Ions are formed in the extraction field of the time-of-flight source. Therefore we can only observe fragments that are formed in the first approximately 100 ns following formation of the original ion. Fragments formed after about 100 ns but within approximately 1.5 μ s appear somewhat later in the TOF spectrum, forming a tail to the fragment mass peak. After about 1.5 μ s all ions have exited the extraction region and any fragments formed in the first leg of the TOF will show up at very different arrival times, while fragments formed in the second leg of the TOF will arrive at the same time as the parent ions. This arrangement practically limits the time window of the observed fragments to the order of about 100 ns. The mass 115 fragment for GGF (Figure 1b) appears as a broad peak, corresponding to fragmentation in one to several tens of microseconds. Other unassigned fragment peaks in the mass spectrum of GGF likely correspond to metastable fragmentation in the first leg of the time-of-flight, covering a period of up to about 50 μ s. These time scales preclude observation of statistical fragmentation. Hu et al. observed photodissociation fragments of thermalized LeuTyr and LeuLeuTyr ions, photodissociated at 579 nm, with rates of $\sim 10^2$ - 10^3 s⁻¹ and calculated that this is consistent with statistical fragmentation. Absorption of two visible photons, providing an extra ~ 2.5 eV of additional internal energy would, according to these calculations, lead to much shorter - sub microsecond - fragmentation times, which is relevant for the discussion of two photon dissociation below. It is thus possible that a fraction of the ions fragments statistically after absorption of one visible photon, but that this fraction remains unobserved in our experiment.

Non-statistical fragmentation suggests absorption to a repulsive state which could depend on conformation since the potential energy landscape depends on conformation. Different potential surfaces could affect Franck-Condon factors as well as different competing channels, such as internal conversion. However non-statistical fragmentation is rather unusual so next we

will consider the possibility of two photon dissociation, which can lead to statistical dissociation at the timescales of our experiment. In this case conformational dependence could occur in the absorption of either the first or the second photon.

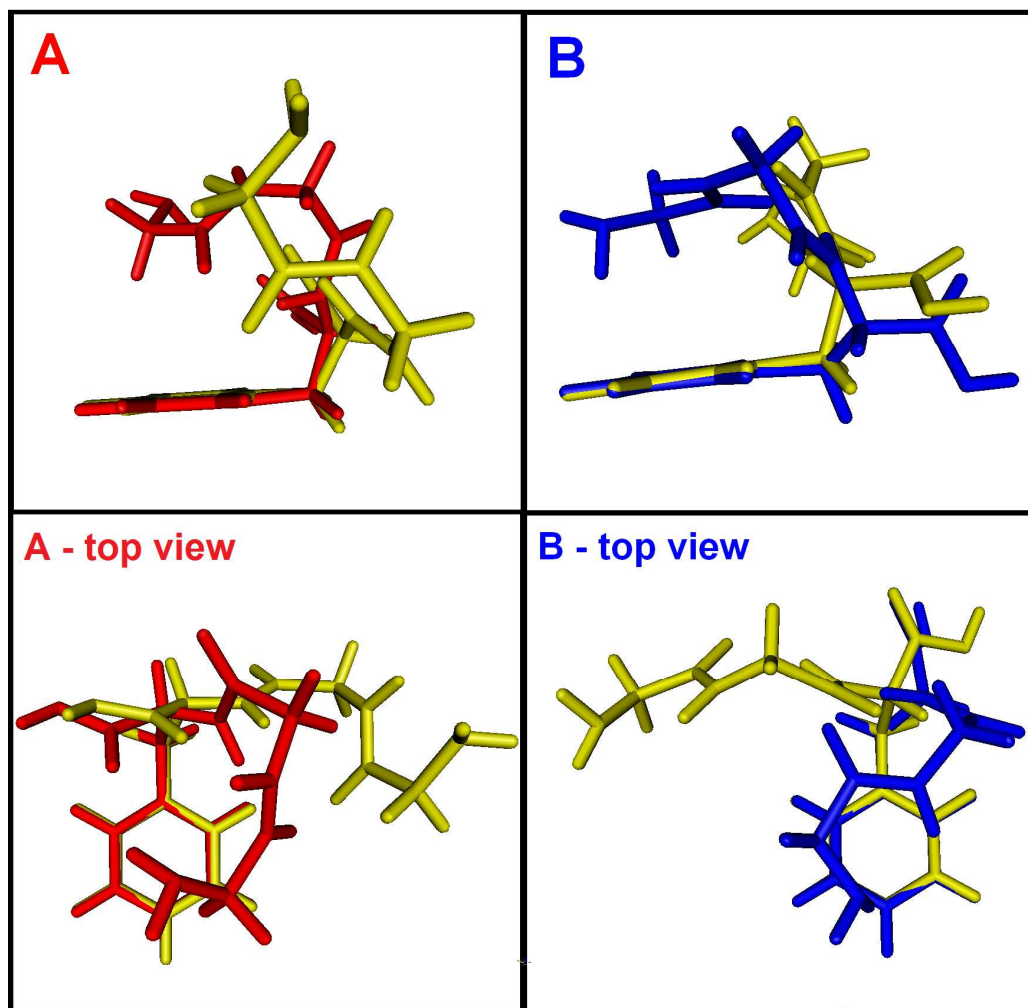


Figure 6. Neutral conformers of GGF (same as those at the top in **Figure 3**) color-coded and labeled according to their respective origin transition and overlaid with their corresponding cation structures (in yellow) calculated at the B3LYP/6-311++G(d,p) level.

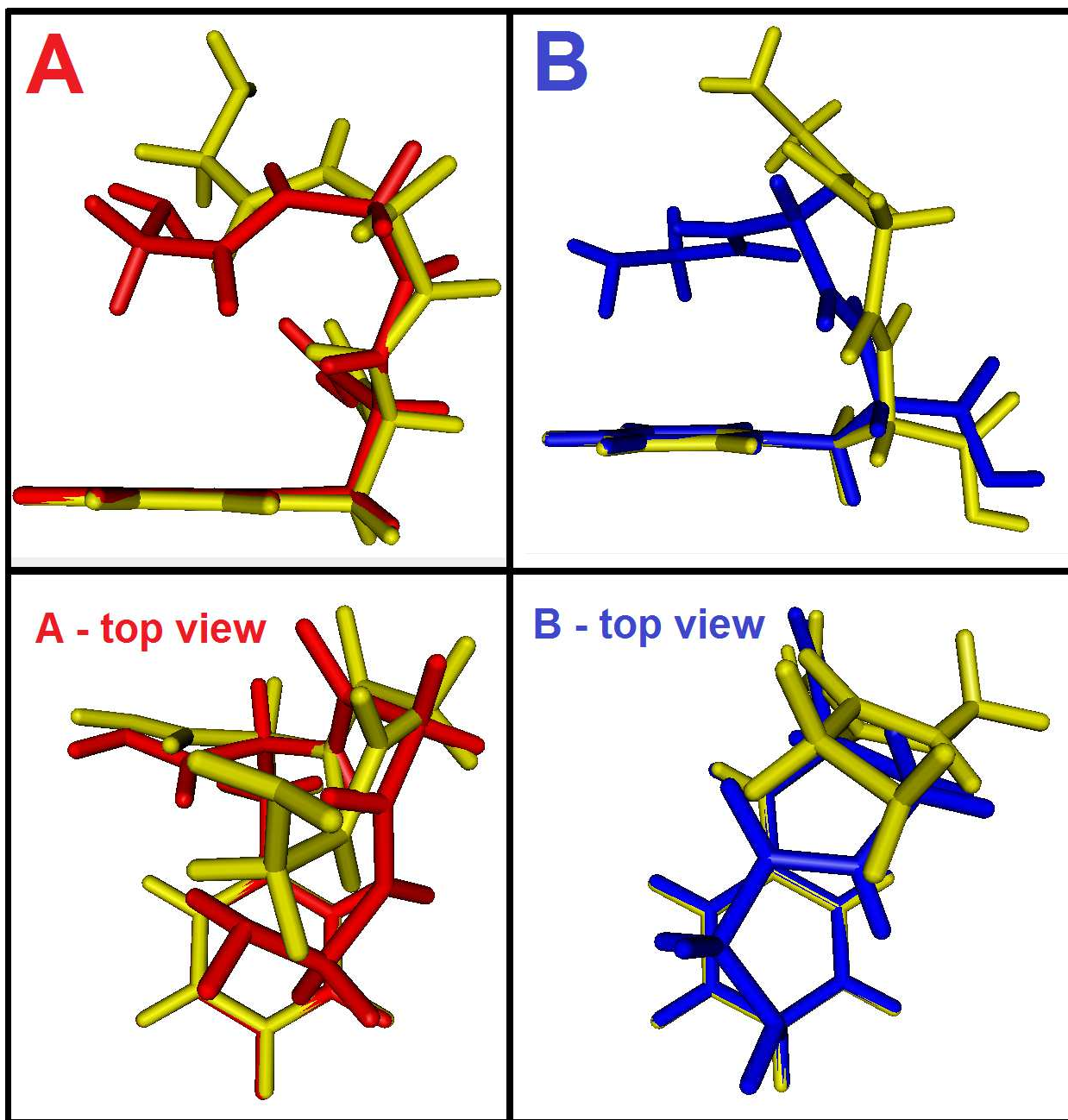


Figure 7. Neutral conformers of GGF (same as those at the top in **Figure 3**) color-coded and labeled according to their respective origin transition and overlaid with their corresponding cation structures (in yellow) calculated at the X3LYP/6-311++G(d,p) level.

5.4.3. Photo-excitation

Visible absorption - and thus visible photodissociation - of the tripeptides can only occur when positive charge is localized on the phenylalanine ring. According to photoelectron spectra of amino acids phenylalanine and glycine, the non-bonding C-terminus oxygen and N-terminus nitrogen lone pair orbitals have vertical ionization potentials of 10.5 eV and 9.3 eV, respectively. The N-terminus lies close in energy to the phenyl π HOMO (vertical IP = 8.9 eV), but since the visible transition focuses on promoting a π electron into the vacancy created by ionization, we approximate the transition strength of a positively charged N terminus by a visible photon as at least 3 orders of magnitude smaller, due to its $n \leftarrow \pi^*$ character, than that of the phenyl cation. The highest occupied orbitals in the peptide backbone, with ionization potential of approximately 9.5 eV, are located on the peptide bonds. All electronic transitions of the peptide backbone groups are greater than 2.3 eV and non-resonant with visible excitation. Therefore the transition to consider for visible absorption by the cation is the excitation of π_1 electron to the π_3 orbital of the phenyl chromophore¹⁸.

Photo-excitation of the ion can depend on conformation in two ways: Either the probability for excitation to occur at all or the probability of absorption of a second photon after initial excitation of the ion can depend on the conformation. Two models put forward in earlier experiments serve as precedents for each of these scenarios. (I) Charge delocalization can reduce the visible absorption cross section of the phenyl chromophore, lessening the probability of absorbing any visible radiation. (II) Charge transfer following absorption of a first photon can occur, rendering the chromophore transparent before a second photon can be absorbed,

preventing fragmentation if two visible photons are required to fragment the ion.

Lee et al. explain their results for phenylalanine in terms of model (I), charge delocalization. They argue that certain conformers are able to split the positive charge between the ring and the N-terminus lone pair. This delocalization of the charge away from the ring decreases the cross section of visible light absorption, potentially quenching the absorption altogether. By this model, conformers that delocalize their charge will fail to capture visible light photons and are not expected to undergo additional fragmentation after exposure to a visible light laser. The converse is also expected, i.e. that structures that maintain the charge on the ring can potentially absorb visible light, and that this absorption can lead to subsequent fragmentation.

Weinkauff et al. explained previous studies by model (II) ^{10,11}. They examined charge transfer and non-statistical dissociation pathways in small peptides such as Leu-leu-leu-Trp and Leu-gly-leu-Trp. In a series of experiments, Weinkauff et al. demonstrated the ability to discern charge transfer pathways, and barriers to charge transfer, by comparing fragmentation products of peptide ions excited with two visible photons versus excitation by one UV photon. By substitution of the amino acid glycine into the peptide chain they observed that sequence affects barriers to charge transfer and suggested that the relative local ionization potentials of amino acids in the chain governed the efficiency of “through bond” charge transfers. A series of theoretical papers followed, summarized in a more recent paper in which Schlag et al. ²³ discuss the alternate decay pathways in energetic peptides that can produce the congruous results of non-statistical fragmentation observed by Weinkauff et al. and statistical dissociation rates and products systematically determined by Hu et al.¹.

We note that the charge transfer mediated model predicts no fragmentation for FGG and GGF because of the relative values of the ionization potentials of phenylalanine and glycine, contrary to our current observations. We also note that both the earlier experiments by Weinkauff et al. and those by Hu et al. were performed at a single ionization wavelength and therefore lack the conformational selection of the present experiments.

In accord with both models of either delocalization or charge transfer, proposed by Lee et al. and Weinkauff et al., respectively, we observe that the secondary structure of the peptide selectively quenches VIS absorption by affecting the stability of charge relative to aromatic ring. The difference in energy between the $S_0 \leftarrow S_1$ transitions for the different conformers is less than 0.02 eV and we assume that the peptides behave similarly in terms of the deposition of excess ionization energy.

In the Weinkauff et al. studies, fragmentation was observed in any peptide sequence where the difference in local ionization energies between neighboring residues presented an insurmountable energetic barrier to the charge. It was postulated that when the relevant molecular orbital positions for adjacent residues cause a situation in which the electron must gain energy in order to fill the electron hole, the charge transfer is barred, causing the electron hole to return to the chromophore, permitting visible absorption. Weinkauff et al. found that ionized tripeptide sequences containing tryptophan as the chromophore and glycine as the nonaromatic residues fragmented with subsequent visible radiation, whereas when tyrosine was used as the chromophore, charge transfer was permitted and thus no fragmentation was observed. If we

attempt to apply this model to our system, then we would expect to find no or little fragmentation for either GGF+ or FGG+.

