Self-assembly of biomolecules at surfaces characterized by NEXAFS[1]

Xiaosong Liu, Fan Zheng, A. Jürgensen, V. Perez-Dieste, D.Y. Petrovykh, N.L. Abbott, and F.J. Himpsel

Abstract: Surface science has made great strides towards tailoring surface properties via self-assembly of nanoscale molecular adsorbates. It is now possible to functionalize surfaces with complex biomolecules such as DNA and proteins. This brief overview shows how NEXAFS (near edge X-ray absorption fine structure spectroscopy) can be used to characterize the assembly of biological molecules at surfaces in atom- and orbital-specific fashion. To illustrate the range of applications, we begin with simple self-assembled monolayers (SAMs), proceed to SAMs with customized terminal groups, and finish with DNA oligonucleotides and Ribonuclease A, a small protein containing 124 amino acids. The N 1s absorption edge is particularly useful for characterizing DNA and proteins because it selectively interrogates the π^* orbitals in nucleobases and the peptide bonds in proteins. Information about the orientation of molecular orbitals is obtained from the polarization dependence. Quantitative NEXAFS models explain the polarization dependence in terms of molecular orientation and structure.

Key words: NEXAFS, bio-interfaces, ribonuclease A, immobilization, orientation.

Résumé : La chimie des surfaces a fait de grands pas dans la direction de la possibilité d'ajuster les propriétés des surfaces par le biais d'autoassemblages d'adsorbats moléculaires à l'échelle nanométrique. Il est maintenant possible de fonctionnaliser des surfaces avec des biomolécules complexes, tels l'ADN et des protéines. Cette brève revue montre comment on peut appliquer la spectroscopie de la structure fine d'absorption des rayons X près du seuil ("NEXAFS") à la caractérisation d'assemblages de molécules biologiques liés à des surfaces de façon spécifiquement atomique et orbitalaire. Pour illustrer l'éventail des applications, on commence avec des monocouches autoassemblées simples (MAA) à partir desquelles on procède à des MAA portant des groupes terminaux personnalisés et on termine avec des oligonucléotides de l'ADN et la ribonucléase A, une petite protéine comportant 124 acides aminés. L'absorption 1s au seuil de l'azote est particulièrement utile pour caractériser l'ADN et les protéines parce qu'il permet d'interroger sélectivement les orbitales π^* des nucléobases et des liaisons peptidiques des protéines. Les informations relatives à l'orientation des orbitales moléculaires sont obtenues à partir de la dépendance de la polarisation. Des modèles quantitatifs de "NEXAFS" permettent d'expliquer la dépendance de la polarisation en fonction de l'orientation moléculaire et de la structure.

Mots-clés : NEXAFS, bio-interfaces, ribonucléase A, immobilisation, orientation.

[Traduit par la Rédaction]

Motivation

The functionalization of surfaces with biomolecules (1) is becoming a significant part of biochemistry and molecular biology, both of which traditionally have focused on phenomena occurring in bulk solution. Among the major drivers of this technology are DNA microarrays ("DNA chips") used for massively parallel analysis and synthesis of DNA (2), detection of proteins involved in signaling networks in cells ("protein chips") (3), as well as solid-state biosensors for rapid detection of pathogens in the field (4). For the study of protein activity and for achieving optimum sensitiv-

Received 18 April 2007. Accepted 7 June 2007. Published on the NRC Research Press Web site at canjchem.nrc.ca on 14 August 2007.

X. Liu and F. Zheng. Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.
A. Jürgensen. Canadian Synchrotron Radiation Facility, Synchrotron Radiation Center, Stoughton, Wisconsin 53589–3097, USA.
V. Perez-Dieste. ALBA Synchrotron Light Facility, Universitat Autònoma de Barcelona, Edifici Ciencies, 08193 Bellaterra, Barcelona, Spain.
D.Y. Petrovykh. Physics Department, University of Maryland, College Park, MD 20742 and Naval Research Laboratory, Washington, DC 20375, USA.
N.L. Abbott. Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.
F.J. Himpsel.[2] Department of Physics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

[1]This article is part of a Special Issue dedicated to Professor G. Michael Bancroft.
[2]Corresponding author (e-mail: fhimpsel@wisc.edu).

 doi:10.1139/V07-079

ity of immunoassays and protein sensors, it is important to control the orientation of proteins adsorbed on a surface (5). Consequently, the development of methods to assess the orientation of proteins on surfaces has been a focus of considerable research activity in recent years. Surface functionalization schemes, such as self-assembled monolayers (SAM) and phospholipid bilayers, make it possible to specifically immobilize proteins on surfaces, control their surface density, increase their orientational order, and prevent them from denaturing.

Biological activity of molecules in solution is typically measured using molecules labeled with tags that can be easily observed and (or) manipulated. These tags can be, for example, fluorescent, radioactive, chemiluminescent, magnetic, and electrochemical. However, in many cases labeled measurements can be difficult or even impossible for biomolecules immobilized on a surface. Surface analysis techniques such as ellipsometry, attenuated total reflection, quartz microbalance, and surface plasmon resonance can be used to measure the wet and dry mass uptake by biologically functionalized surfaces (1, 6). To understand the chemistry and structure of these surfaces, however, it is desirable to obtain element-specific and bond-specific information. This is where sophisticated techniques are required, such as FTIR (Fourier transform IR spectroscopy), XPS (X-ray photoelectron spectroscopy), and NEXAFS (near edge X-ray absorption fine structure spectroscopy) (1, 7, 8).

NEXAFS technique

NEXAFS measures optical dipole transitions from a core level to unoccupied valence orbitals, as shown in Figs. 1 and 2. The binding energy of the core level identifies a specific element and its charge, for example, neutral carbon and positively charged carbon in Fig. 2. The binding energy and its chemical shift can be measured independently by XPS. NEXAFS goes beyond XPS by identifying a specific σ* or π* orbital located at a particular atom (left side of Fig. 2, from ref. 9). Furthermore, the orientation of the orbital can be determined using optical dipole selection rules from the polarization dependence of the transition intensity. Accordingly, NEXAFS measurements require synchrotron radiation because there are no other tunable polarized light sources available in the soft X-ray regime. Typical elements in biomolecules, such as C, N, O, and S, exhibit simple s-to-p transitions, which are dipole-allowed if the electric field vector $\boldsymbol{E}$ of the incident X-rays is parallel to the transition dipole moment, which is perpendicular to the plane of peptide bond in proteins and perpendicular to the plane of the nucleotide bases in DNA. The observed intensity follows a $\cos^2\theta_E$ pattern around that axis. A measurement of the transition intensity makes it possible to determine the amount of certain bonds on a surface.

