J. Phys. Chem. A **2008,** *112,* 11449–11461

Probing Multiple Core−Hole Interactions in the Nitrogen K-Edge of DNA Base Pairs by Multidimensional Attosecond X-ray Spectroscopy. A Simulation Study

Daniel M. Healion,[†] Igor V. Schweigert,[‡] and Shaul Mukamel*,[†]

Department of Chemistry, University of California Irvine, 456 Rowland Hall, Irvine, California 92697, and Department of Chemistry, University of California Irvine, 1102 Natural Sciences II, Irvine, California 92697,

Received: April 30, 2008; Revised Manuscript Received: July 3, 2008

Two-dimensional X-ray correlation spectroscopy (2DXCS) signals of the isolated DNA bases and Watson−Crick base pairs which contain multiple absorbing nitrogen atoms are calculated. Core−hole excited states are calculated using density functional theory with the B3LYP functional and 6-311G** basis set. Sum over states calculations of the signals reveal changes in cross-peak intensities between hydrogen-bonded and stacked base pairs. Nucleobase analogues are proposed for investigating base-stacking and hydrogen-bonding interactions.

I. Introduction

X-ray absorption near-edge structure (XANES) spectroscopy locally probes the unoccupied density of states around an absorbing atom.[1−3] Ultrafast four-wave mixing, wave-vector-resolved nonlinear experiments have been proposed to extend XANES for examining core−hole couplings through valence electronic states.[4−6] This 2DXCS may become feasible with new light sources such as the fourth-generation X-ray free electron laser[7] or table-top high harmonic generation sources.[8] Two-color experiments where the applied pulses are resonant with the absorption edges of nitrogen (~400 eV) and oxygen (~500 eV) core−holes were predicted to show variations in cross-peak intensities with the distance between the absorbing cores.[9] We extend this study to systems containing multiple core atoms of the same type which can be coherently excited with the bandwidth of attosecond pulses, the nucleobases and Watson−Crick base pairs of DNA. In this paper, we extend this approach to the nucleobases of DNA, which contain several nitrogen atoms. We focus on the K-edge of nitrogen, an element found only in the nucleobases, and ignore the surrounding backbone and ionic environments. This is an analogue of homonuclear NMR, where several nuclei are excited coherently within the pulse bandwidths. Specific cross-peaks in the 2DXCS spectra are associated with interatomic core−hole couplings.

In addition to using the nucleobases as a case study for this type of 2DXCS experiment, where several core transitions are excited coherently by a broadband X-ray pulse, the technique may be used as a tool to measure the ultrafast dynamics in a DNA strand. Although the melting curve for a given sequence of DNA can be described empirically or through one of several models, the dynamics of transient hydrogen-bond breaking between paired nucleobases, which takes place on time scales comparable to the period of the hydrogen bond vibration (~13 fs) when a base-paired oligomer is heated almost to its melting point, is currently a topic of great interest.[10] We show that some features of 2DXCS for a basepair in vacuo can be used to assign the relative orientation of the bases. The nearest experimental realization of our model would be a gas phase experiment on single base pairs. If double-stranded B-DNA has localized excited states, it might be possible to describe its 2DXCS as a sum of contributions similar to those calculated in this paper. Calculating the response of a nonlocal valence band to a suddenly created core−hole requires a high level of electronic structure theory.

2DXCS employs sequences of femtosecond to attosecond pulses with central frequencies tuned to different X-ray transitions in order to probe the spatially resolved response of a system by exploiting the elemental specificity of core X-ray resonances.[4,9] Formal expressions for the all-X-ray four-wave mixing response and its variation with the delays between pulses have been derived.[4,9,11] Coherences in the 2DXCS spectra corresponding to absorption from an excited state and couplings between excited states probed with ultrafast X-ray transitions in the ortho-, meta-, and para-isomers of aminophenal were simulated.[9] 2DXCS is an analogue of multidimensional NMR, visible, and infrared techniques.[12] By spreading different components of the molecular response onto different frequency axes, photon echo experiments (the $\mathbf{k}_I$ technique described in this paper) have been used to extract anharmonicities and characterize site-energy fluctuations in coupled vibrational systems. During the time interval between the second and third pulses the molecules which contribute to the response will be in a population, either on the ground or an excited state. Varying this delay measures population transport caused by the coupling between the system and its environment. The material response can be related to a distribution of coupled chromophores in space. In 2DXCS, the chromophores are the core orbitals absorbing atoms, which share a common valence band. The core electron injected into this shared band complicates the interpretation of core excitation pathways in terms of coupled excitons. In photoemission spectroscopy in metals, the excited electron is removed from the system and can be neglected; the system is naturally partitioned into a core and valence part:[13] the valence electrons act as a bath, coupled to a core−hole excitation. For molecular systems with core electrons excited to bound states (as in XANES) this partitioning is not that obvious.

The first XANES study of the nucleobases and polynucleotides of DNA which treated different combinations of singly

---

* Corresponding author: e-mail smukamel@uci.edu; ph (949) 824-6164.

[†] University of California Irvine.

[‡] US Naval Research Laboratory.

10.1021/jp803824a CCC: $40.75

and doubly stranded DNA of various base compositions was reported by Kirtley et al.[14] They considered the nitrogen K-edge spectra for the nucleobase monophosphates, single strands of homopolymer DNA, and double-stranded samples. Later, a comphrensive study by MacNaughton et al.[15] compared theoretical predictions to experimental results for the C, O, and N K-edges of the nucleobases. Energy-loss peaks in the X-ray photoelectron spectra have been used to assign a peak shift in the nitrogen XANES caused by a glycosidic bond to the bases.[16] In another study differences in the X-ray emission spectroscopy (XES) and XANES signals from different samples of DNA prepared as a dried complex on a substrate were assigned to a variable bandgap for different spots on the surface and counterions bound to the sample.[17]

The photon-echo technique has been used in electronic and vibrational spectroscopies in the visible and the infrared[12] to characterize the distribution of spectral frequencies that contribute to broad peaks in linear absorption. For a general model of system−bath coupling, each transition has an intrinsic homogeneous and inhomogeneous line width, averaging over frequencies varying with the environment of each chromophore. Diagonal peak line shapes carry information on the width of the inhomogeneous distribution of each transition, and cross-peak line shapes reveal correlations between the fluctuations of separate transitions, averaged over all chromophores in the sample. An extension of this technique to the 2DXCS of a bulk sample of double-stranded DNA might be able to characterize the fluctuating electronic states of the DNA nucleobases in solid or solvent environments in a single experiment. Characterizing the 2DXCS of individual nucleobases and base pairs is a necessary precursor to addressing large polymer systems.

Extending nonlinear techniques to the X-ray regime poses major challenges. Attosecond pulses will be required to compete with a core−hole lifetime of a few femtoseconds. Intensity and coherence requirements impose demanding experimental technical challenges.[5] New X-ray sources should make high intensities achievable. High-level theoretical methods exist for calculating the X-ray response to all orders,[18,19] but their computational cost in larger systems requires making some approximations.

II. Methods

A. Level of Theory. 2DXCS is calculated using the same level of theory used for XANES. Doubly core excited states probed during multiple X-ray transitions are modeled by approximating the doubly excited block of the electronic Hamiltonian using its diagonal elements. XANES is calculated for the nucleobases and base pairs and compared with prior experiments and simulations.[15]

The nitrogen atoms in each nucleobase can be classified into two types: double-bonded imine and the trivalent amine.[16] A previous restricted active space self-consistent field (RASSCF) study of adenine reported core-level amine nitrogen orbital energies lower than the less-shielded imine nitrogen by up to 2.0 eV.[16,20] In calculations involving only one absorbing atom, the absorption edge can be matched to experiment or calculated using relativistic atomic codes to a high degree of accuracy for low-*Z* atoms.[21] In the present study of multiple absorbing atoms, with chemical shifts arising from valence electronic differences, the contribution from each core was shifted based on the depth of the corresponding core−hole. Expressions for the response are based on energy differences between manifolds of valence-excited states with differently core-ionized atoms. Electronic structure methods directly calculate energy differences between states with the same core occupation. For states with different core occupations, the relative ionization energies were approximated using the core orbital energies in the ground state, with an overall shift to reproduce the experimental core edge at 400.7 eV.

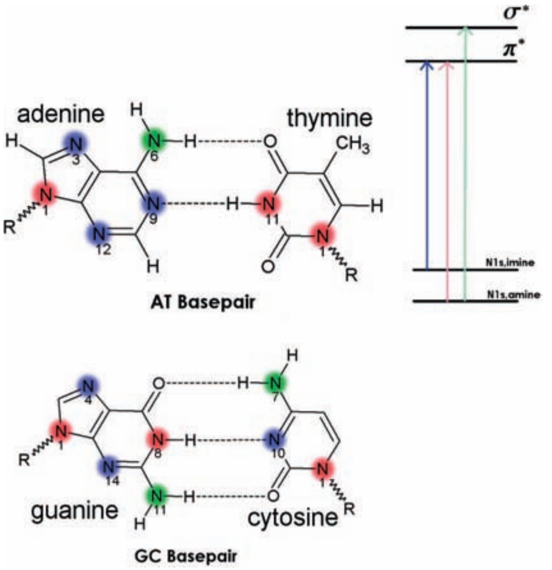


Figure 1. Structures of the Watson−Crick base pairs and a general level scheme for a nucleobase containing imine and amine nitrogens. Nitrogens are colored as imine (blue), amine (red), or $-NH_2$ (green) type. The R group representing the DNA backbone was replaced with a methyl in the ECA simulation.