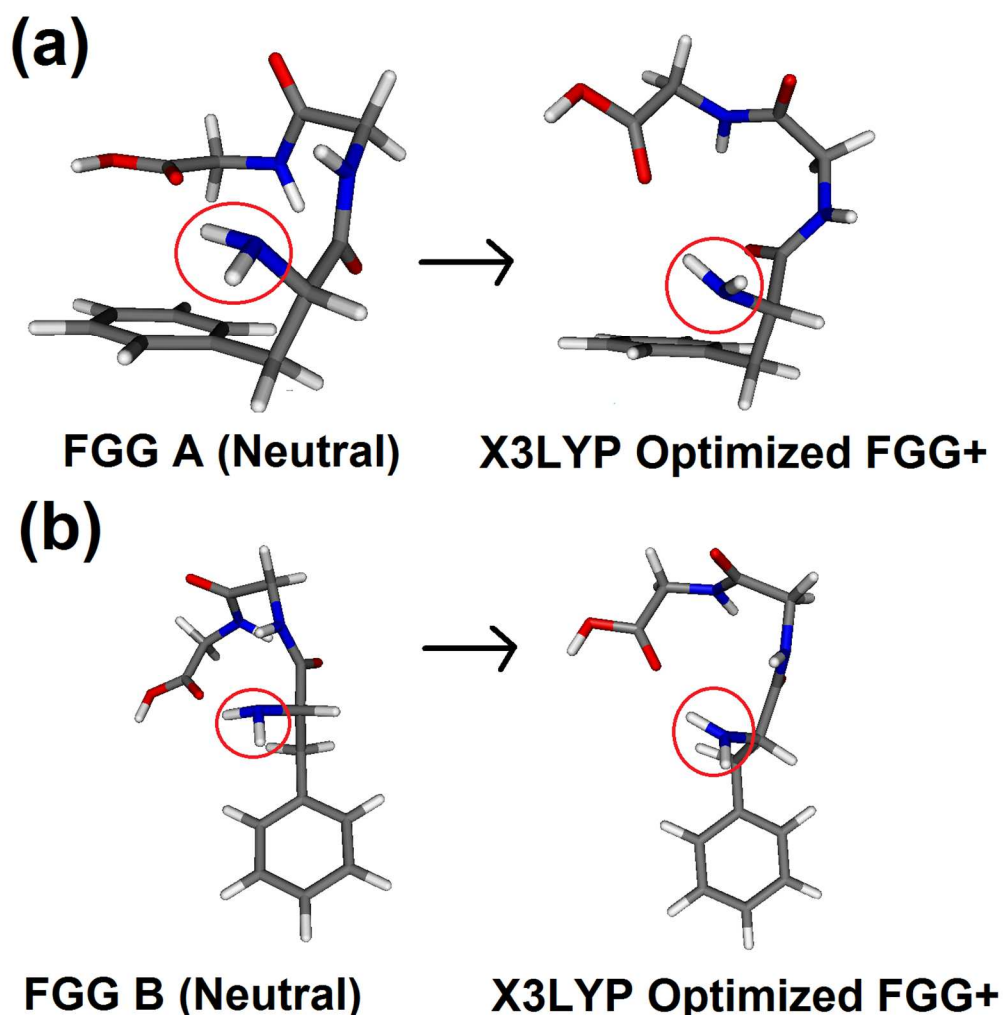


Figure 8. The change in planarity of the amino group for neutral-to-cation transformations starting from (a) the neutral folded conformer A of FGG and from (b) the neutral extended conformer B of FGG. All of the B3LYP and X3LYP-optimized structures exhibit this trend of greater NH₃ planarity in the cation.

Our results are more consistent with the charge delocalization model, in which higher fragmentation is predicted for a peptide conformer that somehow allows significant

delocalization of positive charge away from the phenyl ring. In this model, because the charge flows via delocalization, the energetic barrier of Gly residues does not necessarily pose a problem. As discussed, any evidence of a more delocalized charge distribution for extended conformers supports the validity of applying this model to our data.

In order to explain how vertical ionization energies also provide support for the charge delocalization model, we summarize the arguments of a follow-up for the Lee et al. study²⁴. It was proposed that the difference in fragmentation propensity was due to charge delocalization caused by the formation of hyperconjugative orbitals between the nonbonding (n-MO's) on the amino end and the π -MO's on the phenyl ring in the second conformer subgroup of neutral Phe. It was suggested that hyperconjugative orbitals are able to form because the energy of the n-MO on the nitrogen is close to the energy of the π -orbitals of the phenyl ring. In conformers possessing hyperconjugative orbitals, ionization may occur by the removal of an electron from the hyperconjugative orbitals, resulting in a distribution of positive charge that is divided between the amino terminus and the phenyl ring. The reduction of positive charge localized on the phenyl ring causes a lower absorption cross section for visible radiation, which results in a lower fragmentation rate for these conformers. The first conformer subgroup does not form hyperconjugative orbitals because the hydrogen-bonding scheme in these conformers lowers the energy of the n-MO's of the lone pair on the amino nitrogen so that is no longer close to that of the phenyl π -orbitals. Thus, in ionization, the electron is mainly removed from the phenyl ring and the resulting positive charge is localized there, allowing for visible absorption and subsequent fragmentation.

Lee et al. support the existence of hyperconjugative orbitals through the calculation and experimental measurement of ionization energies and orbital energies. Using R2PI photoionization efficiency curves and photoelectron spectroscopy, the vertical and adiabatic ionization energies were obtained. These were also calculated theoretically with various DFT functionals. They observed that the ionization energies of subgroup 1 conformers were consistently greater than that of subgroup 2. This was consistent with the theoretical ionization energies as well as the theoretical orbital energies as obtained from an NBO analysis.

We calculated charge distributions by the Natural Bond Orbital (NBO) analysis approach. **Table 1** lists the charge values of the amino terminus and of the phenyl moiety for each neutral and cation conformer of GGF and FGG, calculated using four different DFT functionals. The “Diff.” column is the difference between the neutral charge value and that of the cation structure, and it represents the amount of “positive charge” that is located on each group (that is, a more positive value indicates that more of the negative charge was removed from that group) in the cation. We see that, except in a few cases, there is a greater amount of positive charge on the phenyl group for the cations originating from the folded structures (FGG A and GGF B) than for the cations originating from the extended conformers.

The two main arguments for the delocalization model are: (1) We see more planarization of the amine in the extended conformers than we do the folded ones. Since Lee predicts that planarization is a consequence of charge delocalization, and we get less fragmentation in the extended structures, this supports that model. But this only happens for FFG, which makes any conclusions weak. (2) There is no sequence dependence on in our experiments, only folded vs

non-folded. Given the large change in energy for N vs C terminal G, this is damaging to the charge transfer model (although it is not really direct support for delocalization either).

Table S2 lists the vertical ionization energies corresponding to each conformer of FGG and GGF, calculated with four different DFT functionals. In accordance with this, we see that the vertical ionization energy for the folded conformer of GGF (the conformer that experiences a greater degree of fragmentation by visible radiation) is calculated to be higher than that of the extended conformer, with all the functionals used. However, the vertical ionization energy for the folded conformer of FGG does not have a higher value than the extended conformers and so these values are not consistent with the model.

Certain neutral-to-cation geometry changes that we have observed also support the idea of a more phenyl-localized electron hole for folded conformers and a more delocalized charge distribution (i.e. shared between the phenyl ring and the amino terminus) for the extended conformers. As previously mentioned, positive charge in the amino group is usually indicated by a more planar geometry. If the amino group cannot adopt the planar conformation, it is less likely to accept charge. It is the next most favorable place for the charge to reside, so if it cannot reside there, it will go back to the phenyl ring, facilitating absorption and fragmentation. In accord with this prediction, as seen in Figure 10, the amino termini of the extended cation conformers of FGG (from both B3LYP and X3LYP) are planar in geometry (in comparison to that of the folded conformer cation), suggesting the presence of a greater degree of positive charge.

The GGF cations do not have planar amino groups, which is likely due to the multiple hydrogen bonds formed in these structures. This result is consistent with the low amount of

positive charge in the amino group even for the extended conformer. For this conformer, the phenyl ring still experiences charge delocalization, but the positive charge is delocalized to the rest of the chain instead of to the amino terminus.

Another feature of the cation geometry that may serve as support to the delocalization model is the marked elongation of the side chain carbon-carbon bond connecting Phe to the chain in the extended conformers in comparison to that of the folded conformer in FGG as seen in Figure 6. This elongation may be further proof of the proposed charge migration for the following reason: The amino group and the phenyl group are adjacent in the FGG sequence, and so, if hyperconjugative orbitals allow a split charge distribution that is nearly equal in weight for the FGG extended conformers, then negative charge will pool in the carbon-carbon bond between the two groups, causing elongation as discussed in the Lee et al. studies^{7,23}.

We also find that photodissociation by a UV pulse does not replicate the spectrum formed by photodissociation with a VIS pulse. In FGG, photodissociation by a UV photon opens the mass 120 fragmentation channel but not the mass 188 channel. In GGF, the large metastable fragment mass 115 is also absent in the UV photodissociation spectrum however the mass 30 fragment appears. We obtain similar results at both 298 nm and 334 nm UV photodissociation wavelengths. These results suggest possible differences in dynamic barriers or dissociation thresholds in the different channels, however they could also reflect excitation to different locations on the same electronic state or to different electronic states, initiating divergent dynamics.

5.5 CONCLUSION

We observe conformer specific fragmentation in the tripeptides FGG and GGF. We find that the folded structures and extended structures of FGG fragment with fragment-to-parent ratios of 1.4 and 0.3, respectively, and the folded and extended structures of GGF fragment with fragment-to-parent ratios near 3 and 0.5, respectively. We conclude that the dynamics of ionized and excited tripeptides depends on the original neutral structure. Our observations can in principle be rationalized by either of the two models. In each case the local charge distribution of the ion is modified, either before or after absorption of one visible photon. The resulting new charge distribution modifies the visible absorption cross section and depends on the starting conformational structure that was originally photoionized. This suggests two implications: (1) The ion conformational structure, which of course will be different from that of the neutral, retains some memory of the original conformer selection. Thus upon photoionization different neutral conformers yield different ionic conformer populations. (2) Charge redistribution - be it by delocalization or by photoinduced charge transfer - depends on conformational structure.

Our results are most consistent with the charge delocalization model developed by Lee et al., which asserts that charge delocalization is the mechanism by which the visible absorption cross section of the phenyl ring is modulating—greater delocalization out of the phenyl ring causes reduced absorption and therefore less fragmentation. Our theoretical results, including DFT-optimized cation structures, NBO charge distributions, and vertical ionization energies, mostly support the idea of a charge distribution that is localized on the phenyl ring for folded structures and delocalized to the amino terminus or to the peptide backbone for the extended structures.

ame	N	thod	Me	eut. Amino	N	ion Amino	Cat	f.	Dif	ut.	Ne	ion Phenyl	Cat	f.	Dif
GF A	G	LYP	B3	0.12643	-	0.08449	-	4194	0.0	0.01413	-	1752	0.5	3165	0.5
		LYP	X3	0.12658	-	0.07825	-	4833	0.0	0.01591	-	7018	0.3	8609	0.3
		6-2X	M0	0.13065	-	0.0394	1.1	3459	1.2	0.0156	-	5775	0.3	7335	0.3
		D	B97	0.12213	-	0.5261	0.1	7474	0.2	0.01593	-	1114	0.2	2707	0.2
GF B	G	LYP	B3	0.13095	-	0.10367	-	2728	0.0	0.01411	-	3595	0.4	5005	0.4
		LYP	X3	0.13117	-	0.11515	-	1602	0.0	0.01399	-	4551	0.4	595	0.4
		6-2X	M0	0.1354	-	0.11022	-	2518	0.0	0.01492	-	146	0.0	2952	0.0
		D	B97	0.12606	-	0.887	0.0	1476	0.2	0.01577	-	8169	0.1	9746	0.1
GG A	F	LYP	B3	0.11912	-	0.0357	0.1	2269	0.2	0.03244	-	4791	0.4	8035	0.4
		LYP	X3	0.11937	-	0.9664	0.0	1601	0.2	0.03228	-	5851	0.4	9079	0.4
		6-2X	M0	0.12472	-	0.784	0.1	0312	0.3	0.03138	-	5775	0.3	8913	0.3
		D	B97	0.11322	-	0.83	0.0	9622	0.1	0.036	-	8617	0.2	2217	0.3
GG B	F	LYP	B3	0.12141	-	0.5577	0.1	7718	0.2	0.03171	-	5426	0.3	8597	0.3
		LYP	X3	0.12165	-	0.58	0.1	7965	0.2	0.03152	-	5635	0.3	8787	0.3
		6-2X	M0	0.1266	-	0.9806	0.1	2466	0.3	0.03178	-	0791	0.3	3969	0.3
		D	B97	0.1157	-	0.7997	0.0	9567	0.1	0.03502	-	1321	0.3	4823	0.3
GG C	F	LYP	B3	0.12194	-	0.7246	0.1	944	0.2	0.02048	-	6453	0.2	8501	0.2
		LYP	X3	0.12216	-	0.1271	0.2	3487	0.3	0.02031	-	588	0.2	7911	0.2
		6-2X	M0	0.12623	-	0.7234	0.2	9857	0.3	0.02129	-	0706	0.1	2835	0.1
		D	B97	0.11671	-	0.466	0.0	6331	0.1	0.02294	-	9464	0.2	1758	0.3
GG D	F	LYP	B3	0.13068	-	0.5715	0.0	8783	0.1	0.02842	-	6527	0.3	9369	0.3
		LYP	X3	0.13093	-	0.6141	0.0	9234	0.1	0.02825	-	7599	0.3	0424	0.4
		6-2X	M0	0.13488	-	0.1299	0.1	4787	0.2	0.02942	-	0841	0.4	3783	0.4
		D	B97	0.12514	-	0.4576	0.0	709	0.1	0.03104	-	841	0.2	1514	0.3

Table 1. Natural population analysis results for each GGF/FGG conformer, calculated with four different DFT functionals. The natural charge values for each atom in the amino terminus and the phenyl group are added up to obtain a total value, in both the neutral and the cation structure. The difference between the charge value of the neutral and the cation is taken in order to give an indication of the “positive charge” localized on each group.

Figure 12. Vertical ionization energies for each conformer of FGG/GGF, calculated with the four DFT

	B3LYP	X3LYP	M06-	B97D
	V.I.E (eV)	V.I.E (eV)	2X V.I.E (eV)	V.I.E (eV)
FGG				
A	8.29	8.29	8.94	7.70
FGG				
B	8.29	8.29	8.97	7.67
FGG				
C	8.33	8.34	9.06	7.73
FGG				
D	8.30	8.31	9.01	7.69
GGF				
A	8.51	8.51	9.16	7.89
GGF				
B	8.35	8.36	9.06	7.74

functionals, with no zero-point correction. [Will fix these with the zero-point corrections, and add the adiabatic ionization energies when done.]

