

## Supporting Information

### Nucleobase Orientation and Ordering in Films of Single-Stranded DNA on Gold

Dmitri Y. Petrovykh,<sup>\*,†,‡</sup> V. Pérez-Dieste,<sup>§</sup> Aric Opdahl,<sup>||</sup> Hiromi Kimura-Suda,<sup>||</sup> J. M. Sullivan,<sup>#,‡</sup>  
Michael J. Tarlov,<sup>||</sup> F. J. Himpsel,<sup>§</sup> Lloyd J. Whitman<sup>‡</sup>

<sup>†</sup>Physics Department, University of Maryland, College Park, MD 20742; <sup>‡</sup>Naval Research Laboratory, Washington, DC 20375;

<sup>§</sup>Physics Department, University of Wisconsin, Madison, WI 53706; <sup>||</sup>National Institute of Standards and Technology, Gaithersburg, MD 2089; <sup>#</sup>Department of Chemistry, Northwestern University, Evanston, IL 60208

\*E-mail: dmitri.petrovykh@nrl.navy.mil

#### Materials and Methods

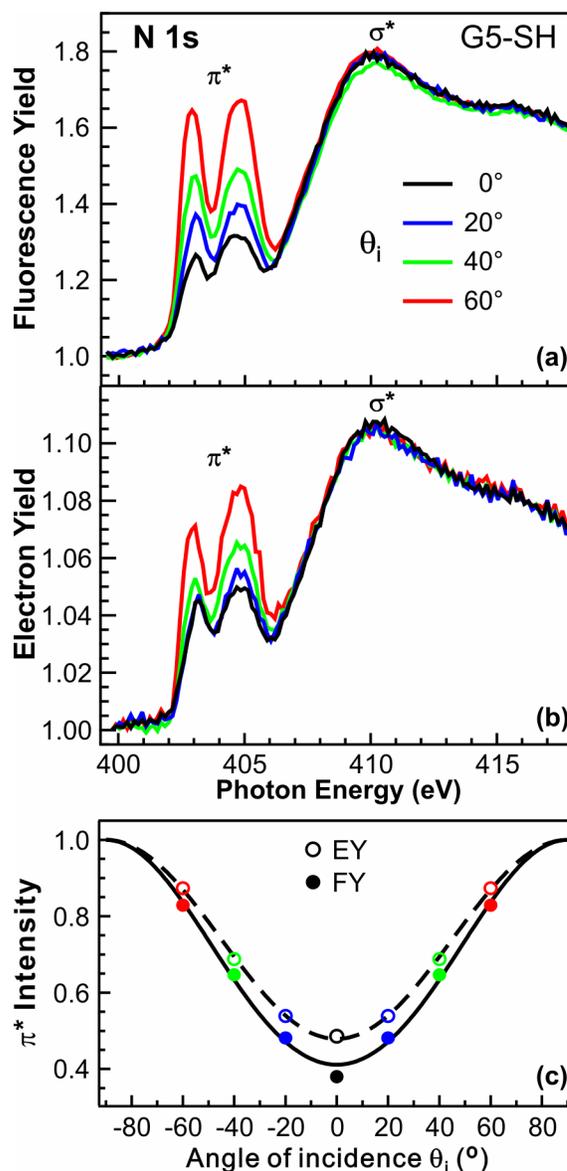
**Materials.** Custom homo-oligonucleotides were synthesized and HPLC purified by the vendors [(dT)<sub>25</sub>-SH from Research Genetics, (dT)<sub>5</sub>-SH and (dT)<sub>5</sub> from IDT] and used as-received without further purification.<sup>a</sup> The 5' alkanethiol modified oligonucleotides [(dT)<sub>25</sub>-SH and (dT)<sub>5</sub>-SH] were used without removing the protective S-(CH<sub>2</sub>)<sub>6</sub>OH group from the 5' end. Buffer solutions were prepared containing 1 x TE (10 mM Tris-HCl, 1 mM EDTA), 1 M NaCl or 1 M CaCl<sub>2</sub>, and were adjusted to pH 7 with HCl.

