Supporting Information

2D IR spectroscopy of the anti-HIV agent KP1212 reveals protonated and neutral tautomers that influence pH-dependent mutagenicity

Chunte Sam Peng[1⊥], Bogdan I. Fedeles[1-3⊥], Vipender Singh[1-3], Deyu Li[1-3], Tiffany Amariuta[2], John M. Essigmann[1-3], Andrei Tokmakoff[4]*

Departments of [1]Chemistry and [2]Biological Engineering, [3]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Department of Chemistry, Institute for Biophysical Dynamics, and James Franck Institute, The University of Chicago, Chicago, IL 60637, USA

[⊥] denotes equal contribution

*E-mail: tokmakoff@uchicago.edu

Contents

Supporting Information

1 Experimental methods

Infrared spectroscopy: Equilibrium 2D IR spectra with perpendicular (ZZYY) polarization were collected using a 2D IR spectrometer described previously (1). The coherence time between the first two pulses was scanned in 4 fs steps from -60 fs to 3.0 ps and 2.4 ps for rephasing and non-rephasing spectra, respectively. The waiting time in all 2D IR spectra presented here was τ_2=150 fs, unless otherwise specified. Linear absorption was divided out along both frequency axes to remove spectral distortions (2). For transient experiments, the 10 ºC T-jump was generated by a 6 ns, 20 Hz Nd:YAG pumped optical parametric oscillator (OPO). The 1.98 μm, 10 mJ output of the OPO heated D_2O by exciting its O-D stretch overtone. The T-jump pulse was focused to a 500 μm waist, which is about five times bigger than the IR beam size, therefore creating a uniform heating within the probing area. Fifty 1 kHz IR pulses at 6 μm following the T-jump pulse were used to measure the transient IR spectra. Parallel (ZZZZ) polarization was used in transient experiments to enhance the signal-to-noise ratio. Transient HDVE spectra were taken using the Fourier-transform spectral interferometry and phase-modulation spectral interferometry methods, in which the local oscillator was stepped from 0 to 25 fs in 5 fs steps from the nonlinear signal (3). The spectra were acquired fixing τ_1=0 fs and the waiting time τ_2=150 fs.

Melting point analysis of DNA duplexes: The 16mer oligonucleotide 5'-GAAGACCTXGGCGTCC-3', where X is KP1212 was synthesized and purified as described previously (4). All other oligonucleotides were from Integrated DNA Technologies (Coralville, IA).Thermal denaturation of DNA duplexes in the presence of a saturating fluorescent dye was carried out on a Roche LightCycler480 machine as previously reported (5), with modifications. The 16mer oligonucleotides were annealed to the 17mer complementary strands by slow cooling (0.1 °C/s) from 80 °C to 4 °C in a 62.5 mM phosphate buffer adjusted to various pHs, which also contained 125 mM potassium ions (adjusted with KCl), 2.5 mM magnesium chloride and 1x LCGreen Plus saturating dye (Idaho Technologies Inc., Salt Lake City, UT). Fluorescence was recorded every second as the samples were heated at 0.1 °C/s from 37 °C to 95 °C. Data was analyzed using the Roche LC480 software. Three replicates were setup and analyzed for each oligonucleotide/complement pair.

In vitro mutagenesis: The M13 single-stranded genomes, containing one KP1212 base at a specific site were constructed and purified as previously reported (4). For the primer extension reactions, 100 fmol extension primer (5'-CGTGATCATGCGCAGACTGACATCATGTGTAAAACGACGGCCAGTGAATTGGA-3') were annealed to 100 fmol M13 genome by heating the mixture at 80 °C for 5 min followed by a slow (0.1 °C/s) cooling to 4 °C. The primer was extended by 10 U of Klenow fragment exo- polymerase (NEB Inc. Ipswitch,MA) in a solution containing 50 mM phosphate buffer (pH as indicated), 100 mM K^+ (adjusted with KCl), 10 mM Mg^{2+}, 1 mM DTT and 125 μM of each of the four dNTPs (NEB), for 4 h at 30 °C. The resulting product was purified using Qiaquick columns (Qiagen) and then PCR amplified with PfuTurbo

polymerase (Agilent) with the primers: 5'-YCAGCTATGACCATGATTCAGTGGAAGAC-3' (fwd) and 5'-YCGTGATCATGCGCAGACTGACATCATGTG-3' (rev), where Y is a phospho-hexylamino linker which prevents 5' phosphorylation during subsequent steps. Amplification was done in 32 cycles of 95/68/72 °C for 30/30/60 s respectively. The PCR products were subsequently purified and analyzed with the REAP assay as previously described (4, 6).

2 SVD analysis of the pH-dependent FTIR spectra

To perform singular value decomposition (SVD) of the pH-dependent FTIR spectra of KP1212, we generate a matrix of raw data, denoted as A. The columns of the A contain the spectra at different pH's. Each spectrum at a specific pH has contributions from the various tautomers of KP1212 at different protonation states. Our goal is to identify these "basis spectra" or the difference spectra between each titratable species in the mixture. We introduce a matrix D whose columns consist of these difference spectra and one reference spectrum. A transformation matrix F specifies the pH dependence of each difference spectrum and the mixing of the columns of D to form A. The columns of F contain the Henderson-Hasselbalch equations for the individual transitions and the last column contains ones for the constant reference spectrum. The transformation follows:

$$DF^T = A \tag{S1}$$

SVD helps us solve for F^T. The raw data, which can be written as the sum of matrix A and the experimental noise E, contains information in the frequency space (rows) and pH space (columns). SVD analysis separates them into individual matrices, denoted as $\mathbb{U}$ (frequency) and $\mathbb{V}$ (pH), following:

$$A + E = \mathbb{U}\mathbb{S}\mathbb{V}^T \tag{S2}$$

The $\mathbb{S}$ matrix generated from the SVD analysis is square and diagonal. The diagonal values of $\mathbb{S}$ are called the singular values and indicate the weights for the corresponding columns of $\mathbb{U}$ and $\mathbb{V}$. The singular values start from a high value and decrease to a low value plateau which indicates the noise level. All of the essential information about the system should be contained in the first few component spectra with high singular values. Fig. S1a and b show the first five components from the $\mathbb{U}$ and $\mathbb{V}$ matrices, along with the corresponding normalized singular values.