References

- (1) Hu, Y. J.; Hadas, B.; Davidovitz, M.; Balta, B.; Lifshitz, C. *J. Phys. Chem. A (USA)* **2003**, *107*, 6507-6514.
- (2) Li, L.; Lubman, D. M. *Analytical Chemistry* **1988**, *60*, 1409-1415.
- (3) Tembreull, R.; Lubman, D. M. *Analytical Chemistry* **1987**, *59*, 1003-1006.
- (4) Antoine, R.; Broyer, M.; Chamot-Rooke, J.; Dedonder, C.; Desfrancois, C.; Dugourd, P.; Gregoire, G.; Jouvet, C.; Onidas, D.; Poulain, P.; Tabarin, T.; van der Rest, G. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1648-1652.
- (5) Weinkauff, R.; Lehrer, F.; Schlag, E. W.; Metsala, A. *Faraday Discuss.* **2000**, 363-381.
- (6) Remacle, F.; Levine, R. D.; Schlag, E. W.; Weinkauff, R. *J. Phys. Chem. A (USA)* **1999**, *103*, 10149-10158.
- (7) Lee, K. T.; Kim, H. M.; Han, K. Y.; Sung, J.; Lee, K. J.; Kim, S. K. *J. Am. Chem. Soc.* **2007**, *129*, 2588-2592.
- (8) Meijer, G.; de Vries, M. S.; Hunziker, H. E.; Wendt, H. R. *Applied Physics B* **1990**, *51*, 395-403.
- (9) Reha, D.; Valdes, H.; Vondrasek, J.; Hobza, P.; Abu-Riziq, A.; Crews, B.; de Vries, M. *S. Chem-Eur J.* **2005**, *11*, 6803-6817.
- (10) Valdes, H.; Pluháčková, K.; Pitonák, M.; Řezáč, J.; Hobza, P. *Phys. Chem. Chem. Phys.* **2008** *10*, 2747-2757.
- (11) K. Kim, K.D. Jordan. *J. Phys. Chem.* **1994** *98*, 10089-10094
- (12) Xu, X. ; Goddard, W.A. *Proc. Natl. Acad. Sci. USA.* **2003** *101*, 2673- 2677.
- (13) Hohenstein, E.G.; Chill, S.T.; Sherrill, C.D. *J. Chem. Theory Comput.* **2008**, *4*, 1996-2000.
- (14) Grimme, S. *J. Chem. Theory Comput.* **2006** *27*, 1787-1789.
- (15) Gaussian 03, Revision C.02, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, Jr., J. A.; 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, O.; 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.; and Pople, J. A.; Gaussian, Inc., Wallingford CT, **2004**.

(16) Gaussian 09, Revision **A.1**, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; 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.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian, Inc., Wallingford CT, **2009**.

(17) Campbell, S.; Marzluff, E.M.; Rodgers, M.T.; Beauchamp; Rempe M.E.; Schwink, K.F.; Lichtenberger, D.L. *J. Am. Chem. Soc.* **1994**, *116*, 5257-5264.

(18) Walter, K.; Weinkauff, R.; Boesl, U.; Schlag, E. W. *Chem. Phys. Lett.* **1989**, *155*, 8-(19)

Weinkauff, R.; Schanen, P.; Yang, D.; Sonkara, S.; Schlag, E. W. *J. Phys. Chem.* **1995**, *99*, 11255-11265.

(20) Weinkauff, R.; Schanen, P.; Metsala, A.; Schlag, E. W.; Burgle, M.; Kessler, H. *J. Phys. Chem.* **1996**, *100*, 18567-18585.

(21) Wilson, K.R.; Jimenez-Cruz, M.; Christophe, N.; Belau, L.; Leone, S.R.; Musahid, A. *J. Phys. Chem. A* . **2006** *110*, 2106-2113.

(22) Lehr, L.; Horneff, T.; Weinkauff, R.; Schlag, E. W. *J. Phys. Chem. A (USA)* **2005**, *109*, 8074-8080.

(23) Schlag, E. W.; Selzle, H. L.; Schanen, P.; Weinkauff, R.; Levine, R. D. *J. Phys. Chem. A (USA)* **2006**, *110*, 8497-8500.

(24) Baek, K.Y.; Hayashi, M.; Fujimura, Y; Lin, S.H.; Kim, S.K. *J. Phys. Chem. A* **2010** *114*, 7583-7589.

6. Low-temperature Formation of Carbonaceous Dust Grains from PAH

6.1. Introduction

Cosmic carbon is fundamental to the formation of complex organic molecules (COMs), the primordial building blocks in prebiotic chemistry. Tracing the origin and evolution of carbon in space is thus necessary to understand the origin of life on Earth. Carbon is first formed in the interiors of stars via the triple-alpha process and scattered into space as molecules and dust by massive stars such as red giants, casting away their carbon-enriched atmospheres into the interstellar medium (ISM) (Zuckerman 1980; Henning & Salama 1998). Carbon then becomes a key ingredient in the formation of the next generation of stars with surrounding planets. Astronomical observations of carbon in space have driven past and current missions to push the limits of sensitivity, spatial and spectral resolution, leading to the detection of COMs surrounding protostars (Öberg et al. 2011), young stars (Lee et al. 2019), and protoplanetary disks (Walsh et al. 2016; Favre et al. 2018). Cosmic carbon has been observed in both the gaseous and condensed phases. The infrared emission features between 3 and 20 μm were first observed with the *Infrared Space Observatory* and later with *Spitzer* toward diverse astronomical objects (H II regions, reflection nebulae, post-AGB stars). The bands at 3.3, 6.2, 7.7, 8.6, and 11.3 μm were first attributed to polycyclic aromatic hydrocarbons (PAHs) (Léger & Puget 1984; Allamandola et al. 1985; see review by Tielens 2008).

Although the structure and individual components of the carriers are still not precisely determined, they are conventionally attributed to a distribution of aromatic carriers (Peeters et al. 2002; Li & Draine 2012) including large PAHs with compact regular structures (Boersma et al.

2009), PAHs with aliphatic sidegroups (Yang et al. 2017), or mixed aromatic aliphatic structures (Kwok & Zhang 2013). The presence of both PAHs and very small grains has also been proposed to explain both the narrow and plateau features in the infrared emission (Peeters et al. 2017).

Infrared absorption bands at 3, 3.28, and 3.4 μm were first observed along a dozen lines of sight toward the Galactic center (Pendleton et al. 1994). The 3.4 μm band has been attributed to saturated aliphatic hydrocarbons in interstellar grains (Pendleton et al. 1994; Dartois et al. 2007), and is also considered the main component of dust in metal-poor carbon stars in the Magellanic Clouds (Sloan et al. 2007). Acke et al. (2010) proposed that grains with a high aliphatic/aromatic CH ratio surround cool Herbig Ae stars, and showed the significant effect of the stellar UV flux on this ratio. A large diversity of materials have been proposed as carriers to these bands, such as quenched- carbonaceous composites, hydrogenated amorphous carbons (HAC), and soots (e.g., Sakata & Wada 1989; Jones et al. 1990; Duley 1994; Mennella et al. 2002; Dartois et al. 2004; Jäger et al. 2009a; Carpentier et al. 2012; Gavilan et al. 2017).

In situ measurements of extraterrestrial materials, e.g., the macromolecular component of organic matter in meteorites (Hayes 1967), have shown that cosmic PAHs can be embedded in solid organic matter. Small PAHs (24 C atoms) have been identified in the insoluble organic matter (IOM) of the Murchison meteorite (carbonaceous chondrite of type CM2) linked by short aliphatic chains (Cronin et al. 1993; Derenne & Robert 2010). PAHs of high molecular weight (>24 C atoms) were later identified in Allende (a carbonaceous chondrite of type CV3) (de Vries et al. 1993), Murchison (Sabbah et al. 2017) and in two stratospherically collected interplanetary dust particles (IDPs Aurelian and Florianus) (Clemett et al. 1993). Cody et al. (2011) proposed that chondritic IOM and cometary refractory organic solids are related chemically and likely were derived from the same precursor. Because the IOM in these materials is enriched in deuterium, an

interstellar origin was proposed.

Several laboratory studies have been performed over recent decades to simulate cosmic grains produced via gas precursors in the interstellar and circumstellar medium (see Jäger et al. 2009c for a review). Such laboratory analogs to cosmic dust span a wide range of sp^2/sp^3 hybridizations and thereby a large range of aliphatic to aromatic ratios. Among the techniques used to prepare grains are laser pyrolysis (Schnaiter et al. 1999; Jäger et al. 2006), laser ablation (Mennella et al. 1999; Jäger et al. 2008), photolysis (Bernstein et al. 1995; Dartois et al. 2004), plasma chemistry (Scott & Duley 1996; Salama et al. 2018), and combustion (Carpentier et al. 2012; Gavilan et al. 2017). Soots are well known for their high aromatic/aliphatic ratio and rich aromatic molecular composition. Gavilan et al. (2017) showed that soots can include polyaromatic units up to 90 rings (≈200 C atoms) while HAC materials can include polyaromatic units up to 10 rings. The formation route to carbonaceous particles is strongly dependent on the temperature in the nucleation zone. Jäger et al. (2009a) distinguished a low-temperature regime (LT < 1700 K) via laser pyrolysis of small hydrocarbon precursors (C_2H_4 , C_2H_2 , C_6H_6) from a high-temperature regime (HT > 3500 K) using laser ablation from graphite. While the LT products are small PAHs (3–5 rings), HT products result in fullerene-like carbonaceous grains (>100 rings). In this study we seek to identify reaction pathways at lower temperatures (<200 K).

Kaiser et al. (2015) reviews theoretical and experimental studies that have addressed the formation of benzene and naphthalene, the smallest aromatic unit and PAH, respectively. Large carbonaceous species, such as fullerenes ($600\text{--}800\text{ g mol}^{-1}$) are more commonly studied in the laboratory via combustion methods (Jäger et al. 2009b; Dunk et al. 2013). The formation of medium-sized PAHs ($150\text{--}400\text{ g mol}^{-1}$) has been less explored and very few studies on the formation of grains at low temperatures have been reported (Contreras & Salama 2013; Fulvio et al. 2017; Salama et al.

2018). To the best of our knowledge, only one study has been reported so far regarding the growth of carbon grain precursors in the gas phase at the low temperatures ($150\text{ K} < T < 200\text{ K}$) relevant to the outer regions of circumstellar outflows (Contreras & Salama 2013). Using the same laboratory facility, our current study now aims to reveal the chemical routes leading to the formation of grains from small PAH precursors ($\sim 20\text{ C atoms}$). In this paper, we will refer to the theoretical mass of a compound in units of g mol^{-1} , and to peaks from mass spectra in units of m/z .

6.2. Experiments

6.2.1. Sample Preparation

For the laboratory simulation of cosmic carbon grain formation we used the COSmIC facility at NASA Ames. This multicomponent system was built with the goal of understanding the properties of carbon-bearing molecules in circumstellar, interstellar, and planetary environments (Salama et al. 2018). Some key properties that are simulated in COSmIC are the low temperatures and densities of such environments, which directly influence the chemical reactivity of species. The COSmIC apparatus includes several components: a pulsed discharge nozzle (PDN), an orthogonal reflectron time-of-flight mass spectrometer (oReTOF-MS, Ricketts et al. 2011), and a cavity ring-down spectroscopy (CRDS) setup (Biennier et al. 2003). The CRDS and oReTOF-MS are used for diagnostics of in situ gas-phase reactivity and were not used in the current study.

The PDN is used as a plasma reactor for the generation of grains. It produces a supersonic jet expansion to cool down a gas and then generates a plasma discharge in the stream of the jet-cooled planar gas expansion. The jet expansion is produced by the injection of a gas at a high flow rate (1500 sccm) at a back pressure of 600 Torr into the PDN copper reservoir, where it accumulates

before being released in pulses into a vacuum chamber through a very thin $127\ \mu\text{m} \times 10\ \text{cm}$ slit. Solenoid valves attached to a poppet that seals the slit inside the copper reservoir drive the escape of the gas through the slit. The solenoids operate at a frequency of 10 Hz and the pulse length is 1280 μs . The resulting pulsed planar jet expansion increases the pressure in the chamber from 8 mTorr (base pressure) to around 60 mTorr. The pressure differential between the back pressure (600 Torr) and the pressure in the vacuum chamber as the gas expands through the slit (60 mTorr) is four orders of magnitude, resulting in a supersonic expansion that adiabatically cools the gas. Two knife-edge electrodes, made of elkonite (10/90 Cu–W alloy) and mostly heat- and ablation-resistant, are mounted on the outside of the copper reservoir.

They are insulated from the grounded copper reservoir of the PDN by an alumina (Al_2O_3) dielectric plate with a slightly wider slit, $400\ \mu\text{m} \times 10\ \text{cm}$. The elkonite electrodes are mounted on each side of the alumina plate's wider slit. The PDN body acts as the anode and the elkonite electrodes as the cathode, where $-700\ \text{V}$ is applied in 600 μs pulses, synchronized and time-centered within the 1280 μs gas pulse.

The plasma discharge hence generated in the stream of the supersonic expansion of the inert carrier gas (usually seeded with a reactive gas) results in the formation of neutrals, ions, free radicals, and electrons (Broks et al. 2005). These are frozen by the expansion (Biennier et al. 2006), providing very efficient rotational cooling over short distances (a few millimeters) from the slit and an ideal simulation tool for astrophysical regions below 300 K. Neutral molecules present in the plasma discharge and in the plasma expansion have been measured to be at astrophysically relevant temperatures. Tan & Salama (2005a) performed CRDS measurements in the visible and compared these to time-dependent density functional calculations, deriving a rotational temperature of $52 \pm 5\ \text{K}$ for perylene in the supersonic expansion, measured at a distance of 4 mm from the slit (Tan

& Salama 2005a), and a rotational temperature of 70 K for benzo[ghi]perylene (Tan & Salama 2005b). Sciamma-O'Brien et al. (2017) measured a neutral temperature in the plasma cavity (within the alumina slit, between anode and cathode) ranging between 150 and 200 K by fitting optical emission spectroscopy of the 2–0 rovibrational band of the first positive system of nitrogen to modeled numerical spectra (Biloiu et al. 2007).

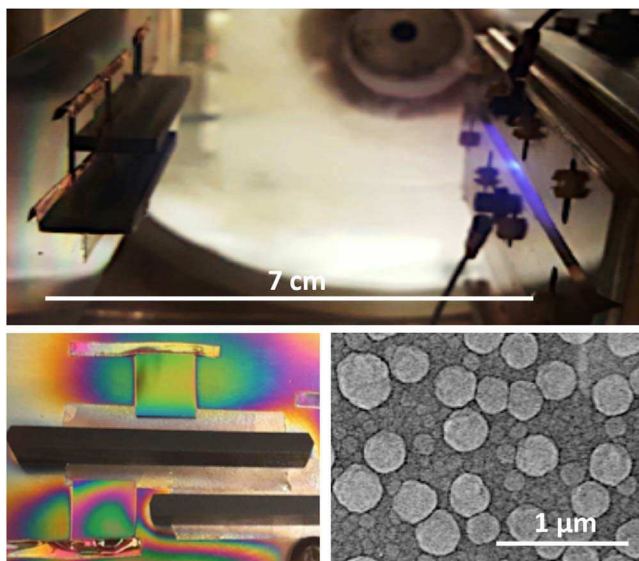


Figure 1. Top: side view of the graphite bars placed on a sample holder facing the supersonic expansion from the PDN within the COSmIC chamber. Bottom left: front view of the grain deposit on the graphite bars following an 8 hr experiment. Bottom right: SEM image of grains deposited during the benzene experiment.