**Preparation of DNA Monolayers on Gold Substrates.** Polycrystalline gold films on single-crystal Si(100) wafers were used as substrates. Prior to the deposition of gold, the wafers were cleaned using a "piranha solution" consisting of 70% H<sub>2</sub>SO<sub>4</sub> and 30% H<sub>2</sub>O<sub>2</sub> (30% H<sub>2</sub>O<sub>2</sub> in H<sub>2</sub>O). Piranha solution must be used with care: it is extremely oxidizing, reacts violently with organics, and should be stored in loosely tightened containers to avoid pressure buildup. After cleaning, a 20 nm Cr adhesion layer was deposited by vapor deposition, followed by 200 nm of Au. Each substrate was again cleaned with piranha solution and rinsed thoroughly with deionized water (18.3 MΩ) immediately prior to adsorption of DNA. Clean gold substrates (≈1 cm<sup>2</sup>) were immersed in 2 mL of 1 μM DNA solution in pH 7 buffer solutions at room temperature for 20 h. The DNA concentration was confirmed by UV absorption measurements. For (dT)<sub>5</sub> and (dT)<sub>5</sub>-SH immobilization NaCl was used as the buffer salt. (dT)<sub>25</sub>-SH was immobilized from a CaCl<sub>2</sub> buffer to increase the molecular packing density for a closer match with that achieved for (dT)<sub>5</sub>-SH in NaCl. Additional (dT)<sub>25</sub>-SH films were immobilized from a NaCl buffer to be used as a coverage reference. After deposition, each sample was rinsed with deionized water to remove buffer salt and loosely bound DNA and blown dry under flowing nitrogen.

**NEXAFS Measurements.** The NEXAFS measurements were done in UHV at the undulator beamline 8.0.1 of the Advanced Light Source. The resolving power of this beamline is better than 8000, providing energy resolution of about 50 meV around the nitrogen K-edge energy range, i.e., less than the intrinsic width of the N 1s core-levels. The custom-built setup for NEXAFS measurements has been previously described in ref 7. Briefly, the measurement chamber allowed the rotation of samples to adjust the (off-normal) angle of incidence  $\theta_i$  of the p-polarized synchrotron light. The fluorescence yield (FY) signal was detected using a channel plate detector with a detection cone of 40° located at a fixed 45° angle above the incident beam port. We used an Al filter to suppress the background fluorescence from the silicon substrate.

Selective FY detection from the nitrogen K-edge is approximately 10<sup>4</sup> times less efficient than electron yield (EY) detection via the sample current. However, the much more effective background suppression in the FY detection compensates for its lower efficiency. In fact, the FY signal-to-background ratio is approximately an order of magnitude higher than the EY signal-to-background ratio

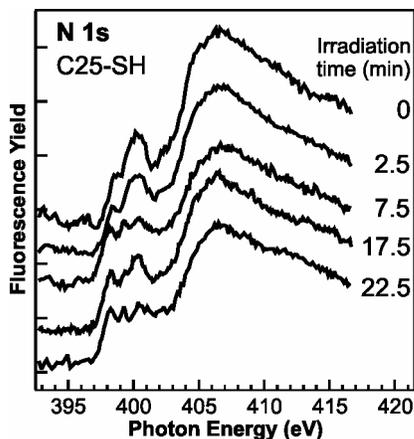
(compare Figure S11a and S11b), despite the fact that the EY consisted mainly of high energy Auger electrons, which are more selective than low energy secondaries. It is important to note that the resulting polarization dependence (i.e., the information used to determine orientation and ordering of nucleobases) is essentially identical whether measured with FY or EY detection (Figure S11c).



**Figure S11.** Comparison of nitrogen K-edge NEXAFS measured with FY and EY detection. The FY (a) and EY (b) spectra were *simultaneously* recorded from a monolayer of (dG)<sub>5</sub>-SH on Au. The FY detection produced about an order of magnitude higher signal-to-background ratio. (c) The modulation of the  $\pi^*$  resonance intensities with the of incidence  $\theta_i$  (color-coded symbols): FY (black line and full symbols) and EY (dashed line and open symbols).

<sup>a</sup> Certain vendors are identified to adequately specify the experimental procedure. In no case does such identification imply endorsement by the National Institute of Standards and Technology or the Naval Research Laboratory.