Transition energies were calculated as differences between the relevant orbital energies for each transition. The contribution of each core to the spectrum was blue-shifted to fit the lowest energy core-to-LUMO transition to an excitation energy of 400.7 eV for the π_i^* transition, except for thymine, where the π_a^* was set to 401.705 eV to match experiment.[15]

We used an equivalent core approximation (ECA) to represent the core−hole valence-electron scattering term in the Mahan−Noziéres−De Dominici (MND) Hamiltonian for different core occupations by performing ground-state SCF calculations on a system with the absorbing core's atomic number incremented by one. Excited states are modeled as all possible singly and doubly excited Slater determinants from these reference states by neglecting quartic terms between valence electronic states.[6] This method will be denoted the equivalent core approximation (ECA)/determinant method. Core−hole lifetime broadenings of 0.4 eV were added to match the experimental XANES spectra of the nucleobases and previous calculations.[15] The expression for the response relies on an expansion of the dipole and the evolution operators in valence many-electron eigenstates of the Hamiltonian.[4] The ECA assumes this Hamiltonian to be diagonal in a singly and doubly substituted Slater determinant basis. We have calculated the nitrogen N 1s XANES spectra of the nucleobases (Figure 1) using the ECA method and compared with other computations and experiment (Figure 2). Electronic structure calculations were performed at the equilibrium geometry using unrestricted DFT, with the UB3LYP functional and the 6-311G** basis set. The orbitals and orbital energies of the equivalent-core states were computed with the Gaussian03 code.[22] Transitions to the continuum were not included. This approximation reproduces many qualitative features of the XANES signal, such as the splitting and intensity difference between the π_i^* and amine peaks, but systematically overestimates the splitting between the π_a^* and σ_a^* peaks (Figure 2).

More rigorous methods for computing the excited states have their own limitations. TDDFT includes excited-state electronic

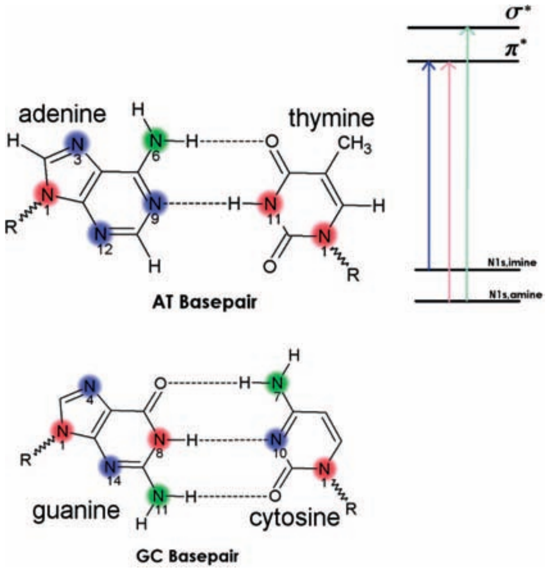


Figure 2. XANES spectra of individual bases: experiment (black),[15] IVO simulations (green)[15] and DFTTP simulations (blue),[15] the present ECA calculations (red). Peak assignments are shown for the adenine experimental spectra and a high-energy $-NH_2$ peak for the calculated spectra of thymine and guanine. Transitions correspond to the level scheme shown in Figure 1. The ECA/determinant technique used to calculated the 2DXCS signal has comparable cost and accuracy to the other techniques, but can be extended to describe doubly core excited states.

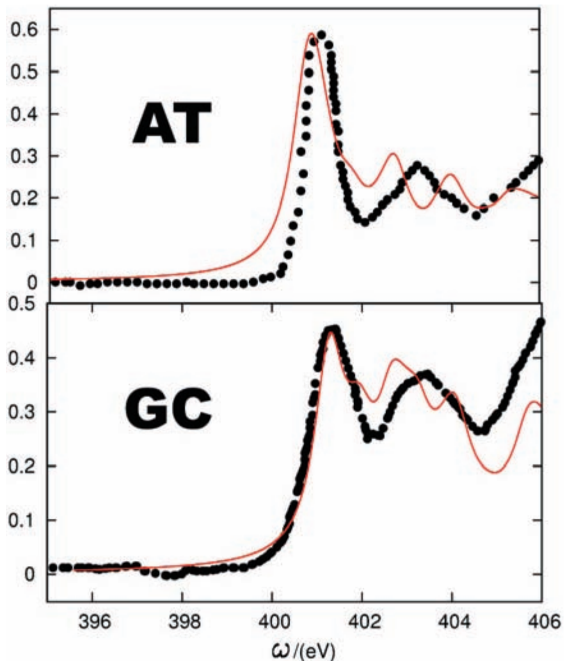


Figure 3. XANES of the Watson−Crick base pairs adenine−thymine (AT) and guanine−cytosine (GC) calculated using the ECA/determinant method (red line) and compared with experiment (black dots).[14] The calculated GC spectra were shifted by +0.57 eV for better comparison with experiment.

exchange and correlation but does not describe the doubly excited states required in nonlinear electronic spectroscopies, whether X-ray or optical.[6] The Bethe−Salpeter equation provides a perturbative method for deriving the time dependence of the density matrix[18] but is computationally intractable for all but the smallest systems. The polarization propogator technique[19] has recently been extended to the X-ray energy range. This approach has already been applied to calculate the XANES spectra of the adenine−thymine and guanine−cytosine base pairs and should be very useful for calculating the nonlinear response for mid-sized to large molecules.[23]

B. Comparison with Prior Results. MacNaughton et al.[15] have presented an exhaustive experimental and theoretical review of the nucleobase carbon, nitrogen, and oxygen XAS and XES. Two computational methods from this study were compared to our calculations. These comparisons are reported in the Supporting Information. The first is an ab initio SCF, improved virtual orbital (IVO) technique, implemented in the GSCF3 code from Kosugi.[24] Starting from the ground-state molecular orbitals and Hamiltonian, a reduced subspace of orbitals is allowed to relax within the core−hole potential. The transition dipoles are then written as one-electron matrix elements between the ground core orbital and the relaxed virtual orbitals. The second is the DFT transition potential method (DFTTP) implemented in the SToBE code.[25] It includes the partial relaxation of final electronic states in the field of the core−hole by constraining the occupation numbers used to form the charge density from the Kohn−Sham orbitals to have a partial occupation of the lowest energy orbital.[26,27] Neither approach describes the nitrogen K-edge of the nucleobases accurately.[15] Both predict two low-energy peaks for thymine, instead of a single, and do not reproduce the splitting between the low-energy peaks of cytosine and guanine (see Figure 2). Base pair calculations with the ECA/determinant method shown in Figure 3 agree very well with experimental results,[14] accurately reproducing the intensity ratios between the imine and amine peaks and the mean position of the π_a^* and σ_a^* peaks. The overestimated splitting between these two peaks is also observed in the base pair spectra.

The lowest energy peaks in the N 1s absorption spectrum can be assigned to the N 1s → $\pi_{i,a}^*$ transitions for imine and amine nitrogens and N 1s → σ_a^* transitions for amine nitrogens only (see Figure 1). Our calculations reproduce the main features of the experimental absorption but show some systematic differences as described previously.[15] Adenine has three imine nitrogens to guanines two, explaining the latter's more intense broad peak at ~403 eV. Thymine has no imine nitrogens and therefore completely lacks the sharp imine peak at ~ 400.7 eV. All three simulations overestimate the splittings between the imine and amine peaks (e.g., guanine and cytosine) and between the π_a^* and σ_a^* contributions from the amine nitrogens. The splitting between the π_a^* and σ_a^* peaks is underestimated in the simulations. The error is most pronounced in thymine, where the sum-over-states calculations predict three peaks corresponding to the π_a^* and σ_a^* peaks on inequivalent nitrogens which is not observed experimentally, being less than the 0.4 eV lifetime broadening of the N 1s core−hole. The same trend of overestimated amine peak energy differences is repeated by the DFTTP and IVO methods.

The splitting between the sharp, initial π_i^* peak and the average of the amine π_a^* and σ_a^* peak positions is also overestimated by the ECA. The discrepancy is largest for cytosine, where the splitting exceeds the experimental splitting. The other two levels of theory have similar systematic discrepancies. These could be caused by one of several approximations. Our systems are the individual nucleobases in vacuo, unlike the powdered solids used in the experiment.[15] Previously reported crystal structures of thymine,[28] guanine,[29] and cytosine[30] include substantial hydrogen-bonding and π-stacking interactions between bases which are neglected in our calculations. The orbital method qualitatively reproduces the observed XAS structure with systematic errors comparable to other approaches.

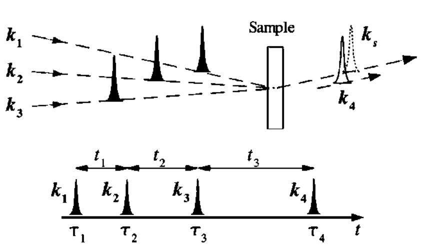


Figure 4. Pulse configurations and time delays for a general 2DXCS experiment with heterodyne detection.

III. Results

The 2DXCS in an isotropic, randomly oriented sample was calculated using the ECA described in section II. The 2DXCS of the nucleobases contain a wealth of information. Some characteristics common to all systems are addressed first, followed by an analysis of hydrogen-bonding and base-stacking signatures for each pair. We conclude by investigating base-stacking interactions in a model system.

A. Theory of 2DXCS. In 2DXCS three pulses with wavevectors $\mathbf{k}_1$, $\mathbf{k}_2$, and $\mathbf{k}_3$ are applied, and the signal is measured in a fourth direction ($\mathbf{k}_s = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$) as a function of the delays between the pulses (Figure 4).[4,9] The signal $P_{\mathbf{k}_I}(\Omega_1, t_2, \Omega_3)$ is displayed vs the frequencies Ω_1 and Ω_3, the Fourier conjugates to the delay between the first and second pulses (t_1), and the last pulse and the time of detection (t_3). It is varied parametrically with t_2, the delay between the second and third pulse. Couplings between core–hole excited states can be directly observed by monitoring cross-peaks in the signal induced by the pulses.[12] We used the expressions for the 2DXCS signal ($P_{\mathbf{k}_I}(\Omega_1, t_2, \Omega_3)$), derived in ref 4 and applied earlier to the nitrogen and oxygen aminophenol K-edges[9] but extended to treat multiple coherently excited core shells.