In the study presented here, all experiments used argon (purity 99.999%, Matheson) as the carrier gas to seed molecular precursors in the supersonic expansion. Argon was chosen over helium because previous calculations showed that helium has a 10–20 K higher temperature than argon (Biennier et al. 2006). In the same reference, a study of naphthalene, phenanthrene, and pyrene seeded in argon gas showed that these three PAHs do not affect the temperature in the expansion because of their small concentrations relative to the carrier gas. Five different aromatic precursors

purchased from Sigma-Aldrich were used: benzene (78.11 g mol^{-1} , purity 99.7%), naphthalene (linear PAH, $128.17 \text{ g mol}^{-1}$, purity 99.7%), anthracene and phenanthrene (linear and branched PAH isomers at $178.23 \text{ g mol}^{-1}$, purity 99.7%), and pyrene (compact PAH, $202.26 \text{ g mol}^{-1}$, purity 98%). Benzene was seeded in the carrier gas by flowing argon through an external cell containing liquid benzene. All other precursors were seeded by placing solid PAH crystals at the bottom of the PDN reservoir and heating the top and bottom of the reservoir (from 50 to 150 °C) to ensure an adequate vapor pressure. Heating of the reservoir has been shown not to alter the expansion temperature of the gas (Tan & Salama 2005a).

To collect the plasma-generated molecular products and solid grains, a sample holder was placed about 7 cm away from the electrodes, as seen in Figure 1. Two graphite bars ($3 \text{ mm} \times 50 \text{ mm} \times 10 \text{ mm}$) were placed on the sample holder parallel to the slit, with the thinner surface facing the plasma expansion. One bar was aligned with the slit, and the other was placed 1 cm below. These bars are specially machined to be used as substrates for laser desorption mass spectrometry. Each experiment was conducted twice in different sequential order to validate reproducibility. Although a single discharge pulse is 600 μs long, the deposition of species runs over 8 hr for each experiment, ensuring that enough solid materials are collected for ex situ diagnostics. We examined the sample produced in the benzene experiment with a scanning electron microscope (SEM, Hitachi S-4800). This SEM image (Figure 1, bottom right) shows that the resulting grains are mostly spherical and with sizes ranging from 5 to 300 nm. The smaller grains aggregate into larger grains, and these all stack and pile up on top of each other, making a macroscopic thin film. The effects of different aromatic molecules on the grain morphology and size distribution will be part of a future study and are beyond the scope of the present work.

6.2.2. Laser Desorption Mass Spectrometry

The graphite bars with solid deposits prepared at the NASA Ames COSmiC facility were analyzed *ex situ* with the laser desorption jet-cooled mass spectrometry (LDJCMS) setup, part of the de Vries lab at the University of California Santa Barbara (UCSB). This setup is described in more detail in Meijer et al. (1990). Two-step laser desorption laser ionization mass spectrometry has been used extensively to probe the molecular composition of solid materials, including extraterrestrial materials such as chondritic meteorites (Clemett et al. 1993; Callahan et al. 2008; Sabbah et al. 2010, 2017) and interplanetary dust particles from the Stardust Mission (Sandford et al. 2006; Spencer & Zare 2007). This technique is favored for its high sensitivity (10^{-18} mol; Deckert & George 1987), low fragmentation, and because, as a solvent-free probe, it minimizes sample contamination and processing. In our study, the graphite bars were placed in front of a 300 μm Amsterdam piezo valve (Irimia et al. 2009) and translated at ~ 0.7 mm minute⁻¹ to steadily provide fresh desorption material. The solids were thermally desorbed from the graphite bar using the attenuated beam of a continuum Minilite I Nd: YAG laser (1064 nm, ~ 3 mJ/pulse, ~ 1 mm², 6 ns pulse width). The desorbed material was entrained in the argon molecular beam (8 atm backing pressure, 35 μs FWHM pulse width) and skimmed before being ionized by an ArF excimer laser (193 nm, ~ 0.5 mJ/pulse, 8 ns pulse width, 2 ns jitter). The UV laser preferentially ionizes the desorbed species at 6.42 eV via (1 + 1) two-photon ionization. Some highly excited states make this a semi-resonant scheme as the laser energy approaches the molecular ionization energy. After photoionization, the resulting ion current was measured from the microchannel plates in the reflectron time-of-flight mass spectrometer. A data acquisition card (Signatec, 400 MHz, 14 bit) was triggered using a photodiode reading the excimer pulse, and the waveform collected for analysis. For each scan, 40 data points were collected, each consisting of an average of 20 shots.

The first 20 data points were collected with the desorption laser shutter closed and the next 20 with the shutter open.

The mass spectrum was calibrated before and after each series of measurements using argon (40.0 g mol^{-1}), and a graphite bar prepared with coronene (300.0 g mol^{-1}) and iron (II) phthalocyanine (568.1 g mol^{-1}). One of the main advantages of this technique is the low fragmentation of the parent species due to the fact that two lasers separated in space and time are used for desorption and ionization, respectively. The focused infrared photons used for desorption induce a heating ramp of 10^{11} K s^{-1} on the graphite substrate (Deckert & George 1987), leading to the efficient desorption of solids into the gas phase. The photoionization laser combined with jet cooling produces no (or very minimal) fragmentation in aromatic and aliphatic molecules (de Vries et al. 1992; de Vries & Hunziker 1996; Nir et al. 1999). Figure 2 shows an example mass spectrum of a pure coronene sample placed on a graphite bar, showing the parent isotope at 300 m/z and the once- and twice-substituted ^{13}C coronene molecules at 301 and 302 m/z , whose relative intensities are consistent with the theoretically expected isotopic abundances.⁵

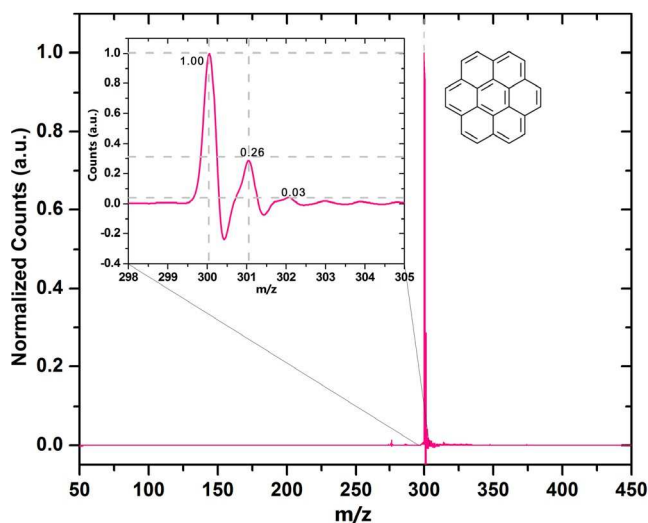


Figure 2. The calibration LDJC mass spectrum of coronene shows minimal fragmentation of the parent molecule $[M]^+$ at 300 m/z . The $[M + 1]^+$ and $[M + 2]^+$ peak abundances correspond to the expected isotopic abundances. The peak at 276 m/z is attributed to benzo[ghi]pyrene impurities.

6.3. Results

In the study presented here, five types of aromatic molecules — benzene, naphthalene, anthracene, phenanthrene, and pyrene — were used as gas-phase precursors in COSmIC to investigate the low-temperature formation of carbonaceous grains in astrophysically relevant environments. Each of these PAH precursors was seeded separately in an argon supersonic expansion and then exposed to a plasma discharge voltage (-700 V). The pool of organic fragments, parent molecules, and plasma-induced chemical products generated in the plasma was deposited for 8 hr on the two graphite bars placed 7 cm away from the slit. Each sample was then analyzed *ex situ* with LDJCMS using a photoionization laser at 193 nm. Although this technique is highly sensitive to aromatic molecules, due to the spread of ionization cross sections of low to medium weight PAHs to 193 nm photons, our LDJCMS analysis cannot provide absolute abundances of the molecular species produced in the COSmIC plasma experiments and captured in the grains. However, it is reasonable to consider that masses with the highest signal-to-noise ratio (S/N) correspond to the most abundant species within the samples. The low intensity of the benzene and naphthalene peaks in their respective experiments is due to their high volatility in the laser desorption analysis because their vapor pressure is higher than the base pressure in the vacuum chamber. The LDJCMS technique is optimized for low-volatility species, which for PAHs corresponds to species with three rings or more.

Table 1 Statistical Distribution of Peaks with $S/N > 10$ in the Five Different Experiments Performed with Aromatic Precursors

Mass range (<i>m/z</i>)	Benz	Np	Ant	Phe	Py
50–99	4	4	2	1	2
100–149	2	4	0	2	1
150–199	1	9	14	15	3
200–249	6	18	36	19	26
250–299	5	8	23	8	11
300–349	1	5	22	1	4
350–399	1	1	28	3	0
400–450	3	1	20	1	15
Total	23	50	145	50	62
Minimum mass	57	57	63	81	63
Maximum mass	436	405	445	411	445

Note. Benzene (Benz), Naphthalene (Np), Anthracene (Ant), Phenanthrene (Phe), and Pyrene (Py). The larger yield in the anthracene experiment partially accounts for the larger number of peaks.

As described above, for each experiment, grains were deposited onto two graphite bars placed at two different positions with respect to the plasma expansion: one aligned with the slit, one placed 1 cm below the slit. The two bars were analyzed separately with LDJCMS, and each bar was measured at different positions (at each end of the bar and at the center) to evaluate differences in their observed mass spectra. For each set of bars prepared in a single experiment, no major differences in composition were detected in their mass spectra, neither from one bar to the other nor at different positions within a bar. However, an overall higher S/N ratio was detected at the center of the bars, due to a higher grain deposition rate at this location. The acquired desorption mass spectra were averages of 20 scans to increase the S/N and average out spurious peaks. A nonlinear baseline was removed from the average mass spectra using OriginLab, to correct for artifacts arising from high ion counts on the microchannel plate (e.g., the Ar⁺ peak at *m/z* = 40). The spectral analysis was conducted using the *mMass* mass spectrometry software (Strohalm et al.

2010).

Only the peaks with $S/N > 10$ in the baseline-subtracted spectra were considered in our analysis. Using this selection criterion, we find that most peaks are detected in the 200–300 m/z mass range for all experiments, which indicates chemical growth in the plasma discharge from the PAH precursors used in this study (78–202 g mol^{-1}) to larger molecular species. The global statistics are given in Table 1. Figure 3 shows the mass spectra for all five experiments, normalized to the maximum peak for each spectrum. Table 1 shows a larger density of peaks between 50 and 150 m/z for the benzene and naphthalene experiments. We also note that the anthracene experiment resulted in the highest production yield (as measured by the average S/N over the whole mass range), followed by the pyrene experiment. The benzene experiment had the lowest production yield.

The mass spectra from $m/z = 1$ to 50 (Figure 4) show very similar peaks across the different experiments (including desorption from a blank graphite bar): Ar^+ (40 g mol^{-1}), Ar^{2+} (20 m/z), Ar^{3+} (13 m/z), Ar^{4+} (10 m/z), H^+ (1 m/z), N^+ (28 m/z), O^+ (32 m/z), O^+ (18 m/z), and the stable Ar isotope $^{36}\text{Ar}^+$ (36 m/z). While the argon peaks are due to the carrier gas in the setup, nitrogen and oxygen are due to residual gas. A peak at 27 m/z is absent from the anthracene experiment. This peak is tentatively attributed to C_2H^+ , the vinyl radical.

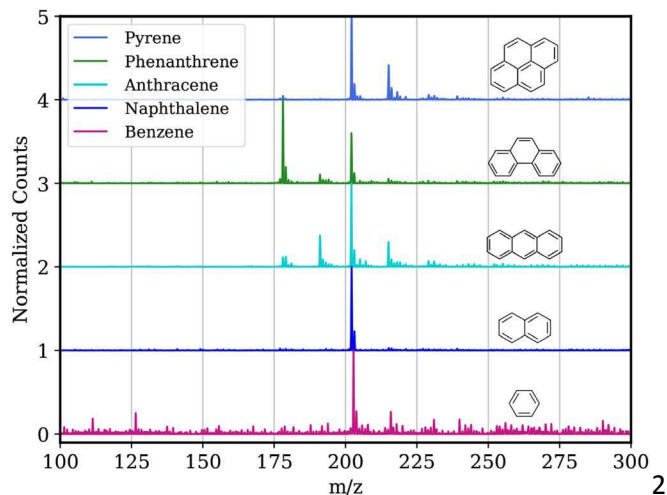


Figure 3. Laser desorption mass spectra normalized to the maximum peak (202 m/z for naphthalene, anthracene, and pyrene, 203 m/z for benzene, and 178 m/z for anthracene) in the 100–300 m/z range. The mass spectra show that pyrene (202.2 g mol⁻¹), methylpyrene (216.2 g mol⁻¹), and/or their isomers are consistent products of grain formation. The molecular products generated in the anthracene and pyrene experiments share a similar molecular diversity at high mass. Benzene and naphthalene show the greatest abundance of peaks for masses less than 150 m/z .

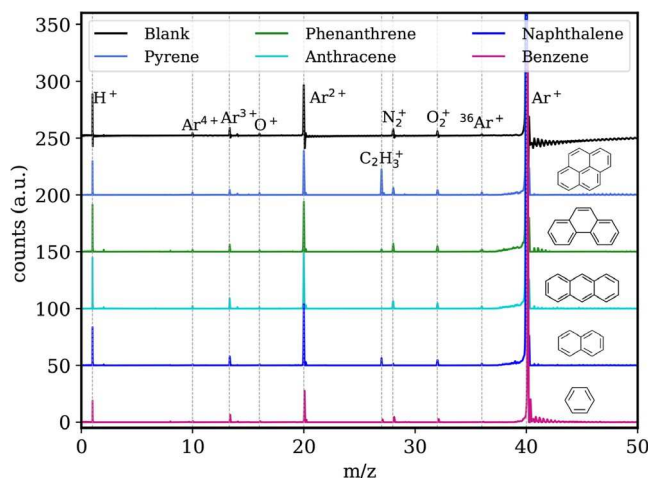


Figure 4. Mass spectra in the 0–50 m/z range, compared to the blank graphite bar. The argon cation peaks present are due to the argon jet cooling species after desorption in the LDJCMS setup while the N^+ , O^+ , and H^+ peaks are due to residual pressure in the chamber. The spectra have been shifted vertically for clarity.