An important consideration in establishing the NEXAFS acquisition parameters was minimizing the potential damage to the DNA caused by the incident radiation. As a test of sample degradation, a series of spectra were acquired as a function of sample irradiation time (Figure S12). Within the first few minutes there was no apparent damage, after 7.5 min of irradiation significant changes in the spectrum occurred. These test spectra were acquired with a photon energy step of 0.2 eV and 2 s per step dwell time. All the data presented in Figure 1a were measured with 0.2 eV photon energy steps, 1 s per step dwell time, and the beam defocused to about  $1 \times 3 \text{ mm}^2$  to ensure that exposure remained below the damage threshold. To further minimize sample damage and to test sample uniformity, spectra were measured from 3-4 different spots on each sample for each incidence angle. These multiple scans for each angle were added together for subsequent analysis.



**Figure S12.** Monitoring DNA sample damage by exposure to synchrotron radiation. The nitrogen K-edge FY NEXAFS spectra were recorded consecutively from the same spot on a monolayer of (dC)<sub>25</sub>-SH on Au.

**NEXAFS data normalization.** Various methods can be used to normalize NEXAFS data (ref 4). We use a normalization which produces N  $\pi^*$  resonance intensities approximately proportional to the N coverage. Another important feature of this normalization is that normalized peak intensities numerically correspond to peak-to-background ratios, i.e., the most physically appropriate representation of NEXAFS data (ref 4).

First, each FY spectrum was normalized to the photon flux by dividing it by the EY measured simultaneously from a gold mesh. Second, this ratio was multiplied by an angle-dependent factor to achieve a common height of the N-edge jump, i.e., the difference between the pre-edge signal and the continuum states above the near-edge. This makes the signal proportional to the absorption per N atom. Third, the absolute scale was set for each sample by a common scale factor for all angles which was chosen such that the pre-edge background intensity became 1 at  $\theta = 60^\circ$  (close to the magic angle). The spectra for the other angles were rigidly shifted by a constant to have a pre-edge intensity of 1 as well. This procedure provides a comparison of the absolute N coverage between different samples, assuming that the pre-edge background from the substrate is the same for each sample.

It is well-known that the intensity of  $\sigma^*$  resonances is more affected by the uncertainty associated with a specific choice of a background function than the intensity of  $\pi^*$  resonances (refs 4,5). It is also known that in aromatic molecules delocalization of molecular orbitals can lead to more complicated relative intensities of  $\pi^*$  and  $\sigma^*$  resonances than in the case of simple diatomic molecules (refs 4,5). Furthermore, there is strong empirical evidence that in multilayers of aromatic molecules with heteroatom substitutions the apparent polarization-dependence of  $\sigma^*$  resonances can be weaker

than for  $\pi^*$  resonances (refs 10,12). Accordingly, we rely on the polarization dependence of  $\pi^*$  rather than  $\sigma^*$  resonances to determine orientation and ordering of nucleobases in ssDNA monolayers.

**XPS measurements** XPS measurements were performed in a commercial UHV system equipped with a monochromatic Al K $\alpha$  source, a hemispherical electron energy analyzer ( $58^\circ$  angle between monochromator and analyzer), and a magnetic electron lens. Detailed descriptions of the quantitative analysis of DNA adsorbed on gold have been presented in refs 2,3. Briefly, normal emission angle-integrated high-resolution scans with 15 eV to 20 eV windows and 20 eV pass energy (0.36 eV nominal analyzer energy resolution) were acquired for Au 4f and 4d, N 1s, P 2p, C 1s, and O 1s core-levels. The reported binding energies (BE) were based on the analyzer energy calibration (Au 4f<sub>7/2</sub> BE = 84.0 eV and Au 4d<sub>5/2</sub> BE = 335.2 eV, for all samples); no charge compensation was necessary. Spectra of the N 1s region were accumulated for 30 min to 60 min, depending on the sample coverage, to obtain an adequate signal-to-noise ratio. The reference Au signals, used to calibrate the photoelectron attenuation and thickness of DNA overlayers, were measured from Au films cleaned by Ar ion sputtering. A convolution of Lorentzian and Gaussian line shapes was used to fit the individual peaks, with typical ratios between 10/90 and 20/80. A linear combination of Shirley and linear functions was used to model the background, with the corresponding coefficients fit simultaneously with the peaks.

