

Using Intrinsic X-ray Absorption Spectral Differences To Identify and Map Peptides and Proteins

Jacob Stewart-Ornstein,[†] Adam P. Hitchcock,^{*,†} Daniel Hernández Cruz,[†] Peter Henklein,[‡] Joerg Overhage,[§] Kai Hilpert,[§] John D. Hale,[§] and Robert E. W. Hancock[§]

BIMR and Department of Chemistry, McMaster University, Hamilton, Ontario, Canada L8S 4M1, Institut für Biochemie, Universitätsklinikum Charité, Humboldt-Universität, 10117 Berlin, Germany, and Department of Microbiology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1Z4

Received: March 15, 2007; In Final Form: April 29, 2007

The intrinsic variation in the near-edge X-ray absorption fine structure (NEXAFS) spectra of peptides and proteins provide an opportunity to identify and map them in various biological environments, without additional labeling. In principle, with sufficiently accurate spectra, peptides (<50 amino acids) or proteins with unusual sequences (e.g., cysteine- or methionine-rich) should be differentiable from other proteins, since the NEXAFS spectrum of each amino acid is distinct. To evaluate the potential for this approach, we have developed X-SpecSim, a tool for quantitatively predicting the C, N, and O 1s NEXAFS spectra of peptides and proteins from their sequences. Here we present the methodology for predicting such spectra, along with tests of its precision using comparisons to the spectra of various proteins and peptides. The C 1s, N 1s, and O 1s spectra of two novel antimicrobial peptides, Indolicidin (ILPWKWPWWPWR-NH₂) and Sub6 (RWKWIWVIR-WWR-NH₂), as well as human serum albumin and fibrinogen are reported and interpreted. The ability to identify, differentiate, and quantitatively map an antimicrobial peptide against a background of protein is demonstrated by a scanning transmission X-ray microscopy study of a mixture of albumin and sub6.

1. Introduction

Although protein detection relative to other biological macromolecules (DNA, polysaccharides, lipids, etc.) by spatially resolved near-edge X-ray absorption fine structure (NEXAFS) spectromicroscopy is well-developed,^{1–4} to date there has been only one example⁵ in which two different proteins were differentiated and mapped using the NEXAFS spectra of the unlabeled species. In principle, with sufficiently accurate spectra, peptides (usually defined as having <50 amino acids) or proteins with unusual sequences (e.g., cysteine- or methionine-rich) should be differentiable since the NEXAFS spectrum of each amino acid is distinct.^{6,7} We are interested in exploiting these intrinsic spectral differences to map specific peptides or proteins in various biological environments. In particular, the motivation for the work described in this paper is to develop and use this capability to locate cationic antimicrobial peptides (CAPs) in biofilms, thereby assisting efforts to investigate the mechanisms of their actions and the ability to penetrate bacterial cells.

Cationic antimicrobial peptides are small (10–50 amino acids), hydrophobic (>30% hydrophobic amino acids), positively charged (net charge between +2 and +9 at neutral pH) peptides produced by a diverse range of organisms including humans, insects, and plants.⁸ Hundreds of natural CAPs having been identified, and more have been designed through the engineering of existing CAPs that aim to produce peptides with increased potency.^{9,10} The majority of CAPs work by interacting and disrupting the integrity of the cytoplasmic membrane in a susceptible bacterial cell.¹¹ However, many examples exist where alternative targets have been identified.¹² The ability to

visualize CAPs in biofilms could help in understanding their mode of action and thereby assist in refining tools that allow prediction of peptide sequences with even greater activity. The rate of advance in this field could be significantly increased if it was possible to determine the location and local concentrations of CAPs within individual cells or in complex microbial communities.

We are using scanning transmission X-ray microscopy (STXM)^{13–15} to record spectra of pure peptides and proteins in the solid state and also to investigate complex biological samples.^{1–3,16,17} The spectra of peptides and proteins on polymer surfaces is being examined by photoemission electron microscopy (X-PEEM).^{4,18} These techniques compliment more common methods for mapping proteins such as laser scanning microscopy of fluorescent labeling proteins,¹⁹ which has high specificity but limited spatial resolution, and immunoelectron microscopy, which has high spatial resolution but requires a chemically fixed sample.²⁰

The X-ray absorption spectra of proteins are quite distinct from that of other biological materials, especially at the C 1s edge. This is the basis for mapping locations and concentrations of proteins relative to other biological materials or carbon-based polymers.^{1–5} Differentiation among proteins is, in contrast, quite difficult and relies on spectral differences caused by differing amino acid composition or, potentially, structure. To begin to locate specific peptides or proteins in a biological sample, it is necessary to understand and exploit these spectral differences, which are often very small. It is impractical to record spectra of all proteins or peptides of interest. In most cases it is likely they would prove too similar to other proteins to provide analytically useful spectral differentiation. Thus, it would be useful to have a means of predicting the NEXAFS spectra of peptides or proteins to evaluate the feasibility of a proposed

[†] McMaster University.

[‡] Universitätsklinikum Charité.

[§] University of British Columbia.