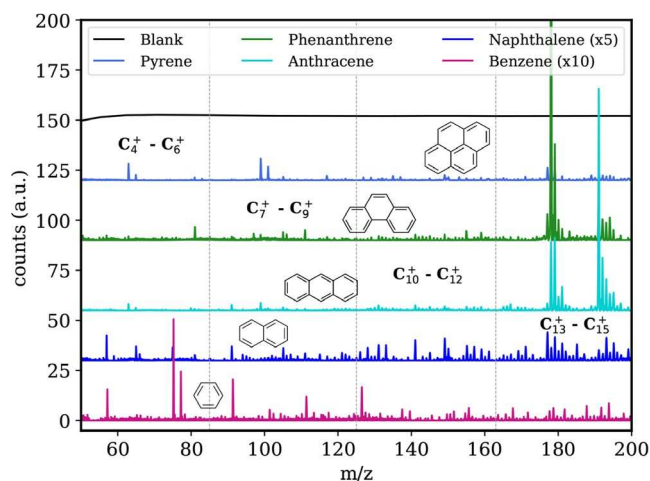


Figure 5. Mass spectra in the 50–200 m/z range, showing the varying molecular complexity for grains formed with aromatic precursors of increasing ring number. The benzene and naphthalene experiments produced more molecules of lower masses (<150 m/z), while all other experiments produced more molecules of higher molecular mass (>175 m/z). The spectra have been shifted vertically for clarity.

For the different precursors, the mass spectra from $m/z = 50$ to 200 (Figure 5) exhibit common trends. In all cases, peaks are more abundant in the 200–250 m/z region. The number of peaks at the lower masses, between 50 and 150 m/z , is higher for the benzene and naphthalene experiments. This can be explained by the more favorable fragmentation of smaller aromatic molecules in the plasma, an effect of their lower average carbon binding energy, E_C (Bauschlicher & Ricca 2009). The experimental configuration in COSmIC (supersonic expansion where the species are accelerated to high velocity in short pulses) can also partly explain this observation because the residence time of the seeded molecular precursors is less than 4 μs , allowing for only the first steps of the chemistry to occur (Sciamma-O'Brien et al. 2017; Raymond et al. 2018). Small hydrocarbons including C_4H_9 (57 m/z), C_5H_3 (63 m/z), C_5H_5 (65 m/z), C_5H_9 (69 m/z), and C_6H_3 (75 m/z) are observed in the mass spectra for most experiments. These radical fragments were produced either during the plasma discharge or as photochemical products of the LDJCMS technique.

For the anthracene experiment, high-S/N envelopes are observed around 178 m/z and 190 m/z , corresponding to functional additions to the parent molecule. For the pyrene experiment, on the other hand, the envelope observed around 178 m/z is likely a result of plasma-induced fragmentation of the parent molecule. As grains are formed in the COSmIC setup, not all precursor molecules will fully react and some may be directly incorporated in the solid via physisorption or chemisorption processes. As a consequence, the mass spectra of the grain deposits can include the parent molecules without fragmentation. The parent molecule is not detected in the benzene experiments, but the cations at 75 m/z and 77 m/z could be photofragments of the original unreacted molecule stabilized within the grain. In all other experiments, the parent molecule is also found in the mass spectra. All peaks with $S/N > 10$ in the mass range up to 199 m/z are

tabulated and tentatively identified in Table 2 in the Appendix. Because of the mass resolution ($m/\Delta m = 500$) of the LDJCMS experimental setup at UCSB, we cannot discriminate between isomers. In all cases, we assume that peaks are due to hydrocarbon C_xH_y products from the parent molecule. We propose possible identifications from the literature (experimental and theoretical studies on PAH-related chemical processes including combustion). In cases where no identification is possible we propose a stable aromatic molecule from a standard database such as NIST⁶ or commercially available PAHs from vendors such as Sigma-Aldrich. In cases where peaks at odd masses are observed, we attribute these to dehydrogenation (H loss) due to the UV photons used for photoionization (in which case the $[M-H]^+$ cation is reported), mostly in the low mass range ($<150 m/z$). Dehydrogenation

⁶ <https://pah.nist.gov/>

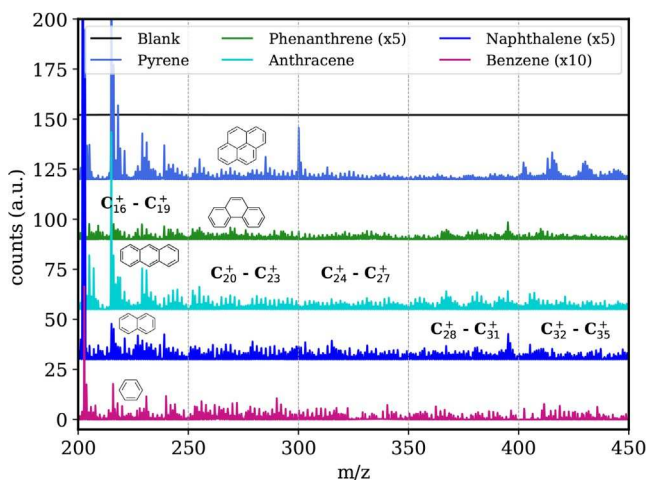


Figure 6. Mass spectra in the 200–450 m/z range. Molecular peaks beyond 200 m/z include pyrene and methylpyrene for all experiments, and are tentatively attributed to hydrocarbons with up to 35 carbon atoms. Coronene and/or its isomers are formed efficiently in the pyrene experiment. The spectra have been shifted vertically for clarity.

(1H, 2H, or 3H loss), or radical formation, is a potential result of the 193 nm photoionization laser. High dehydrogenation is expected for species of small molecular weight

(Sabbah et al. 2017), which results in the peak intensity of $[M-H]^+$ being H greater than that of M^+ . Jochims et al. (1994) showed that for many PAH structures (from benzene to coronene) 1H loss is the dominant fragmentation channel for all PAHs with the lowest appearance energy (the minimum energy that must be supplied to a molecule to produce a detectable ion). For the larger masses ($>150 m/z$) we expect isotopic effects to contribute to the odd-mass peaks (^{13}C substitution of PAHs), due to the increasing cumulative effect of ^{13}C isotopic substitution as the number of carbon atoms in a PAH molecule grows.

For the different precursors, the mass spectra from 200 to 450 m/z (Figure 6) exhibit a major similarity, the presence of very high-S/N envelopes at 202 m/z , 216 m/z , and 230 m/z . The envelope starting at 202 m/z is attributed to pyrene and/or pyrene isomers, while the next envelope is attributed to the methyl-addition series of pyrene. The anthracene experiment shows high-S/N envelopes at 242, 258, 360, 370, 380, and 402 m/z , while the pyrene experiment shows significant envelopes at 242, 258, 300, 402, 416, and 432 m/z . This is consistent with the expected growth of both anthracene and pyrene parent molecules, which are expected to produce larger aromatic fragments than smaller parent molecules such as naphthalene in the COSmIC experimental setup. Peaks with $S/N > 10$ in the 200–450 m/z range are tabulated and tentatively identified in Tables 3–6. Beyond 300 m/z , identifications are less certain because many possible isomers could carry species of high molecular weight.

In Figure 7 the mass spectrum of solid deposits produced in the phenanthrene experiment is compared to that produced in the anthracene experiment. The goal of the phenanthrene experiment was to test whether the structure of two PAH isomers has an impact on the yield and molecular composition of solid grains in the COSmIC experiments. The major difference in the mass spectra of the solid deposits resulting from these two experiments is the fact that the largest

peak for the phenanthrene experiment is at mass 178 m/z (i.e., the mass of the seeded molecule) as opposed to 202 m/z (pyrene and/or its isomer) for the anthracene experiment. This indicates a much lower reactivity of the parent molecule for the phenanthrene experiment, because the anthracene experiment produces much more pyrene and/or its isomer. This finding is further confirmed by the much larger solid yield obtained in the anthracene experiment. We note, however, that the position of the resulting peaks from 180 to 240 m/z is strikingly similar for both isomers, signaling similar molecular growth routes.

6.4. Analysis

6.4.1. Aromatic/Aliphatic Structure

In the following analysis, we only consider the tabulated peaks with $S/N > 10$ and even masses, and assume these molecules are hydrocarbons of unknown structure (C_xH_y). We can then calculate their double-bond equivalent (DBE) or level of unsaturation. DBE is widely used in the characterization of organic matter, and its calculation is a well-established tool in mass spectrometry (Marshall & Rodgers 2008; Sabbah et al. 2017; Schuler et al. 2017). DBE can be defined as the sum of the number of rings plus double bonds involving carbon; each ring or double bond results in the loss of two H atoms. In the case of simple hydrocarbons, this is defined as

$$DBE = C - H/2 + I, \quad (1)$$

where C is the number of carbons and H is the number of hydrogens. A benzene ring will have $DBE = 4$ while polycyclic aromatic molecules will have a nonlinearly increasing DBE number as the sum of rings and π -bonds increases. We remark that double bonds can also be

present in hydrocarbons such as cumulenes (three or more cumulative double bonds). In this analysis, we assume that most compounds are made of aromatic rings because these are generally more stable.

We first compare the DBE for all experiments presented in this study, shown in Figure 8. We notice that all precursors generally follow the same DBE trend: linear growth toward greater DBE numbers as a function of carbon number. These generally follow the DBE/C numbers of reference PAHs including benzene (C_6H_6 ; DBE = 4), naphthalene ($C_{10}H_8$; DBE = 7), pyrene ($C_{16}H_{10}$; DBE = 12), coronene ($C_{24}H_{12}$; DBE = 19), and ovalene ($C_{32}H_{14}$; DBE = 26) and indicate the formation of fused aromatic rings. Data points are sparse for molecules with carbon numbers below 10 and between 22 and 27 atoms. We also note that the pyrene experiment results in more data points for the largest carbon numbers. Both the pyrene and anthracene experiments result in similar DBE values, which cluster around DBE = 12 and 22. DBE points with differing numbers of carbon atoms are likely members of an alkylation series, and are attributed to methyl-substituted species. In the anthracene experiment, data points cluster at DBE = 8, 10, 12, 14 but also at 20, 22, and 24. All experiments (except phenanthrene) converge at DBE = 12, corresponding to pyrene and its isomers, which is the peak with the highest S/N seen in all mass spectra. Figure 8 also includes the DBE obtained for the phenanthrene experiment. Despite the much larger solid grain yield and number of peaks observed in the anthracene experiment compared to the phenanthrene experiment, the masses detected and the DBE values show a similar tendency of molecular growth.

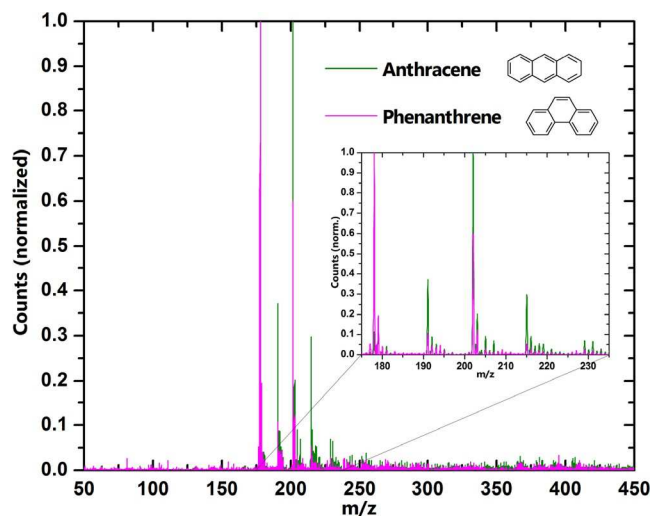


Figure 7. Mass spectra in the 50–450 m/z range for the anthracene and phenanthrene experiments. The inset shows that molecular products with identical masses are generated in both experiments, signaling a very similar growth mechanism for the two isomers, irrespective of their dissociation (plasma) or photoionization (LDJCMS) cross sections.

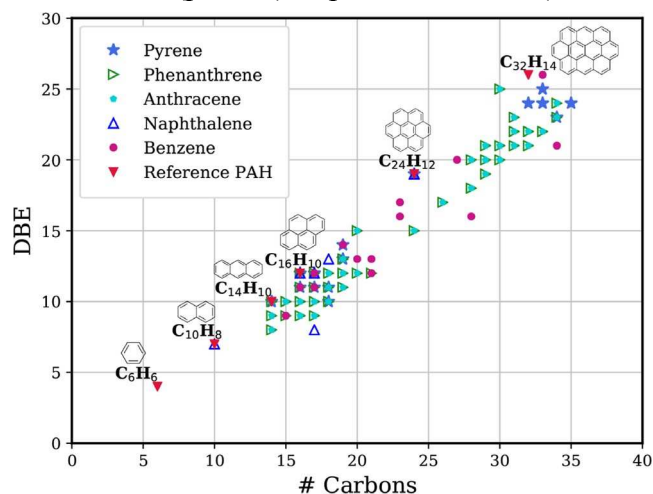


Figure 8. Double-bond equivalent (DBE; defined in Section 4) vs. number of carbon atoms for all experiments. Molecules with a similar DBE present the same number of rings, but differ in alkyl group. DBE vs. C follows a linear trend comparable to reference PAHs, indicating the growth of the precursor molecules into more complex condensed polycyclic structures. The DBE comparison for the anthracene vs. phenanthrene experiments shows a significant overlap in their molecular diversity.

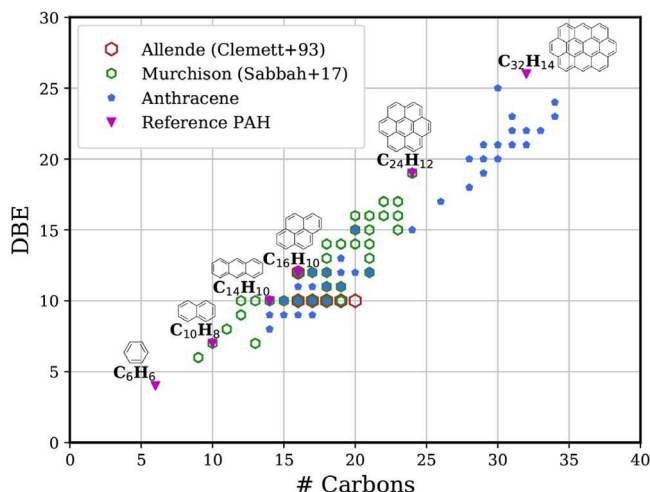


Figure 9. DBE comparison for the anthracene vs. data from the Murchison meteorite (Sabbah et al. 2017) and the Allende meteorite (Clemett et al. 1993). Murchison overlaps the DBE growth pattern of anthracene for hydrocarbons with 15 to 20 carbon atoms.