The response function for a series of well-separated pulses can be written as a sum of terms where the system interacts with three photons with frequencies chosen from the pulse envelopes in the frequency domain.[31] The wavevector of the signal generated by the induced polarization must lie in specific directions satisfying the conservation of field momentum, $\mathbf{k}_I = -\mathbf{k}_1 + \mathbf{k}_2 + \mathbf{k}_3$. The response function can be written as a sum of three terms

$$S_I^{(3)}(\Omega_1, t_2, \Omega_3) = S_I^{\mathrm{GSB}}(\Omega_1, t_2, \Omega_3) + S_I^{\mathrm{ESE}}(\Omega_1, t_2, \Omega_3) + S_I^{\mathrm{ESA}}(\Omega_1, t_2, \Omega_3) \quad (1)$$

The double-sided Feynman diagrams corresponding to these terms are shown in Figure 5. The excited-state absorption (ESA) pathways in the figure were expanded to show contributions from separate ordering of core excitation. The other pathways represent excited-state emission (ESE) and ground-state bleaching (GSB). The electric field is decomposed into positive and negative frequency parts:

$$E(\mathbf{r}, t) = \boldsymbol{E}^{+}(t) \exp[-i\mathbf{k}_p \cdot \mathbf{r} + i\omega_p t] + \boldsymbol{E}^{-}(t) \exp[i\mathbf{k}_p \cdot \mathbf{r} - i\omega_p t] \quad (2)$$

The temporal envelopes $\mathbf{E}^{\pm}(t)$ have Fourier transforms

$$\boldsymbol{E}^{\pm}(\omega) = \int_{-\infty}^{\infty} \mathrm{d}t\, \mathrm{e}^{-i\omega t} \boldsymbol{E}^{\pm}(t) \quad (3)$$

All pulses have the same central frequency ωp (405.7 eV) and square pulse envelopes with a 10 eV bandwidth

$$\boldsymbol{E}^{\pm}(\omega) = \theta(\omega + 5\ \mathrm{eV})\theta(5\ \mathrm{eV} - \omega) \quad (4)$$

setting an energy cutoff for the active space of excitations. Denoting the ground, singly, and doubly core excited states as g, e_i, and f, respectively,[9] we get

$$S_I^{\mathrm{GSB}}(\Omega_1, t_2, \Omega_3) = \sum_{e_1,e_2,g} \frac{\boldsymbol{E}^{-}(\omega_p - \omega_{e_1g_0})\boldsymbol{E}^{+}(\omega_{e_1g} - \omega_p)\boldsymbol{E}^{+}(\omega_{e_2g} - \omega_p)\boldsymbol{E}^{-}(\omega_p - \omega_{e_2g})}{(\Omega_1 - \omega_{g_0e_1} + i\Gamma_{g_0e_1})(\Omega_3 - \omega_{e_2g} + i\Gamma_{e_2g})} \times V_{g_0e_1}V_{e_1g}V_{e_2g_0}V_{ge_2} \exp[-i\omega_{g_0g}t_2 - \Gamma_{g_0g}t_2]$$

$$S_I^{\mathrm{SE}}(\Omega_1, t_2, \Omega_3) = \sum_{e_1,e_2,g} \frac{\boldsymbol{E}^{-}(\omega_p - \omega_{e_1g_0})\boldsymbol{E}^{+}(\omega_{e_2g_0} - \omega_p)\boldsymbol{E}^{+}(\omega_{e_1g} - \omega_p)\boldsymbol{E}^{-}(\omega_p - \omega_{e_2g})}{(\Omega_1 - \omega_{g_0e_1} + i\Gamma_{g_0e_1})(\Omega_3 - \omega_{e_2g} + i\Gamma_{e_2g})} \times V_{g_0e_1}V_{e_2g_0}V_{e_1g}V_{ge_2} \exp[-i\omega_{e_2e_1}t_2 - \Gamma_{e_2e_1}t_2]$$

$$S_I^{\mathrm{ESA}}(\Omega_1, t_2, \Omega_3) = \sum_{e_1,e_2,f} \frac{-\boldsymbol{E}^{-}(\omega_p - \omega_{e_1g_0})\boldsymbol{E}^{+}(\omega_{e_2g_0} - \omega_p)\boldsymbol{E}^{+}(\omega_{fe_2} - \omega_p)\boldsymbol{E}^{-}(\omega_p - \omega_{fe_1})}{(\Omega_1 - \omega_{g_0e_1} + i\Gamma_{g_0e_1})(\Omega_3 - \omega_{fe_1} + i\Gamma_{fe_1})} \times V_{g_0e_1}V_{e_2g_0}V_{fe_2}V_{e_1f} \exp[-i\omega_{e_2e_1}t_2 - \Gamma_{e_2e_1}t_2] \quad (5)$$

$t_{1,2,3}$ are the delay times between the system interactions with the field and for short pulses can be experimentally controlled by changing the delay between pulses. Ω_1 and Ω_3 are the Fourier conjugates of t_1 and t_3, respectively. V_{ab} is the dipole operator between the states a and b. All frequencies and dipole matrix elements are calculated using the level of theory described in section II.A. The excited states e_1, e_2, and f include many competing pathways over core edges resonant with the central frequency of the applied pulse.

In the current study the dephasing rate Γ_{ab} is taken to be the same (0.4 eV) for all coherences. In low-Z atoms this dephasing rate is dominated by Auger processes. In all calculations we took $t_2 = 0$. In this limit the four Liouville space pathways of Figure 5 collapse to two ($a = b$ and $c = d$). The $\mathbf{k}_I$ spectra of the nucleobases are laid out on a grid built from the peaks describing the transitions in Figures 1 and 2. Bases containing an $-NH_2$ group may have a peak at 405 eV. In the following discussion, "amine" refers to trivalent nitrogens bound to at least two carbons; the $-NH_2$ group will be stated explicitly.

Earlier studies on aminophenol[9] addressed systems containing two absorbing atoms of different elements and applied pulses tuned to these frequencies. In the current calculations all applied pulses are tuned to the same central frequency (405.7 eV), the nucleobases contain multiple absorbing atoms, and the response is a coherent sum of many terms for each pair of core–hole atoms. $\sum_{nm}$ in eq 5 must lie within the pulse bandwidths. First, some of the common characteristics of 2DXCS will be demonstrated, the dependence of the 2DXCS on specific Liouville space pathways, and the polarization dependence of the response.

The contributions of individual Liouville space pathways (see Figure 5) to the XXXX 2DXCS of adenine are displayed in Figure 6. All pulses are assumed parallel polarized along X. In the ESA contribution (bottom row) higher energy peaks both above the ground state π_i^* resonance (reflecting a blue-shifting of the possible Ω_3 imine nitrogen transitions due to either Fermi repulsion with the Ω_1 excited core electron to low energy virtual orbitals or a change in the valence band density of states due to the creation of the Ω_1 core–hole)

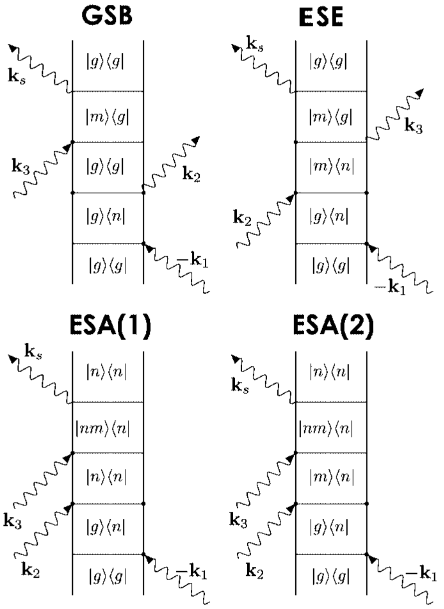


Figure 5. Double-sided Feynman diagrams contributing to $P_{\mathbf{k}_I}(t_3, t_2, t_1)$. $|g\rangle$, $|n\rangle$, and $|nm\rangle$ represent valence electronic states with no core–hole, with a core–hole on the nth atom, and with two core–holes on the nth and mth atom, respectively. Indices labeling valence excitations are omitted for clarity. The pulses are assumed to have a small bandwidth to forbid double core ionization at the same atom (i.e., $|nn\rangle$). $S_1^{nm}(t_3, t_2, t_1)$ represents a ground-state bleaching (GSB) contribution to the response. $R_{\mathrm{ESE}}^{nm}(t_3, t_2, t_1)$ is referred to as the excited-state emission (ESE) term. The last two pathways correspond to excited-state absorption (ESA1,2).

and below (possibly representing a lowering of the valence band core edge due to the contact potential of the first core–hole). However, the intensity of the ESA response is an order of magnitude lower than the GSB contribution. As the challenging nature of the proposed experiments may make measuring small intensity differences impossible, these features of the 2DXCS were not further used in this study. The full pathway decomposition for all pulse configurations is not presented in this paper; however, an interesting consequence of XXYY, XYYX, or XYXY experiments is that they remove the equivalence of the GSB(ESA1) and SE(ESA2) terms for $t_2 = 0$ (see Figure 5).

The applied electric fields and the induced polarization are vector quantities, providing a wide range of possible experimental control parameters. Expressions for the polarization-dependent response have been used to calculate the nonlinear spectroscopy of vibrational excitons in proteins.[32,33] Three independent pulse polarization configurations, XXYY, XYXY, and XYYX, contribute to the response in isotropic samples. Each polarization can highlight certain Liouville pathways. For instance, the XXYY signal is proportional to

$$\varepsilon_{xxyy} \propto 4(\vec{\mu}_1 \cdot \vec{\mu}_2)(\vec{\mu}_3 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_3)(\vec{\mu}_2 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_4)(\vec{\mu}_2 \cdot \vec{\mu}_3) \quad (6)$$

where $\vec{\mu}_n$ are the transition dipoles for the nth interaction with the field. All of the GSB terms (Figure 5) contribute to the first term in eq 6 because the first and second and the third and fourth interactions with the field are constrained to be with the same