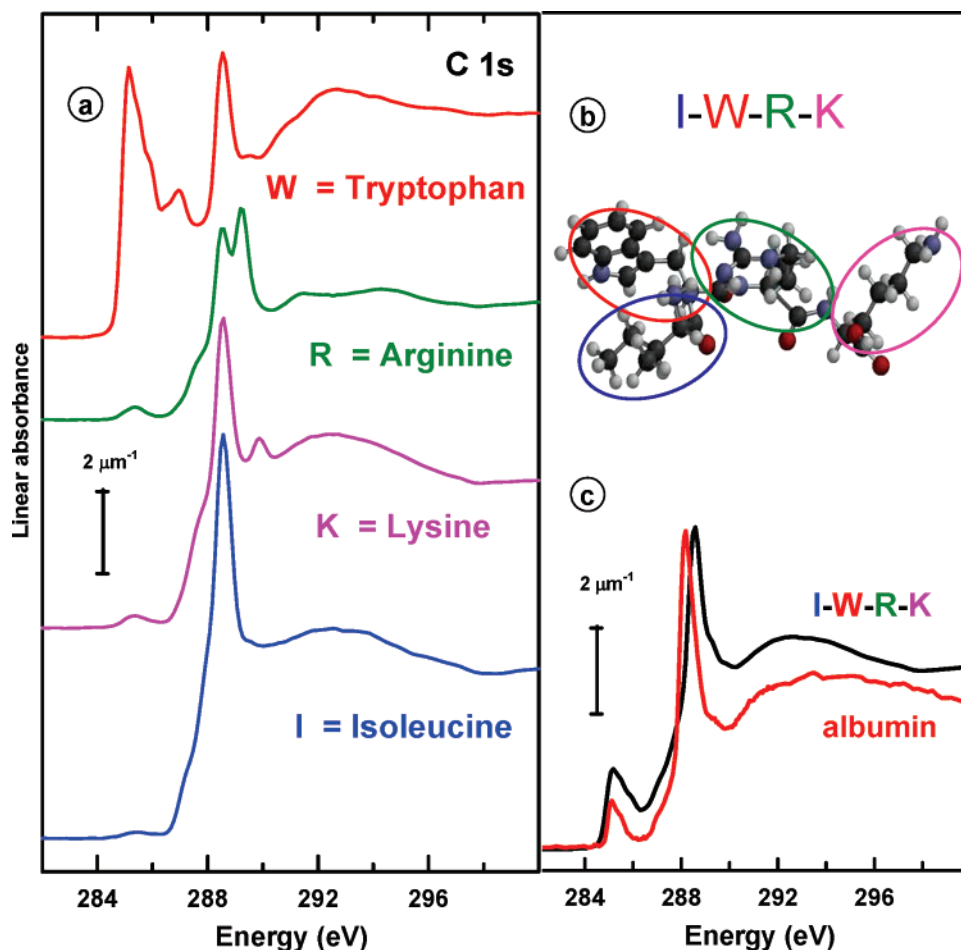


Figure 1. Illustration of the NEXAFS building block model with a tetrapeptide (IWRK-NH₂): (a) C 1s spectra of the individual amino acids; (b) ball and stick model of IWRK-NH₂; (c) simple sum of C 1s spectra of the four amino acids, compared to the experimental C 1s spectrum of human serum albumin.

detection and visualization scheme. Fortunately, the spectra of large molecules often resemble a sum of the functional groups present—the building block hypothesis.²¹ Though this is typically applied to small molecules (as in a recent comparison of the spectra of phenylalanine to those of benzene and alanine²²), studies of small peptides (two to three monomer units) have shown that their spectra are similar to the sum of the spectra of their amino acid constituents.^{23,24} The recent publication of high-quality C 1s, N 1s, and O 1s spectra of the 20 common amino acids⁷ in their zwitterionic neutral form makes it possible to predict the spectra of any peptide or protein using a weighted sum of the spectra of its constituent amino acids. To evaluate the potential for this approach, we have developed a software tool for quantitatively predicting the C 1s, N 1s, and O 1s NEXAFS spectra of peptides and proteins from their sequences, within a modified building block approach. Specifically our code generates a sum based on the sequence and then adds a distortion which mimics the spectral modifications arising from the structural changes associated with peptide bond formation.^{24,25} With the X-ray absorption spectral simulator (X-SpecSim) we are able to explore the applicability of NEXAFS spectral differentiation to any size of peptides and even to proteins made up of hundreds of amino acids. Here we present the methodology for predicting such spectra, explain how it deals with the spectral changes associated with peptide bond formation, and present tests of its precision by comparing its predictions to the experimental spectra of two novel antimicrobial peptides and two blood proteins, human serum albumin (HSA) and fibrinogen. To demonstrate that intrinsic differences in their

NEXAFS spectra enable identification and mapping, we conclude by presenting a successful quantitative analysis of a spatially heterogeneous peptide–protein mixture by STXM using the intrinsic C 1s spectral differences.

Other recent work to simulate the NEXAFS spectra of proteins includes a very elegant study by Liu et al.²⁶ who took this concept in some sense one step further, by analyzing the full 3D structure of a ribonuclease to predict the polarization dependence of the N 1s → π^* _{amide} and S 1s → σ^* _{S-S} transitions in the NEXAFS spectra of an oriented sample. They were able to show that polarization-dependent NEXAFS could be used to differentiate oriented and unoriented ribonuclease molecules.

2. Experimental Section

2.1. Sample Sources and Preparation. Sub6 (RWWKI-WVIRWWR-NH₂),²⁷ an optimized variant of Bac2A, which in turn is a linearized version of bactenecin, a naturally occurring bovine peptide,²⁸ was synthesized by P. Henklein using 9-fluorenylmethyl carbamate (Fmoc) solid-phase synthesis. Indolicidin (ILPWKWPWWPWR-NH₂) was synthesized by P. Owen at the Peptide Synthesis Facility, Biomedical Research Centre, UBC, using tertiary butyloxycarbonyl (tBOC) solid-phase synthesis.^{29,30} Both peptides were purified by HPLC and their respective masses confirmed by mass spectrometry.

Human serum albumin (HSA) was obtained from Behringwerke AG, Marburg, Germany, and found to be homogeneous as judged by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Plasminogen-free human plasma

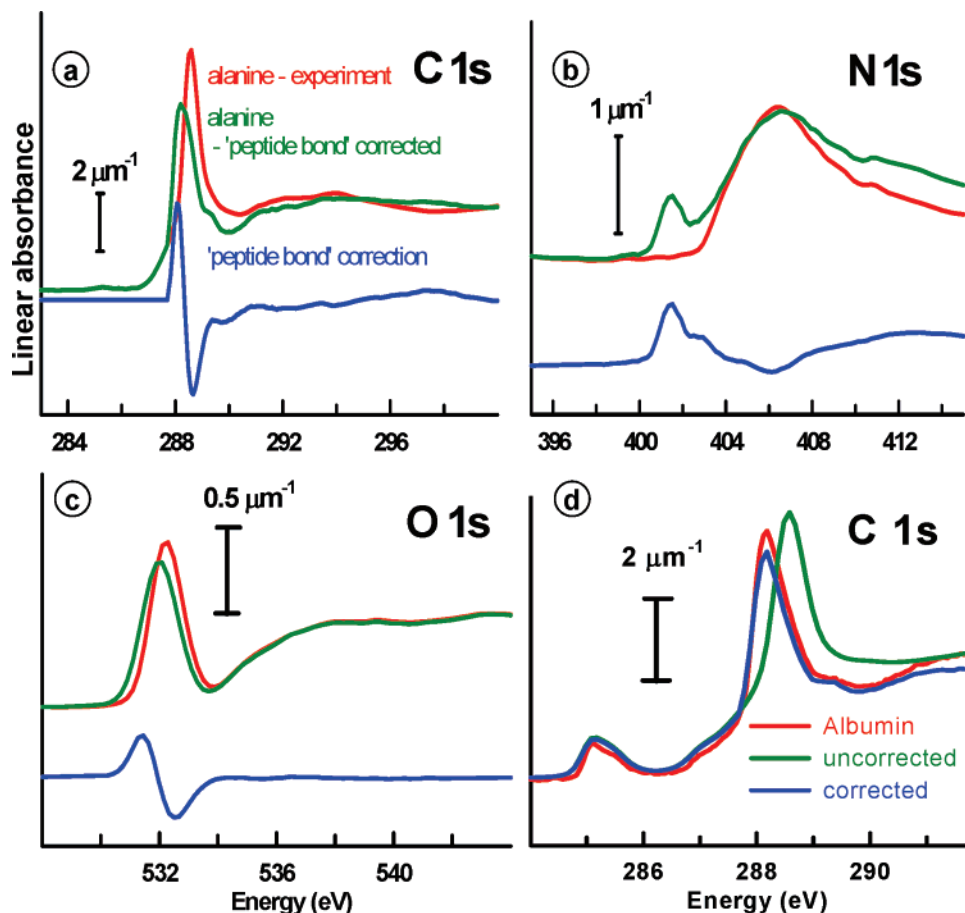


Figure 2. Demonstration of correction for the peptide shift: (a) experimental C 1s spectrum of alanine (red) compared to the peptide-bond corrected spectrum (green) (i.e., the spectrum expected of polyalanine) and the C 1s “peptide bond spectrum”, which is the correction applied; (b) experimental N 1s spectrum of alanine compared to the peptide-bond corrected spectrum and the N 1s peptide bond spectrum; (c) experimental O 1s spectrum of alanine compared to the peptide-bond corrected spectrum and the O 1s peptide bond spectrum; (d) the C 1s spectrum of human serum albumin, predicted with and without the peptide bond correction, compared to the experimental spectrum.

