Innershell Absorption Spectroscopy of Amino Acids

K. Kaznacheyev,* A. Osanna, and C. Jacobsen

Department of Physics and Astronomy, State University of New York, Stony Brook, New York 11794

O. Plashkevych,† O. Vahtras, and H. Ågren

Theoretical Chemistry, Royal Institute of Technology, Stockholm, S-10044, Sweden

V. Carravetta

ICQEM-CNR, via Moruzzi 1, 56124 Pisa, Italy

A. P. Hitchcock

Department of Chemistry, McMaster University, Hamilton, Ontario L8S 4M1, Canada

Received: September 5, 2001

We present comprehensive measurements of the C (carbon) K edge near-edge X-ray absorption (NEXAFS) spectra of all 20 amino acids commonly occurring in nature. Qualitative trends among the spectra of amino acids with similar chemical character are identified and spectral features are compared with extensive ab initio calculations. The contributions of individual units and substitutional groups have been determined to explore their fingerprinting character using the building block concept. Several such units are found. Two that give particularly clear features in the C 1s NEXAFS spectra are the carboxyl group (which can be clearly identified by a pronounced structure due to the C 1s$\rightarrow\pi^*_{C=O}$ transition with maximum at 288.65(5) eV) and modified phenol rings in aromatic amino acids (which give sharp C 1s$\rightarrow\pi^*_{C=C}$ structures). The latter transitions are located around 285 eV, and their shape is specific for each aromatic amino acid. Other building blocks, such as the CNH_n group and the CH, CC, CO, CN pair bonds, are also identified, although their characteristic features are less pronounced in the C K edge spectra than the carboxylic and aromatic structures. This study provides the basis for rigorous assignment of the NEXAFS spectra of the amino acids, and will be helpful in developing X-ray absorption spectroscopy for quantitative analysis of proteins.

1. Introduction

This work extends early experimental[1] and computational[2,3] studies of the near edge X-ray absorption (NEXAFS) spectra of amino acids and small polypeptides. These previous measurements demonstrated the ability of NEXAFS spectroscopy to identify particular amino acids and to use their characteristic features for chemical mapping in soft X-ray spectromicroscopy. It was found that amino acids with aromatic side chains have distinct spectra with signatures that do not change during the peptide bond formation. Subsequent computational analysis[2−4] helped to identify the origin of individual peaks and motivated a more extended analysis of experimental data and their spectroscopic assignments. In this work we present a comprehensive set of significantly improved experimental spectra, including now all 20 major amino acids, and give a detailed spectral interpretation on the bases of experimentally observed trends for each group of amino acids and on the bases of ab initio computations.

Studies applying innershell excitation spectroscopy to many molecules[5] have proven that NEXAFS spectroscopy is a powerful probe of molecular electronic and geometric structures. This has led to many applications including the determination of the electronic structure and orientation of molecules in the solid state and absorbed on surfaces.[6] Qualitatively, it is possible to correlate specific NEXAFS features with functional groups and in some cases, individual bonds, such that the total spectrum can be considered as a linear combination of elementary spectra.[6] This conceptual approach, named the "building block principle", provides a useful starting point for the interpretation of the spectra of very complicated molecules. However, there are clearly limits to the building block approach. Delocalization of electronic charge across multiple functional groups leads to new molecular orbitals (MO) combining the properties of the conjugated groups. In such cases the NEXAFS features of the individual groups can be changed significantly and new features may appear.[7,8] One theme of this work is to explore these limits in the case of amino acids and to determine which "building blocks" can be considered to be the best for descriptions of the NEXAFS spectra of amino acids and of larger protein complexes.

It is well-known that amino acids exist in different charge states depending on the local environment, or pH in the case of aqueous solution. We have explored these effects and a detailed study of NEXAFS dependence on ion-state formation in the case of glycine will be presented elsewhere[9] in conjunction with a study of how NEXAFS reflects peptide bond formation. It

* To whom the correspondence should be addressed. Canadian Light Source, 107 North Road, Saskatoon, SK S7N 5C6 Canada. E-mail: kaznache@cls.usask.ca.