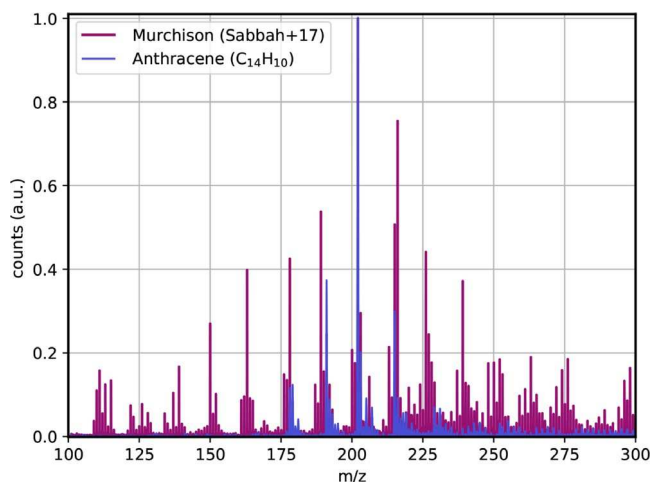


Figure 10. Comparison of Murchison mass spectrum (Sabbah et al. 2017) to the mass spectrum of the anthracene experiment. The production of the major peak, pyrene, suggests a possible low-temperature PAH growth process in the formation history of organics in Murchison.

We also compare the DBE values for the anthracene experiment (as the PAH experiment with largest S/N) to DBE values found in previous studies with two-step laser desorption laser ionization mass spectrometry (L2MS) on meteorites, i.e., Allende (Clemett et al. 1993) and Murchison (Sabbah et al. 2017), both carbonaceous chondrites. This comparison is shown in

Figure 9. The data points from Murchison follow a higher DBE/C tendency than the anthracene experiment, which can be indicative of a larger number of condensed aromatic structures. Allende data points cluster at DBE = 10 and 12, corresponding to $C_{14}H_{10}$ (e.g., anthracene, phenanthrene, etc.) and $C_{16}H_{10}$ (e.g., pyrene) and their alkylated series members. In Figure 10 the mass spectrum of the Murchison meteorite obtained via L2DMS of a powder sample (Sabbah et al. 2017) is compared to the mass spectrum of the solid samples in the anthracene experiment. In both cases, pyrene is the most abundant peak, and envelopes around mass 178 (anthracene), 190, 216 m/z are found. On the other hand, envelopes at 112, 126, 136, 150, and 260 m/z seen in Murchison are not as prominent in the laboratory sample. This can be partly explained by the design of our experimental setup, where the short residence time of species exposed to the plasma enables the study of chemical pathways associated with the first steps of growth of a specific molecular precursor. As a consequence, the lower and higher mass peaks observed in the Murchison sample are less likely to be generated in our anthracene experiments. The routes of complex molecular growth in meteorites seem consistent with the growth observed in our laboratory samples via stable PAHs such as anthracene and pyrene (or their isomers). Meteoritic organics may undergo additional processing (high temperature, non-thermal, etc.) that could account for the larger abundance of lower and higher mass spectra.

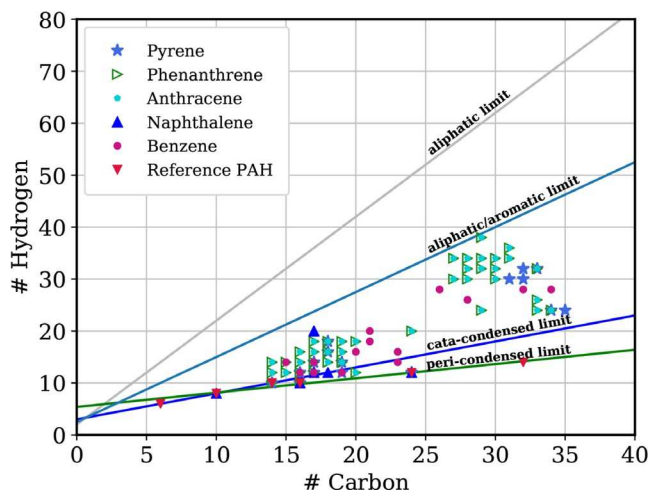


Figure 11. Number of H vs. C atoms in compounds detected at even masses with $S/N > 10$. Empirical limits are also shown for aliphatic hydrocarbons (gray line), aliphatic/aromatic hydrocarbons (light blue line), cata-condensed PAHs (blue line), and peri-condensed PAHs (green line) (Cain et al. 2014).

We can use an additional method to evaluate the possible structure of hydrocarbons based on the C and H number of molecular compounds probed in soots (Cain et al. 2014). In their study, Cain et al. assigned empirical limits between different hydrocarbon classes going from peri-condensed to purely aliphatic. A plot of H versus C (number of hydrogen versus number of carbon atoms) for all even peaks seen in our mass spectra is shown in Figure 11. As described in Cain et al. (2014), empirical limits for different hydrocarbon classes are shown: aliphatic ($H = 2C + 2$), aliphatic/aromatic ($H = 1.25C + 2.5$), cata-condensed (non-compact, no more than two rings have a single C atom in common) PAHs ($H = 0.5C + 3$), and peri-condensed (compact, three or more rings share common C atoms) PAHs ($H = 6C^{0.5}$). Most species in our study are found within the cata-condensed and aliphatic/ aromatic limits, indicating the presence of alkylated aromatics as shown by the DBE analysis.

6.5. Discussion

6.5.1. Evidence for PAH Growth within Grains

The laser desorption mass spectra of the solid grains for all precursors reveal a rich molecular complexity, where the precursor has been either fully or partially fragmented in favor of growth into larger molecules. Grains effectively trap COMs formed in the PDN because these are aggregated as a solid. The absence of major peaks beyond 400 m/z reveals that our grains are not a refractory type of organic matter made of macromolecules but rather an organic matrix likely composed of medium-sized aromatic molecules (150–400 g mol^{-1}) held together via weaker (aliphatic) bonds. One of the most interesting results from this series of experiments is the large relative abundance of a peak at 202 m/z , likely corresponding to pyrene and/or its isomers. Pyrene isomers include stable PAHs resulting from the progressive addition of functional groups to smaller hydrocarbons.

Several routes for the growth of PAHs have been proposed by the combustion and astrochemical communities (Reizer et al. 2019). One of the more cited pathways is the hydrogen abstraction–acetylene addition (HACA) mechanism (Frenklach & Wang 1991). As its name indicates, the formation of a radical by hydrogen abstraction is the first step in the sequential addition of acetylene, usually occurring in combustion where gas temperatures exceed 1000 K. Parker et al. (2014) experimentally showed that HACA is key to the growth of small PAHs such as naphthalene in combustion-like conditions. However, limitations to HACA have been singled out, such as the underpredicted soot yields in flames. Calculations by Kislov et al. (2013) also showed that HACA led mostly to growth of five-membered rings. Finally, HACA has a limited application to a narrow high-temperature window, which is not suitable for interstellar environments where temperatures are lower than 1000 K.

Alternative models of PAH growth via the addition of larger radicals have also been

proposed. In the phenyl-addition- cyclization (PAC) mechanism, the addition of a phenyl radical followed by hydrogen abstraction and H₂ loss leads the growth (Shukla et al. 2008; Shukla & Koshi 2010). Later on, these authors also proposed the methyl-addition-cyclization (MAC) mechanism, which involves the addition of two or three methyl radicals on PAHs. The methyl radical is one major product in aromatic fuels, and plays a role in the formation of propargyl (H–C[∘]C–CH₃) radicals, the basic building block of benzene. Experimental evidence for this was shown from a toluene pyrolysis experiment, where the mass spectra showed a mass difference of 14 *m/z* (Shukla et al. 2010). Calculations by Reizer et al. (2019) show that MAC is an efficient route for the formation of benzo(a)pyrene (252 g mol⁻¹) from chrysene and benzo(a)anthracene (228 g mol⁻¹). More recently, a new mechanism for low-temperature PAH growth, known as hydrogen abstraction–vinylacetylene addition (HAVA) has also been explored for the formation of naphthalene (Parker et al. 2012; Zhao et al. 2018b). One of the major advantages of HAVA is that it is barrierless at low temperatures (Kaiser et al. 2015).

To investigate these possible PAH growth routes we use the mass spectra with highest S/N and largest number of peaks produced in the anthracene experiment. Additionally, we only consider even-mass peaks as isotopic peaks; dehydrogenated/ radical fragments appearing at odd masses are a result of the L2DMS technique rather than a signature of the inherent molecular composition of the grains. Because of this, we cannot probe growth mechanisms where even masses would result in odd masses, i.e., we cannot discard if HACA, HAVA, or PAC growth occurred and led to the assembly of complex molecules within the grains. HACA would show up as a mass difference of (–H + C₂H₂) 25 *m/z*, HAVA would show up as a mass difference of (–H + CH₂=CH–C[∘]CH) 51 *m/z*, while PAC would show up as a mass difference of (C₆H₅ – H₂) 75 *m/z*. However, we can probe for MAC, where the mass differential (–H + CH₃) is 14 g mol⁻¹,

indicative of methyl addition on a radical. We find that several peaks in the anthracene mass spectrum fall within this growth pattern, as shown in Figure 12. This mechanism seems to be efficient for species of low mass (178–258 m/z) and high mass (366–420 m/z).

6.5.2. Effect of Structure on PAH Growth

We performed two separate experiments to explore the effect of PAH structure on the molecular growth within grains. These were performed using the C₁₄H₁₀ isomers, anthracene and phenanthrene, both found at mass 178.23 g mol⁻¹. These are prototypes of linear and cata-condensed polycyclic benzenoids respectively. A comparison of their mass spectra is shown in Figure 7 and their DBE versus C content in Figure 8. These show that, while the grain formation yield is much greater for the anthracene experiment (~6 times greater mean number of counts), the molecular diversity within the grains is very similar for both experiments. Mass spectra show that more of the parent molecule remains in phenanthrene than in the anthracene experiment, indicating that dissociation is more efficient for anthracene in the plasma discharge. This could be explained by the fact that the first ionization potential of phenanthrene (IE = 7.903 eV) is 0.4 eV higher than that for anthracene (IE = 7.439 eV) (Boschi et al. 1974; Zakrzewski et al. 1996; Dabestani & Ivanov 1999). In addition, the enhanced stability of cata-condensed PAHs compared to linear PAHs is well established (Clar 1964, 1972; Dabestani & Ivanov 1999; Randić 2003; Poater et al. 2007; Solà 2013) and there is a general consensus that phenanthrene is kinetically more stable than its linear isomer. This could be a result of its higher thermochemical stability due to its larger π -bonding, i.e., higher aromaticity (Kato et al. 2002; Poater et al. 2007; Gutman & Stanković 2007). Indeed, according to Clar's rule, which defines aromaticity by aromatic π -sextets (six π -electrons localized in a single benzene-like ring separated from adjacent rings by formal CC single bonds),

phenanthrene has two sextets and hence a higher aromaticity than anthracene (one sextet) (Clar 1972; Solà 2013). PAHs with a higher number of π -sextets have in general larger HOMO–LUMO gaps (>2.1 eV), which enhance their stability (Dias 2004; Solà 2013).

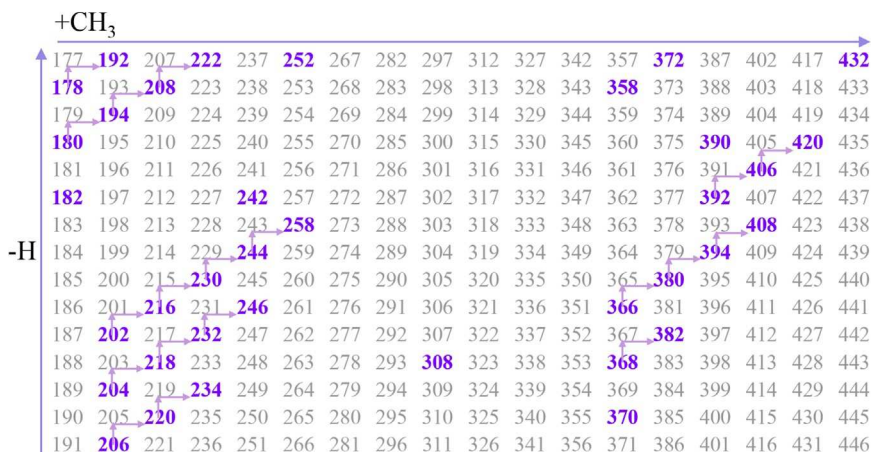


Figure 12. Schematic showing the molecular growth within grains prepared in the anthracene experiment. The staircase pattern corresponds to the abstraction of one hydrogen and the addition of one methyl group, i.e., the methyl-addition-cyclization or MAC mechanism (Shukla et al. 2010).

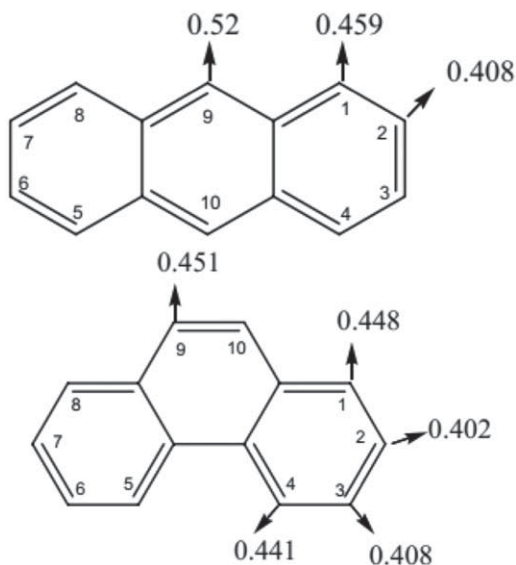


Figure 13. Free valence indices of anthracene and phenanthrene (Abrahamson 2017). These indices impact the carbonization reactivity. Position 9 in anthracene has the highest value of either PAH (Walker 1990).

The higher yield for the formation of grains in the anthracene experiment could be due to the fact that it is more readily dissociated in the plasma than phenanthrene, and therefore the formation of larger PAHs and COMs in general is more efficient because it begins with more reactive fragments. The free valence indices of anthracene and phenanthrene are shown in Figure 13. We note that, because of the symmetry of anthracene, only the free valence indices of certain positions are given. The free valence indices are the same for positions 9 and 10 (0.52), positions 2, 3, and 7 (0.408), and positions 1, 4, 5, and 8 (0.459). The meso position (positions 9 and 10) is the most reactive in anthracene, where carbon atoms display the highest free valence indices (Gore 1956; Walker 1990). Hydrogen loss and free radical formation preferentially occur at these sites of highest reactivity. Additionally, the free valence indices of the central carbon atoms are higher than those in phenanthrene, which could lead to the higher reactivity of the linear PAH. It has been proposed that the meso position of anthracene undergoes hydrogen-abstraction reactions, forming intermediate species such as 9,9'-bianthracene (detected in our experiment, see Table 6), and then stabilizes on the bisanthene structure (eight aromatic rings) (Kinney et al. 1957). These results are a first exploration of the effects of PAH structure on the formation of carbonaceous grains and the pathways to PAH growth within grains.