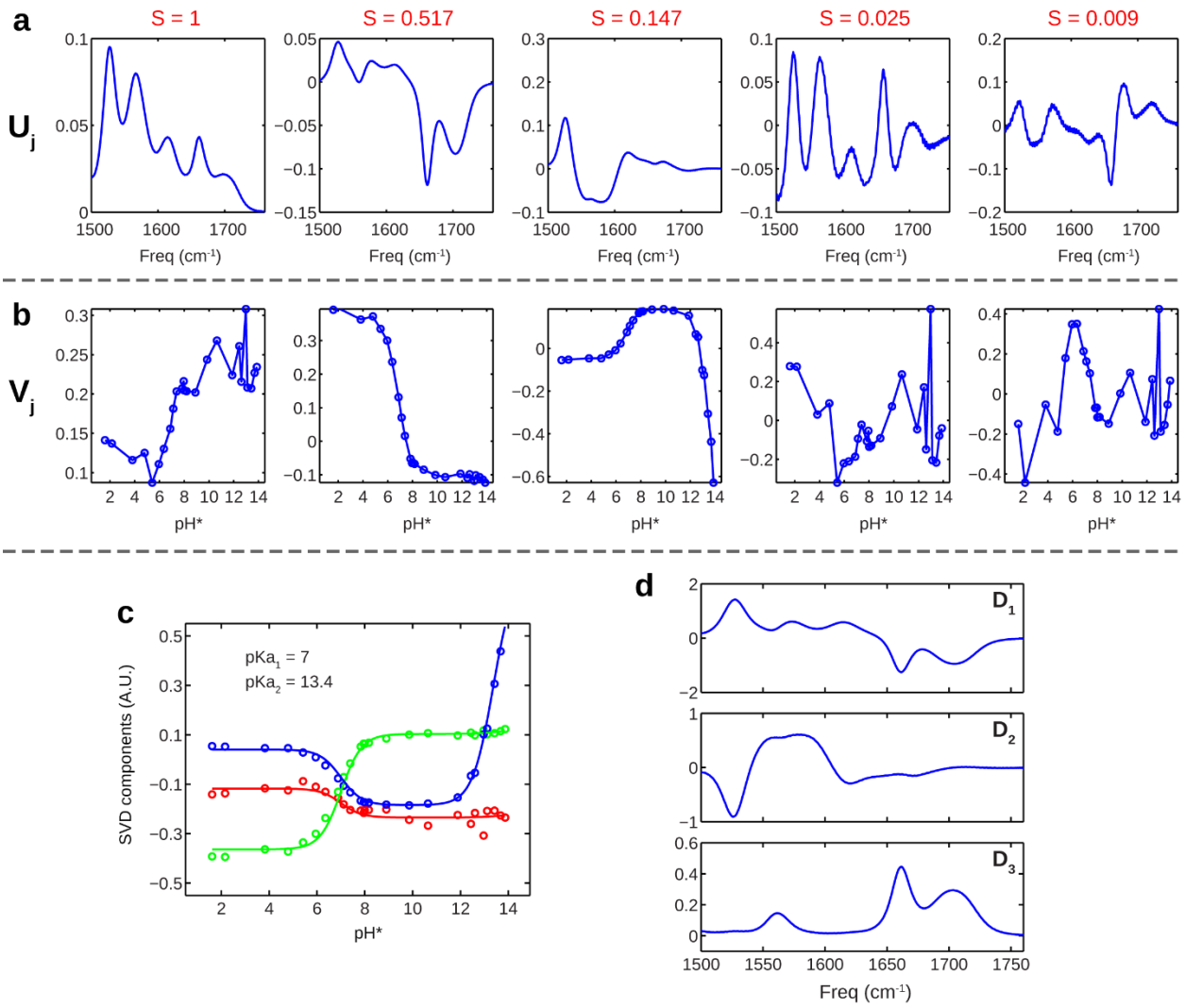

Fig. S1: SVD analysis of the pH-dependent FTIR spectra shown in Fig. 2. The first five components from the $\mathbb{U}$ and $\mathbb{V}$ matrices are shown in panels (a) and (b), respectively. The corresponding singular values are listed on the top. (c) The first three SVD components with fits to Henderson-Hasselbalch equations. (d) SVD reconstructed basis spectra. The D components represent one reference spectrum and two difference spectra.

However, the SVD component spectra are just the result of mathematic manipulations, and they do not necessarily represent physically relevant spectra such as the individual tautomer spectra. In order to reconstruct the basis spectra using the SVD components, we first pick out the significant components based on the singular values and the continuity of the $\mathbb{V}$ components. The first three singular values are much higher than the rest, and that the $\mathbb{V}$ components become very noisy starting from the fourth component, therefore, we choose to describe the data set with a reduced 3x3 matrix. After reducing the $\mathbb{S}$ matrix to 3x3, and the $\mathbb{U}$ and $\mathbb{V}$ to three columns, we can combine Equation (S1) and (S2) and rewrite as:

$$DF^T = A = USV^T \tag{S3}$$

where U, S, and V represent the reduced SVD components. The goal is to solve for D which contains the true difference spectra between the individual tautomer spectra. Multiplying both sides of the equation by the pseudoinverse of F^T gives

$$D = USV^T(F^T)^+ = USH \tag{S4}$$

where

$$H = V^T(F^T)^+. \tag{S5}$$

Now we fit the experimentally determined V components to find the F matrix, whose first two columns consist of the two Henderson-Hasselbalch equations with pKa_1 and pKa_2, respectively, and the third column is a column of ones to represent the background spectrum. The calculated V matrix is written as:

$$\tilde{V} = F\tilde{H}^T = F(F^+V) \tag{S6}$$

where ~ denotes that the values are based on initial guess. The Hendersen-Hasselbach terms in the F components are expressed as functions of pH,

$$C_j[1+10^{(pKa_j - pH)}]^{-1} \tag{S7}$$

where C_j is a constant fitting parameter. We fit the V components simultaneously by minimizing the sum of the squares of the residuals R

$$R = V - \tilde{V} \tag{S8}$$

During the minimization process, each column of V and $\tilde{V}$ is weighted by the appropriate square of the singular value. The fit results of the V components are shown in Fig. S1c with $pKa_1 = 7.0$ and $pKa_2 = 13.4$.

Finally after the best F is found, the D matrix can be reconstructed as follows,

$$D = A(F^T)^+ \tag{S9}$$

These three D components displayed in Fig. S1d represent one reference spectrum, and two difference spectra. It is clear that D3 should be the reference spectrum due to the absence of negative peaks. The blue-shifted peak at 1704 cm^{-1} suggests that it arises from the C=O stretch of the protonated keto tautomer, therefore D_3 is assigned to the "protonated" spectrum. Further inspection informs us that D1 represents the difference spectrum between the "protonated" and "neutral" spectra, and that D2 represents the difference spectrum between the "neutral" and "deprotonated" spectra.

Finally, to obtain the spectra for protonated, neutral, and deprotonated KP1212, we apply a transformation matrix W on D and F.

$$A = DF^T = (DW)\left(W^{-1}F^T\right) = BY^T \tag{S10}$$

where the columns of B (Fig. 1E-G) contain the "protonated", "neutral", and "deprotonated" spectra, and Y (Fig. 1D) shows the pH dependence of these three populations.

A caveat that should be noted is that the pKa value determined directly from the measurements in D_2O solution is not the same as the pKa in H_2O solutions. The pH of a H_2O solution can be accurately measured using standard glass electrode; however, the measurement of pD of a D_2O solution is not as straightforward. In practice, a quantity pH* is often used, which is the direct pH reading in a D_2O solution using the standard pH meter calibrated in H_2O solutions. An empirical relationship between pD and pH* is established as (7): pD = pH* + 0.4. The pKa value obtained in D_2O can be approximated to get the pKa value in H_2O following (8): $pKa(H_2O) = pKa(D_2O)*0.929 + 0.42$. In our case, the pKa in H_2O would be calculated to be 6.9. However, since the constant terms in the two conversions nearly cancel out each other, the pKa value in H_2O is often simplistically reported as the pKa determined in D_2O solutions using pH*. This is called the "cancel-out" approach (7), which is what we used in this study.

3 2D IR spectra of protonated CMP and KP1212

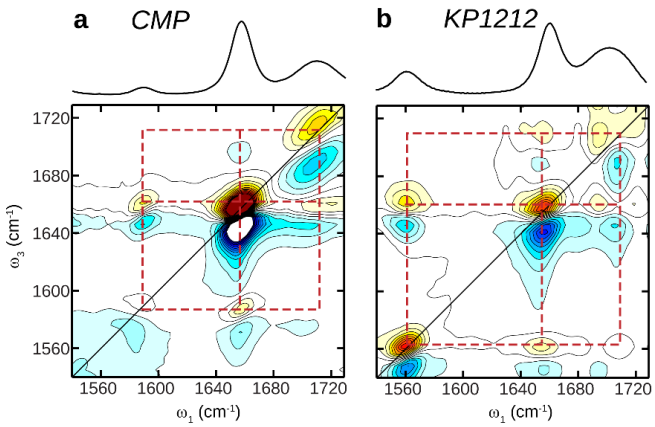

Fig. S2: FTIR (top) and 2D IR (bottom) spectra of CMP (left) and KP1212 (right) at pH* 1.0 at 25 °C.