**FTIR Measurements.** IR absorption spectra were obtained using an FTIR spectrometer equipped with a wire grid infrared polarizer (p-polarized) and a variable angle specular reflectance accessory ( $75^\circ$  grazing incidence angle). FTIR measurements were performed in a nitrogen purged environment on freshly prepared samples and using a piranha cleaned gold substrate as a reference.

To determine relative dT coverages from FTIR data,  $\nu(\text{C}=\text{O})$  peaks were integrated over the  $1600\text{-}1850 \text{ cm}^{-1}$  range (refs 2,3). The  $\nu(\text{C}=\text{O})$  peaks for (dT)<sub>5</sub>-SH and (dT)<sub>25</sub>-SH overlap almost perfectly after this normalization.

**Model of polarization-dependent NEXAFS.** In a NEXAFS experiment orientation and ordering information is obtained by measuring intensities of a  $\pi^*$  resonance as a function of the angle of incidence of x-rays  $\theta$ , e.g., as shown in Figure 1a-b. In general, this  $\pi^*$  intensity modulation can be described by two parameters: a modulation amplitude  $A_\pi$ , and an intensity minimum angle  $\theta_{\text{min}}$ . For a fully ordered system the orientation of molecular orbitals can be determined exactly from the polarization-dependence data (refs 4,5).

However, the complexity and flexibility of biological molecules such as ssDNA mean that perfect ordering can no longer be expected and thus should not be assumed in modeling and interpreting the data. In fact, all the caveats noted in the original paper on determination of molecular orientations by NEXAFS apply to ssDNA monolayers: a distribution of  $\pi^*$  orbital orientations, apparent tilt angles close to the magic angle, and delocalization of molecular orbitals in aromatic systems (ref 5). Moreover, for a system that contains a distribution of orbital orientations, interpretation of the polarization-dependent measurements in terms of a unique apparent tilt angle in general *will not* result in a value close to the average of that distribution, as has been previously reported for a model of second-harmonic generation (SHG) measurements (ref 15) and as we demonstrate in Figure S13 for a NEXAFS model described below.

In the dipole approximation, the experimental cross-section for a particular transition is proportional to the square of the dot-product of the polarization vector with the orbital vectors of the  $\pi^*$  orbitals

$$I(\theta, \omega, \phi) \propto \left[ \hat{\epsilon}(\theta) \cdot \hat{O}(\omega, \phi) \right]^2,$$

where  $\hat{\varepsilon}$  is the polarization vector of the incident radiation, and  $\hat{O}$  is the  $\pi^*$  orbital vector described by polar angle  $\omega$  with respect to the surface normal and azimuthal angle  $\phi$  in the surface plane. We model the total cross-section for the experiment as a statistical average over the contributions from many different orbital orientations. We assume that the orbital vectors have random azimuthal orientation but that the angle they take with respect to the surface normal is distributed according to a Gaussian distribution centered at some preferred tilt angle  $\tau$ , and with distribution of likely tilt angles given by the width  $\delta$  (Figure SI3d). The width of the distribution,  $\delta$ , effectively serves as an orientational disorder parameter in this model. The NEXAFS intensity from such a system excited by an x-ray at an incidence angle  $\theta_i$  is given by:

$$I_{\text{model}}(\theta_i, \tau, \delta) = \int_0^{2\pi} d\phi \int_0^{\pi} d\omega \sin(\omega) P(\omega, \tau, \delta) I(\theta_i, \omega, \phi).$$