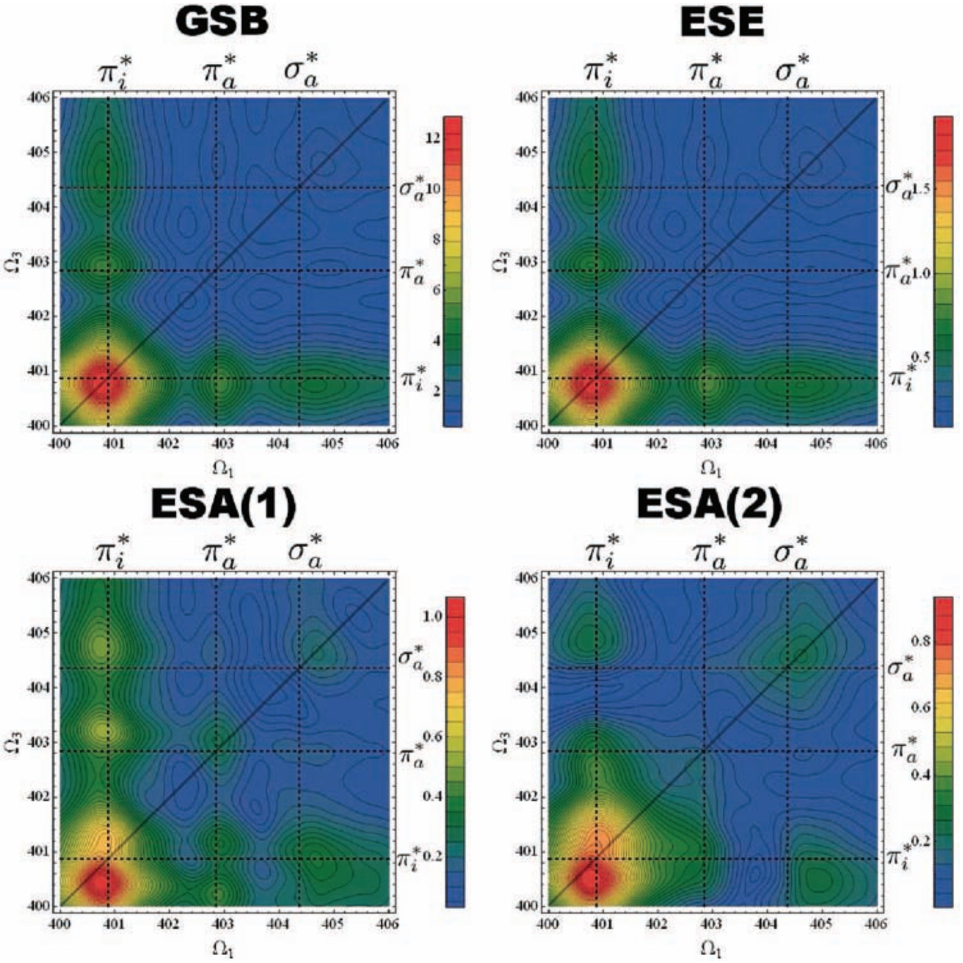


Figure 6. Liouville space pathway decomposition (see Figure 5) for the adenine $|P_{\mathbf{k}_I}(\Omega_1, t_2 = 0, \Omega_3)|$ signal for the XXXX pulse configuration.

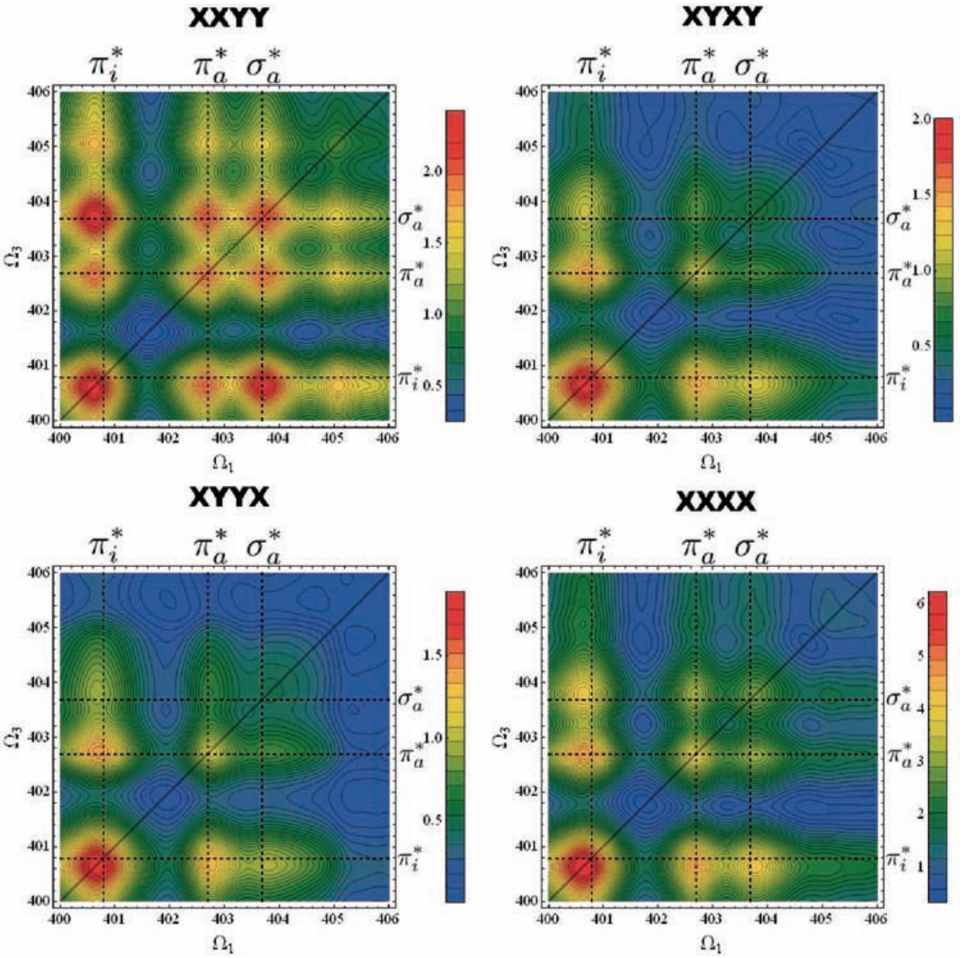


Figure 7. $S(-\Omega_1, t_2 = 0, \Omega_3)$ signal (absolute value) for guanine calculated for the three independent polarization tensor components. The XXXX signal is the sum other three. All peaks in the XANES (Figure 8) contribute to the diagonal peaks and cross-peaks in this response. This result is typical for the nucleobases. The XXYY polarization resembles a product of the linear absorption in Ω_1 and Ω_3, which the other two polarizations decrease the intensity of the diagonal π_i^* peak and higher energy amine nitrogen peaks.

state. The XXYY signal therefore should have a strong GSB character, resembling the product of the XANES along the Ω_1 and Ω_3 axes. The 2DXCS for each of the four polarizations is shown in Figure 7. The guanine XXYY signal resembles the product of the XANES spectra on $-\Omega_1$ and Ω_3, more than the other two polarizations. For this polarization, every pair of peaks in the XANES contributes a diagonal peak or cross-peak to the 2DXCS spectra. For the other polarizations, higher energy amine peaks are lower in intensity relative to the π_i^* peak. The XXXX signal is the sum of the three independent configurations.

The tensor components described above give the entire nonlinear response within the dipole approximation. Going beyond the dipole approximation to first order in $\mathbf{k}\cdot\mathbf{r}$ other signals are possible (XXXY, XXYX, and XYXX). The first wave vector dependent contributions to the measured XXXX signal are second order. The wavelength corresponding to the nitrogen K-edge is ~30 Å; the maximum distance between any two nitrogens in the guanine−cytosine base pair is ~9 Å. On this length scale, a linear approximation to nondipole effects in the 2DXCS may be sufficient. Further study will be needed to clarify the limits of this approximation.

B. 2DXCS of the Nucleobases. We now turn to measurable signatures of base orientations in 2DXCS of the nucleobase pairs. We first examine the 2DXCS of the individual nucleobases. The $|P_{\mathbf{k}_I}(\Omega_1, t_2, \Omega_3)|$ signal for the all parallel XXXX polarization direction for each of the nucleobases and the basepairs is shown in Figure 9. A previous study had pointed out the similarity of the 2DXCS to a product of the Ω_1 and Ω_3 XANES in the same energy range.[5] A diagonal $-\Omega_1 = \Omega_3$ slice through the 2DXCS should resemble the square of the linear absorption. The intensity of peaks and cross-peaks can be estimated by multiplying the corresponding peaks in the linear absorption spectra. As a consequence, the maximum peak intensity of the adenine spectra is ~10 times that of thymine. If the purines and pyrimidines were plotted on the same z-axis, the latter would not be visible. The frames were independently scaled for clarity.

The GC base pair 2DXCS contains a diagonal contribution that deviates significantly from the square of its XANES spectra (compare the π_a^* peak intensity for GC in Figure 9 to the XANES peak in Figure 8). Peak positions are unchanged, but for the diagonal slice π_a^* is 58% the intensity of the main π_i^* peak, whereas in the square of the XANES, this ratio is 79% (see Figure 10). The XANES is

$$S^{(1)}(\omega) = \sum_{e_1} |V_{e_1 g_0}|^2 \delta(\omega - \omega_{ge_1}) \qquad (7)$$

By comparing the square of this expression to the 2DXCS shown in eq 5, we see that terms containing double excitations are absent. This is probably the source of the difference in Figure 10.

Both adenine and guanine, shown in Figure 9, display strong cross-peaks between the imine nitrogens and the $-NH_2$ nitrogen.