fibrinogen (Calbiochem) was used. It is reported to be >95% clottable by thrombin and pure as judged by SDS-PAGE.

Samples of antimicrobial peptides and human serum albumin were prepared by dissolving lyophilized material in deionized water to a concentration of ~ 1 mg/mL. A few microliters of this solution was deposited onto a Si_3N_4 window and allowed to air-dry. The resulting nonuniform film was examined for regions of appropriate thickness for STXM by optical microscopy. In the case of fibrinogen, initial attempts to dissolve and solvent-cast gave unreliable results, probably due to limited solubility and the presence of highly soluble lyophilization salts. This problem was circumvented by placing small amounts of the solid on a Si_3N_4 window and then solubilizing and removing most of the material by a micropipette, until areas thin enough for STXM (~ 100 nm) were left. A combined sample with Sub6 on HSA was prepared by first solvent-casting HSA onto a Si_3N_4 window, then sprinkling Sub6 powder on that surface, and blowing off the excess.

2.2. Scanning Transmission X-ray Microscopy. STXM³¹ at the dedicated polymer STXM beamline 5.3.2³² of the Advanced Light Source (ALS) was used to measure the spectra of the pure peptides and proteins. STXM is a microscopy-based version of NEXAFS spectroscopy, which combines good energy resolution (<0.1 eV) with a spatial resolution (~ 40 nm) intermediate between optical techniques and electron or scanning probe microscopies. The high spatial resolution was used to find micrometer-sized regions of suitable thickness (~ 100 nm) in the solvent-cast samples. Typically spectra are recorded using

image sequences³³ because the samples are only homogeneous over small areas, and the image sequence can be carefully aligned postacquisition. Image sequence spectra also minimize the radiation dose. Checks were made after each measurement to ensure the radiation damage was negligible.

2.3. Method of Predicting Spectra. The key tenet of the building block model²¹ is that the spectral signature of a functional group does not change from the reference compound to the final chemical compound or that if there is a change, it is uniform and well-quantified. The polymerization of amino acids to form peptides results in the removal of two functional groups—the carboxylic acid and amine groups—and the creation of a new group—the amide. There are modifications to the C 1s, N 1s, and O 1s NEXAFS spectra associated with this change.²⁴ To produce reasonable simulations of the spectra of peptides and proteins, these changes must be reproduced by the software. The spectral modification associated with the polymerization of amino acids (the formation of the peptide bond) is accomplished by addition of a “peptide bond spectrum” to the spectrum of the amino acid constituents. The peptide bond spectrum consists of addition of signal associated with the amide group and subtraction of signal associated with the amine and carboxylic acid groups.

2.3.1. X-ray Spectral Simulator (X-SpecSim). X-SpecSim is a program written in Interactive Data Language.³⁴ Its input is an amino acid sequence, which can be obtained from the synthesis or analysis of peptides, or for proteins, from standard databases such as the protein data bank (<http://>

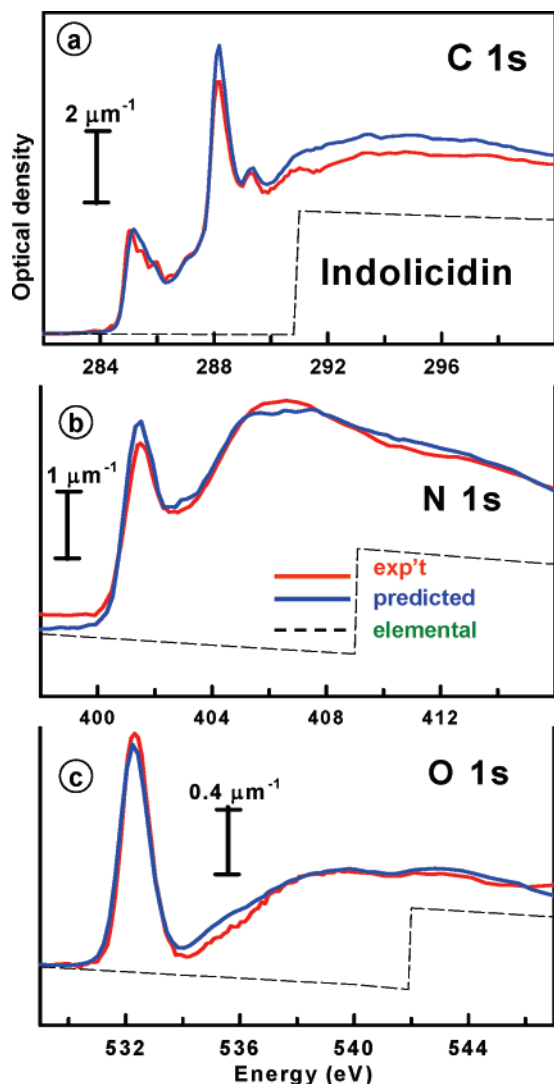


Figure 3. Measured versus predicted C 1s, N 1s, and O 1s spectra of indolicidin, a tryptophan-rich antimicrobial peptide (ILPWKWPWW-PWRR-NH₂). The elemental line represents the absorption of the elemental composition of the peptide predicted from tabulated data.³⁶ It is the expected spectrum without contributions from molecular structure and is used to set the intensity scale of the spectrum.

www.pdb.org/).³⁵ X-SpecSim uses a library of C 1s, N 1s, and O 1s spectra of the amino acids to compute the expected NEXAFS spectra of that sequence. The default library is the spectra reported by Zubavichus et al.,⁷ which were measured by total electron yield from powdered samples and thus avoided a solvent contamination issue that plagued an earlier study of the solid amino acids.⁶ This library is stored in tab-delimited text files. It can be modified easily as more accurate data becomes available or if different edges (e.g., S 1s or S 2p) are required. The output of the predicted spectrum is written to a file in a format similar to the input. The user can choose the energy scale and point spacing from the graphical interface, or a stored spectrum can be used to set these parameters. The amino acid spectra reported by Zubavichus et al.⁷ are sampled at approximately 0.1 eV in key spectral regions.