† Department of Bioorganic Chemistry, Box 581, Uppsala University, 75123 Uppsala Sweden.

10.1021/jp013385w CCC: $22.00
Published on Web 03/08/2002

was found that C 1s NEXAFS spectra are sensitive to protonation/deprotonation of the carboxyl and amine groups, even though such proton transfers do not directly bond to the carbon skeleton. Because most of the NEXAFS measurements presented here have been performed on dried films, where the degree of protonation depends on the preparation method, we present some of the glycine data to explain the charge structure which is formed with our particular sample preparation method.

Initial comparisons of new experimental data with the earlier calculations[2] revealed limitations of the theoretical approximations used in the initial calculations. This stimulated the application of a new model,[3] based on a polarized ion static exchange (STEX) technique. The new calculation method has helped to resolve discrepancies between experiment and some of our previous simulations.[2] Here, we present new theoretical data, as well as the scheme of the calculations, reflecting the importance of relaxation effects and indicating the limits of the static exchange approximation.

This article is organized in the following manner. After description of the experimental techniques and approximations used in the calculations, we present and analyze the NEXAFS spectra of glycine for different charge states. Because control of the charge state is done by tuning solution pH, this section includes a description of the sample preparation methods for wet specimens. The main method of preparation of the dried films is reported in a separate section. The C 1s experimental NEXAFS spectra of all the amino acids are grouped according to "biological functionality". Within each group, we present not only arguments which can lead to rigorous assignment of the NEXAFS spectrum of a particular amino acid, but also focus on common spectroscopic trends for the group. Wherever possible the experimental data are discussed in comparison with the results of our theoretical calculations. The "common" properties of the amino acids as reflected in NEXAFS spectra are discussed briefly, and the capability of NEXAFS spectroscopy to identify specific amino acids is evaluated. A short summary is given in the last section of the paper.

2. Experimental and Theoretical Methods

2.1. Experimental Methods and Sample Preparation. The amino acids used in this study were obtained from Sigma-Aldrich in the form of crystalline powders with minimum purity of 99%. They were used without further purification. Most of the measurements were performed on naturally dried specimens. For thin film preparations the powder was dissolved in a solvent (water or trifluoracetic acid, TFA) and then naturally dried on a Si_3N_4 window[10] under ambient conditions. The 100 nm thick Si_3N_4 windows (half-transparent for X-rays in the investigated energy region) were not treated beyond the KOH etch, acid rinse, and water rinse steps of the window fabrication process. Empty windows did not show any spectroscopic features across the C K edge. The concentration of the solutions (typically ~1 mg/mL) was adjusted to produce thin films that partly absorb X-ray light. Two different solvents were explored. Amino acids deposited from aqueous solutions form a highly textured structure with typical grain sizes of the order of micrometers, and with the orientation of crystallites affected by the chemical properties of the surface ad-layer.[11] In such samples we found a strong orientational dependence of the NEXAFS spectra, which varied for differently oriented crystallites. Such crystallites were too thick for accurate C 1s NEXAFS measurements in transmission.

Trifluoroacetic acid (TFA) dries faster and produces more homogeneous films of amino acids than aqueous solutions do.[1] With TFA, no polarization dependence of the NEXAFS spectra was recorded. However, we found that there was a small residue of TFA which leads to additional NEXAFS structure at 293−297 eV.[1] Since the films prepared from TFA solution were superior for spectroscopy, we have adopted the TFA preparation method for all the samples reported here, notwithstanding the spectral contamination above 293 eV. Additional discussion of the charge state of amino acids in solvent cast films from aqueous or TFA solution and their influence on the NEXAFS spectra is given in section 3.1 which describes the NEXAFS spectra of different glycine preparations in detail.