Usually, the optical absorption is measured indirectly by detecting decay products of the core hole, such as the total electron yield (TEY), the fluorescence yield (FY), or the Auger electron yield (Fig. 1). All of these decay products are proportional to the number of core holes created in the absorption process. The probing depth, however, varies. Electron detection is surface sensitive (typically 1–2 nm), while fluorescence detection probes deeper (typically

Fig. 1. Schematic of various methods to measure X-ray absorption via decay products of the core hole that is created in the absorption process. Detecting energy-selected fluorescence photons provides sensitivity to dilute species. Detecting Auger electrons, photoelectrons, and secondary electrons provides surface sensitivity.

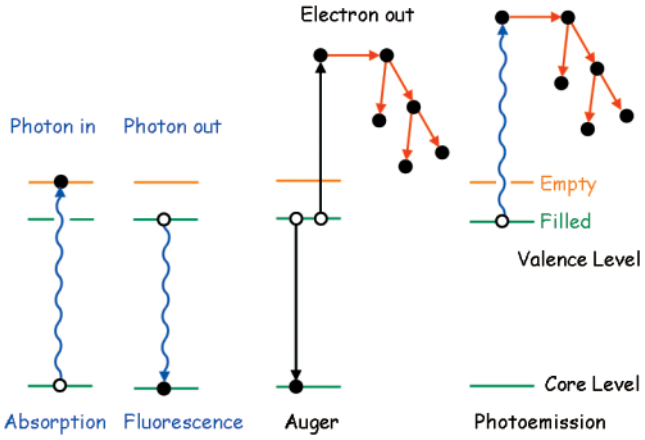
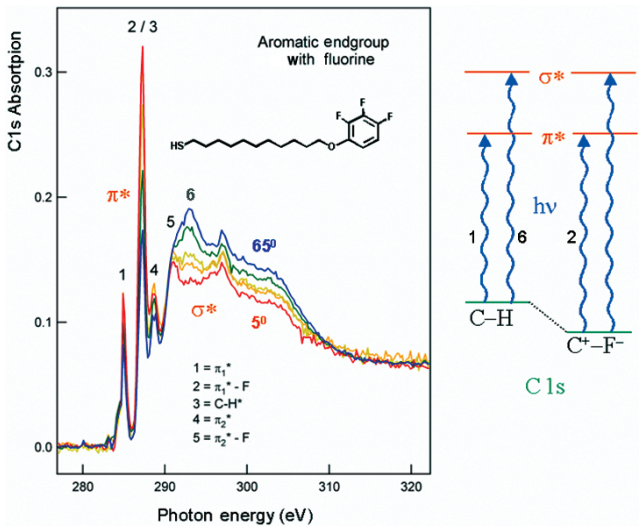


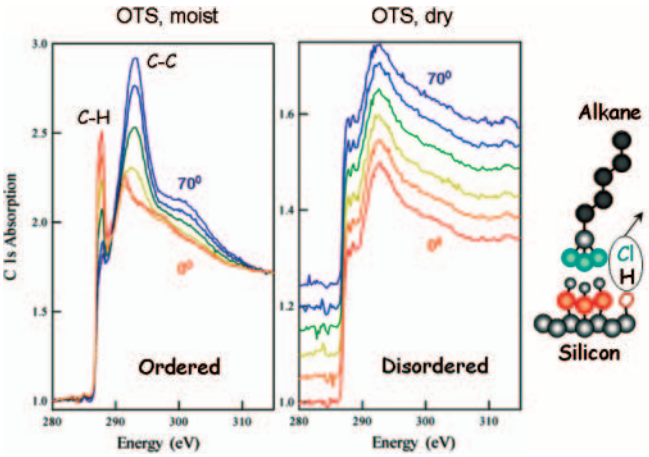
Fig. 2. Chemical selectivity of NEXAFS, demonstrated by a self-assembled monolayer of alkane chains with aromatic terminal groups (9). Transitions from the C 1s core level to unoccupied valence orbitals are observed. Specific atoms can be addressed by selecting the core level and by using its chemical shift to determine the oxidation state (here neutral C^0 in C-H versus C^{1+} in C-F). Specific valence orbitals are addressed via the final state, such as σ* and π*. The orientation of an orbital is obtained from its polarization dependence when changing the angle of incidence θ (measured from the surface normal).



100 nm to several μm, depending on the absorption length at the photon energy of the emitted photons). By simultaneously measuring electron and fluorescence yield, it is possible to obtain a coarse depth profile of the sample.

Electron detection produces a much higher signal than fluorescence, since most of the core holes in the 100–1000 eV binding energy range decay by an Auger process. Fluorescence decay occurs 10^4–10^5 times less frequently. The decay of an Auger electron into multiple secondary electrons am-

Fig. 3. Functionalization of an oxidized silicon surface via siloxane chemistry (11). The coverage can be inferred from the height of the C 1s edge jump relative to the pre-edge background (which is normalized to 1). The orientation of the alkane chains is obtained from the polarization dependence. Both depend strongly on the preparation, with a dry environment leading to low coverage and random orientation.



plifies the signal further in the total electron yield mode (see Fig. 1). Nevertheless, fluorescence detection makes it possible to filter out photons characteristic of the substrate, thereby enhancing the signal-to-noise ratio for dilute systems, such as a fraction of a monolayer of DNA or protein immobilized at a surface. This trade-off between electron and photon detection will be demonstrated in the section on DNA and proteins. A channel plate with an Al filter is able to discriminate against a Si substrate, because Al absorbs the red-shifted Si fluorescence optimally ($Z_{Al} = Z_{Si} - 1$). Unfortunately, there is no good filter material to absorb C 1s fluorescence while transmitting the N 1s edge, which is of interest for DNA and proteins. In that case, a more sophisticated Si–Li detector or a multilayer band pass reflector is called for. Auger electron detection via an electron spectrometer allows some filtering as well. In Auger detection, however, the NEXAFS features are superimposed on the photoemission spectrum of the occupied valence orbitals sweeping through the fixed Auger energy window.