The peptide shift correction signals, which take the loss of amine and carboxylic acid groups and the gain of the amide group into account, are computed by subtracting the spectra of some known peptides and proteins from their amino acid constituents, as described in more detail below. A further issue requiring attention is the correct absolute scaling of the predicted spectrum. The intensity scales of NEXAFS spectra recorded

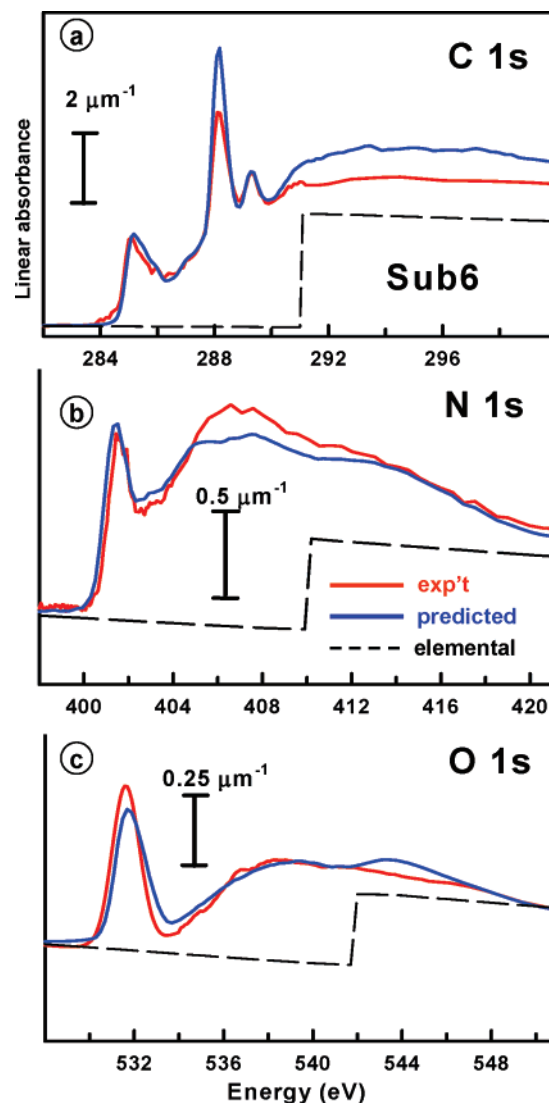


Figure 4. Measured versus predicted C 1s, N 1s, and O 1s spectra of sub6, a tryptophan-rich antimicrobial peptide (RWWKIWVIRWWR-NH₂).

by STXM are generally shown in units of optical density (OD) or absorbance. Outside of the strongly structured near-edge region (-10 to $+30$ eV relative to the ionization limit, typically), the linear absorbance of a material (optical density per unit thickness) is determined solely by its elemental composition and density.³⁶ For quantitative analysis we find it useful to place X-ray absorption reference spectra on an “OD1” scale, which is the optical density of a nanometer thickness of a material at its normal density. The final predicted spectrum is thus constructed by (i) generating a sum of spectra of the amino acids, weighted according to their frequency in the sequence; (ii) modifying the spectrum to correct for spectral changes associated with peptide bond formation; and (iii) rescaling the summed spectrum to match the elemental absorption of the final peptide or protein predicted from the sequence using tabulated data.³⁶ The final scaling step is necessary as peptide bond formation—specifically the loss of H₂O—changes the ratios of C, N, O, S, and H in proteins relative to their amino acid constituents. X-SpecSim is integrated into aXis2000,³⁷ a publicly available program for analysis of spectromicroscopy data. Additionally, the SF utility³⁸ is used to compute the expected OD1 elemental response of the peptide or protein.

An important consideration is the ionic state of the peptide or protein. The spectrum of glycine has been shown to undergo

TABLE 1: Energies and Assignments of Spectral Features in the C 1s, N 1s, and O 1s Spectra of Indolicidin (ILPWKWPWWPWR-NH₂) and Sub6 (RWWKIWVIRWWR-NH₂)

energy (eV)		assignment (final orbital)	residue
indolicidin	sub-6		
C 1s			
285.15	285.13	$\pi^*_{C=C(\text{aromatic})} (\nu = 0)$	Trp
285.55	285.58	$\pi^*_{C=C(\text{aromatic})} (\nu_{C-H} = 1)$	Trp
286.05	286.02	$C 1s(C-R) \rightarrow \pi^*_{C=C}$	Trp
287.0 (sh)	287.1 (sh)	σ^*_{C-H}	all
288.24(5) ^a	288.22(5) ^a	π^*_{amide}	all
289.36	289.37	$\pi^*_{C=N}$	Arg
291.1	291.0	π^*_{C-C}	all
294.4 (br)	294.3 (br)	π^*_{C-C}, π^*_{C-O}	all
298(1) (br)	300(1) (br)	$\pi^*_{C=C}$	Trp
302(1) (br)	302(1) (br)	$\pi^*_{C=C}$	Trp
N 1s			
401.20(8) ^b	401.17(8) ^b	π^*_{amide}	all
403.0	402.9	$\pi^*_{C=N}$	Arg
406.2	406.6	σ^*_{C-N}	all
412(1) (sh, br)	413(1) (sh, br)	$\sigma^*_{C=N}$	Arg
O 1s			
532.17(10) ^c	532.15(10) ^c	π^*_{amide}	all
535.4 (sh)	535.74 (sh)	$O 1s \rightarrow \pi^*_{COO}$	Asp
539.0(2)	539.0(3)	σ^*_{C-O}	all
543 (br, sh)	543 (br, sh)	σ^*_{C-O}	all
548(1) (br)	548(1) (br)	$\sigma^*_{C=O}$	all

^a C 1s calibration: $-4.50(6)$ relative to C 1s \rightarrow 3s ($\nu = 0$) in CO₂.⁴⁵ ^b N 1s calibration: $-4.98(8)$ relative to N 1s \rightarrow 3s ($\nu = 0$) in N₂.⁴⁶ ^c O 1s calibration: $-6.80(10)$ relative to O 1s \rightarrow 3s in CO₂.⁴⁷