4 SVD analysis of temperature-dependent FTIR spectra

4.1 pH* 7.9 data set

In order to extract thermodynamic data such as enthalpy, entropy, and equilibrium constants, we performed SVD analysis of the temperature-dependent FTIR spectra. With two significant SVD components, a two-state model was applied to reconstruct two physically-relevant spectra that can represent two species. Although we have identified the presence of more tautomers from 2D IR spectra, the population changes of rare tautomers are on the order of experimental noise as indicated by the SVD analysis. Therefore we considered only the broad categories of keto and enol species for spectral reconstruction.

The reconstructed spectra are shown in Fig. S3a and b, and their corresponding populations are shown in Fig. S3c. The blue spectrum accounts for the dominant species at this pH, and can be assigned mostly to the **EI** tautomer because of its characteristic triplet peak structure. The red spectrum represents the sum of the different remaining tautomers including the keto forms based on the broad C=O peak observed ~1650 cm^{-1}. The possibility of including the **EA** tautomer in Fig. S3b can contribute to the intensity observed in the 1530-1570 cm^{-1} region. With the populations of the two species obtained, a van't Hoff plot (Fig. S3d) can be used to extract the thermodynamic variables listed in Table S1.

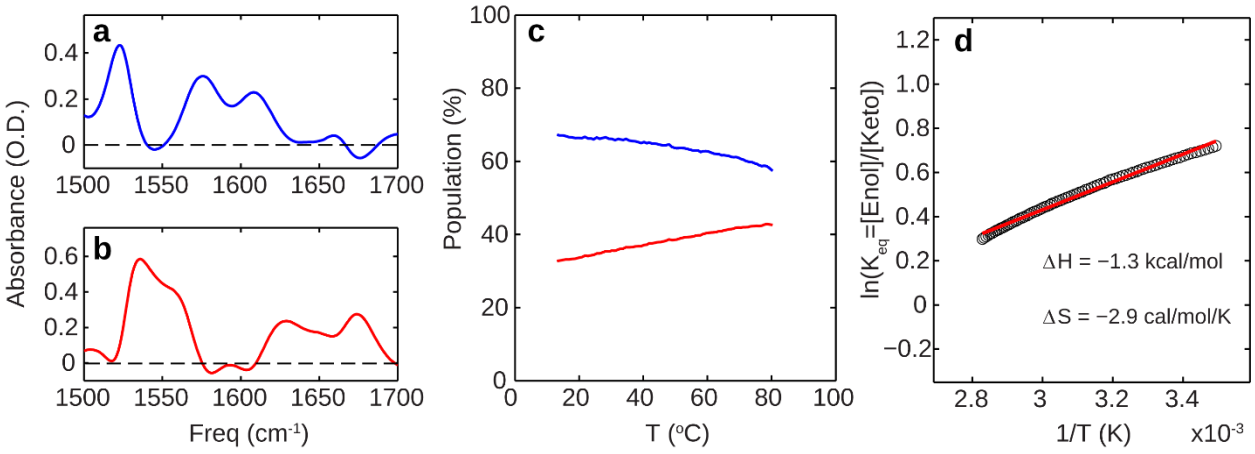

Fig. S3: The results from the SVD reconstruction using a two-state model: (a) the reconstructed spectrum representing the enol tautomer; (b) the reconstructed spectrum representing the keto tautomer, and possibly other minor enol tautomers. (c) The populations of the two species shown in (a) and (b) as a function of temperature. (d) The van't Hoff plot.

In the present analysis, we did not account for the contribution of thermally-induced frequency shift, which can also result in spectral features of peak gain and loss (such as in the 1530-1570 cm^{-1} region).

However, this contribution is believed to be small as small frequency shifts are observed in the temperature-dependent FTIR spectra of CMP (Fig. 3A).

In order to provide further support the populations obtained from the two-state SVD analysis, we compared the FTIR spectra of KP1212 to that of 5-aza-2'-deoxycytidine (5-aza-dC, structure shown in Fig. S4b). 5-aza-dC is the precursor of KP1212 in the keto-amino form, we therefore assume that the keto C=O stretch of 5-aza-dC and KP1212 have similar dipole strength. The keto population of KP1212 can then be estimated by comparing the integrated C=O peak intensity (shown as blue dash curves). It was found to be ~30% at 25 °C, which is close to 35% estimated from the two-state SVD analysis. This result suggests the validity of the SVD analysis.

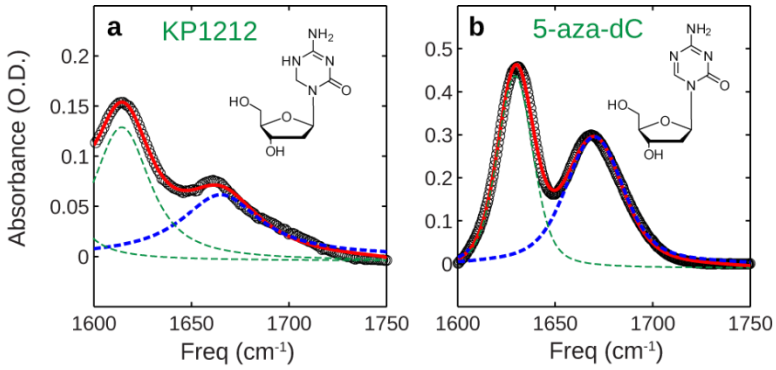

Fig. S4: FTIR spectra of 20 mg/ml of KP1212 (a) and 5-aza-dC (b). The experimental spectra (black circles) were fitted to the sum of Voigt profiles shown in red (individual peak in dash blue curves).

Table S1. Thermodynamic variables obtained from the SVD two-state analysis of the temperature-dependent FTIR spectra.

	pH* 7.9 keto → enol	**pH* 6.6 protonated keto → neutral enol**
ΔH [kcal/mol]	-1.3	5.9
ΔS [cal/mol/K]	-2.9	16.8
ΔG^o (37 °C) [kcal/mol]	-0.4	0.69
K_{eq} (37 °C)	1.78	0.33

4.2 pH* 6.6 data set

At pH* 6.6, KP1212 is substantially protonated and thus there should be at least three species involved: protonated keto, neutral keto and neutral enol. However, since the spectral features of neutral keto overlap with the protonated keto, we are restricted to the two-state model by grouping the two keto species together.

The contribution from the neutral keto tautomer should be much smaller than the protonated species. The reconstructed spectrum plotted in blue in Fig. S5a represents the spectrum for protonated keto tautomer, and can be compared to the low pH FTIR spectrum. The spectrum in red (Fig. S5b) represents the neutral EI spectrum, and it is likely to contain contributions from the other tautomers since it does not match perfectly with the EI tautomer spectrum obtained from the SVD analysis of the pH* 7.9 data set. The obtained K_{eq} = [protonated keto]/[neutral enol] shown in Fig. S5d can then be used to estimate temperature-dependent pK_a.