We compare the results of our study to those of Fulvio et al. (2017), the only reported experimental study on low-temperature carbonaceous dust formation. They studied the formation of carbon grains in solid cages on low-temperature substrates ($T \sim 10$ K). The precursor molecules were directed into a solid rare-gas argon matrix as opposed to the free-flying cold molecular precursors in the gas phase used in the study reported here. To the best of our knowledge, the only study of

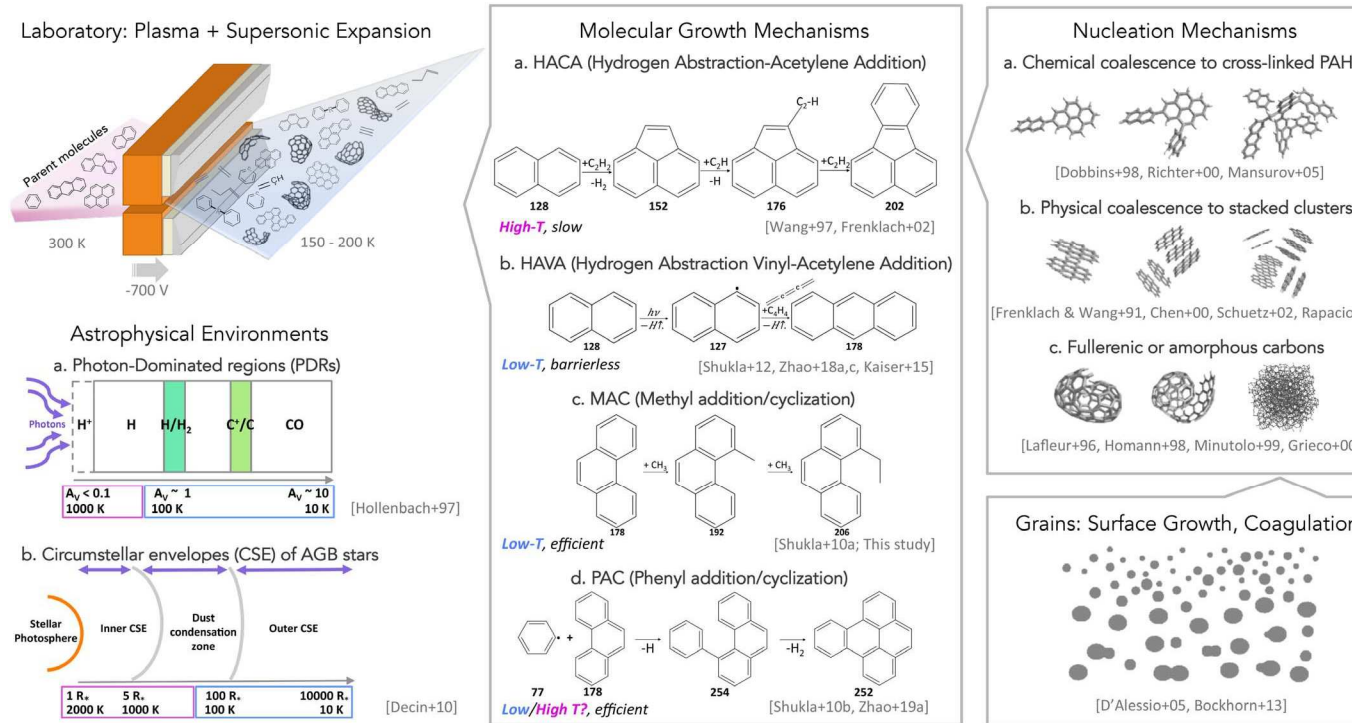


Figure 14. Laboratory simulation of carbonaceous dust formation in astrophysical environments (PDRs and CSEs) in the context of known PAH growth mechanisms. Left: the pulsed discharge nozzle (PDN) of the COSmIC setup where parent molecules are cooled in a supersonic expansion exposed to a plasma discharge, inducing the dissociation, formation, and growth of simple PAHs into more complex molecules and solids. The laboratory low-temperature (150–200 K) conditions are applicable to the colder or ionized/molecular transitions in photon-dominated regions (Hollenbach & Tielens 1997) and circumstellar envelopes (Decin et al. 2010). Center: mechanisms of PAH growth, including HACA (Wang & Frenklach 1997; Frenklach 2002), HAVA (Shukla & Koshi 2012; Zhao et al. 2018a, 2018b; Kaiser et al. 2015), MAC (Shukla et al. 2010), and PAC (Shukla & Koshi 2010; Zhao et al. 2019b).

Right: nucleation mechanisms from the aggregation of PAHs (Dobbins et al. 1998; Richter & Howard 2000; Mansurov 2005), the formation of clusters (Frenklach & Wang 1991; Chen & Dobbins 2000; Schuetz & Frenklach 2002; Rapacioli et al. 2005), the formation of amorphous and fullerenic nanograins (Lafleur et al. 1996; Homann 1998; Minutolo et al. 1999; Grieco et al. 2000), and the growth into larger grains (D'Alessio et al. 2005; Bockhorn 2013).

the growth of carbon grain precursors in the gas phase at the low temperatures ($150 \text{ K} < T$)

< 200 K) relevant to the outer regions of circumstellar outflows was reported in Contreras & Salama (2013). Although the precursors are cooled before reacting in both cases, the recombination processes and the resulting energy budgets involved can be quite different because of the strong perturbation induced by the solid Ar matrix (cage effect). In Fulvio et al. (2017), the solid rare-gas matrix is heated at $1\text{--}2\text{ K minute}^{-1}$ up to 40 K to promote the

diffusion and reactivity of these precursors within the solid matrix cage. Unlike their method, the precursors in our PDN are rotationally cooled before and during the plasma chemistry, which occurs entirely in the gas phase, mimicking more closely the conditions encountered by molecules and ions in a circumstellar envelope. Fulvio et al. (2017) identified the molecular species in the icy matrices via UV and mid-infrared spectroscopy as carbon chains ($\text{C}_2\text{--C}_{13}$ atoms). They claim that the condensation process is governed by carbon chains that first condense into small carbon clusters and finally into more stable carbonaceous compounds. It cannot be excluded, however, that the formation of carbon chains is favored in the solid phase. In our study, the grains are formed in the gas phase and we use laser desorption mass spectrometry to identify higher-weight molecular compounds with an estimated $\text{C}_4\text{--C}_{35}$ atoms. Finally, in their study samples were warmed up to room temperature to analyze the residues *ex situ*. High-resolution transmission electron microscopy analysis showed fullerene-like carbon material characterized by curved graphene layers.

The structural characteristics of their solids are comparable to high-temperature gas-phase condensation products (Jäger et al. 2009b). Similarly, our SEM images of grains from the benzene experiment show 3D spherical structures composed of clusters of smaller grains. Both studies reveal that soot-like and fullerenic-like grain morphologies are possible in a low-temperature formation process.

6.5.3. Astrophysical Implications

Our experimental results are presented in the context of astrophysical environments in Figure 14. The low-temperature (150–200 K) reactivity that is simulated in the supersonic expansion of the COSMIC chamber is applicable to the colder outer shells of circumstellar envelopes (CSEs) of AGB stars (Decin et al. 2010) but also to the ionized/molecular transitions in photon-dominated regions (PDRs) (Hollenbach & Tielens 1997). Dust is formed in the middle to outer CSEs, where temperature gradients from thousands to tens of kelvin lead to molecular freezout (Ziurys 2006). Allamandola et al. (1989) first suggested acetylene as a building block of aromatic hydrocarbons in CSEs. To explain the formation of grains from molecular precursors, PAH growth mechanisms studied by the combustion community were initially suggested (Frenklach et al. 1985; Cherchneff et al. 1992). Frenklach & Feigelson (1989) suggested HACA as the major mechanism of PAH growth in CSEs and derived dust masses and PAH yields, finding that a narrow temperature window of 900–1100 K was necessary for grain formation. Cherchneff & Cau (1999) found that the conversion of single rings to PAHs begins at $1.7 R_*$ at a temperature of ~ 1700 K. PAHs found in presolar grains extracted from meteorites present further evidence that PAHs can be synthesized in such outflows (Messenger et al. 1998; Bernatowicz et al. 2006).

The first observations of benzene toward the protoplanetary nebula CRL 618 pointed at possible non-HACA formation routes for the smallest aromatic unit (Cernicharo et al. 2001). Although PAHs have been mostly observed in emission near an exciting source, the recent detection of benzonitrile by McGuire et al. (2018) in TMC-1, a cold molecular cloud, has opened the door to observational searches for PAHs in the cold ISM. The inclusion of PAHs in models of TMC-1 has led to a better agreement between models and observations (Wakelam & Herbst 2008). *Spitzer*

provided the first detections of the μm PAH absorption feature toward stars that are not associated with circumstellar dust (Malsberger et al. 2009). This implies that PAHs are widespread components of cold ISM dust, similar to the well-studied aliphatic hydrocarbon band at 3.4 μm .

Models have shown that PAHs can be destroyed more rapidly than they are supplied in the ISM due to electrons, photons, cosmic rays, and shocks (Tappe et al. 2006; Micelotta et al. 2010a, 2010b, 2011; Kaneda et al. 2011; Hervé et al. 2019). These authors proposed the need for a re-formation mechanism of carbonaceous grains taking place directly in the ISM. Alternatively, Tielens (2013) suggested that only the most stable PAHs, e.g., circumcoronene, are expected to survive the harsh conditions of the ISM.

Low-temperature routes to PAH formation and growth could resupply the missing PAH reservoir. Recent experiments have shown that vinylacetylene-mediated formation of PAHs (i.e., the HAVA mechanism) is barrierless at low temperatures (Zhao et al. 2018a, 2018b). HAVA could be expanded to more complex PAHs in the ISM (Kaiser et al. 2015). Additional mechanisms including MAC (Shukla et al. 2010) and PAC (Shukla & Koshi 2010; Zhao et al. 2019b) could provide further routes to PAH growth.

We provide experimental evidence that PAHs and grains can grow at low temperatures and we propose this as a potential re-formation mechanism of interstellar grains. By growing within solid grains, small and medium-sized PAHs (50–250 g mol^{-1}) could be shielded against harsh energetic processing in the ISM. PAHs in solid grains could then be released into the gas phase via ablation or grain sputtering mechanisms.

6.6. Conclusions

We have produced dust grain analogs to interstellar carbonaceous dust by using a

supersonic jet exposed to a plasma discharge, within the PDN of the COSmIC facility. The samples were prepared in separate experiments involving aromatic precursors of increasing ring number (one to four rings). Phenanthrene was also included to evaluate the impact of PAH structure in the formation of grains and was compared to its isomer, anthracene. These solid samples were characterized via laser desorption mass spectrometry. Mass spectra of these samples exhibited many differences, and the quantitative analysis of the peaks with $S/N > 10$ allowed us to draw the following conclusions.

For all aromatic precursors, the formation of pyrene and/ or its isomers and its methylated series is ubiquitous. The formation of pyrene (or isomers) signals a stable route in the growth of larger PAHs and COMs. The highest dust yield was obtained in the anthracene experiment, while the lowest yield was obtained in the benzene experiment. This could be partially due to the short residence time of parent molecules in the plasma discharge and could also be explained by the fact that anthracene is less dissociated in the PDN and therefore the formation and growth starts with larger building blocks that are already aromatic and highly reactive.

For the anthracene experiment, we have shown that molecular growth is possible via the MAC mechanism at low temperature, but we cannot disregard the possibility of other growth processes such as PAC, HACA, and HAVA, a limitation of our probing technique.

Because of the rapid passage of molecules across the discharge (a few microseconds), the formation and molecular growth within grains is mostly representative of the first steps of PAH growth. The absence of high- mass compounds ($>400 m/z$) indicates that the organic solids produced consist of small aromatic units held together by aliphatic bridges.

We explored the effect of PAH structure by studying anthracene and phenanthrene, isomers with linear and cata-condensed structures respectively. Our results show that the lower grain yield

of phenanthrene matches the current understanding of its greater stability, explaining its lower dissociation and reactivity in the PDN compared to anthracene. Their mass spectra, however, show that their molecular diversity is very similar, suggesting that once fragmented they follow similar growth routes.

Comparison to L2DMS measurements on the Allende and Murchison meteorites confirm the formation of stable PAHs including anthracene and pyrene in their organic reservoir. Mass spectra of the anthracene and Murchison meteorite show that the formation of pyrene isomers and methylated series is ubiquitous. Based on our experimental results, we propose low-temperature PAH growth as a possible mechanism in the formation and evolution of organic compounds found in meteorites.

References

- Abrahamson, J. P. 2017, PhD thesis, Pennsylvania State Univ. Acke, B., Bouwman, J., Juhász, A., et al. 2010, [ApJ](#), **718**, 558
- Adamson, B. D., Skeen, S. A., Ahmed, M., & Hansen, N. 2018, [JPCA](#), **122**, 9338
- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1985, [ApJL](#), **290**, L25
- Allamandola, L. J., Tielens, A. G. G. M., & Barker, J. R. 1989, [ApJS](#), **71**, 733 Bauschlicher, Charles W. J., & Ricca, A. 2009, [ApJ](#), **698**, 275
- Bernatowicz, T. J., Croat, T. K., & Daulton, T. L. 2006, in *Origin and Evolution of Carbonaceous Presolar Grains in Stellar Environments*, ed.
- D. S. Lauretta & H. Y. McSween, Jr (Tucson, AZ: Univ. Arizona Press), 109
- Bernstein, M. P., Sandford, S. A., Allamandola, L. J., Chang, S., & Scharberg, M. A. 1995, [ApJ](#), **454**, 327

- Biennier, L., Benidar, A., & Salama, F. 2006, *CP*, 326, 445
- Biennier, L., Salama, F., Allamandola, L. J., & Scherer, J. J. 2003, *JChPh*, 118, 7863
- Biloiu, C., Sun, X., Harvey, Z., & Scime, E. 2007, *JAP*, 101, 073303 Bockhorn, H. 2013, *Soot Formation in Combustion: Mechanisms and Models* (Berlin: Springer)
- Boersma, C., Mattioda, A. L., Bauschlicher, C. W. J., et al. 2009, *ApJ*, 690, 1208
- Boschi, R., Clar, E., & Schmidt, W. 1974, *JChPh*, 60, 4406
- Broks, B. H. P., Brok, W. J. M., Remy, J., et al. 2005, *PhRvE*, 71, 036409
- Cain, J., Laskin, A., Kholghy, M. R., Thomson, M. J., & Wang, H. 2014, *PCCP*, 16, 25862
- Callahan, M. P., Abo-Riziq, A., Crews, B., Grace, L., & de Vries, M. S. 2008, *AcSpA*, 71, 1492
- Callahan, M. P., Gerakines, P. A., Martin, M. G., Peeters, Z., & Hudson, R. L. 2013, *Icar*, 226, 1201
- Carpentier, Y., Féraud, G., Dartois, E., et al. 2012, *A&A*, 548, A40 Cavallotti, C., & Polino, D. 2013, *Proc. Combust. Inst.*, 34, 557
- Cernicharo, J., Heras, A. M., Tielens, A. G. G. M., et al. 2001, *ApJL*, 546, L123
- Chen, H., & Dobbins, R. 2000, *CST*, 159, 109
- Cherchneff, I., Barker, J. R., & Tielens, A. e. G. G. M. 1992, *ApJ*, 401, 269 Cherchneff, I., & Cau, P. 1999, in *IAU Symp. 191, Asymptotic Giant Branch Stars*, ed. T. Le Bertre, A. Lebre, & C. Waelkens (Cambridge: Cambridge Univ. Press), 251
- Choi, S.-S., & So, H.-Y. 2005, *Bull. Korean Chem. Soc.*, 26, 1711 Clar, E. 1964, *Polycyclic Hydrocarbons* (London: Academic)

Clar, E. 1972, *The Aromatic Sextet in Mobile Source Emissions Including Polycyclic Organic Species* (New York: Wiley), 49

Clemett, S. J., Maechling, C. R., Zare, R. N., Swan, P. D., & Walker, R. M.