TABLE 2: Energies and Assignments of Spectral Features in the C 1s, N 1s, and O 1s Human Serum Albumin and Fibrinogen

energy (eV)		assignment (final orbital)	residues
albumin	fibrinogen		
C 1s			
285.15	285.13	$\pi^*_{C=C(\text{aromatic})} (\nu = 0)$	Phe, Tyr, Trp, His
285.55	285.52	$\pi^*_{C=C(\text{aromatic})} (\nu_{C-H} = 1)$	Phe, Tyr, Trp, His
287.0 (sh)	287.1 (sh)	σ^*_{C-H}	all
288.20(5) ^a	288.20(5) ^a	π^*_{amide}	all
289.3	289.4	$\pi^*_{C=N}$	Arg
291.2 (sh, br)	291.2 (sh, br)	π^*_{C-C}	all
293.6 (br)	293.8 (br)	π^*_{C-C}, π^*_{C-O}	all
302(1) (br)	302(1) (br)	$\pi^*_{C=C}$	Phe, Tyr, Trp, His
N 1s			
398.6(w)	398.4 (w)	$\pi^*_{C=N}$	damaged protein
399.9 (w)	399.5 (w)	$\pi^*_{C=N}$	damaged protein
401.20(8) ^b	401.2(8) ^b	π^*_{amide}	all
403.0	402.9	$\pi^*_{C=N}$	Arg
406.4	406.2	σ^*_{C-N}	all
412(1) (sh, br)	413(1) (sh, br)	$\sigma^*_{C=N}$	Arg
O 1s			
532.2(1) ^c	532.2(1) ^c	π^*_{amide}	all
535.4 (sh)	535.74 (sh)	$O 1s \pi^*_{COO}$	Asp
540.0(3)	539.5(3)	σ^*_{C-O}	all
547 (br, sh)	545 (br, sh)	$\sigma^*_{C=O}$	all

^a C 1s calibration: $-4.54(6)$ relative to C 1s \rightarrow 3s ($\nu = 0$) in CO₂.⁴⁵ ^b N 1s calibration: $-4.95(8)$ relative to N 1s \rightarrow 3s ($\nu = 0$) in N₂.⁴⁶ ^c O 1s calibration: $-6.8(1)$ relative to O 1s \rightarrow 3s in CO₂.⁴⁷

significant changes in all edges as the pH of its environment changes and the molecule transitions between cationic, zwitterionic, and anionic states.^{6,39} Less well-studied, but probably equally significant, is the affect of ionization state on the spectra of charged amino acid side chains such as arginine or lysine. All spectra presented here are from molecules at neutral pH; the amino acids are predominantly zwitterionic, lysine, arginine, and the amino terminals of peptides positively charged, and aspartic acid, glutamic acid, and the carboxyl terminal of peptides negatively charged. The sub6 peptide has a net charge of +5. Caution needs to be exercised when applying spectra

calculated with the standard amino acid library to samples in low- or high-pH environments.

2.3.2. Illustrating X-SpecSim for Predicting Peptide Spectra. Previous applications of the building block model to NEXAFS spectra have been on a relatively small scale, typically involving molecules consisting of a few dozen atoms at most, not the hundreds of atoms which make up even relatively small peptides. Thus, automation of the adding and scaling processes was required. In addition to sheer scale, peptides offer two other significant challenges. They are diverse polymers, with 20 common monomers, and as they polymerize they undergo

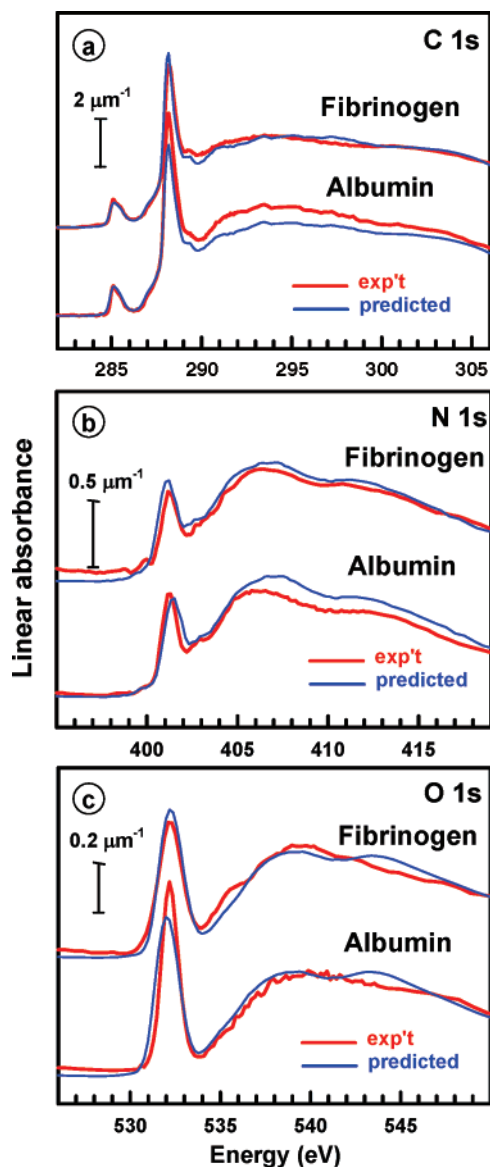


Figure 5. Measured versus predicted C 1s, N 1s, and O 1s spectra of human serum albumin and fibrinogen.

significant chemical change. Figure 1a plots the C 1s spectra of four amino acids to illustrate the diversity of amino acid spectra. This diversity is, however, nearly absent in protein spectra, due to averaging over hundreds of residues. In smaller peptides, or proteins with relatively unusual sequences, the contribution of individual amino acids or classes of amino acids can be observed. The amino acids with readily distinguishable spectral features include the following: (i) aromatic amino acids, which have a strong, relatively narrow peak at 285.2 eV (phenylalanine, tyrosine, and tryptophan); (ii) S-containing amino acids (methionine, cysteine) (relatively little work to date has used S 2p or S 1s spectroscopy, a notable exception being that of Liu et al.²⁶); (iii) amino acids containing unsaturated C=N bonds (histidine, arginine, and tryptophan).

The challenge of spectral changes associated with peptide bond formation is also shown in Figure 1c, which compares the spectrum of the hypothetical peptide IWRK-NH₂, simulated according to the “classic” building block approach (i.e., without correction for peptide bond formation), to that of human serum albumin, a protein with a relatively typical amino acid content in terms of the ratio of aromatic to nonaromatic amino acids. There is a clear difference in the 288 eV region. Albumin

has a C 1s $\rightarrow \pi^*_{\text{C=O}}$ peak at 288.2 eV, a position which is consistent with an amide bond. It is red-shifted relative to the 288.5 eV C 1s $\rightarrow \pi^*_{\text{C=O}}$ peak of the carboxylic acid of an amino acid. This shift is well-documented²⁴ and is the principal change in the C 1s spectra of peptides compared to amino acids. The change at the N 1s edge is much more dramatic, with introduction of a low-lying N 1s $\rightarrow \pi^*_{\text{amide}}$ peak at 401.2 eV,²⁴ a feature entirely absent in the N 1s spectrum of amino acids. The change at the O 1s edge includes a ~ 0.3 eV shift from a O 1s(C=O) $\rightarrow \pi^*_{\text{C=O}}$ peak to the O 1s $\rightarrow \pi^*_{\text{amide}}$ peak, and loss of the O 1s(C-O) $\rightarrow \pi^*_{\text{C=O}}$ peak at 535.5 eV, as the carboxylic acid functional group is transformed to an amide.