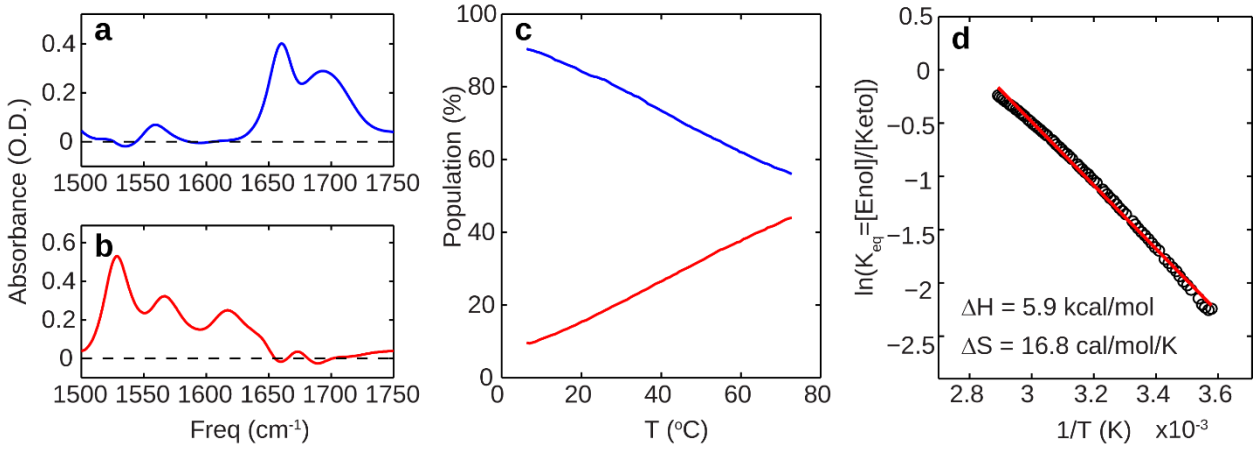

Fig. S5: The results from the SVD reconstruction of the pH* 6.6 spectra using a two-state model: (a) the reconstructed spectrum representing the protonated keto tautomer; (b) the reconstructed spectrum representing the enol tautomer, and possibly other minor enol tautomers. (c) The populations of the two species shown in (a) and (b) as a function of temperature. (d) The van't Hoff plot.

5 DFT calculations

5.1 Explicit solvent effect

In order to examine the solvation effect on the vibrational spectra, we performed calculations in vacuum and with explicit D_2O molecules in close proximity of a strong hydrogen bond donor/acceptor for example the carbonyl and amino groups. All labile protons were deuterated to match the experimental condition. Calculated spectra *in vacuo* (Fig. S6a) show only two significant peaks in the spectral region of interests for each of the five tautomers, which clearly cannot explain the pronounced cross-peaks between the three low frequency modes observed in the 2D IR spectra. As more explicit water molecules were placed near the hydrogen bond donor and acceptor sites to solvate KP1212, the calculated spectra change dramatically (Fig. S6b), especially for the enol tautomers. Fig. S6c shows the progression of the calculated spectra for the *cis*-enol-imino (**cEI**) tautomer as the number of explicit water molecules increases. Upon solvation, the coupling of the O-D stretch of D_2O and the KP1212 ring modes gives rise to three significant peaks in the 1500–1600

cm^{-1} spectral region that resemble the triplets seen in the experimental spectra. This observation demonstrates the importance of microsolvation in simulating IR spectra in the condense phase where hydrogen bond interaction plays an essential role in altering vibrational couplings. This finding should also invoke further theoretical studies to investigate the physical origins of these spectral variations, such as intermolecular coupling or electrostatic interactions. For the peak assignments in this work, we relied on just the qualitative peak patterns rather than the absolute peak frequencies and intensities.

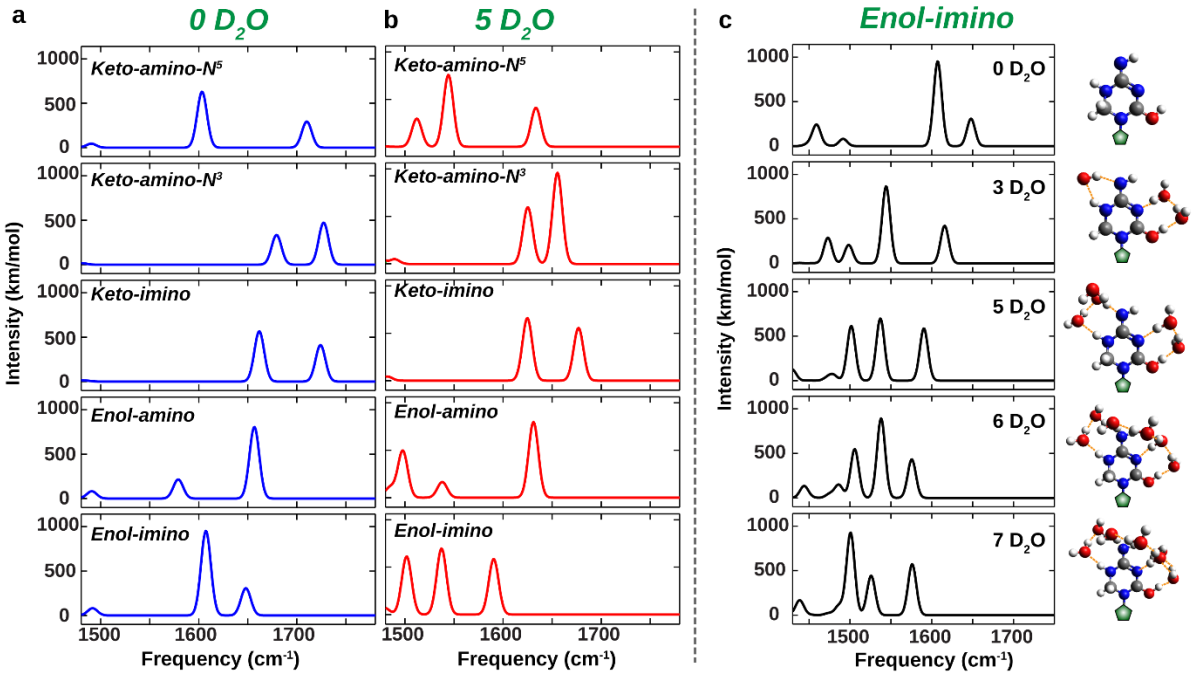

Fig. S6: Calculated IR absorption spectra for the five possible tautomers of KP1212 without (a) and with (b) five explicit water solvent molecules. The *cis* isomers of the imino tautomers are shown. (c) Calculated IR absorption spectra for the *cis*-enol-imino tautomer with different microsolvation environments.

5.2 Cis-trans isomerization

In principle, the imino tautomers can exist in both the *cis* or *trans* isomers. For the keto-imino tautomer, the *cis* and *trans* isomers have anti-correlated frequency shifts for the C=O and ring modes (Fig. 4). For the enol-imino tautomer, the *trans* isomer exhibit slight frequency red-shift for the first two ring modes. The **cEI** spectrum resembles the experimental spectrum better and therefore we attributed the experimentally observed species to **cEI**. However, it is less clear which **KI** spectrum corresponds to the experiment. An improved and more quantitative calculation would be required to make the definitive assignment.

5.3 pH dependence

We also found very good agreements between the experimental and calculated spectra for protonated and deprotonated species (Fig. S7), which provides further support for the assignment on the neutral KP1212. Protonated KP1212 exists in the keto-amino (**KA**) form whereas deprotonated KP1212 are in the enol-amino (**EA**) and enol-imino (**EI**) forms. For deprotonated KP1212, the combination of the doublet peak of **EA** and the intense peak of **EI** can explain the broad peak in the experimental spectrum.