$P(\omega, \tau, \delta)$  is a Gaussian distribution normalized over the unit sphere

$$P(\omega, \tau, \delta) = G\left(\frac{\omega - \tau}{\delta}\right) / N(\tau, \delta),$$

where  $G(x) = 16^{-x^2}$  and with normalization

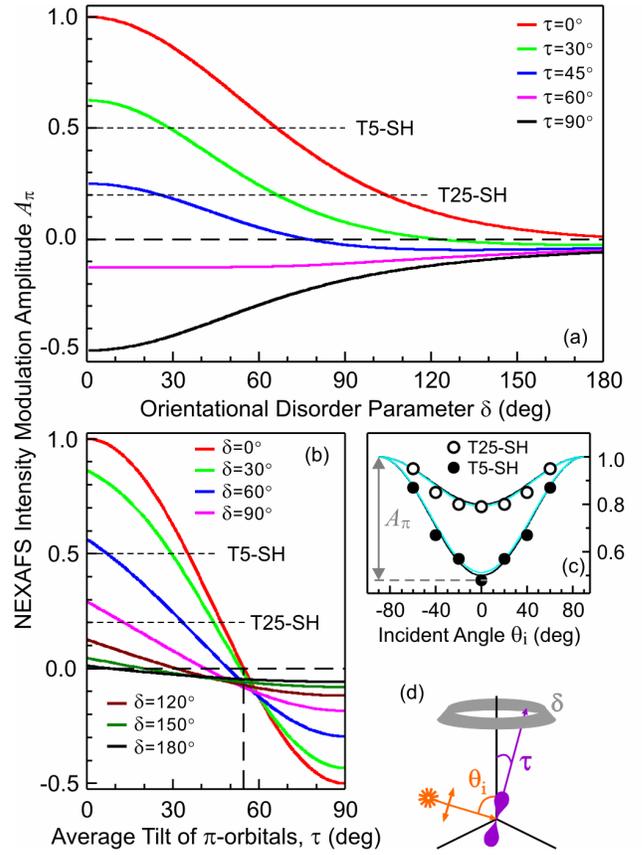
$$N(\tau, \delta) = \int_0^{2\pi} d\phi \int_0^{\pi} dx \sin(x) G\left(\frac{x - \tau}{\delta}\right).$$

Our model agrees with the standard results in the limiting cases where analytical solutions are known (ref 4). Specifically, the standard angular dependence of the NEXAFS intensity is recovered in the *limit of a fixed tilt angle* ( $\delta=0^\circ$ , red line in Figure SI3b). In the *limit of random orientation*, the total cross-section for the experiment is independent of the incidence angle, as expected.

The purpose of this model is to offer a simple means of *interpreting* experimental NEXAFS data for disordered systems. Accordingly, the *parametrization* of NEXAFS data that we use is exactly the same as that used for the traditional interpretation – the polarization-dependent modulation of NEXAFS intensity is characterized by its amplitude  $A_\pi$  and the position of the intensity minimum  $\theta_{\min}$  (Figures 1b, SI3c).

The behavior of the intensity minimum in our model is identical to that in the standard case of substrates with threefold or higher symmetry (refs 4,5), i.e., the intensity minimum can only occur at an incidence angle  $\theta_{\min}=0^\circ$  or  $\theta_{\min}=90^\circ$  corresponding to an average tilt angle  $\tau < 55^\circ$  or  $\tau > 55^\circ$ , respectively. This bi-valued  $\theta_{\min}$  result is a consequence of integrating orbitals with random azimuthal orientation projected onto a fixed x-ray polarization vector, and is in contrast with the result for a system with fixed tilt and azimuthal angles, for which  $\theta_{\min}$  is uniquely determined by the orientation of the measured orbital.

The main and the only distinction between our analysis and the traditional approach is the *interpretation* of the experimental values of the NEXAFS intensity modulation amplitude  $A_\pi$  (Figures 1b, SI3c). Figure SI3c illustrates the need for such an alternative interpretation – in general the presence or degree of orientational disorder can not be established based on NEXAFS data (open and full symbols) alone, as shown by the essentially identical fits using the standard  $a+b\sin^2\theta$  function, which assumes a unique tilt angle (black lines), and our model, which includes disorder (cyan lines). Furthermore, the common assertion that the tilt angle obtained assuming a perfectly uniform system will reflect at least the average value even for a disordered system is actually in general incorrect, e.g., as demonstrated for polarization-dependent SHG in ref 15. For NEXAFS, in Figure SI3a we show that for a fixed average tilt angle  $\tau$ , the observed intensity modulation amplitude  $A_\pi$  decreases with increasing orientational disorder  $\delta$  (Cf. Figure 1 in ref 15).