To correctly estimate the NEXAFS spectra of peptides, it is critical to correct for the spectral changes associated with peptide bond formation. X-SpecSim achieves this by adding a “peptide bond spectrum” to the simple summation of amino acids. The peptide bond spectrum is a differential correction which incorporates changes associated with the loss of the amine and carboxylic groups and the addition of the amide group. This correction and its influence on the resulting spectrum is illustrated in Figure 2. The peptide bond spectrum that is added to the amino acid spectra to account for polymerization is shown in Figure 2a–c, for the case of converting the spectrum of alanine into that of polyalanine. The major features of the C 1s peptide bond spectrum (Figure 2a) are the negative peak at 288.5 eV, which arises from loss of the carboxylic acid group, and the positive peak at 288.2 eV that arises from formation of the amide group. The principal effects of this correction, a shift in the location, mild loss in intensity, and slight broadening of the 288.2 eV peak, can be observed in Figure 2a, where the C 1s spectrum of polyalanine is shown before and after the peptide bond correction. The N 1s peptide bond spectrum (Figure 2b) has three distinct spectral features: the N 1s $\rightarrow \pi^*_{\text{amide}}$ peak at 401.2 eV; a mild loss of intensity from the broad N 1s $\rightarrow \sigma^*_{\text{N-C}}$ peak at 406 eV; and the addition of a broad N 1s $\rightarrow \sigma^*_{\text{C-N}}$ peak at 412 eV, due to the formation of an additional N–C bond. The effect of these changes is illustrated in Figure 2b, which compares the N 1s spectrum of polyalanine before and after the peptide bond correction. The O 1s peptide bond spectrum (Figure 2c) is a differential shape consisting of the shift in the location of the O 1s $\rightarrow \pi^*_{\text{C=O}}$ peak, which is at 532.3 eV in amino acids and at 532.0 eV in peptides. This shift can be seen in Figure 2c, where the peptide bond spectrum has a negative peak at 532.2, showing the loss of the carboxylic acid, and a positive peak at 532.0 eV, corresponding to the formation of the amide group. The effect of these changes is illustrated in Figure 2c, which compares the O 1s spectrum of polyalanine before and after the peptide bond correction.

Figure 2d presents the X-SpecSim spectrum of albumin, predicted with and without the correction, in comparison to the experimental C 1s spectrum of albumin. Away from the 288 eV region, both spectra show good agreement; for example, the height of the 285 eV aromatic peak is nearly identical in the predicted and experimental spectra. This is comforting, as this region should be unaffected by peptide bond formation according to the building block model. The corrected spectrum deviates somewhat from the experimental spectrum in several slowly varying regions, but it fits well at both major peaks.

3. Results and Discussion

3.1. Antimicrobial Peptides. The antimicrobial peptides we have studied have unusual amino acid compositions. Both

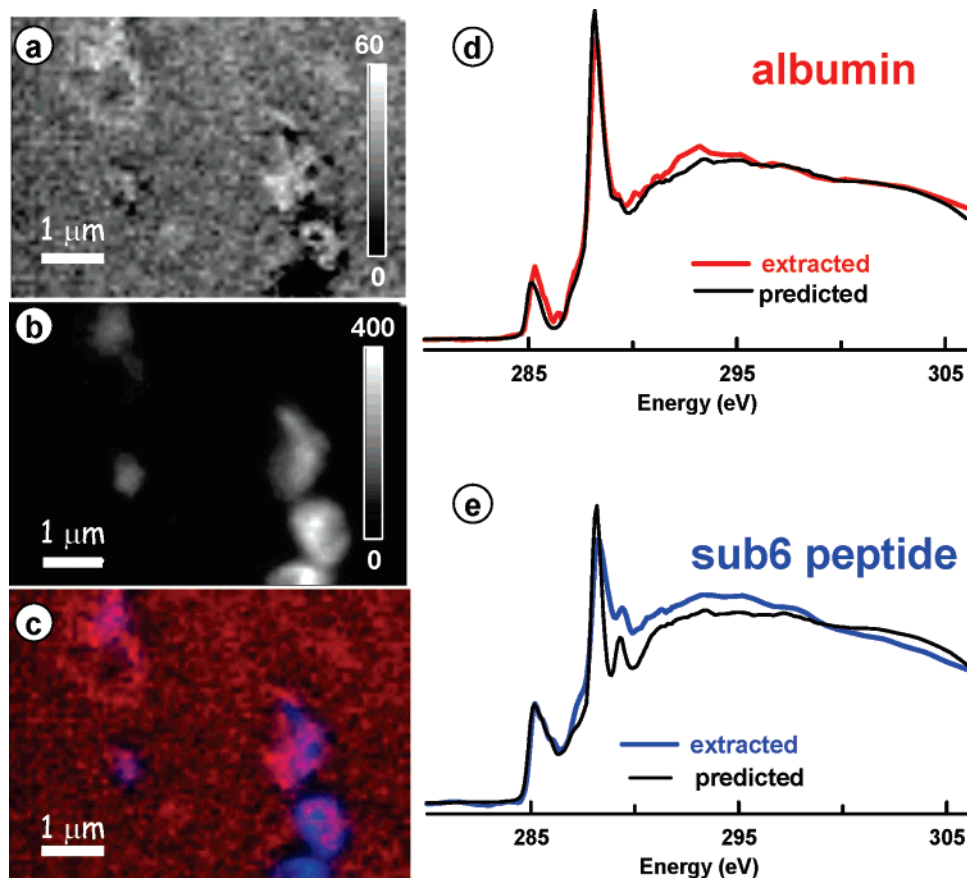


Figure 6. Demonstration of the ability to differentiate sub6 and human serum albumin in a heterogeneous system: (a) component map of albumin (the gray scale is thickness in nanometers); (b) component map of sub6, derived from fit to a C 1s image sequence (280–320 eV); (c) color-coded composite map (red = albumin, blue = sub6). Extracted C 1s spectra of (d) HSA, and (e) sub6, based on threshold masking the high-intensity regions of the component maps, compared with the spectra of the pure materials.

indolicidin and sub6 are rich in tryptophan and arginine residues. Sub6 is 35% arginine and 42% tryptophan, while indolicidin is 38% tryptophan and 15% arginine. The tryptophan aromatic residue affects the strength of the 285.2 eV peak, and the guanidine group of arginine produces a peak at 289.3 eV. These features provide spectral handles which may enable detection of these peptides against a background of signal from proteins in a biological environment such as a cell. Thus, they provide a good test of our prediction method. Figure 3 compares measured versus predicted C 1s, N 1s, and O 1s spectra for indolicidin, a 13 residue, tryptophan-rich peptide, which has antimicrobial properties.^{9,10} Note that both the energy and intensity scales are absolute. There is very good agreement between the experimental and predicted spectra, indicating that, with inclusion of the peptide shift correction, the spectral simulation works. Table 1 reports the energies and assignments of the spectral features. As parts a–c of Figure 3 show, there is good general agreement between the measured and predicted NEXAFS spectrum of indolicidin at the C 1s, N 1s, and O 1s edges. At the C 1s edge the aromatic and guanidine peaks agree nearly perfectly with the expected values; the $\pi^*_{C=O}$ peak and the σ^*_{C-C} region of the predicted spectrum agree well in shape and location, but there are minor intensity differences as compared to the measured spectrum. The better definition of the near-edge fine structure (especially on the high-energy side of the 285 eV peak) in the experimental spectrum of indolicidin is probably due to the higher experimental resolution in our STXM measurement (~ 150 meV full width at half-maximum (fwhm)) as compared to that of 300 meV used by Zubavachus et al.⁷ in

their measurements of the constituent amino acids. The N 1s and O 1s spectra are an excellent match. The predicted and measured C 1s, N 1s, and O 1s spectra of sub6 are shown in Figure 4, in comparison to the spectra predicted by X-SpecSim. As with indolicidin, there is good agreement.