Self-assembled monolayers (SAMs)

Self-assembled monolayers of alkanes are widely used as the first step of chemically functionalizing surfaces. The two prevailing methods of attaching alkanes to a surface utilize thiol groups (on noble metals) or siloxanes (on oxidized silicon) (10). Figure 2 shows the NEXAFS spectrum of an alkanethiol on gold (10), and Fig. 3 of a siloxane on oxidized Si (OTS = octadecyltrichlorosilane) (11). The siloxane chemistry can be applied directly to a silicon wafer but requires stringent control over preparation conditions to achieve reproducible and well-defined surface functionalization. A well-ordered, densely packed SAM requires long soaking in solution (Fig. 3, left side), while quick drying in a nitrogen atmosphere produces a randomly oriented layer with much lower coverage (Fig. 3, right side).

The orientation of molecules in SAMs is inferred from the polarization dependence of NEXAFS. Rotating the sample changes the polar angle of incidence and therefore the angle θ of the electric field vector $\boldsymbol{E}$ with respect to the surface normal (for p-polarized light). The spectra on the left of Fig. 3 exhibit a strong θ-dependence, while the spectra on the right are identical to each other to the extent that they can only be distinguished by a vertical offset. On average, the C-H orbitals are oriented perpendicular to the alkane chain and the C-C orbitals parallel to it. Therefore, the intensity maximum of the C-H signal and the intensity minimum of the C-C orbital at $\theta = 90°$ both imply that the chains are, on average, oriented perpendicular to the surface. The fact that the intensity modulation is not quite 100% is usually taken as a sign that the chains are slightly tilted away from the normal (7), forming many domains with a well-defined polar tilt angle α but randomly distributed azimuthal tilt angles ϕ. Independent information from other techniques supports a tilted geometry for alkane-based SAMs. From the NEXAFS spectra alone, it is difficult to distinguish such a situation from a finite spread of tilt angles α around the surface normal. Examples are single-stranded DNA chains (section on DNA and proteins) and the orientation of the terminal group in Fig. 2 (10). These situations will be discussed more quantitatively in the quantitative analysis section.

The coverage can be estimated by normalizing the NEXAFS spectra to their pre-edge region, which is dominated by the substrate. While the spectra of ordered OTS on the left of Fig. 3 rise by 2 units (from 1 to 3), the spectra of disordered OTS on the right increase by only 0.5 units (from 1 to 1.5, see the red (bottom-most) spectrum with zero offset). Thus, the coverage is four times larger for ordered adsorption. At the highest possible coverage, the alkane chains come within a van der Waal's bond distance of each other and are forced to adopt a tilted geometry with dense packing.

Alkane-passivated surfaces can be further functionalized by adding specific terminal groups on top of a SAM. Two examples of functionalized SAMs are shown in Figs. 2 and 4 (from refs. 10 and 12, respectively). Spectroscopically, the terminal group can be distinguished from the supporting alkane chain by its valence orbital or its core level.

Particularly useful for selecting terminal groups are the sharp π^* orbitals, whose cross-section is large compared with the C-H or C-C orbitals of alkane chains. In Fig. 2, the π^* orbitals of the C=C-H groups in the phenyl group (Peak 1) appear below all the other orbitals. The π^* orbitals of the C=C-F groups in the fluorinated part of the phenyl group (Peak 2) overlap with the C-H orbitals of the alkane chain (Peak 3), but their higher cross section makes them dominant. This can be inferred from the π^* intensity ratio of Peaks 2 and 1, which should be 4:2 (counting =C-O- with =C-F) in this system.

In Fig. 4, the terminal group contains Fe, which can be probed selectively at the sharp Fe 2p absorption edge. The 3d transition metals are involved in many biological systems, such as enzymes and photosystems (12). Their sharpest edge is the 2p edge. It is ideal for probing the oxidation state from the multiplet structure induced by the interaction of the 2p core hole with the 3d valence electrons (13) as well as covalent bonding effects (14, 13). The 1s edges are much broader, but they are needed for probing highly dilute metallo-proteins in solution. The functionalization of alkane

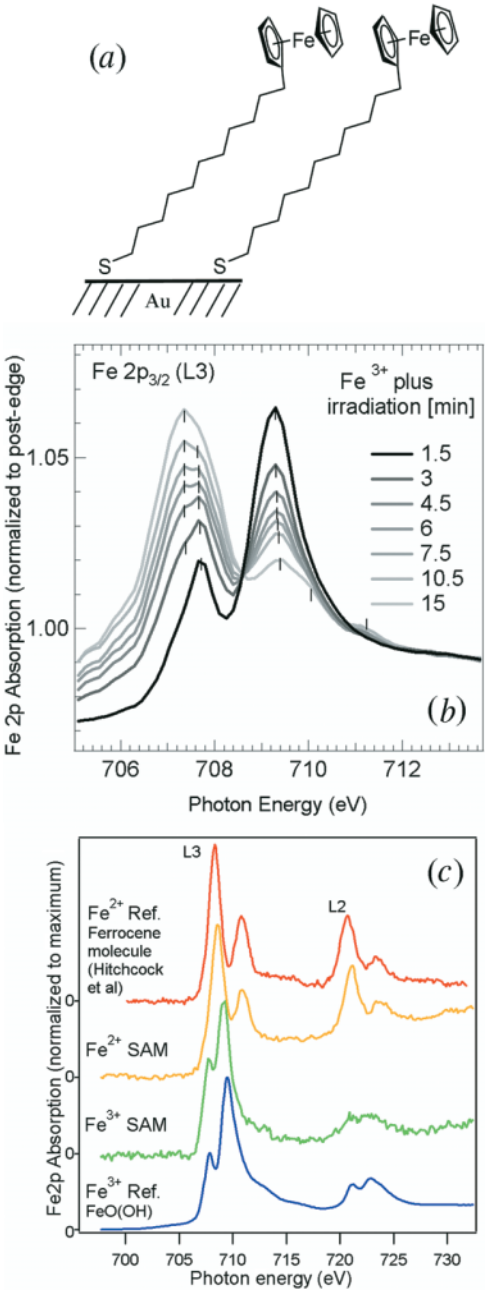
chains by ferrocene in the 2+ and 3+ oxidation states allows for prototypical experiments that investigate electrochemical and photochemical switching of the oxidation state. The multiplet structure of Fe^{2+} and Fe^{3+} and their energy shift are obtained via reference spectra, using molecular ferrocene (14) and a Fe^{3+} mineral (Fig. 4*c*, top and bottom spectra). They compare well to the spectra of two SAMs with ferrocene terminal groups in the 2+ and 3+ oxidation states, which were prepared by electrochemical means (15) (Fig. 4*c*, spectra at the center). Irradiation with intense undulator radiation reduces Fe^{3+} back to Fe^{2+} and further (Fig. 4*b*).