**Figure SI3.** Modeling the effects of orientational disorder and the average tilt of  $\pi$ -orbitals on the NEXAFS intensity modulation. Polarization-modulation amplitude  $A_\pi$  as a function of: (a) the orientational disorder parameter  $\delta$ , (b) the average tilt angle  $\tau$  of  $\pi$ -orbitals. In (a) and (b) colored lines correspond to the indicated discrete values of  $\tau$  and  $\delta$ , respectively. (c) Polarization dependence of the  $\pi^*$  resonance intensity (symbols = NEXAFS data from Figure 1b). The fits using the model with fixed tilt angle (black line) and the model with disorder (cyan line) are nearly identical. Experimental amplitudes  $A_\pi$  for T5-SH and T25-SH from (c) are indicated in (a) and (b) with dashed lines. (d) Schematic representation of the model parameters:  $\pi$ -orbitals with random azimuthal orientation and a Gaussian distribution of tilt angles, the average tilt angle  $\tau$ , the width of the distribution  $\delta$ , and the x-ray angle of incidence  $\theta_i$ .

The definition of the intensity modulation amplitude  $A_\pi$  that we use in this work is shown in Figure SI3c. The fixed positions of the intensity maxima and minima allow us to simply define  $A_\pi$  as  $I(\pi/2, \tau, \delta) - I(0, \tau, \delta)$ , such that  $A_\pi > 0$  for  $\tau < 55^\circ$  and  $A_\pi < 0$  for  $\tau > 55^\circ$ , which makes this particular definition convenient for numerical analysis. Other common empirical parametrizations of the standard sine-squared form of polarization-dependent NEXAFS data, e.g.,  $\langle \cos^2\theta \rangle$  or  $I(\theta=0^\circ)/I(\theta=70^\circ)$ , can be readily converted to  $A_\pi$  values via multiplication by an appropriate scaling factor.

In the interpretation commonly used for NEXAFS data from SAMs, a value of a tilt angle is extracted from the measured  $A_\pi$  value assuming a unique correspondence between  $A_\pi$  and  $\tau$  shown by the red line ( $\delta=0^\circ$ ) in Figure SI3b. Including disorder in the model removes the uniqueness of such correspondence, because a  $A_\pi(\tau)$  curve can be calculated for each value of  $\delta$ ; a few representative examples are shown as colored lines in Figure SI3b. Without additional external constraints establishing a unique  $A_\pi(\tau)$  correspondence becomes therefore impossible for a system with an unknown degree of orientational disorder.

We can still, however, offer a quantitative probabilistic interpretation of the experimental NEXAFS modulation amplitudes by observing that for each value of  $A_\pi$  only a certain range of possible combinations of the tilt angle,  $\tau$  and disorder parameter  $\delta$  will yield model amplitudes consistent with the experimental data. In the graphical representations in Figures SI3a and SI3b this effect is reflected in the fact that only some of the (colored) model curves intersect the horizontal dashed lines corresponding to the experimentally measured  $A_\pi$  values.

The set of solutions that yield the measured amplitude lie along a curve in the  $\tau - \delta$  plane which is, to a very high degree of accuracy, parabolic in the disorder parameter,  $\tau \approx \tau_o \left(1 - \delta^2 / \delta_o^2\right)$ , where  $\tau_o$  is the orbital ordering angle for the fully ordered case, and  $\delta_o$  is the maximum amount of disorder consistent with the experimental data. Values for  $\tau_o$  and  $\delta_o$  thus determine all possible solutions which are consistent with the experimental data. One may determine  $\tau_o$  from the familiar expression for the NEXAFS amplitude for a fully ordered monolayer averaged over all azimuthal orientations,  $\text{Sin}^2(\tau_o) = 2(1 - A_\pi)/3$ , where  $A_\pi$  is the measured amplitude. Determining  $\delta_o$  is more complicated and involves numerical solution based on the model amplitude for  $\tau = 0$  shown in Figure SI3a. However, a very accurate estimate is obtained from the analytic form  $\delta_o \approx 67.5^\circ \sqrt{\text{Log}_2(1/A_\pi)}$  for amplitudes,  $A_\pi$ , greater than zero. Thus we evaluate  $(\tau_o, \delta_o)$  for the films with NEXAFS amplitudes of 0.5 and 0.2 to be  $(35.3^\circ, 67.5^\circ)$  and  $(46.9^\circ, 102.8^\circ)$  respectively and plot the cumulative probability distribution functions in Figure 3c.