1993, *Sci*, 262, 721

Cody, G. D., Heying, E., Alexander, C. M. O., et al. 2011, *PNAS*, 108, 19171 Contreras, C. S., & Salama, F. 2013, *ApJS*, 208, 6

Cronin, J. R., Pizzarello, S., Epstein, S., & Krishnamurthy, R. V. 1993, *GeCoA*, 57, 4745

Dabestani, R., & Ivanov, I. N. 1999, *Photochem. Photobiol.*, 70, 10 D'Alessio, A., Barone, A., Cau, R., D'Anna, A., & Minutolo, P. 2005, *Proc.*

Combust. Inst., 30, 2595

Dartois, E., Geballe, T. R., Pino, T., et al. 2007, *A&A*, 463, 635 Dartois, E., Muñoz Caro, G. M., Deboffle, D., & d'Hendecourt, L. 2004, *A&A*,

423, L33

de Vries, M. S., & Hunziker, H. E. 1996, *ApSS*, 106, 466

de Vries, M. S., Hunziker, H. E., Meijer, G., & Wendt, H. R. 1992, *Proc. SPIE*, 1638, 102

de Vries, M. S., Reihs, K., Wendt, H. R., et al. 1993, *GeCoA*, 57, 933 Decin, L., Agúndez, M., Barlow, M. J., et al. 2010, *Natur*, 467, 64 Deckert, A. A., & George, S. M. 1987, *SurSc*, 182, L215

Derenne, S., & Robert, F. 2010, *M&PS*, 45, 1461 Dias, J. R. 2004, *J. Chem. Inf. Comput. Sci*, 44, 1210

Dobbins, R., Fletcher, R., & Chang, H.-C. 1998, *CoFl*, 115, 285 Duley, W. W. 1994, *ApJL*, 430, L133

Dunk, P. W., Adjizian, J. J., Kaiser, N. K., et al. 2013, *PNAS*, 110, 18081 Dzepina, K.,

Arey, J., Marr, L. C., et al. 2007, [IJMSp](#), **263**, 152

Favre, C., Fedele, D., Semenov, D., et al. 2018, [ApJL](#), **862**, L2 Frenklach, M. 2002, [PCCP](#), **4**, 2028

Frenklach, M., Clary, D. W., Gardiner, W. C., & Stein, S. E. 1985, in 20th Symp. (Int.) on Combustion (Amsterdam: Elsevier), 887

Frenklach, M., & Feigelson, E. D. 1989, [ApJ](#), **341**, 372

Frenklach, M., & Wang, H. 1991, in 23rd Symp. (Int.) on Combustion (Amsterdam: Elsevier), 1559

Fulvio, D., Góbi, S., Jäger, C., Kereszturi, Á., & Henning, T. 2017, [ApJS](#), **233**, 14

Gavilan, L., Le, K. C., Pino, T., et al. 2017, [A&A](#), **607**, A73 Gore, P. 1956, [J. Org. Chem.](#), **22**, 135

Grieco, W. J., Howard, J. B., Rainey, L. C., & Sande, J. B. V. 2000, [Carbon](#), **38**, 597

Gutman, I., & Stanković, S. 2007, [Maced. J. Chem. Chem. Eng.](#), **26**, 111 Hayes, J. M. 1967, [GeCoA](#), **31**, 1395

Henning, T., & Salama, F. 1998, [Sci](#), **282**, 2204

Herring, C. L., Faiola, C. L., Massoli, P., et al. 2015, [AerST](#), **49**, 1131 Hervé, M., Castellanos, P., Karras, G., et al. 2019, [EPJ Web Conf.](#), **205**, 06012 Hollenbach, D. J., & Tielens, A. G. G. M. 1997, [ARA&A](#), **35**, 179

Homann, K. 1998, [Angew. Chem. Int. Ed.](#), **37**, 2434

Hsia, C.-H., Sadjadi, S., Zhang, Y., & Kwok, S. 2016, [ApJ](#), **832**, 213 Irimia, D., Dobrikov, D., Kortekaas, R., et al. 2009, [RSci](#), **80**, 113303

Jäger, C., Huisken, F., Mutschke, H., Jansa, I. L., & Henning, T. 2009a, [ApJ](#), **696**, 706

Jäger, C., Huisken, F., Mutschke, H., Jansa, I. L., & Henning, T. 2009b, [ApJ](#), **696**, 706

Jäger, C., Krasnokutski, S., Staicu, A., et al. 2006, [ApJS](#), **166**, 557

Jäger, C., Mutschke, H., Henning, T., & Huisken, F. 2008, [ApJ](#), **689**, 249 Jäger, C., Mutschke, H., Henning, T., & Huisken, F. 2009c, in ASP Conf. Ser. 414, Cosmic Dust—Near and Far, ed. T. Henning, E. Grün, & J. Steinacker (San Francisco, CA: ASP), 319

Jochims, H. W., Ruhl, E., Baumgartel, H., Tobita, S., & Leach, S. 1994, [ApJ](#), **420**, 307

Jones, A. P., Duley, W. W., & Williams, D. A. 1990, [QJRAS](#), **31**, 567 Kaiser, R. I., Parker, D. S. N., & Mebel, A. M. 2015, [ARPC](#), **66**, 43

Kaneda, H., Onaka, T., Sakon, I., et al. 2011, in EAS Publications Series 46, PAHs and the Universe, ed. C. Joblin & A. G. G. M. Tielens (Les Ulis: EAS), 157

Kato, T., Yoshizawa, K., & Hirao, K. 2002, [JChPh](#), **116**, 3420

Kinney, C. R., Nunn, R. C., & Walker, P. L. 1957, [Ind. Eng. Chem](#), **49**, 880 Kislov, V. V., Sadovnikov, A. I., & Mebel, A. M. 2013, [JPCA](#), **117**, 4794 Kwok, S., & Zhang, Y. 2013, [ApJ](#), **771**, 5

Lafleur, A. L., Taghizadeh, K., Howard, J. B., Anacleto, J. F., & Quilliam, M. A. 1996, [JASMS](#), **7**, 276

Lee, J.-E., Lee, S., Baek, G., et al. 2019, [NatAs](#), **3**, 314 Léger, A., & Puget, J. L. 1984, [A&A](#), **500**, 279

Li, A., & Draine, B. T. 2012, [ApJL](#), **760**, L35

Madison, J., & R, R. 1958, [Ind. Eng. Chem.](#), **50**, 237

Malsberger, R., Chiar, J. E., Tielens, A. G. G. M., & Sloan, G. C. 2009, AAS Meeting Abstracts, [213](#), 412.02

Mansurov, Z. A. 2005, [CESW](#), 41, 727

Marshall, A. G., & Rodgers, R. P. 2008, [PNAS](#), 105, 18090

McGuire, B. A., Burkhardt, A. M., Kalenskii, S., et al. 2018, [Sci](#), 359, 202 Meijer, G., de Vries, M. S., Hunziker, H. E., & Wendt, H. R. 1990, [ApPhB](#), 51, 395

Mennella, V., Brucato, J. R., Colangeli, L., & Palumbo, P. 1999, [ApJL](#), 524, L71

Mennella, V., Brucato, J. R., Colangeli, L., & Palumbo, P. 2002, [ApJ](#), 569, 531 Messenger, S., Amari, S., Gao, X., et al. 1998, [ApJ](#), 502, 284

Micelotta, E. R., Jones, A. P., & Tielens, A. G. G. M. 2010a, [A&A](#), 510, A36 Micelotta, E. R., Jones, A. P., & Tielens, A. G. G. M. 2010b, [A&A](#), 510, A37

Micelotta, E. R., Jones, A. P., & Tielens, A. G. G. M. 2011, [A&A](#), 526, A52 Miller, J. A. 2003, [JPCA](#), 107, 7783

Minutolo, P., Gambi, G., DAlessio, A., & Carlucci, S. 1999, [AtmEn](#), 33, 2725 Nir, E. E., Hunziker, H., & de Vries, M. 1999, [AnaCh](#), 71, 1674

Öberg, K. I., Boogert, A. C. A., Pontoppidan, K. M., et al. 2011, [ApJ](#), 740, 109 Parker, D. S. N., Kaiser, R. I., Troy, T. P., & Ahmed, M. 2014, [Angew. Chem. Int. Ed.](#), 53, 7740

Parker, D. S. N., Zhang, F., Kim, Y. S., et al. 2012, [PNAS](#), 109, 53

Peeters, E., Bauschlicher, C. W. J., Allamandola, L. J., et al. 2017, [ApJ](#), 836, 198

Peeters, E., Hony, S., Van Kerckhoven, C., et al. 2002, [A&A](#), 390, 1089 Pendleton, Y. J., Sandford, S. A., Allamandola, L. J., Tielens, A. G. G. M., & Sellgren, K. 1994, [ApJ](#), 437, 683

Poater, J., Visser, R., Solà, M., & Bickelhaupt, F. M. 2007, *J. Org. Chem.*, **72**, 1134

Randić, M. 2003, *ChRv*, **103**, 3449

Rapacioli, M., Calvo, F., Spiegelman, F., Joblin, C., & Wales, D. J. 2005, *JPCA*, **109**, 2487

Raymond, A. W., Sciamma-O'Brien, E., Salama, F., & Mazur, E. 2018, *ApJ*, **853**, 107

Reizer, E., Csizmadia, I., Palotás, B., Viskolcz, B., & Fiser, B. 2019, *Molecules*, **24**, 1040

Richter, H., & Howard, J. 2000, *PrECS*, **26**, 565

Ricketts, C. L., Contreras, C. S., Walker, R. L., & Salama, F. 2011, *IJMSP*, **300**, 26

Sabbah, H., Bonnamy, A., Papanastasiou, D., et al. 2017, *ApJ*, **843**, 34 Sabbah, H.,
Morrow, A. L., Jenniskens, P., Shaddad, M. H., & Zare, R. N.
2010, *M&PS*, **45**, 1710

Sadjadi, S., Zhang, Y., & Kwok, S. 2017, *ApJ*, **845**, 123

Sakata, A., & Wada, S. 1989, in IAU Symp. 135, *Interstellar Dust*, ed.
L. J. Allamandola & A. G. G. M. Tielens (Dordrecht: Kluwer), **191** Salama, F., Sciamma-
O'Brien, E., Contreras, C. S., & Bejaoui, S. 2018, in IAU
Symp. 332, *Astrochemistry VII: Through the Cosmos from Galaxies to*

Planets, ed. M. Cunningham, T. Millar, & Y. Aikawa (Cambridge: Cambridge Univ.
Press), **364**

Sandford, S. A., Aléon, J., Alexander, C. M. O.'D., et al. 2006, *Sci*, **314**, 1720 Schnaiter,
M., Henning, T., Mutschke, H., et al. 1999, *ApJ*, **519**, 687 Schuetz, C. A., & Frenklach, M. 2002,
Proc. Combust. Inst., **29**, 2307 Schuler, B., Fatayer, S., Meyer, G., et al. 2017, *Energy Fuels*, **31**,
6856 Sciamma-O'Brien, E., Upton, K. T., & Salama, F. 2017, *Icar*, **289**, 214 Scott, A., & Duley,
W. W. 1996, *ApJL*, **472**, L123

- Shukla, B., & Koshi, M. 2010, [PCCP](#), **12**, 2427
- Shukla, B., & Koshi, M. 2012, [CoFI](#), **159**, 3589
- Shukla, B., Miyoshi, A., & Koshi, M. 2010, [JASMS](#), **21**, 534
- Shukla, B., Susa, A., Miyoshi, A., & Koshi, M. 2008, [JPCA](#), **112**, 2362 Sloan, G. C., Jura, M., Duley, W. W., et al. 2007, [ApJ](#), **664**, 1144
- Solà, M. 2013, [FrCh](#), **1**, 22
- Spencer, M. K., & Zare, R. N. 2007, [Sci](#), **317**, 1680
- Strohalm, M., Kavan, D., Novak, P., Volny, M., & Havlicek, V. 2010, [AnaCh](#), **82**, 4648
- Tan, X., & Salama, F. 2005a, [JChPh](#), **122**, 084318 Tan, X., & Salama, F. 2005b, [JChPh](#), **123**, 014312
- Tappe, A., Rho, J., & Reach, W. T. 2006, [ApJ](#), **653**, 267 Tielens, A. G. G. M. 2008, [ARA&A](#), **46**, 289
- Tielens, A. G. G. M. 2013, [RvMP](#), **85**, 1021
- Wakelam, V., & Herbst, E. 2008, [ApJ](#), **680**, 371
- Walker, P. 1990, [Carbon](#), **28**, 261
- Walsh, C., Loomis, R. A., Öberg, K. I., et al. 2016, [ApJL](#), **823**, L10 Wang, H., & Frenklach, M. 1997, [CoFI](#), **110**, 173
- Yang, X. J., Li, A., Glaser, R., & Zhong, J. X. 2017, [ApJ](#), **837**, 171
- Zakrzewski, V. G., Dolgounitcheva, O., & Ortiz, J. V. 1996, [JChPh](#), **105**, 8748
- Zhao, L., Kaiser, R. I., Xu, B., et al. 2018a, [NatAs](#), **2**, 413 Zhao, L., Kaiser, R. I., Xu, B., et al. 2018b, [NatAs](#), **2**, 973 Zhao, L., Kaiser, R. I., Xu, B., et al. 2019a, [NatCo](#), **10**, 1510
- Zhao, L., Prendergast, M., Kaiser, R., et al. 2019b, [Angew. Chem. Int. Ed.](#), **58**, 17442

Zhao, L., Prendergast, M. B., Kaiser, R. I., et al. 2019c, *Chem. Phys. Chem.*, 20, 1437

Ziurys, L. M. 2006, *PNAS*, 103, 12274

Zuckerman, B. 1980, *ARA&A*, 18, 263