DNA and proteins

Immobilization of DNA and proteins at surfaces is a fast-growing area of research. DNA microarrays are revolutionizing biotechnology by allowing the simultaneous identification of many DNA sequences. They are a critical component of advanced systems for medical diagnostic and pathogen detection. The full potential of chip-based sensors in genomics and proteomics can only be realized if the fragile biomolecules survive the attachment process intact. Furthermore, they must be properly oriented to perform their biological function, for example, binding to a complementary DNA strand in a DNA chip and to a specific protein in an immunosensor. The ability to characterize the chemistry and structure of biomolecules at surfaces facilitates the systematic and controlled development of future generations of DNA chips and protein biosensors.

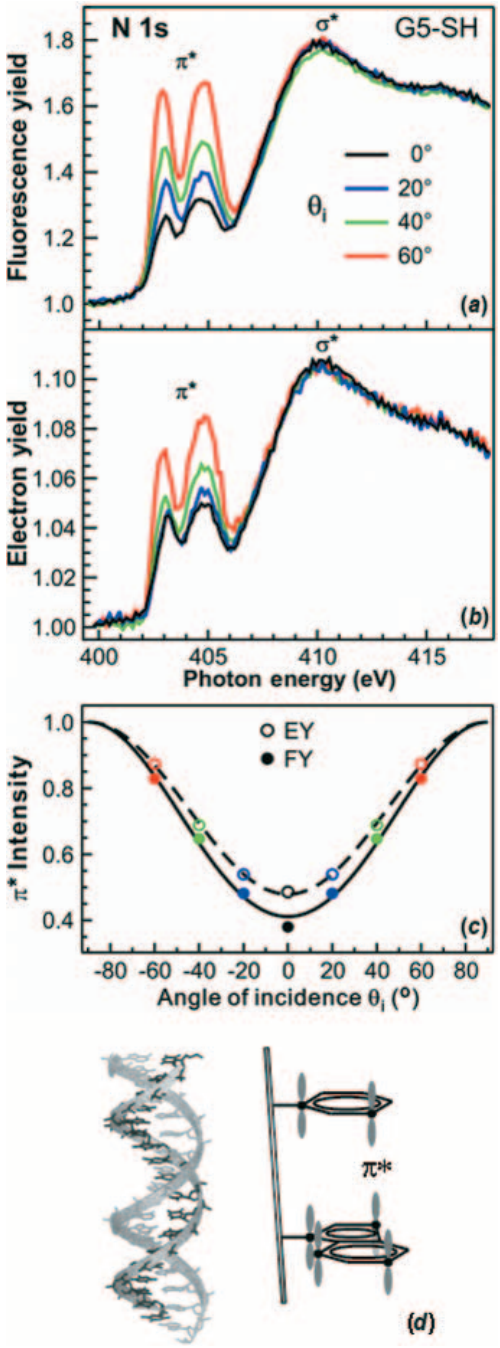
Despite the complexity of biomolecules, NEXAFS is able to provide information about the attachment process and the orientation of immobilized DNA and proteins. The C 1s edge tends to be difficult to analyze because of the multitude of carbon atoms in different chemical environments. However, the N 1s edge singles out specific atoms and bond orbitals. For both DNA and proteins, the lowest unoccupied N valence orbital is a π*, as shown in Figs. 5 and 6. These orbitals are sharp and have a high cross section, which makes them easy to detect. In DNA, the N π* orbitals are located at the base pairs, and they are all oriented parallel to the axis of the double-helix in double-stranded DNA (Fig. 5*d*). In proteins, the N π* orbitals are located at the peptide bonds, which link amino acids into peptides and eventually proteins. This has been demonstrated very clearly in (16), where the N π* orbital is observed only when amino acids are linked together, not in individual amino acids. Fig. 6 shows this peptide bond as the first major peak at the N 1s absorption edge. It also includes the wave function of the peptide bond orbital, which extends over four atoms (Cα, C, N, O) and forces them into a planar arrangement. Even more selective are the absorption edges of minority elements in proteins. S is the most prominent example. It is only part of a few amino acids, such as cysteine, but these are important in cross-linking proteins and attaching them to a surface. Fig. 6 shows the S 1s edge spectrum. As a second row element, sulfur has little tendency towards π bonding. Therefore, the S 1s edge does not exhibit a π* orbital, only a σ*. However, it is sensitive to the formation of S-S crosslinks, for example between two cysteines (17).

Fig. 4. Determination of the oxidation state of Fe in a self-assembled monolayer functionalized by a ferrocene terminal group (15). The multiplet structure of the Fe 2p to 3d transitions reveals the oxidation state. Spectra for molecular ferrocene (from



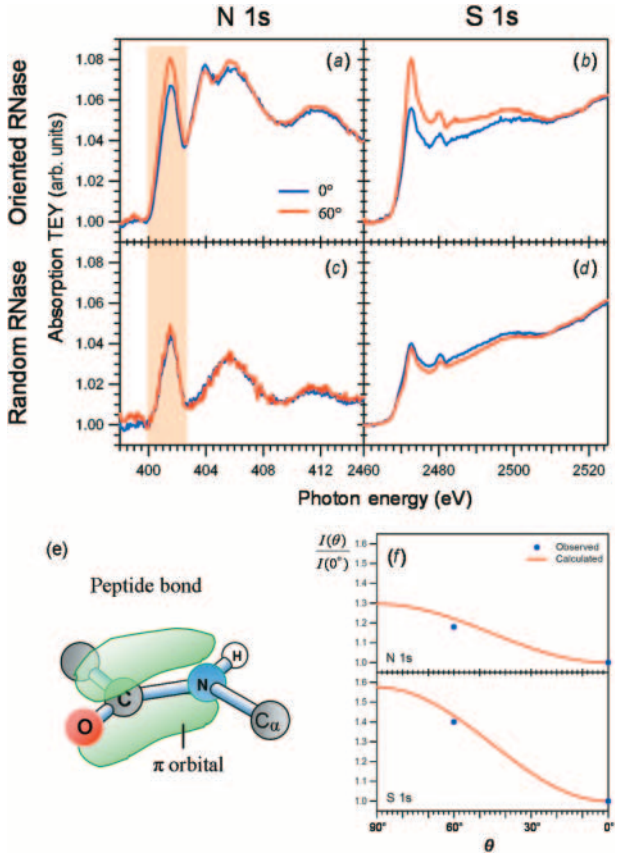
The information from NEXAFS becomes more and more selective by going from C to N, S, and other minority elements, such as transition metals in metallo-proteins (12, 13). NEXAFS of the 2p to 3d transitions of 3d transition metals provides very specific information about the oxidation state and spin state (see Fig. 4). The electron yield signal of dilute elements becomes swamped by background absorption from the C 1s core level, the C 2s, 2p valence orbitals, and the substrate. This problem is solved by fluorescence detection, where the background photons from the C 2s, 2p valence orbitals, and from the substrate can be filtered out. Figure 5 demonstrates the effect of a simple Al filter in front of a