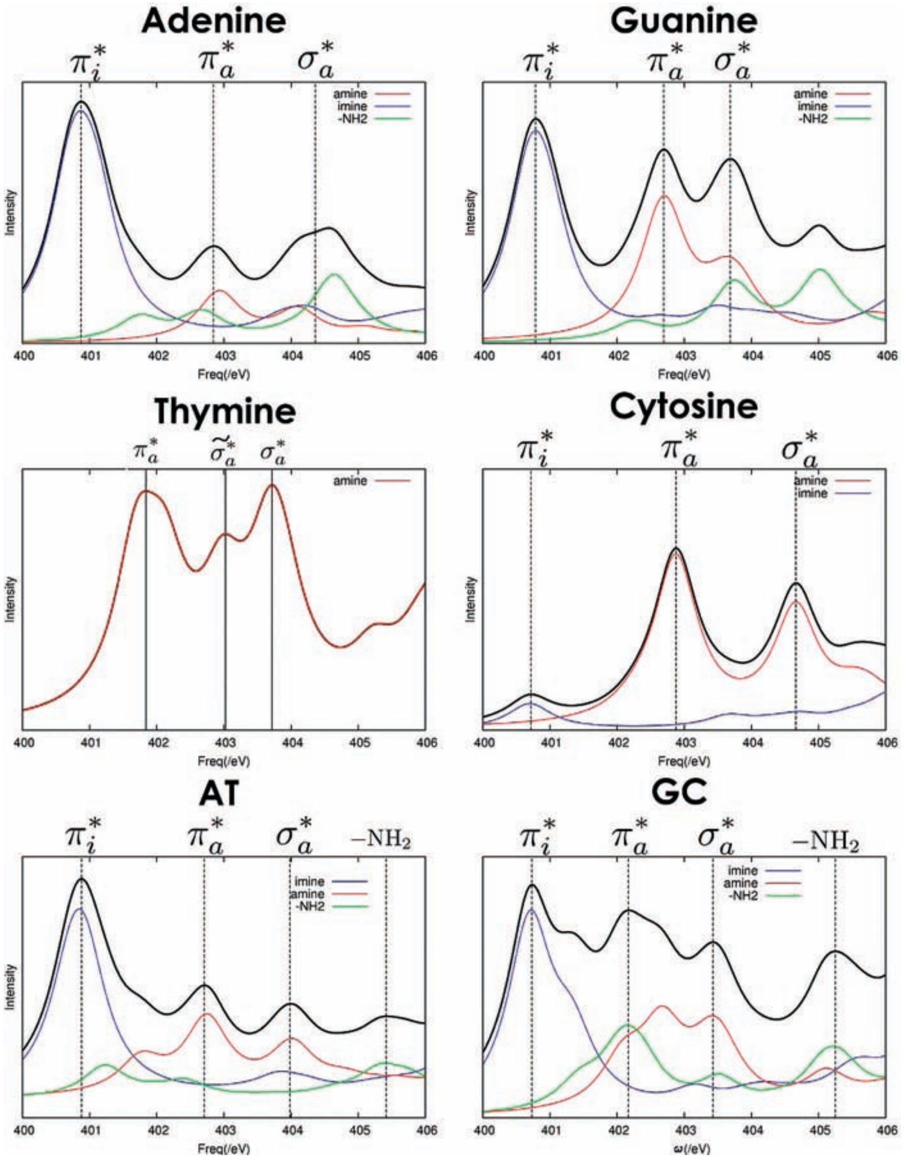


Figure 8. XANES for the nucleobases and Watson−Crick base pairs from Figures 2 and 3, decomposed into contributions from imine (blue), amine (red), and $-NH_2$ (green) nitrogens, as shown in Figure 1. The vertical lines at the peak maxima are a visual aid.

The thymine XANES forms a pattern of three peaks, which does not correspond to the peak assignments given in Figure 2. In other bases, the glycosidic nitrogens contribute some small intensity to a peak around 405 eV. In thymine this contribution is apparently blue-shifted into the same frequency range as the π_a^* and σ_a^* peaks. The cross-peaks mirror the XANES peak positions, but the strongest peaks are slightly red-shifted, indicating that the main contribution to these peaks is from the N11 nitrogen between the two carbonyl groups. The cytosine spectra are dominated by the contribution of the glycosidic nitrogen (N1) (see Figure 8). The accuracy of the XANES calculations for this system were discussed in section II. The imine nitrogen between the $-NH_2$ group and the carbonyl contributes an uncharacteristically small π_i^* peak which couples with the strong π_a^* peak from the N1 nitrogen. The weak $-NH_2$ peak at ~405.5 eV is also visible as a cross-peak with the main π_i^* peak. The 2DXCS strongly resembles the product of the linear absorption

$$S^0(\Omega_1, \Omega_3) = I(-\Omega_1)I(\Omega_3) \tag{8}$$

where $I(\Omega)$ is the XANES. A detailed comparison of the base-paired and separate bases 2DXCS with this linear absorption product is required to identify signatures of their interaction.

C. 2DXCS of Hydrogen-Bonded Base Pairs. The Watson−Crick base pair geometry was optimized using the UB3LYP functional and the 6-21G basis set. The $\mathbf{k}_I$ signal for the AT pair in Figure 9 is very similar to that of adenine, whose five nitrogens tend to overlay thymine's two. The ~402.6 eV peak is clearly assignable to amine-type nitrogens. The cross-peak between the weak $-NH_2$ peak at ~405.5 eV and the strong π_i^* peak is as intense as in the adenine spectra. The GC signal shows structure in its π_a^* and σ_a^* peaks, indicating a sum of contributions from both bases, with a strong (π_i^*, $-NH_2$) cross-peak.

To explore whether the hydrogen-bound base pairs resembled a sum of their monomer contributions, we have calculated the difference between the absolute value of the $\mathbf{k}_I$ spectra for the base pair and the sum of the $\mathbf{k}_I$ spectra of its consituents (these are shown in the top row of Figure 11)

$$S_{\text{diff}}(-\Omega_1, \Omega_3) = |P_{\text{AT}}^{\mathbf{k}_I}(-\Omega_1, \Omega_3)| - |P_{\text{A}}^{\mathbf{k}_I}(-\Omega_1, \Omega_3)| - |P_{\text{T}}^{\mathbf{k}_I}(-\Omega_1, \Omega_3)| \tag{9}$$

where $P_i^{\mathbf{k}_I}(-\Omega_1, \Omega_3)$ represents the complex $\mathbf{k}_\text{I}$ 2DXCS for system i. For the AT base pair, where all the imine nitrogen atoms are located on one base pair, hydrogen bonding leads to an increase in cross-peak intensity between the main π_i^* peak and the amine peaks, at the expense of the diagonal π_i^* peak. The GC base pair has a mix of imine and amine nitrogens on each subunit, and therefore no clear signature of base pairing in the 2DXCS signal exists (Figure 11). For the AT base pair, the maximum intensity of the difference spectra is roughly one-sixth the maximum intensity of the base pair 2DXCS shown in Figure 9, and only one-third the intensity of the imine cross-peaks in the same region of the spectra, providing an ultrafast spectral feature corresponding to hydrogen bonding for a single base pair in vacuo.

To show that this feature cannot be attributed to changes in the XANES between the bases and the monomers, the 2D linear absorption product difference between each basepair and its monomers

$$S^0_{\mathrm{diff}}(-\Omega_1,\Omega_3) = S^0_{\mathrm{AT}}(-\Omega_1,\Omega_3) - (S^0_{\mathrm{A}}(-\Omega_1,\Omega_3) + S^0_{\mathrm{T}}(-\Omega_1,\Omega_3)) \quad (10)$$

is shown in the bottom row of Figure 11. For both pairs, the difference spectra for the 2DXCS is much more intense and contains completely different features than this product. The imine cross-peak intensity increase observed for the 2DXCS is

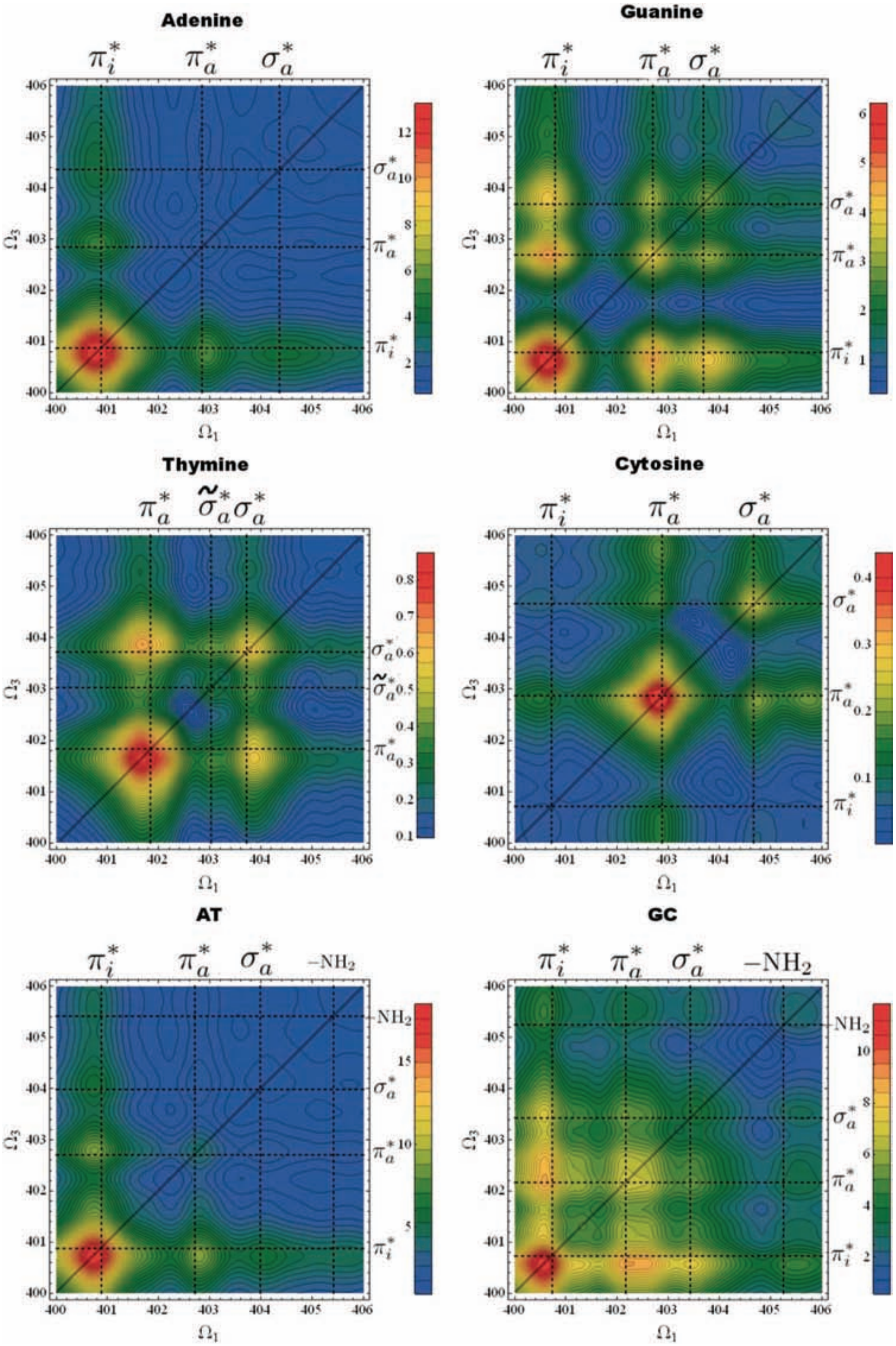


Figure 9. The XXXX tensor component of the $\mathbf{k}_I$ absolute magnitude signal $S(-\Omega_1, t_2 = 0, \Omega_3)$ for the various bases and base pairs. Dashed lines correspond to the peak positions shown in Figure 8.

not present. The similarity of the base pair XANES to the sum of its monomers (see Figure 4 in ref 14) is responsible for the weak intensity of the $S^0_{\mathrm{diff}}(-\Omega_1, \Omega_3)$.