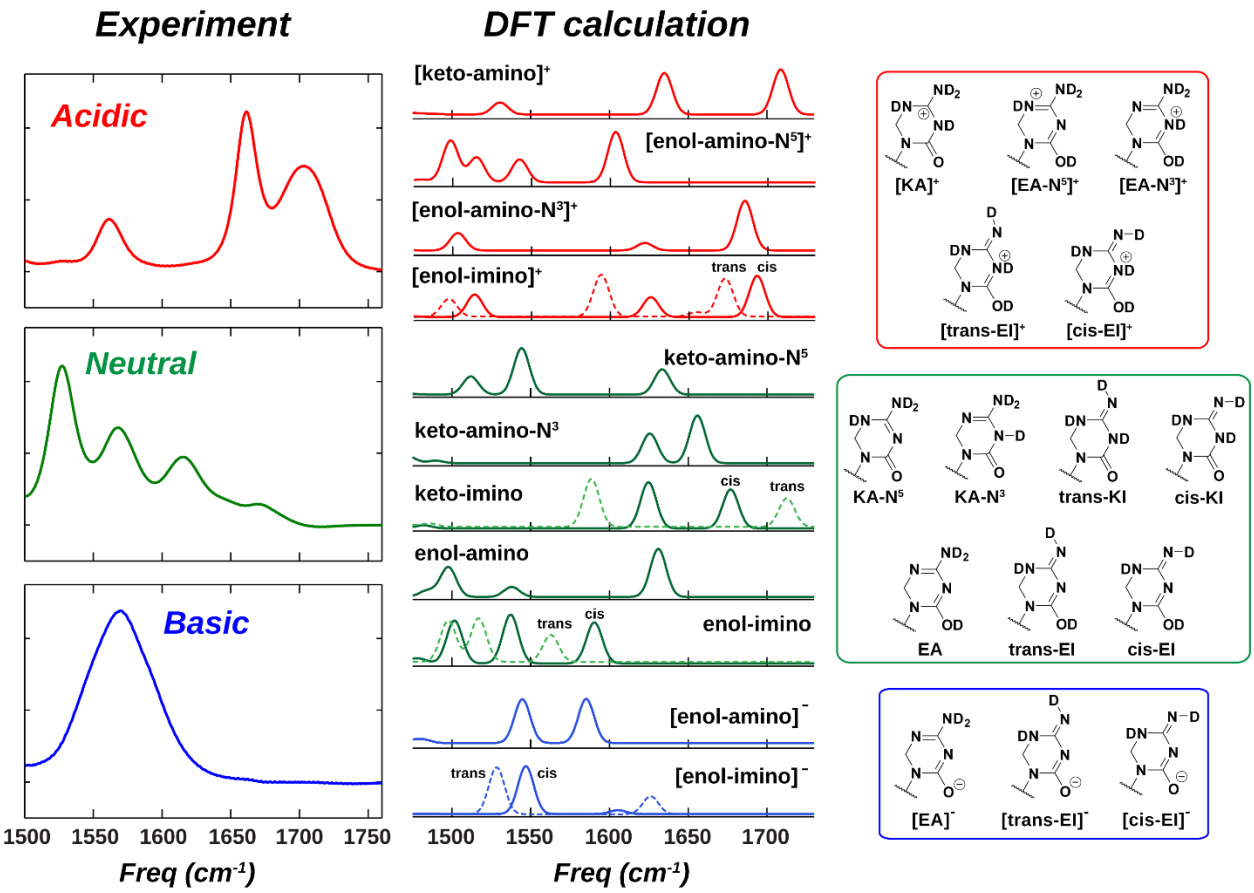

Fig. S7: Comparison between the experimental (left) and DFT calculated (middle) absorption spectra under different ionic state. The protonated, neutral, and deprotonated spectra are highlighted in red, green, and blue, respectively. The chemical structures are shown on the right. Five explicit D_2O molecules were included.

Table S2. DFT calculated vibrational frequencies (scaled by 0.9614) of the various tautomers.

KA-N^5	KA-N^3	cis-KI	trans-KI	EA	cis-EI	trans-EI
1634	1657	1678	1712	1631	1591	1563
1544	1626	1626	1589	1538	1537	1518
1512				1498	1502	1498

$[KA]^+$	$[EA\text{-}N^5]^+$	$[EA\text{-}N^3]^+$	$[cis\text{-}EI]^+$	$[trans\text{-}EI]^+$	$[EA]^-$	$[cis\text{-}EI]^-$	$[trans\text{-}EI]^-$
1710	1603	1685	1693	1674	1585	1607	1626
1635	1542	1622	1626	1594	1545	1548	1529
1530	1515	1503	1514	1498			
	1498						

6 Assignments of the minor tautomers

6.1 Keto tautomers in water

The absence of **KA-N⁵** is concluded by the lack of cross-peaks between a C=O stretch mode to low frequency ring modes in the 1500-1600 cm^{-1} region. The broad distribution of C=O frequencies at high temperatures from 1650 to 1675 cm^{-1} (FWHM = 46 cm^{-1}) provides evidence for the existence of multiple keto tautomers. The DFT calculations of the **KA-N³** and **KI** tautomers suggest they both possess a doublet peak structure: two diagonal and two cross-peaks. However, since the two peaks of **KA-N³** are close in frequency, the *pure* **KA-N³** spectrum may appear to be just one broad peak. In order to identify these expected cross-peak patterns, we fit the experimental spectrum with the sum of tilted 2D Gaussian peaks. The experimental spectrum used in the fit (Fig. S8a) was taken at 1 ps waiting time in order to enhance the cross-peak structure. We fixed a single peak at 1612 cm^{-1} (Fig. S8f) in order to represent the ring mode from the **EI** tautomer, which does not have any cross-peaks in this spectral region. In the first model, we assumed that there is only one keto tautomer with one C=O stretching vibration. The best fit using two uncoupled peaks (Fig. S8b) does not show a ridge along ω_3 = 1675 cm^{-1}.

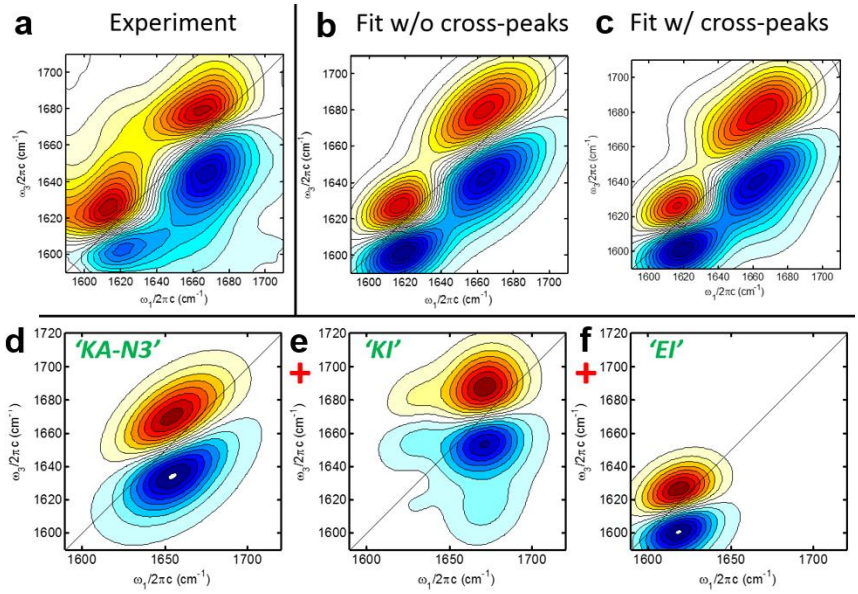

Fig. S8: (a) Carbonyl stretch region of the experimental 2D IR spectrum of KP1212 at pH* 8.9 and 80 °C. Waiting time was set to 1 ps in order to enhance the weak cross-peak features. Spectrum with 150 fs waiting time is shown in the SI. (b, c) Fits to the experimental spectrum using 2D Gaussian profiles with the absence (b) and presence (c) of cross-peaks. In both fits, a peak at 1612 cm^{-1} is fixed for the **EI** vibration shown in (f). In (b), only one peak is included to simulate the broad C=O lineshape. In (c), two different keto tautomers each with two vibrations are considered, as shown in (d) and (e).