Fig. 5. The π* orbitals of the nucleotide bases in DNA detected at the N 1s edge. The spectra in (*a*)–(*c*) are for single-stranded DNA consisting of 5 guanine nucleotides and attached to a gold surface via a thiol group (18). Fluorescence detection in (*a*) provides a much better signal to background ratio than electron detection in (*b*). The polarization dependence shows that the π* orbitals are oriented perpendicular to the surface (on average). Similar spectra are obtained for double-stranded DNA (19), where the π* orbitals of all base pairs point along the axis of the double helix, as indicated in (*d*).



channel plate detector. After normalizing the data to the pre-edge background, the electron yield in panel 5*b* provides a low signal-to-background ratio of 8%, which increases the noise substantially and makes a reliable background subtraction difficult. The filtered fluorescence measurement in panel 5*a* improves this ratio to 65%, i.e., signal and background become comparable. This ratio can be improved further by using an energy-dispersive Si-Li or Ge detector (not shown). However, it is difficult to maintain the large acceptance angle of a channel plate, and the maximum count rate suffers as the energy resolution of the detector is improved. Detector efficiency is a critical factor with biological samples because radiation damage imposes severe limits on the exposure time.

Fig. 6. Polarization-dependent NEXAFS of a protein (RNase A, from (21)). The N 1s edge reveals the π* orbital of the peptide bond that polymerizes amino acids into a protein (shaded in (*a*), and (*c*)). Despite the summation over 124 amino acids in RNase A, there is substantial polarization dependence at the N 1s and S 1s edges (see (*a*) and (*b*)). A parameter-free calculation is consistent with the data in (*f*). The polarization dependence can be enhanced by choosing a minority species, such as sulfur. The *y*-axis scales are same in panels (*a*) to (*d*). (*a*) and (*b*) are for oriented immobilization, (*c*) and (*d*) for random immobilization (21).



The orientation of DNA nucleobases can be determined from the polarization dependence of the N 1s NEXAFS spectra (18–20). At the C 1s edge, the polarization effects are averaged out because of the many inequivalent C atoms with nearly random bond orientations. Figures 5*a*–5*c* show the polarization dependence for single-stranded DNA oligomers consisting of 5 guanine nucleotides attached to a Au surface via a thiol bond (from ref. 18). Similar measurements have been made for double-stranded DNA (19). The maximum of the π* intensity occurs at grazing incidence, i.e., with the

electric field vector ***E*** perpendicular to the surface. The dipole selection rules lead to the conclusion that the π* orbitals of the bases are preferentially oriented perpendicular to the surface (parallel to ***E***). Therefore, the aromatic rings of the nucleobases are parallel to the surface. A more quantitative evaluation is shown in panel 5*c* for both electron and fluorescence detection. Since the fluorescence measurement is affected by an additional dipole selection rule in the emission process, one has to make sure that the detection geometry does not distort the polarization dependence. This can be achieved by having a large acceptance angle or by placing the detector near the magic angle, where the orientation dependence vanishes. For 100% orientation, one should observe a $\cos^2\theta_E$ dependence of the NEXAFS intensity according to the matrix element for a s to p dipole transition. Experimentally, the $\cos^2\theta_E$ modulation is only 60% of the maximum, which indicates only partial ordering. A quantitative analysis will by given in the section Quantitative analysis. The degree of orientation can be changed (18) by varying the length of the DNA segment, its base sequence, and the preparation conditions. The σ* orbitals generally have a much smaller polarization dependence than the π* orbitals, which can be explained by the fact that an aromatic ring contains σ* orbitals in many in-plane directions, while all the π* orbitals are parallel to each other and perpendicular to the plane.

Proteins are even more complex and less regular than DNA, and yet it has been possible to observe a polarization dependence in RNase A (21), a protein containing 124 amino acids and more than 10^3 atoms (not counting H). The π* orbital at the N 1s edge exhibits a 18% modulation between θ = 0° and 60°, while the S 1s edge is modulated by 41%. This follows the expectation that a less frequent element, such as S, experiences less averaging. The higher σ* orbitals have a much weaker polarization dependence than the π* orbitals, similar to the DNA results. As a further cross-check, RNase A can be immobilized by a different technique that gives random orientation (21). In that case the polarization dependence of the peptide band π* orbital is comparable to the experimental error of a few percent.

The polarization dependence at the N 1s edge is surprisingly strong in view of the 123 peptide bonds contributing to the spectrum. A purely random arrangement of these bonds within one protein molecule (together with perfect orientation of the molecules relative to each other) would give a modulation of $1/\sqrt{123} = 9\%$. Therefore, we have developed a program that sums the dipole matrix elements over all 123 peptide bonds using the atomic coordinates in the protein data base and a specific orientation of the molecules based on the immobilization method. (The protein is bound to the supporting SAM via a S–S bond to the cysteine at the 19th residue, and that bond is assumed to be perpendicular to the surface (21).) The calculated modulation (red curve in Fig. 6*f*) is surprisingly close to the experimental points (blue dots), and the remaining difference goes in the right direction. The data are modulated less than the calculation because the molecules are not perfectly oriented and the polarization of the synchrotron light is not perfect (93% in this experiment). A similar calculation can also be performed for the 12 sulfur atoms in RNase A, and it represents the data equally well. Thus, we conclude that RNase has substantial internal correlations between the orientations of the peptide bonds, which double the statistical polarization dependence. This is quite reasonable in view of the internal structure of proteins, which contain highly regular elements, such as the α-helix and the β-sheet. RNase A, for example, contains 22% α-helices and 46% β-sheets.