D. 2DXCS of Stacked Nucleobase Analogues. DNA nucleobases interact with each other through hydrogen bonding and stacking interactions. So far we focused on the former. To identify signatures of base-stacking, the 2DXCS spectra of base pair analogues were calculated. Two types of nitrogens can be clearly resolved directly in the 2DXCS spectra. Imine and $-\mathrm{NH_2}$ nitrogens provide clear peaks at ~400.7 and ~405 eV, respectively. On the Watson−Crick base pairs, each of these markers is located on the same base. By replacing the methane group on thymine with adenine's $-\mathrm{NH_2}$ group, we constructed an artificial model system for examining 2DXCS signatures of base-stacking (Figure 12). Base-stacking interactions were recently investigated in a model system containing nucleobase analogues linked by a short alkane chain. Quenching of absorption at a characteristic UV frequency as the temperature of the sample was decreased was used as a measure of base-stacking between the linked chromophores.[34] A 2DXCS experiment on the same type of system may reveal the electronic coupling between these bases in their stacked geometry.

We considered two pair geometries, one with the bases stacked (b) and the other separated by ~10 Å (a). For the separated case, geometry minimization was performed on each monomer separately at the 6-311g**/B3LP level of theory, and each base was then translated manually. Both base monomers

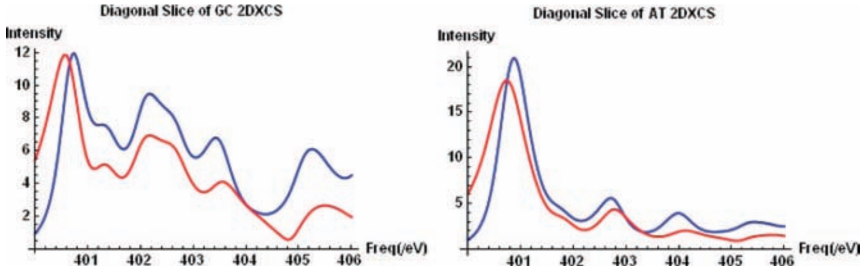


Figure 10. Comparison of diagonal slices through the 2DXCS (red) for the GC and AT base pairs and the square of their XANES (blue). The ratio of intensity in the amine peaks of the GC base pair relative to the main imine peak is 58% for the 2DXCS and 79% for the square of the XANES.

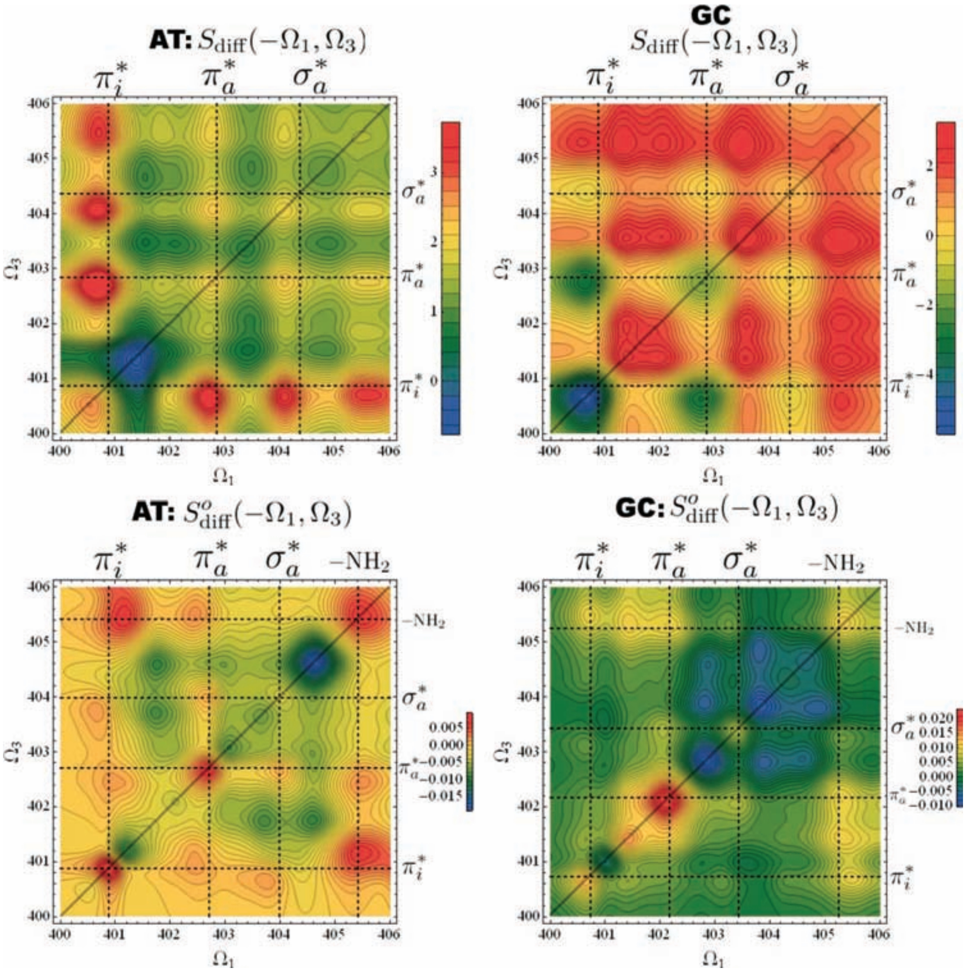
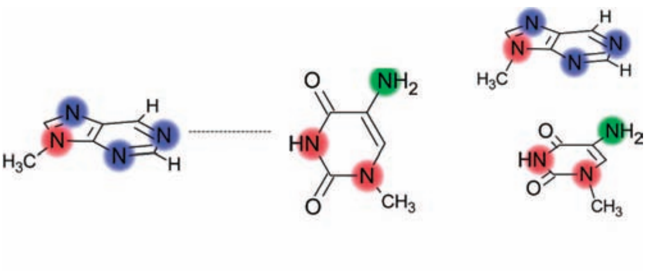


Figure 11. (top row) Difference between the base pair and the additive spectra of two noninteractive monomers. (left AT and right GC) $|P_{\mathbf{k}_I}(\text{pair})| - (|P_{\mathbf{k}_I}(\text{monomer 1})| + |P_{\mathbf{k}_I}(\text{monomer 2})|)$. (bottom row) Difference of the linear absorption products, S^0_{diff} for each base pair. The change in the 2DXCS signal which accompanies hydrogen bonding (Figure 11) cannot be explained as a difference in the XANES. Dashed lines represent the peaks in the base pair XANES.



(a) separated (b) stacked

Figure 12. Structures of the nucleobase analogues used to investigate base-stacking interactions.

are neutral, with dipole moments lying in the plane of the nucleobase. To minimize the electrostatic coupling between the bases, they were placed orthogonally to each other. For the stacked bases, it was assumed that solvent and alkane chain effects constrain the relative orientation of the two bases. The geometry used in the electronic structure calculation was manually fitted to the crystal structure of a short segment of DNA.

The simulated XANES and contributions from different nitrogen species are shown in Figure 13. In the separated geometry the σ_a^* peak and the $-NH_2$ peak are separated by over 1 eV. For the stack, the $-NH_2$ and σ_a^* peaks lie within 0.5 eV of each other, much closer than in the separated system. The 2DXCS of each system is shown in Figure 13. As with the base pairs, the 2DXCS difference spectra between the stacked and separated analogues was calculated and is shown in Figure 15. The cross-peak between the $-NH_2$ and π_i^* on different bases is visible for both geometries, but slightly more intense for the stacked analogues. The main features of base-stacking are a decrease in the intensity of the π_i^* peak and a corresponding increase in the intensity of the $-NH_2$ peak at around 405.5 eV.

To test whether these differences could be explained by a change in the linear absorption of the stacked analogues, $S_{\text{diff}}^0(-\Omega_1, \Omega_3)$ was compared to $S_{\text{diff}}(-\Omega_1, \Omega_3)$ in Figure 15. Besides an overall difference in the intensity, it is not obvious that the 2DXCS gives new information on the system; however, closer inspection of the imine cross-peaks of $S_{\text{diff}}(-\Omega_1, \Omega_3)$ reveals a double-lobed structure not present in $S_{\text{diff}}^0(-\Omega_1, \Omega_3)$, indicating that the width of the π_i^* cross-peaks are larger for the base-stacked analogue pair than the separated case, as can be visually confirmed from Figure 14.

These results suggest that cross-peak widths and intensities in the N K-edge ofthe nucleobases can be used to assign the relative geometry of two nucleobases. XANES is relatively insensitive to this information. If the electronic states on each base are highly localized, the 2DXCS of a double strand could be represented as a sum of the signals generated from each pair of bases. Melting in a DNA strand should be observed as a decrease in intensity and a sharpening of the imine nitrogen cross-peak. A higher level of theory than that adopted in this paper will be required to model the extended electronic states in DNA.

IV. Discussion

A level of theory that accurately predicts the XANES and can affordably calculate the 2DXCS spectra for the nitrogen K-edge XAS spectra of a DNA strand has yet to be developed.[35] Higher levels of theory exist which can reproduce the XANES of smaller organic molecules,[18,19] but their expense precludes applying them to the XANES of systems larger than one or two base pairs. The nonlinear response imposes an additional computational challenge. This study aimed at simulating the 2DXCS spectra for the individual DNA nucleobases using techniques that could be applied to larger strands of these bases. The orbital technique employed here is inexpensive yet qualitatively reproduces the major π_i^*, π_a^*, and σ_a^* peaks characteristic of the XANES spectra.

The responses of individual bases and base pairs in vacuo were examined. The extended electronic structure of DNA should differ substantially from this simple model system.[36] To treat the 2DXCS of a system of double-stranded DNA, the electronic structure of at least the nearest neighbors of the absorbing base must be included in the calculation. The nearest physical realization of the systems studied here would be gas phase clusters selected by mass spectroscopy to contain only pairs of bases. For these very small systems, features in the 2DXCS response have been linked to hydrogen bond formation and base-stacking interactions. If similar spectral features could be observed in double-stranded DNA, these simulations suggest that 2DXCS could act as an ultrafast probe of bubbles forming in an poly(dAdT) strand of DNA. Short pulses can provide ~10 fs snapshots of nuclear motions. The sample damage from irradiation with an X-ray pulse with an intensity required to perform the type of experiment described here will be substantial, but the time scale of the third-order response can be chosen short enough to avoid large-amplitude nuclear motions, including damage.