3.2. Proteins. The same approach for predicting peptide spectra from their constituent amino acids can be applied to proteins of hundreds or thousands of amino acids. Figure 5 plots the C 1s, N 1s, and O 1s spectra of HSA and fibrinogen in comparison to the predicted spectra. These spectra are newly recorded and differ slightly from previously reported spectra of fibrinogen¹⁸ and albumin.⁴⁰ Spectral features and assignments are given in Table 2. Figure 5 also displays the predicted spectra, based on the known sequences of HSA⁴¹ and fibrinogen.^{42,43} There is excellent agreement between the measured and predicted spectra. Figure 5 illustrates one of the challenges of using natural spectral contrast for protein identification. Because of the large numbers of amino acids, there is a high degree of spectral averaging, making proteins with “typical” ratios of aromatic to nonaromatic amino acids almost indistinguishable. One exception to this “indistinguishability by averaging” is proteins composed of a repeating motif, such as many structural proteins. Proteins with a relatively short repeated sequence of amino acids are potentially very different spectrally from proteins with more typical amino acid compositions. Finally, in cases where there is a high degree of regularity (such as proteins with large fractions of β -sheets) and these regions are aligned across different proteins, spectral differences can arise from linear dichroic effects.^{5,44}

3.3. Experimental Test of Distinguishing Peptide and Protein by NEXAFS. As a test of the ability of the intrinsic differences in the NEXAFS spectra to provide a contrast mechanism and thus provide a means to quantitatively map peptides against a protein background, STXM was used to measure a C 1s image sequence of a sample consisting of a Si₃N₄ window coated with a uniform layer of HSA with a small amount of sub6 deposited on top. The results of a fit of this C 1s image sequence to the reference spectra for albumin and sub6 are shown in Figure 6. Figure 6a is the component map for the albumin indicating the film is relatively uniform, and ~40 nm thick. Figure 6b is the component map for the sub6 peptide, indicating much thicker, but irregular, deposits, up to 400 nm thick. Figure 6 demonstrates clearly that it is possible to discriminate between the uniform background of HSA and the nonuniform sub6 deposits using the differences in their C 1s spectra and to derive quantitative maps. Parts d and e of Figure 6 plot the spectra extracted from two regions selected by threshold masks of the high-intensity pixels of the component maps. These extracted spectra are found to be very good matches to the spectra of pure HSA and sub6. Although this is an artificial two-component mixture, it indicates that differentiation in a more complex sample should be possible, if sufficient concentrations of the peptide or protein of interest are present within the spatially resolved analytical zone, which can be as small as a 30 nm diameter circle in state-of-the-art STXMs.

4. Summary

We have developed X-SpecSim, a tool that allows prediction of the NEXAFS spectra of peptides and proteins of arbitrary sequence from a data bank of the spectra of the constituent amino acids.⁷ This methodology can be readily extended to predictions of the NEXAFS spectra of any substance consisting of repeat units, such as block copolymers, polysaccharides, and nucleic acids, etc. In addition we have demonstrated the effectiveness of X-SpecSim by comparing its predictions to the experimental spectra of previously unmeasured peptides and several proteins. Finally we have demonstrated that the intrinsic spectral differences are sufficiently large, at least in some cases, to enable mapping of a peptide against a background of proteins. Efforts are presently underway to use this capability to track these and other antimicrobial peptides dosed at sublethal levels in bacterial biofilms. If this works, we hope to gain insights into mechanisms of antimicrobial resistance.

Acknowledgment. We especially thank Drs. Zubavichus, Zharnikov, and Grunze for permission to use their amino acid spectra in the spectral simulator program. Research was supported by Advanced Food and Materials Network (AFMnet), NSERC, Canada Foundation for Innovation, and the Canada Research Chair program. STXM was performed at beamline 5.3.2 at the Advanced Light Source, Berkeley, CA, which is supported by the U.S. Department of Energy under Contract DE-AC03-76SF00098. We thank David Kilcoyne and Tolek Tyliczszak for their excellent support of the STXM instrumentation at the ALS.