Quantitative analysis

NEXAFS has been mostly used for qualitative analysis of the chemical reactions that occur during the immobilization process of biomolecules at surfaces. However, quantitative NEXAFS analysis can be used for determining both coverage and orientation of molecules. Here we mainly address the orientational information that can be obtained from the polarization dependence of the NEXAFS signal. A widely used formula for the polarization dependence (7) assumes a layer of tilted molecules consisting of domains with perfect orientation within each domain. The domains all have the same polar tilt angle α, but a random-distribution of azimuthal angles φ. This leads to the formula,

$$[1] \qquad I(\theta) / I(0°) = 1 + P[(2/\sin^2\alpha) - 3]\sin^2\theta$$

where θ is the angle of incidence from the sample normal ($\theta = 90° - \theta_E$), α is the tilt angle of a p-orbital from normal, and *P* is the degree of polarization of the soft X-rays. These assumptions are reasonable for simple SAMs that are known to be nearly perfectly ordered. However, in more complex systems, such as SAMs with tailored terminal groups, DNA, or proteins, one encounters two phenomena that are not captured by eq. [1].

First, there can be more than a single angle determining the geometry of the system. A terminal group is often able to rotate around the bond that attaches it to the supporting alkane chain. An extreme case is the orientation of a protein at a surface, which requires three angles to be fully characterized. If one allows for different folding patterns, the number of variables increases rapidly.

Second, larger molecules may exhibit a distribution of bond angles instead of being perfectly oriented. That introduces a finite spread σ of the bond angles. Moreover, for a system that contains a distribution of bond orientations, the interpretation of a polarization-dependent measurement in terms of a unique angle, in general, will not result in a value close to the average of that distribution (18, 22, 23).

In all cases, one faces the dilemma that NEXAFS provides at best two experimental numbers for each orbital that characterize the polar and azimuthal polarization dependence of the peak height. Therefore, one needs to carefully select the free parameters when fitting the data and freeze the remaining variables by using other experiments or making judicious assumptions. For example, it makes little sense to assume a fixed polar tilt angle (as in eq. [1]) for characterizing the orientation of floppy molecules, such as single-stranded DNA (18). A more reasonable assumption is a distribution of tilt angles with an average orientation α_0 relative to the surface normal. Using a Gaussian distribution with standard deviation σ leads to the following polarization dependence (24),

[2] $I(\theta) / I(0°) = 1 + P[8z/(3z - 1 - 2\cos(2\alpha_0)) - 3]\sin^2\theta$, where $z = \exp(4\sigma^2)$

With this approach, the free parameter becomes σ, and the average angle α_0 needs to be obtained from symmetry arguments (e.g., $\alpha_0 = 0°$ for DNA (18, 19)), or from other measurements (e.g., X-ray diffraction of the bulk phase (24)).

Another situation occurs in proteins, where the π* signal at the N 1s edge contains contributions from many nitrogen atoms in peptide bonds with different orientations. In that case one needs to sum over the angular dependence of all individual orbitals. Each overall orientation (Θ, Φ) of the protein corresponds to a well-defined polarization dependence $I(\theta, \phi, \Theta, \Phi)$ with respect to the angle of incidence (θ, ϕ). The polarization dependence is calculated by summing over the individual contributions $I_n(\theta, \phi, \Theta, \Phi)$ from the *n*th peptide bond. Each of these follows a $\cos^2\theta_n$ law, where θ_n is the angle between the electric field vector $\boldsymbol{E}$ and the direction $\boldsymbol{e}_n^{\mathrm{N}}(\Theta, \Phi)$ of the π* orbital in the *n*th peptide bond. The vectors $\boldsymbol{e}_n^{\mathrm{N}}$ are obtained from the cross product of the vectors $\boldsymbol{r}_n^{\mathrm{CO}}$ and $\boldsymbol{r}_n^{\mathrm{CN}}$ that define the plane of the peptide bond (see Fig. 6*e*). They connect the carbon atom in the *n*th peptide bond with its neighboring oxygen and nitrogen atoms. For the S 1s edge, the unit vector $\boldsymbol{e}^{\mathrm{S}}{}_m$ connects the *m*th S atom with the adjacent C atom along a σ bond. The atomic coordinates are obtained from the protein databank available from www.pdb.org.

The overall orientation angle Θ of RNase in Figs. 6*a*, 6*b*, and 6*f* is determined by its attachment to the supporting SAM via a S–S bond to the cysteine at the 19th residue. Assuming that this S–S bond is perpendicular to the surface, one needs to rotate the coordinates in the protein data bank such that the vector $\boldsymbol{r}_{19}^{\mathrm{SH}}$ becomes parallel to the surface normal. This is achieved by calculating the transformation matrix for this rotation and applying it to all vectors $\boldsymbol{e}_n$. The azimuthal orientation Φ of the RNase molecules around the S–S bond is assumed to be random. That leads to an average over the azimuthal orientation Φ, which is equivalent to an average over the angle of incidence ϕ with fixed orientation Φ. This approach eliminates the angles Θ, Φ, and ϕ from the problem. The normalized polarization dependence of the NEXAFS signal $I(\theta)/I(0°)$ becomes a function of the angle of incidence θ from the surface normal, without adjustable parameters. For the N 1s and S 1s edges one obtains,

[3*a*]
$$\boldsymbol{e}_n^{\mathrm{N}} = \frac{\boldsymbol{r}_n^{\mathrm{CO}} \times \boldsymbol{r}_n^{\mathrm{CN}}}{\left|\boldsymbol{r}_n^{\mathrm{CO}} \times \boldsymbol{r}_n^{\mathrm{CN}}\right|}$$
$$\frac{I^{\mathrm{N}}(\theta)}{I^{\mathrm{N}}(0°)} = \frac{\int \sum_n [\boldsymbol{E}(\theta)\, \boldsymbol{e}_n^{\mathrm{N}}]^2 \, \mathrm{d}\phi}{\int \sum_n [\boldsymbol{E}(0°)\, \boldsymbol{e}_n^{\mathrm{N}}]^2 \, \mathrm{d}\phi}$$