In the second model which is guided by the DFT calculations, we considered both the **KA-N³** and **KI** tautomers. The **KA-N³** tautomer has a C=O stretch at 1664 cm^{-1} and a ring mode at 1645 cm^{-1}; however, as expected, the two modes and their cross-peaks overlap to give rise a broad peak shown in Fig. S8d. The **KI** tautomer (Fig. S8e) has a C=O stretch at 1675 cm^{-1} and a ring mode at 1630 cm^{-1}. The best fit shown in Fig. S8c is the superposition of Fig. S8d-f, with a 7:1 **KA-N³** to **KI** ratio. A better qualitative agreement between the experimental and calculated lineshapes was achieved with this model. The small **KI** population explains why the cross-peak appears as a ridge instead of a distinct peak. Although we cannot guarantee the uniqueness of the spectral decomposition, the fit presented is consistent with the DFT calculations and offers a reasonable explanation to the experimental spectrum.

6.2 Enol tautomers in water

Based on the comparison between the experimental and calculated spectra, the pronounced triplet feature in the 1500–1600 cm^{-1} region is attributed to the cis-**EI** tautomer. In order to identify the **EA** tautomer, we measured 2D IR spectra at pH* 8.9 and 4 °C (Fig. S9) where the enol tautomers are favored. We also used all-parallel polarization to enhance the diagonal peak features. In between the broad C=O peak ~1660 cm^{-1} and the ring mode of **EI** at 1621 cm^{-1}, a new diagonal feature was observed at 1638 cm^{-1} (marked as feature 1). When we increased the waiting time τ_2 to allow spectral diffusion to symmetrize the peaks, feature 1 persists and exhibits spectral diffusion, indicating that it is a distinct peak.

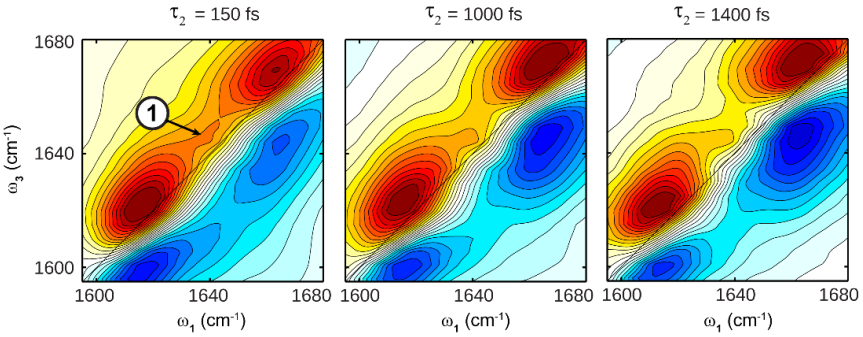

Fig. S9: 2D IR spectra at selected waiting times of KP1212 at pH* 8.9 and 4 °C.

We also observed the same feature in the pH* 7.9 spectrum at 4 °C (Fig. S10a). Feature 2 at (1638 cm^{-1}, 1529 cm^{-1}) is suggestive a cross-peak. A slice along ω_1 at ω_3 = 1529 cm^{-1} (Fig. S10b) shows a bump at 1638 cm^{-1}, which is suggestive of a cross-peak feature to the 1638 cm^{-1} diagonal peak. The peak pattern (1529 and 1638 cm^{-1}) of this new species matches well with the calculated spectrum of **EA**, which has a three peak pattern of ring vibrational modes, but with two modes overlapping with the **EI** modes at 1498 and 1538 cm^{-1},

and a ring mode at 1631 cm^{-1}, which is blue-shifted compared to that of **EI**. Because of the low population of **EA**, the diagonal peak at 1638 cm^{-1} is weak and can only be observed at lower temperatures.

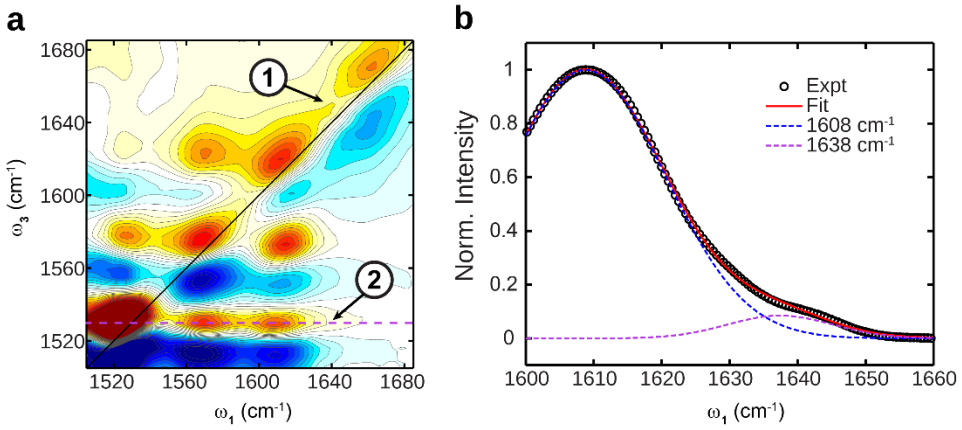

Fig. S10: (a) 2D IR spectrum at pH* = 7.9 and 4 °C. (b) ω_3 = 1529 cm^{-1} slice (purple line in (a)). The slice was fitted to the sum of two Gaussian functions shown as dash curves.

To further support our assignment, we simulated 2D IR spectra to illustrate how the weakly populated **EA** tautomer may manifest itself as the spectral features described above. 2D IR spectra for **EA** and **EI** were generated using 2D Gaussian functions and shown in Fig. S11. The relative peak positions were guided by the DFT calculations and the exact frequencies were set to match the experiment. Both **EA** and **EI** have two ring vibrations at 1529 cm^{-1} and 1572 cm^{-1}, and **EA** has a ring mode at 1638 cm^{-1} that is blueshifted compared to the **EI** ring mode at 1612 cm^{-1}. The peak intensities and linewidths are arbitrary chosen to roughly match the experiment. The **EA** and **EI** spectra were added with a 1:9 ratio, and the resulting spectrum displays similar spectral features identified in Fig. S10 (the corresponding features are marked). The ω_3 = 1529 cm^{-1} slice is plotted on top of the spectrum and it shows a weak bump at 1638 cm^{-1}.

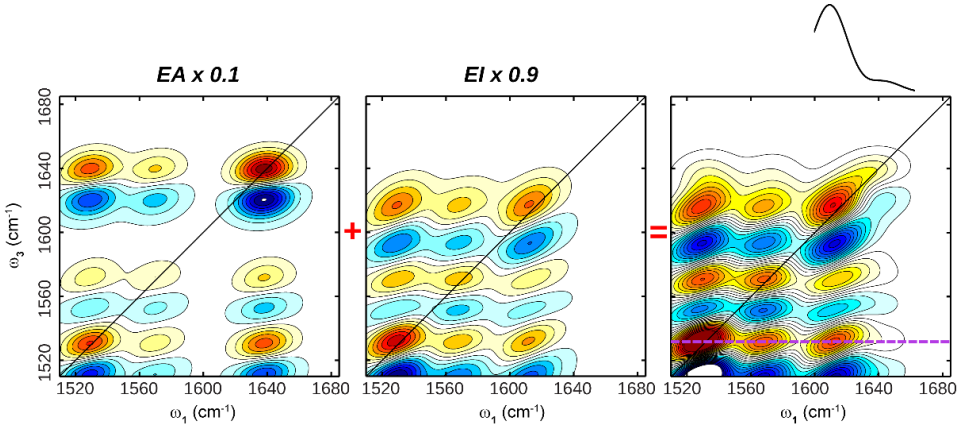

Fig. S11: The simulated 2D IR spectra using tilted 2D Gaussian functions for the **EA** (left) and **EI** (middle) tautomers, and the superposition of the two spectra (right). The ω_3 slice is plotted on top of the spectrum.