References and Notes

- Zhang, X.; Balhorn, R.; Mazrimas, J.; Kirz, J. *J. Struct. Biol.* **1996**, *116*, 335–344.
- Hitchcock, A. P.; Morin, C.; Heng, Y. M.; Cornelius, R. M.; Brash, J. L. *J. Biomater. Sci., Polym. Ed.* **2002**, *13*, 919–938.
- Lawrence, J. R.; Swerhone, G. D. W.; Leppard, G. G.; Araki, T.; Zhang, X.; West, M. M.; Hitchcock, A. P. *Appl. Environ. Microbiol.* **2003**, *69*, 5543–5554.
- Li, L.; Hitchcock, A. P.; Robar, N.; Cornelius, R.; Brash, J. L.; Scholl, A.; Doran, A. *J. Phys. Chem. B* **2006**, *110*, 16763–16773.
- Hernández Cruz, D.; Hitchcock, A. P.; West, M. M.; Rousseau, M.-E.; Pézolet, M. *Biomacromolecules* **2006**, *7*, 836–843.
- Kaznatcheyev, K.; Osanna, A.; Jacobsen, C.; Plashkevych, O.; Vahtas, O.; Ågren, H.; Carravetta, V.; Hitchcock, A. P. *J. Phys. Chem. A* **2002**, *106*, 3153–3168.
- Zubavichus, Y.; Shaporenko, A.; Grunze, M.; Zharnikov, M. *J. Phys. Chem. A* **2005**, *109*, 6998–7000.
- Hancock, R. E.; Sahl, H. G. *Nat. Biotechnol.* **2006**, *24*, 1551–1557.
- Hilpert, K.; Elliott, M. R.; Volkmer-Engert, R.; Henklein, P.; Donini, O.; Zhou, Q.; Winkler, D. F.; Hancock, R. E. *Chem. Biol.* **2006**, *13*, 1101–1107.
- Zhang, L.; Scott, M. G.; Yan, H.; Mayer, L. D.; Hancock, R. E. *Biochemistry* **2000**, *39*, 14504–14514.
- Brogden, K. A. *Nat. Rev. Microbiol.* **2005**, *3*, 238–250.
- Jenssen, H.; Hamill, P.; Hancock, R. E. *Clin. Microbiol. Rev.* **2006**, *19*, 491–511.
- Kirz, J.; Jacobsen, C.; Howells, M. *Q. Rev. Biophys.* **1995**, *28*, 33–130.
- Ade, H. X-Ray Spectromicroscopy. In *Experimental Methods in the Physical Sciences*, Vol. 32; Samson, J. A. R., Ederer, D. L., Eds.; Academic Press: New York, 1998; pp 225–261.
- Ade, H.; Urquhart, S. G. NEXAFS Spectroscopy and Microscopy of Natural and Synthetic Polymers. In *Chemical Applications of Synchrotron Radiation*; Sham, T. K., Ed.; World Scientific: Singapore, 2002; pp 285–355.
- Hitchcock, A. P.; Morin, C.; Zhang, X.; Araki, T.; Dynes, J. J.; Stöver, H. D. H.; Brash, J. L.; Lawrence, J. R.; Leppard, G. G. *J. Electron Spectrosc. Relat. Phenom.* **2005**, *144–147*, 259–269.
- Hitchcock, A. P.; Stöver, H. D. H.; Croll, L. M.; Childs, R. F. *Aust. J. Chem.* **2005**, *58*, 423–432.
- Morin, C.; Hitchcock, A. P.; Cornelius, R. M.; Brash, J. L.; Urquhart, S. G.; Scholl, A.; Doran, A. *J. Electron Spectrosc. Relat. Phenom.* **2004**, *137–140*, 785–794.
- Lippincott-Schwartz, J.; Patterson, G. H. *Science* **2003**, *300*, 87–91.
- Roth, J.; Heitz, P. U. *Ultrastruct. Pathol.* **1989**, *13*, 467–484.
- Stöhr, J. *NEXAFS Spectroscopy*; Springer Tracts in Surface Science; Springer: Berlin, 1992; Vol. 25.
- Cooper, G.; Gordon, M.; Tulumello, D.; Turci, C. C.; Kaznatcheev, K.; Hitchcock, A. P. *J. Electron Spectrosc. Relat. Phenom.* **2004**, *137–140*, 795–799.
- Boese, J.; Osanna, A.; Jacobsen, C.; Kirz, J. *J. Electron Spectrosc. Relat. Phenom.* **1997**, *85*, 9–15.
- Gordon, M. L.; Cooper, G.; Morin, C.; Araki, T.; Turci, C. C.; Kaznatcheev, K.; Hitchcock, A. P. *J. Phys. Chem. A* **2003**, *107*, 6144–6159.
- Zubavichus, Y.; Zharnikov, M.; Schaporenko, A.; Grunze, M. *J. Electron Spectrosc. Relat. Phenom.* **2004**, *134*, 25–33.
- Liu, X.; Jang, C.-H.; Zheng, F.; Jurgensen, A.; Denlinger, J. D.; Dickson, K. A.; Raines, R. T.; Abbott, N. L.; Himpel, F. J. *Langmuir* **2006**, *22*, 7719–7725.
- Hilpert, K.; Volkmer-Engert, R.; Walter, T.; Hancock, R. E. *Nat. Biotechnol.* **2005**, *23*, 1008–1012.
- Romeo, D.; Skerlavaj, B.; Bolognesi, M.; Gennaro, R. *J. Biol. Chem.* **1988**, *263*, 9573–9575.
- Hilpert, K.; Hancock, R. E. *Nat. Protoc.*, in press.
- Hilpert, K.; Winkler, F. H. D.; Hancock, R. E. *Nat. Protoc.* **2007**, *2*, 1333–1349.
- Kilcoyne, A. L. D.; Tyliczszak, T.; Steele, W. F.; Fakra, S.; Hitchcock, P.; Franck, K.; Anderson, E.; Harteneck, B.; Rightor, E. G.; Mitchell, G. E.; Hitchcock, A. P.; Yang, L.; Warwick, T.; Ade, H. *J. Synchrotron Radiat.* **2003**, *10*, 125–136.
- Warwick, T.; Ade, H.; Kilcoyne, A. L. D.; Kritscher, M.; Tyliczszak, T.; Fakra, S.; Hitchcock, A. P.; Hitchcock, P.; Padmore, H. A. *J. Synchrotron Radiat.* **2002**, *9*, 254–257.
- Jacobsen, C.; Wirick, S.; Flynn, G.; Zimba, C. *J. Microsc.* **2000**, *197*, 173–84.
- IDL is a scientific computing language developed by Research Systems Inc, currently part of ITT Visual Imaging Solutions (<http://www.itvis.com/idl/>).
- Berman, H. M.; Henrick, K.; Nakamura, H. *Nat. Struct. Biol.* **2003**, *10*, 980.
- Henke, B. L.; Gullikson, E. M.; Davis, J. C. *At. Data Nucl. Data Tables* **1993**, *54*, 181–342.
- aXis2000 is written in Interactive Data Language (IDL). It is available free for noncommercial use from <http://unicorn.mcmaster.ca/aXis2000.html>.
- SF is a precursor to the program on the CXRO web site (http://www.cxro.lbl.gov/optical_constants/) which computes elemental X-ray absorption. The version implemented in aXis2000 was programmed for IDL by Billy Loo.

(39) Zubavichus, Y.; Shaporenko, A.; Grunze, M.; Zharnikov, M. *J. Phys. Chem. B* **2006**, *110*, 3420–3427.

(40) Loo, B. W., Jr.; Sauerwald, I. M.; Hitchcock, A. P.; Rothman, S. *J. Microsc.* **2001**, *204*, 69–86.

(41) Minghetti, P. P.; Ruffner, D. E.; Kuang, W. J.; Dennison, O. E.; Hawkins, J. W.; Beattie, W. G.; Dugaiczak, A. *J. Biol. Chem.* **1986**, *261*, 6747–6757.

(42) Rixon, M. W.; Chan, W. Y.; Davie, E. W.; Chung, D. W. *Biochemistry* **1983**, *22*, 3237–3244.

(43) Rixon, M. W.; Chung, D. W.; Davie, E. W. *Biochemistry* **1985**, *24*, 2077–2086.

(44) Rousseau, M. E.; Hernández-Cruz, D.; West, M. M.; Hitchcock, A. P.; Pézolet, M. *J. Am. Chem. Soc.* **2007**, *129*, 3897–3905.

(45) Ma, Y.; Chen, C. T.; Meigs, G.; Randall, K.; Sette, F. *Phys. Rev. A* **1991**, *44*, 1848–1858

(46) Chen, C. T.; Ma, Y.; Sette, F. *Phys. Rev. A* **1989**, *40*, 6737–6740.

(47) Ishii, I.; Hitchcock, A. P. *J. Electron Spectrosc. Relat. Phen.* **1987**, *46*, 55–84.