[3*b*]
$$\boldsymbol{e}_m^{\mathrm{S}} = \frac{\boldsymbol{r}^{\mathrm{SC}}}{\left|\boldsymbol{r}_m^{\mathrm{SC}}\right|}$$
$$\frac{I^{\mathrm{S}}(\theta)}{I^{\mathrm{S}}(0°)} = \frac{\int \sum_m [\boldsymbol{E}(\theta)\, \boldsymbol{e}_m^{\mathrm{S}}]^2 \, \mathrm{d}\phi}{\int \sum_m [\boldsymbol{E}(0°)\, \boldsymbol{e}_m^{\mathrm{S}}]^2 \, \mathrm{d}\phi}$$

Applying these formulas to the N 1s and S 1s data of oriented RNase A in Fig. 6 gives the correct magnitude of the polarization dependence (see panel (*f*)). This calculation could be refined further by including the degree of polarization *P* and a distribution of orientation angles (compare eqs. [1] and [2]). Both would bring the calculation closer to the experimental values.

Apart from the polarization dependence, there are several other quantitative results that can be extracted from NEXAFS data of proteins. Spatially resolved spectromicroscopy (25, 26) maps the distribution of proteins in biological samples.[3] The spectrum of a protein can be modeled (27) by using reference spectra for individual amino acids (28) and taking the changes due to the formation of the peptide bond into account. For the determination of relative concentrations, it is useful to normalize the spectra to the edge jump between the pre-edge and the continuum absorption 20–30 eV above the edge. This normalization provides a signal proportional to the density of C, N, O, and S atoms.

Summary and outlook

In summary, we demonstrate how the element and orbital sensitivity of NEXAFS spectroscopy can be used to investigate the binding properties of biomolecules at surfaces. Simple alkane passivation layers serve as tutorial examples. The complexity is increased systematically by adding customized terminal groups. Eventually, DNA and proteins are tackled. The orientation of the immobilized protein is detected from the polarization dependence of the NEXAFS intensity. The results suggest that NEXAFS can be used to investigate fairly complex biomolecules at surfaces, such as DNA oligomers and proteins containing more than a hundred amino acids. The agreement between the results of complementary NEXAFS, XPS, and FTIR measurements (18) strongly suggests that these ex situ techniques capture salient properties of biomolecules as they are immobilized in solution. Such element-specific surface-sensitive diagnostic methods help to monitor and control the multistep immobilization process of biomolecules.

To have an impact on day-to-day biochemical research, it will be necessary to simplify the access to the required synchrotron light sources and the analysis of the results. Protein crystallography can serve as a model, where automation in sample handling and data processing has become routine. Even remote control data acquisition has become possible, as long as there is a synchrotron staff member available to load the samples. Due to the vacuum requirements for NEXAFS, the sample introduction becomes more complicated. However, it is possible to build sample cells with thin Si or Si_3N_4 windows that separate the sample atmosphere from the detector and the synchrotron, while transmitting enough light to operate in the fluorescence detection mode. In the "water window" between the C 1s and O 1s edges, the mean free path of soft X-rays is only a few μm in liquid water (4 μm at the N 1s edge). This length increases to 0.3 m in saturated water vapor (18 Torr at 100% humidity; 1 torr = 133.322 4 Pa). That makes sample handling much easier. In addition, the radiation damage from radicals created by

[3] C.J. Johnsen, R.M. Olabisi, R.A. Metzler, B.Gilbert, B.H. Frazer, D. McKenzie, J.M. Aiken, and P.U.P.A. Gilbert. Manuscript in preparation.

photodecomposition of the liquid water film between sample and window is avoided. Developing practical "wet" or "humid" sample cells is an important step in bringing NEXAFS closer to a mainstream analytical technique for surface biochemistry.

Looking farther into the future, one could go beyond self-assembled monolayers and consider more complex objects, such as phospholipid membranes, supported by molecular stilts (29). Like a cell wall, such a membrane may contain ion channels, photosynthetic centers, molecular motors, etc. Compared to a natural cell, this is a more controllable environment that can be used for selectively investigating a specific part of the incredibly complex cell machinery.

Acknowledgment

This work was supported by the National Science Foundation (NSF) under awards DMR-0520527 (MRSEC) and DMR-0084402 (SRC), and by the DOE under contracts No. DE-FG02–01ER45917 and No. DE-AC03–76SF00098 (ALS). Financial support for CSRF is provided by Natural Sciences and Engineering Research Council of Canada (NSERC) and NRC–CRNC (National Research Council Canada – Conseil National de Recherches Canada). Work at NRL (DYP) was supported by the Air Force Office of Scientific Research.