We also examine the possibility to separate *cis* and *trans* **EI** tautomers. The 2D IR lineshapes of the three **EI** ring modes are all inhomogeneously broadened, which is in sharp contrast to the homogeneous lineshapes in the CMP spectrum. Fig. S12 displays the zoom-in view of the three **EI** modes. The inhomogeneity of the 1568 and 1529 cm^{-1} modes are clearly illustrated as marked by features 3 and 4. Both features do not appear to be the result of one broadened peak because their corresponding cross-peaks do not have the same linewidths. For example, the (1566 cm^{-1}, 1625 cm^{-1}) cross-peak appears to be much narrower and aligns to the red-side of the 1568 cm^{-1} diagonal peak. Comparing to the DFT calculations, feature 3 may be assigned to the 1563 cm^{-1} *trans*-**EI** vibration that is in between the two highest frequency modes of *cis*-**EI**.

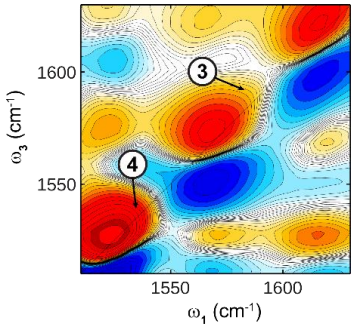

Fig. S12: 2D IR spectra at pH* 8.9 and 4 °C, with hyperbolic sine contours to enhance the weaker features.

7 Base-pairing geometries

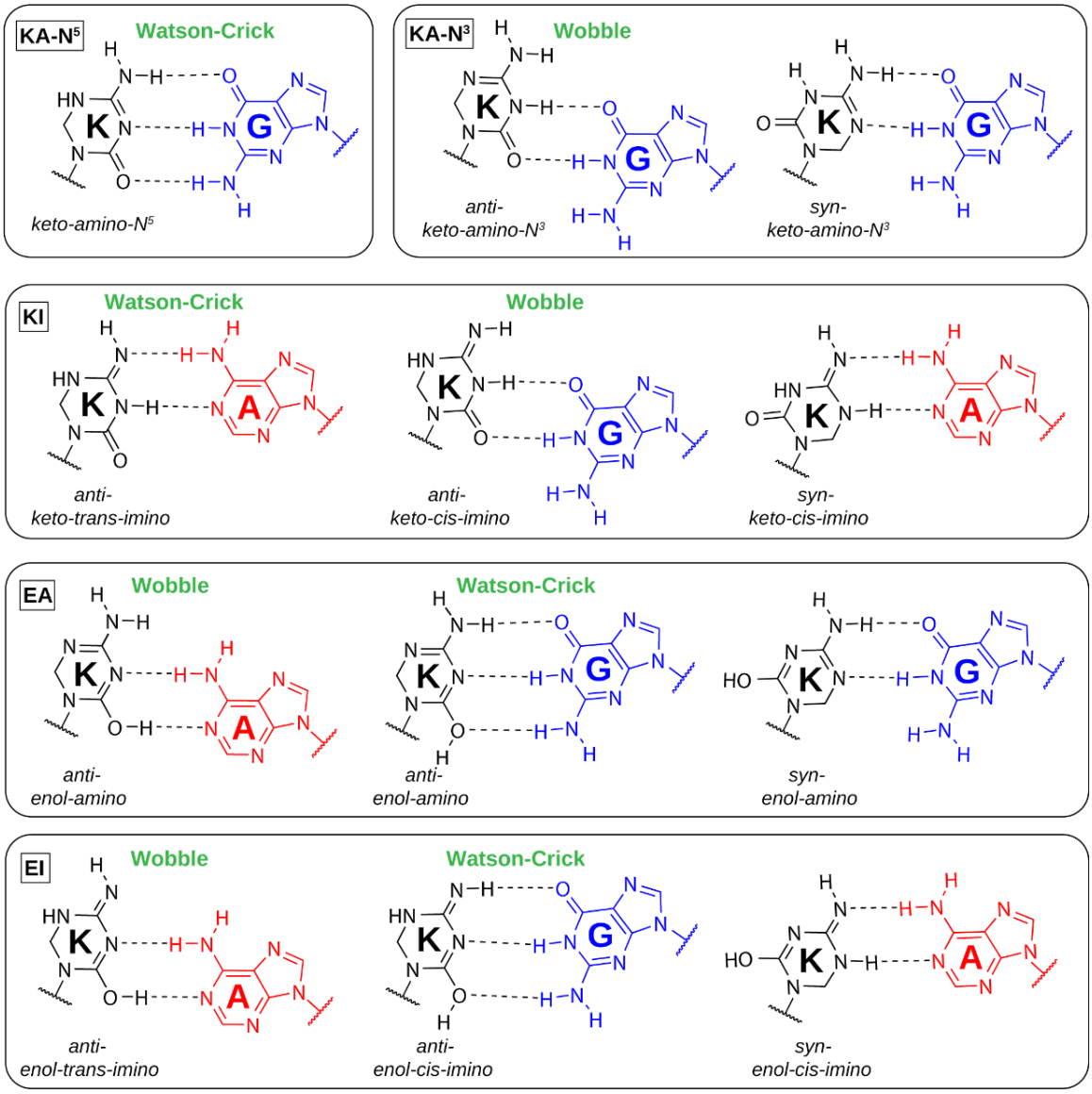

Fig. S13: Possible base-pairing geometries between the various KP1212 tautomers with either G or A. These base-pairs can adopt Watson-Crick, wobble geometries, or involve the *syn*-conformer of KP1212. The top panel shows the protonated keto-amino KP1212 and the rest show the tautomers at neutral state. The glycosidic conformation is the more common *anti*- form unless otherwise specified. For the *anti*-enol tautomers (bottom two panels), the hydroxyl proton can point towards or away from the opposing base, leading to different base-pairs. However, there may be steric hindrance with the hydroxyl proton pointing towards the deoxyribose. The two keto-amino (**KA**) tautomers can only base-pair with G, whereas the keto-imino (**KI**), enol-amino (**EA**), and enol-imino (**EI**) tautomers can base-pair with either A or G.

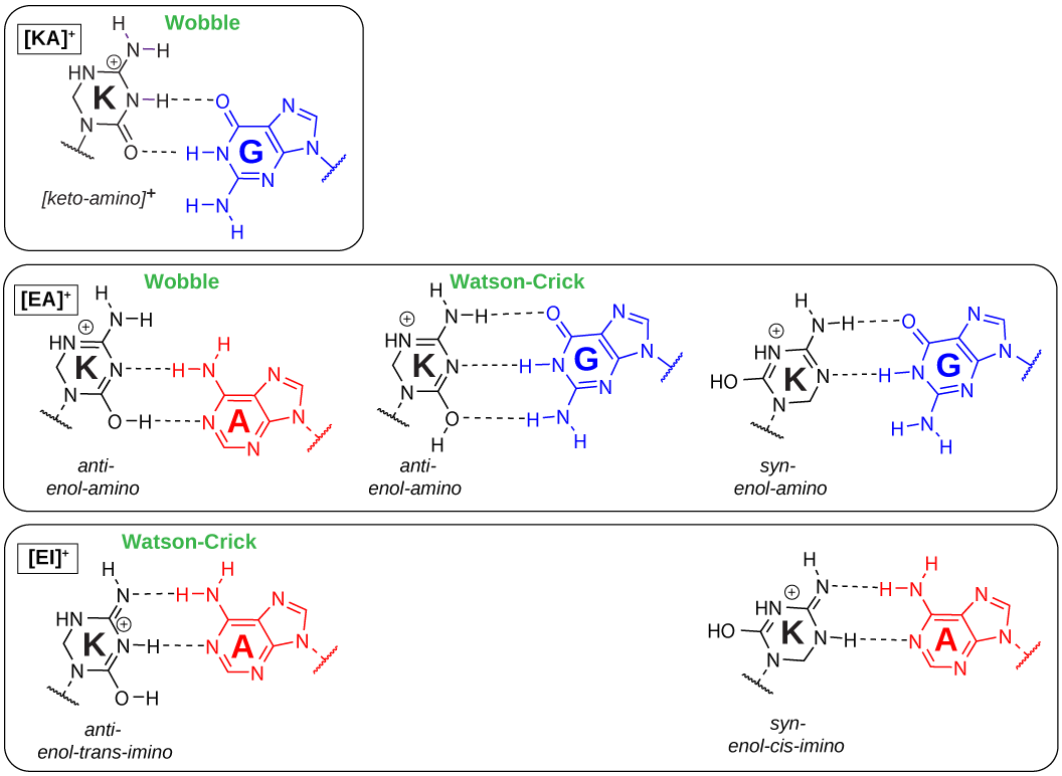

Fig. S14: Possible base-pairing geometries between the various protonated KP1212 tautomers with either G or A.