Cumulative probability is a natural statistical representation in cases when the direct model prediction is expressed as a limit rather than a specific value (the cumulative probability is defined as the probability that a variable will attain a value less than or equal to each value that the variable can take on). The main advantage of this representation for our model is that a narrower cumulative probability distribution of  $\delta$  indicates a system with a smaller degree of disorder, i.e., a more ordered system. In general increasing ordering correlates with stronger interactions, thus this approach provides a practical route to determination and comparison of molecular interactions at biointerfaces.

#### Full References

- (1) Castner, D. G.; Ratner, B. D. *Surf. Sci.* **2002**, *500*, 28-60.
- (2) Petrovykh, D. Y.; Kimura-Suda, H.; Tarlov, M. J.; Whitman, L. J. *Langmuir* **2004**, *20*, 429-440.
- (3) Petrovykh, D. Y.; Kimura-Suda, H.; Whitman, L. J.; Tarlov, M. J. *J. Am. Chem. Soc.* **2003**, *125*, 5219-5226.
- (4) Stöhr, J. *NEXAFS Spectroscopy*; Springer-Verlag: Berlin, 1992.
- (5) Stöhr, J.; Outka, D. A. *Phys. Rev. B* **1987**, *36*, 7891-7905.
- (6) Kimura-Suda, H.; Petrovykh, D. Y.; Tarlov, M. J.; Whitman, L. J. *J. Am. Chem. Soc.* **2003**, *125*, 9014-9015.
- (7) Pérez-Dieste, V.; Crain, J. N.; Kirakosian, A.; McChesney, J. L.; Arenholz, E.; Young, A. T.; Denlinger, J. D.; Ederer, D. L.; Callcott, T. A.; Lopez-Rivera, S. A.; Himpsel, F. J. *Phys. Rev. B* **2004**, *70*, 085205.
- (8) Both initial state (core-level shift) and final state (metallic screening) effects contribute to the red-shift of resonances for dT chemisorbed on Au.
- (9) Kirtley, S. M.; Mullins, O. C.; Chen, J.; Vanelp, J.; George, S. J.; Chen, C. T.; Ohalloran, T.; Cramer, S. P. *Biochim. Biophys. Acta* **1992**, *1132*, 249-254.
- (10) Fujii, K.; Akamatsu, K.; Yokoya, A. *J. Phys. Chem. B* **2004**, *108*, 8031-8035.
- (11) Crain, J. N.; Kirakosian, A.; Lin, J. L.; Yuedong, G.; Shah, R. R.; Abbott, N. L.; Himpsel, F. J. *J. Appl. Phys.* **2001**, *90*, 3291-3295.
- (12) Carravetta, V.; Contini, G.; Plashkevych, O.; Agren, H.; Polzonetti, G. *J. Phys. Chem. A* **1999**, *103*, 4641-4648.
- (13) Dipole selection rules allow s-p transitions when the electric field vector is parallel to p-orbitals, which for  $\pi^*$  are normal to planes of nucleobases.
- (14) DNA films are *structurally similar* to multilayers of nucleobases, but electrostatic repulsion prevents formation of DNA multilayers in solution.
- (15) Simpson, G. J.; Rowlen, K. L. *J. Am. Chem. Soc.* **1999**, *121*, 2635-2636.
- (16) The nearest-neighbor distance in a (dT)<sub>5</sub>-SH film is 1.1 nm (i.e., comparable to that in a DNA duplex), so electrostatic repulsion and entropy may also affect the ordering.
- (17) Rant, U.; Arinaga, K.; Fujita, S.; Yokoyama, N.; Abstreiter, G.; Tornow, M. *Langmuir* **2004**, *20*, 10086-10092.