References

1. B. Kasemo. Surf. Sci. **500**, 656 (2002); D.G. Castner and B.D. Ratner. Surf. Sci. **500**, 28 (2002).
2. J.D. Hoheisel. Nat. Rev. Genet. **7**, 200 (2006); M.C. Pirrung. Angew. Chem., Int. Ed. **41**, 1277 (2002); S. Draghici, P. Khatri, A.C. Eklund, and Z. Szallasi. Trends Genet. **22**, 101 (2006).
3. H. Zhu, M. Bilgin, R. Bangham, D. Hall, A. Casamayor, P. Bertone. N. Lan, R. Jansen, S. Bidlingmaier, T. Houfek, T. Mitchell, P. Miller, R.A. Dean, M. Gerstein, M. Snyder. Science, **293**, 2101 (2001); G. MacBeath, S.L. Schreiber. Science, **289**, 1760 (2000); C.-H. Jang, M.L. Tingey, N.L. Korpi, G.J. Wiepz, J.H. Schiller, P. Bertics, N.L. Abbott. J. Am. Chem. Soc. **25**, 8912 (2005); A. Guo and X. Zhu. Int. J. Nanosci. **6**, 109 (2007).
4. S.A. Soper, K. Brown, A. Ellington, B. Frazier, G. Garcia-Manero, V. Gau, S.I. Gutman, D. F. Hayes, B. Korte, J.L. Landers, D. Larson, F. Ligler, A. Majumdar, M. Mascini, D. Nolte, Z. Rosenzweig, J. Wang, and D. Wilson. Biosens. Bioelectron. **21**, 1932 (2006); J.J. Gooding. Anal. Chim. Acta, **559**, 137 (2006); V.K. Gupta, J.J. Skaife, T.B. Dubrovsky, N.L. Abbott, Science, **279**, 2077 (1998).
5. Y.Y. Luk, M.L. Tingey, D.J. Hall, B.A. Israel, C.J. Murphy, P.J. Bertics, N.L. Abbott. Langmuir, **19**, 1671 (2003); Y.Y. Luk, M.L. Tingey, D.J. Hall, K.A. Dickson, R.T. Raines, N.L. Abbott. J. Am. Chem. Soc. **126**, 9024 (2004).
6. M. Bally, M. Halter, J. Voros, and H.M. Grandin. Surf. Interface Anal. **38**, 1442 (2006); A. Janshoff, H.-J. Galla, and C. Steinem. Angew. Chem., Int. Ed. **39**, 4004 (2000).
7. J. Stöhr. NEXAFS Spectroscopy. Springer-Verlag, New York. 1996.
8. D.Y. Petrovykh, H. Kimura-Suda, L.J. Whitman, and M.J. Tarlov. J. Am. Chem. Soc. **125**, 5219 (2003); D.Y. Petrovykh, H. Kimura-Suda, M.J. Tarlov, and L.J. Whitman. Langmuir, **20**, 429 (2004).
9. Y.-Y. Luk, N.L. Abbott, J.N. Crain, and F.J. Himpsel. J. Chem. Phys. **120**, 10792 (2004).
10. F. Luderer and U. Walschus. Top. Curr. Chem. **260**, 37 (2005); J.C. Love, L.A. Estroff, J.K. Kriebel, R.G. Nuzzo, and G.M. Whitesides. Chem. Rev. (Washington, DC), **105**, 1103 (2005).
11. R.D. Peters, P.F. Nealey, J.N. Crain, and F.J. Himpsel. Langmuir, **18**, 1250 (2002).
12. S.P. Cramer, J. Chen, S.J. George, J. Vanelp, J. Moore, O. Tensch, J. Colaresi, M. Yocum, O.C. Mullins, C.T. Chen. Nucl. Instrum. Methods Phys. Res., Sect. A, **319**, 285 (1992); X. Wang, M.M. Grush, A.G. Froeschner, S.P. Cramer. J. Synchrotron Rad. **4,** 236 (1997); M.C. Corbett, F.A., Tezcan, O. Einsle, M.Y. Walton, D.C. Rees, M.J. Latimer, B. Hedman, K.O. Hodgson. J. Synchrotron Rad. **12**, 28 (2005).
13. F. De Groot. Coord. Chem. Rev. **249**, 31 (2005).
14. A.P. Hitchcock, A.T. Wen, E. Rühl. Chem. Phys. **147**, 51 (1990).
15. F. Zheng, V. Pérez-Dieste, J.L. McChesney, Y.-Y. Luk, N.L. Abbott, and F.J. Himpsel. Surf. Sci. **587**, L191 (2005).
16. M.L. Gordon, G. Cooper, C. Morin, T. Araki, C.C. Turci, K. Kaznatcheev, A.P. Hitchcock. J. Phys. Chem. A, **107**, 6144 (2003).
17. A. Rompel, R.M. Cinco, M.J. Latimer, A.E. McDermott, R.D. Guiles, A. Quintanilha, R.M. Krauss, K. Sauer, V.K. Yachandra, and M.P. Klein. Proc. Natl. Acad. Sci. USA, **95**, 6122 (1998).
18. D.Y. Petrovykh, V. Perez-Dieste, A. Opdahl, H. Kimura-Suda, J.M. Sullivan, M.J. Tarlov, F.J. Himpsel, and L.J. Whitman. J. Am. Chem. Soc**. 128**, 2 (2006).
19. J.N. Crain, A. Kirakosian, J.-L. Lin, Y. Gu, R.R. Shah, N.L. Abbott, and F.J. Himpsel. J. Appl. Phys. **90**, 3291 (2001).
20. K. Fujii, K. Akamatsu, and A. Yokoya. J. Phys. Chem. B, **108**, 8031 (2004); C.Y. Lee, P. Gong, G.M. Harbers, D.W. Grainger, D.G. Castner, and L.J. Gamble. Anal. Chem. **78**, 3316 (2006); N.T. Samuel, C.Y. Lee, L.J. Gamble, D.A. Fischer, and D.G. Castner. J. Electron Spectrosc. Relat. Phenom. **152**, 134 (2006).
21. X. Liu, C.H. Jang, F. Zheng, A. Jürgensen, J.D. Denlinger, K.A. Dickson, R.T. Raines, N.L. Abbott, and F.J. Himpsel. Langmuir, **22**, 7719 (2006).
22. G.J. Simpson and K.L. Rowlen. J. Am. Chem. Soc. **121**, 2635 (1999).
23. F. Zheng, J.L. McChesney, X. Liu, and F.J. Himpsel. Phys. Rev. B, **73**, 205315 (2006).
24. F. Zheng, B.-N. Park, S. Seo, P.G. Evans, and F.J. Himpsel. J. Chem Phys. **126**, 154702 (2007).
25. L. Li., A.P. Hitchcock, N. Robar, R. Cornelius, J.L. Brash, A. Scholl, and A. Doran. J. Phys. Chem. B, **110**, 16763 (2006).
26. C.S. Chan, G. De Stasio, S.A. Welch, M. Girasole, B.H. Frazer, M.V. Nesterova, S. Fakra, and J.F. Banfield. Science, **303**, 1656 (2004).
27. J. Stewart-Ornstein, A.P. Hitchcock, D. Hernandez-Cruz, P. Henklein, J. Overhage, K. Hilpert, J.D. Hale, and R.E.W. Hancock. J. Phys. Chem. B, **111**, 7691 (2007).
28. Y. Zubavichus, A. Shaporenko, M. Grunze, and M. Zharnikov. J. Phys. Chem. A, **109**, 6998 (2005).
29. B.A. Cornell, V.L.B. Braach-Maksvytis, L.G. King, P.D.J. Osman, B. Raguse, L. Wieczorek, and R.J. Pace. Nature **387**, 580 (1997); M. Tanaka and E. Sackmann. Phys. Status Solidi A, **203**, 3452 (2006); E.T. Castellana and P.S. Cremer. Surf. Sci. Rep. **61**, 429 (2006).