XANES is insensitive to the degree of hydrogen bonding in a sample. Our results indicate the off-diagonal π_i^* intensity of the 2DXCS signal may carry this information. The 2DXCS was analyzed by expressing it as a sum over Liouville-space pathways. Some polarization configurations show a preference for the ground-state bleaching. Cross-peaks in the $P_{\mathbf{k}_I}$ signal corresponding to valence electronic coupling between hydrogen-bonded Watson−Crick base pairs were identified. Base-stacking interactions between nucleobase analogues were also measurable, as a broadening of the π_i^* cross-peaks for the stacked analogues. Since this signature does not depend on the separation between the imine and $-NH_2$ nitrogens between the separated bases, it may also carry over to base-stacking between the nucleobases.

2DXCS allows an ultrafast sampling of nonstationary electronic states. Weak electronic couplings between hydrogen-bonded DNA nucleobases were identified, and stronger stacking interactions between nucleobase analogues were simulated using this technique. It will be interesting to develop model systems of the electronic transport in DNA capable of scaling up to systems describing long-range interactions. Extending energy-transport simulations on models of DNA to include bleaching of electronic transitions will allow optical/X-ray experiments on DNA strands to be simulated. The current work on the individual nucleobases represents a first step in this direction. The ability to take attosecond snapshots of evolving electronic wavepackets would resolve many controversies in the study of low-dimensional energy transport.

The size of electronic components and the wires connecting them approach molecular length scales as lithography improves. DNA is an attractive candidate for a nanoscale wire due to the biochemical toolbox available for characterizing, synthesizing, and manipulating single- and double-stranded DNA. Experimental and simulation studies of energy transfer in DNA have yielded contradictory results, reporting insulating, metallic, and even superconducting behavior.[36] Excited-state dynamics in DNA strands is an interesting basic

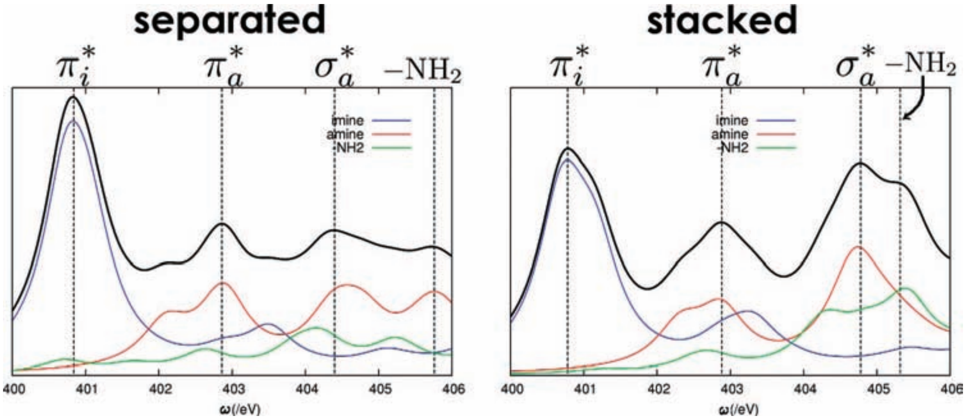


Figure 13. XANES spectra of the nucleobase analogues in Figure 12, expressed as a sum of contributions from imine, amine, and $-NH_2$ nitrogens.

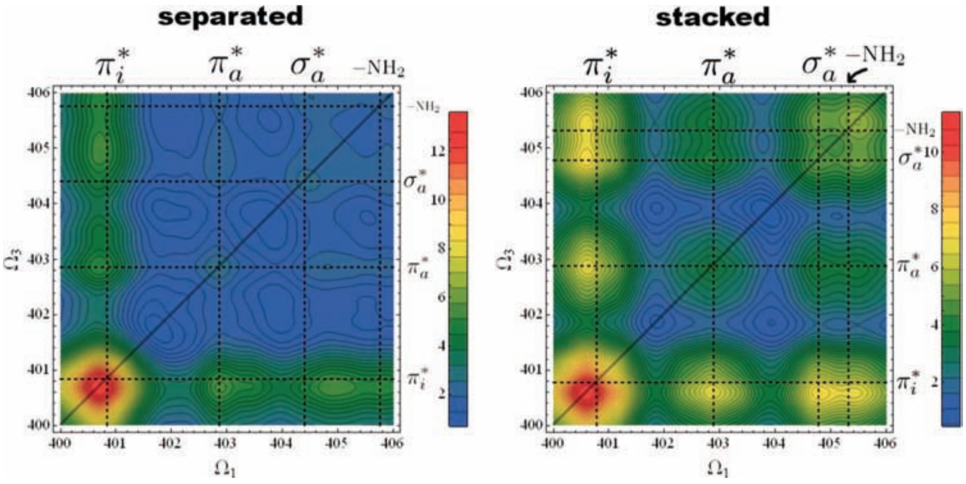


Figure 14. Comparison of the $S_{\mathbf{k}_I}(-\Omega_1, t_2 = 0, \Omega_3)$ XXXX spectra for the separated and stacked analogues shown in Figures 12 and 13.

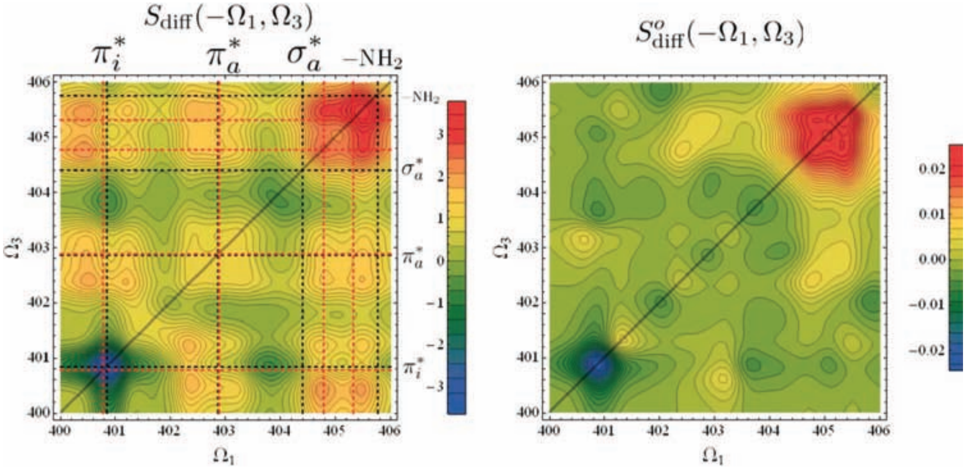


Figure 15. (a) Difference in absorption between the stacked and the separated base analogues, $|P_{\mathbf{k}_I}(\text{stacked})| - |P_{\mathbf{k}_I}(\text{separated})|$. The black and red dashed lines correspond to the peaks in the separated and stacked analogues from Figure 13. (b) Difference in the 2D linear absorption of the stacked and separated analogues, $I_{\text{stack}}(-\Omega_1)I_{\text{stack}}(\Omega_3) - I_{\text{sep}}(-\Omega_1)I_{\text{sep}}(\Omega_3)$, where $I(\Omega)$ is the linear absorption shown in Figure 13.

research topic in its own right; mechanisms for damping electronic excitations may have stabilized early genetic material against photolysis. 2DXCS could take local snapshots of the projection of an excited-state electronic wavepacket onto the core orbital of a specific atom and might play a valuable role in testing the validity of some of these models.

Acknowledgment. The support of the Chemical Sciences, Geosciences and Biosciences Division, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy, is gratefully acknowledged.

Supporting Information Available: Real, imaginary, and absolute values of the 2DXCS signal for each of the nucleobases, the base pairs, and the nucleobase analogues as well as XANES spectra with an increased energy range. This material is available free of charge via the Internet at http://pubs.acs.org.

Appendix A. Orientational Averaging

We have assumed that the DNA strand is randomly oriented in space.[32] We will use the following convention with indices $a, b, c, \ldots$. referring to components of a vector in the laboratory frame, $\alpha, \beta, \gamma, \lambda, \ldots$. to tensor components in the molecular frame, and $i, j, k, \ldots$ to eigenstates and outer products of eigenstates. In the semiclassical approximation for the coupled material-field system, the third-order response for an isolated, oriented system can be written as a path integral in Liouville space[12]

$$P_4^{(3)}(t) = \int \mathrm{d}t_1 \int \mathrm{d}t_2 \int \mathrm{d}t_3 \sum_{\nu_1\nu_2\nu_3} (\vec{\varepsilon}_4 \cdot \vec{\mu}_4)(\vec{\varepsilon}_3 \cdot \vec{\mu}_3)(\vec{\varepsilon}_1 \cdot \vec{\mu}_2)(\vec{\varepsilon}_1 \cdot \vec{\mu}_1) R_{\nu_1\nu_2\nu_3}(t_3, t_2, t_1) E_3(t - t_3) E_2(t - t_3 - t_2) E_1(t - t_3 - t_2 - t_1) \tag{A1}$$

$\vec{\varepsilon}_a$ are the polarization vectors of the applied (1−3) and measured (4) electric fields, μ_i are the dipole matrix elements between each transition

$$\mu_i = \langle\langle \nu_{i-1}^{\dagger}[\hat{\mu}, \nu_i]\rangle\rangle \tag{A2}$$

where ν_i are vectors in the outer product space of the eigenstates of the unperturbed Hamiltonian (i.e., $\nu_i = |a\rangle\langle b|$, $a, b \in \mathrm{Eig}(H_0)$), with corresponding frequencies $\omega_i = \varepsilon_a - \varepsilon_b$. In this sum-over-states approach, the factor $R_{\nu_1}\nu_2\nu_3(t_3, t_2, t_1)$ is a product of damped oscillatory functions of ω_{ν_1}, ω_{ν_2}, and ω_{ν_3}.