8 Temperature-jump relaxation rates

In order to obtain the time scale for the exchange rates between KP1212 tautomers, we performed temperature-jump experiments (9). The transient heterodyne disperse vibrational echo (t-HDVE) spectra of KP1212 at pH* 8.1 with temperature jumping from 27 °C to 37 °C are plotted Fig. S15a. The spectral changes in the C=O stretch region around 1660 cm^{-1} resembles the temperature-dependent FTIR spectra shown in Figure 3C and report on the tautomer exchange rates. Single-exponential fits to the time traces at ω_3 = 1659 cm^{-1} and 1685 cm^{-1} result in relaxation time constants of 21.3±6.1 ns and 25.3±16.7 ns, respectively. These values are on the same order of magnitude as those obtained for pyridone derivatives (9). It should be noted that deducing the molecular mechanism and the rate constants between the various tautomers require further studies. The convolution between protonation and tautomerization should also be considered in future experiments.

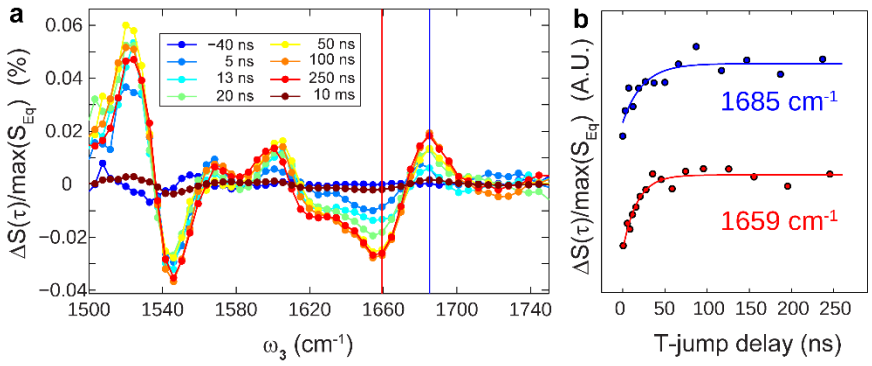

Fig. S15: (a) t-HDVE spectra of KP1212 at pH* 8.1 and T-jump from 27 °C to 37 °C. Only spectra at selected delays are shown. (b) Normalized time evolution of the T-jump difference signal at ω_3 = 1659 cm^{-1} (red) and 1685 cm^{-1} (blue). Single-exponential fits are shown in solid curves.

9 KP1212 oligomer

To test whether the DNA phosphate backbone alters the tautomeric states of KP1212, we measured the FTIR spectrum of a DNA oligomer consisting of alternating thymine and KP1212, with a sequence of 5'-TKTKTKTKTKT-3' (abbreviated as $d(TK)_5T$). We avoided a sequence of continuous KP1212 because its base stacking is not optimal due to sp^3 carbon center at the C6 position. Fig. S16 shows the FTIR spectrum of $d(TK)_5T$, along with the spectra of the KP1212 and TMP bases. It is clear that the oligomer spectrum is the superposition of the two. The low frequency modes at 1612, 1568, and 1527 cm^{-1} indicate the presence of the **EI** tautomer of KP1212 in the context of a DNA oligomer.

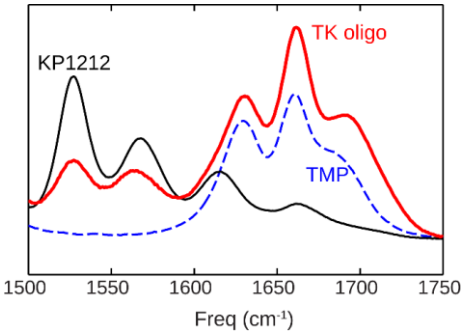

Fig. S16: FTIR of KP1212 (black), TMP (blue dotted), and $d(TK)_5T$ (red).

10 Solvent dependence

In order to examine the solvent effect on the tautomeric equilibrium of KP1212, we measured the FTIR spectrum in DMSO (Fig. S17) and found that the C=O stretch intensity increases significantly, indicating a higher population of keto tautomers.

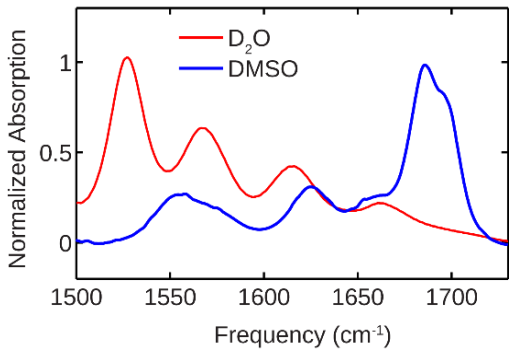

Fig. S17: Normalized FTIR spectra of KP1212 in D_2O/phosphate buffer at pH* 7.9 (red) and DMSO (blue). The increased peak intensity above 1650 cm^{-1} indicates the predominance of the keto tautomers in DMSO.

11 References

1. Chung HS, Khalil M, Smith AW, & Tokmakoff A (2007) Transient two-dimensional IR spectrometer for probing nanosecond temperature-jump kinetics. *Rev. Sci. Instrum.* 78(6):063101.
2. Jones KC, Ganim Z, Peng CS, & Tokmakoff A (2012) Transient two-dimensional spectroscopy with linear absorption corrections applied to temperature-jump two-dimensional infrared. *J. Opt. Soc. Am. B* 29(1):118-129.
3. Jones KC, Ganim Z, & Tokmakoff A (2009) Heterodyne-detected dispersed vibrational echo spectroscopy. *J. Phys. Chem. A* 113(51):14060-14066.
4. Li D*, et al.* (2014) Tautomerism provides a molecular explanation for the mutagenic properties of the anti-HIV nucleoside 5-aza-5,6-dihydro-2'-deoxycytidine. *Proc. Natl. Acad. Sci. U.S.A.* 111(32):E3252-E3259.
5. Singh V*, et al.* (2014) Mechanism of repair of acrolein- and malondialdehyde-derived exocyclic guanine adducts by the alpha-ketoglutarate/Fe(II) dioxygenase AlkB. *Chem. Res. Toxicol.* 27(9):1619-1631.
6. Delaney JC & Essigmann JM (2006) Assays for Determining Lesion Bypass Efficiency and Mutagenicity of Site-Specific DNA Lesions In Vivo. *Methods Enzymol.*, eds Judith C & Paul M (Academic Press), Vol Volume 408, pp 1-15.
7. Scheller KH, Scheller-Krattiger V, & Martin RB (1981) Equilibriums in solutions of nucleosides, 5'-nucleotides, and diethylenetriaminepalladium(2+). *J. Am. Chem. Soc.* 103(23):6833-6839.
8. Krężel A & Bal W (2004) A formula for correlating pK_a values determined in D_2O and H_2O. *J. Inorg. Biochem.* 98(1):161-166.
9. Peng CS, Baiz CR, & Tokmakoff A (2013) Direct observation of ground-state lactam–lactim tautomerization using temperature-jump transient 2D IR spectroscopy. *Proc. Natl. Acad. Sci. U.S.A.* 110(23):9243-9248.