For an isotropic collection of randomly oriented molecules, an integral over all possible orientations must be performed

$$P_4^{(3)}(t) = \int \mathrm{d}t_1 \int \mathrm{d}t_2 \int \mathrm{d}t_3 \int \mathrm{d}\theta \int \mathrm{d}\phi \int \mathrm{d}\varphi \sum_{\nu_1\nu_2\nu_3} (\vec{\varepsilon}_4 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_4) \times (\vec{\varepsilon}_3 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_3)(\vec{\varepsilon}_2 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_2)(\vec{\varepsilon}_1 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_1) R_{\nu_1\nu_2\nu_3}(t_3, t_2, t_1) \times E_3(t - t_3) E_2(t - t_3 - t_2) E_1(t - t_3 - t_2 - t_1) \tag{A3}$$

where $\hat{M}^{\theta\phi\varphi}$ is the unitary rotational matrix in $\mathbb{R}^3$ parametrized by the angles θ, ϕ, and φ. If we expand the electric field unit vectors and dipole moments in their tensor components, we get the following expression:

$$\begin{aligned} &\int \mathrm{d}\theta \int \mathrm{d}\phi \int \mathrm{d}\varphi \sum_{\nu_1\nu_2\nu_3} (\vec{\varepsilon}_4 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_4)(\vec{\varepsilon}_3 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_3) \times (\vec{\varepsilon}_2 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_2)(\vec{\varepsilon}_1 \cdot \widehat{M}^{\theta\phi\varphi}\vec{\mu}_1) \\ &= \int \mathrm{d}\theta \int \mathrm{d}\phi \int \mathrm{d}\varphi \sum_{abcd;\alpha\beta\gamma\lambda} \varepsilon_{4d}\varepsilon_{3c}\varepsilon_{2b}\varepsilon_{1a} M_{a\alpha}^{\theta\phi\varphi} \times M_{b\beta}^{\theta\phi\varphi} M_{c\gamma}^{\theta\phi\varphi} M_{d\lambda}^{\theta\phi\varphi} \mu_{4\lambda}\mu_{3\gamma}\mu_{2\beta}\mu_{1\alpha} \end{aligned} \tag{A4}$$

The tensor operator linking the laboratory and molecular frames has symmetry properties which allow the following analytic solution.[32] In the semiclassical approximation, there are only three experimental degrees of freedom associated with the polarization directions of the fields when measuring an isotropic dipole distribution. The indices refer to core and valence electronic states. For each term in these sums, there is a vector contribution to $\vec{\mu}_{\mathrm{eff}}$, depending on the projection of each transition dipole onto the others.

$$\begin{aligned} &\textstyle\sum_{abcd;\alpha\beta\gamma\lambda} \varepsilon_{4d}\varepsilon_{3c}\varepsilon_{2b}\varepsilon_{1a} \int \mathrm{d}\theta \int \mathrm{d}\phi \int \mathrm{d}\varphi M_{a\alpha}^{\theta\phi\varphi} M_{b\beta}^{\theta\phi\varphi} M_{c\gamma}^{\theta\phi\varphi} M_{d\lambda}^{\theta\phi\varphi} \mu_{4\lambda}\mu_{3\gamma}\mu_{2\beta}\mu_{1\alpha} = \\ &\underbrace{\frac{1}{30}\begin{pmatrix} (\vec{\varepsilon}_1 \cdot \vec{\varepsilon}_2)(\vec{\varepsilon}_3 \cdot \vec{\varepsilon}_4) \\ (\vec{\varepsilon}_1 \cdot \vec{\varepsilon}_3)(\vec{\varepsilon}_2 \cdot \vec{\varepsilon}_4) \\ (\vec{\varepsilon}_1 \cdot \vec{\varepsilon}_4)(\vec{\varepsilon}_2 \cdot \vec{\varepsilon}_3) \end{pmatrix}} \times \underbrace{\begin{pmatrix} 4 & -1 & -1 \\ 1 & 4 & 1 \\ -1 & -1 & 4 \end{pmatrix} \begin{pmatrix} (\vec{\mu}_1 \cdot \vec{\mu}_2)(\vec{\mu}_3 \cdot \vec{\mu}_4) \\ (\vec{\mu}_1 \cdot \vec{\mu}_3)(\vec{\mu}_2 \cdot \vec{\mu}_4) \\ (\vec{\mu}_1 \cdot \vec{\mu}_4)(\vec{\mu}_2 \cdot \vec{\mu}_3) \end{pmatrix}} \\ = \quad &\vec{\varepsilon}^{\,\mathrm{eff}} \cdot \vec{\mu}^{\,\mathrm{eff}} \end{aligned} \tag{A5}$$

For an experiment, with a set of chosen $\varepsilon_{1,2,3,4}$, the signal can be written as a linear combination of three terms, $\vec{\varepsilon}^{\mathrm{eff}}$. These are labeled by their polarizations in a system in which all pulses are collinear: XXYY, XYXY, and XYYX.

$$\begin{aligned} \varepsilon_{xxyy} &= 4(\vec{\mu}_1 \cdot \vec{\mu}_2)(\vec{\mu}_3 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_3)(\vec{\mu}_2 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_4)(\vec{\mu}_2 \cdot \vec{\mu}_3) \\ \varepsilon_{xyxy} &= 4(\vec{\mu}_1 \cdot \vec{\mu}_3)(\vec{\mu}_2 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_2)(\vec{\mu}_3 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_4)(\vec{\mu}_2 \cdot \vec{\mu}_3) \\ \varepsilon_{xyyx} &= 4(\vec{\mu}_1 \cdot \vec{\mu}_4)(\vec{\mu}_2 \cdot \vec{\mu}_3) - (\vec{\mu}_1 \cdot \vec{\mu}_3)(\vec{\mu}_2 \cdot \vec{\mu}_4) - (\vec{\mu}_1 \cdot \vec{\mu}_2)(\vec{\mu}_3 \cdot \vec{\mu}_4) \end{aligned} \tag{A6}$$

The $\vec{\mu}_i$ are the transition dipoles of the system's interaction with the ith pulse.

References and Notes

(1) Stohr, J. *NEXAFS Spectroscopy*; Springer: New York, 1996.
(2) Rehr, J. J.; Albers, R. C. *Rev. Mod. Phys.* **2000**, *72*, 621.
(3) De Groot, F.; Kotani, A. *Core Level Spectroscopy of Solids*; CRC Press: Boca Raton, FL, 2008.
(4) Mukamel, S. *Phys. Rev. B* **2005**, *72*, 235110.
(5) Schweigert, I. V.; Mukamel, S. *Phys. Rev. Lett.* **2007**, *99*, 163001.
(6) Schweigert, I. V.; Mukamel, S. *Phys. Rev. A* **2007**, *76*, 0125041.
(7) Evans, J. *Phys. Chem. Chem. Phys.* **2006**, *8*, 3045.
(8) Kapteyn, H.; Cohen, O.; Christov, I.; Murnane, M. *Science* **2007**, *317*, 775.
(9) Schweigert, I.; Mukamel, S. Submitted to *Phys. Rev. B*.
(10) Peyrard, M.; López, S. C.; Angelov, D. *Eur. Phys. J. Spec. Top.* **2007**, *147*, 173.
(11) Tanaka, S.; Mukamel, S. *J. Chem. Phys.* **2002**, *116*, 1877.
(12) Mukamel, S. *Principles of Nonlinear Optical Spectroscopy*; Oxford University Press: New York, 1995.
(13) Roulet, B.; Gavoret, J.; Nozieres, P. *Phys. Rev.* **1969**, *178*, 1072.
(14) Kirtley, S. M.; Mullins, O. C.; Chen, J.; van Elp, J.; George, S. J.; Chen, C. T.; O'Halloran, T.; Cramer, S. P. *Biochim. Biophys. Acta* **1992**, *1132*, 249.
(15) MacNaughton, J.; Moewes, A.; Kurmaev, E. Z. *J. Phys. Chem. B* **2005**, *109*, 7749.
(16) Harada, Y.; Takeuchi, T.; Kino, H.; Fukushima, A.; Takakura, K.; Hieda, K.; Nakao, A.; Shin, S.; Fukuyama, H. *J. Phys. Chem. A* **2006**, *110*, 13227.
(17) MacNaughton, J. B.; Moewes, A.; Lee, J. S.; Wettig, S. D.; Kraatz, H. B.; Ouyang, L. Z.; Ching, W. Y.; Kurmaev, E. Z. *J. Phys. Chem. B* **2006**, *110*, 15742.
(18) Shirley, E. L. *J. Electron. Spectrosc. Relat. Phenom.* **2005**, *144−147*, 1187.
(19) Jiemchooroj, A.; Ekström, U.; Norman, P. *J. Am. Chem. Soc.* **2007**, *127*, 165104.
(20) Mochizuki, Y.; Koide, H.; Imamura, T.; Takemiya, H. *J. Synchrotron Radiat.* **2001**, *8*, 1003.
(21) Campbell, L.; Mukamel, S. *J. Chem. Phys.* **2004**, *121*, 12323.
(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03, revision c.02*; Gaussian, Inc.: Wallingford, CT, 2004.
(23) Ekström, U.; Norman, P.; Carravetta, V.; Agren, H. *Phys. Rev. Lett.* **2006**, *97*, 143001.
(24) Kosugi, N.; Kuroda, H. *Chem. Phys. Lett.* **1980**, *74*, 490.
(25) Shukla, M. K.; Leszczynski, J. *J. Comput. Chem.* **2004**, *25*, 768.
(26) Triguero, L.; Pettersson, L. G. M.; Ågren, H. *Phys. Rev. B* **1998**, *58*, 8097.

(27) Williams, A. R.; Degroot, R. A. *J. Chem. Phys.* **1975**, *63*, 628.

(28) Ozeki, K.; Sakabe, N.; Tanaka, J. *Acta Crystallogr.* **1969**, *B25*, 1038.

(29) Guille, K.; Clegg, W. *Acta Crystallogr.* **2006**, *C62*, o515.

(30) Barker, D. L.; Marsh, R. E. *Acta Crystallogr.* **1964**, *17*, 1581.

(31) Schweigert, I.; Mukamel, S. *Phys. Rev. A* **2008**, *77*, 33802.

(32) Andrews, D. L.; Thirunamachandran, T. *J. Chem. Phys.* **1977**, *67*, 5026.

(33) Abramavicius, D.; Mukamel, S. *J. Chem. Phys.* **2005**, *122*, 134305.

(34) Itahara, T.; Imaizumi, K. *J. Phys. Chem.* **2007**, *111*, 2025.

(35) Schweigert, I.; Mukamel, S. Manuscript in preparation.

(36) Endres, R. G.; Cox, D. L.; Singh, R. R. P. *Rev. Mod. Phys.* **2004**, *76*, 195.

JP803824A