The SUNY−Stony Brook scanning transmission X-ray microscope (STXM)[12,13] at the X-1A beam line[14] of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory (BNL) was used to measure the C 1s NEXAFS spectra of the 20 common amino acids at room temperature. Although the microscope is capable of imaging thin specimens in transmission with a spatial resolution of about 50 nm, only the specimens dried from aqueous solution with normal pH required study at high spatial resolution. All other samples were measured in a defocused mode with a beam size of the order of a micron in order to minimize radiation damage (the resulting dose of about 10^6 Gray is below the threshold for observable mass loss in dry biological specimens[15]). Some samples were checked for radiation induced damage or chemical modification by imaging the specimen before and after the NEXAFS spectrum was recorded. Bleaching or other changes in the specimen structure were not observed for the radiation doses used in these experiments.

The C 1s spectra presented here were recorded with an energy resolution of 150 meV full width at half-maximum (FWHM), as determined by measurements of the 3p/4p Rydberg transitions of CO_2.[16] Energy scales were calibrated by adjusting the instrumental scale to place the 3s Rydberg transition in CO_2 at 292.8 eV[17] in conjunction with each measurement to ensure high accuracy (50 meV) and reproducibility.

2.2. Calculations. Among ab initio methods commonly used for calculations of electronic excitation, the static-exchange approximation (STEX)[18] occupies a special place. It is a separate state approach in which the excited state is approximated by the coupling of a target ionic state and an excited orbital optimized in the static field of the molecular ion. In a sense, the STEX method is also the lowest order in a hierarchy of propagator methods. Excitation energies and transition probabilities are computed by a single calculation for discrete and continuum spectra with respect to the ground and core-ionized reference states that are calculated at the Hartree−Fock level. The X-ray photoabsorption spectrum is obtained from singly excited states described in an independent electron picture using the virtual orbitals of the core-ionized system. These orbitals are the eigenvectors of a one-particle Hamiltonian that describes the motion of the excited electron in the field of the remaining molecular ion corresponding to a specific core-hole. In the direct implementation of the STEX approximation used in the present investigation,[19] the Hamiltonian matrix is constructed directly from one- and two-electron integrals computed in the atomic orbital basis set. A standard orbital basis set for the calculation of the reference states is extended by several additional diffuse functions centered at the ionized site (the so-called double basis set technique). Electronic relaxation around the core hole is taken into account by the ΔSCF (self-consistent field) procedure and is assumed to be independent of the excited electron, with the excitation levels converging to a common ionization limit. The inter-channel coupling between different core excitation channels

is neglected because of the significant energy and/or spatial separation between core holes in most of the molecular systems.

The major limitations of the static exchange approximation are related to electron correlation effects and the effect of screening of the excited electron. The former can eventually be reduced to the correlation between the target and the excited electron, but the residual relaxation of the bound electron due to the screening cannot be considered negligible for the lowest excited states. STEX is evidently more suitable for those excited states that can be reasonably described by "single excitations", and, among those, for high-energy excited states of Rydberg or continuum character. For the lowest core excited states of valence (often π^*) character, the missing screening results in energy errors of about 1−2 eV. It was recently shown that a simple modification of the static-exchange potential can be effective in the short range around the excitation site, while retaining the correct asymptotic behavior for the potential tail.[3] We have applied this screened static-exchange potential approach in the calculation of the NEXAFS spectra of glycine.

Geometry optimizations and initial SCF calculations of the reference ground and core-ionized states were carried out employing triple-ζ basis sets with polarizing and diffuse functions. There are of "cc-pVTZ" type with (10s,5p,2d)/(5s,2p)→(4s,3p,2d)/(3s,2p) functions[20] but without inclusion of *f*-functions. The STEX Hamiltonian was then calculated by adding to this basis set a large number of diffuse functions (20s,18p,20d) to give an augmented basis set. The optimized geometries of the amino acids have been obtained for the zwitterionic forms and are found to be very close to the computed geometries reported previously.[21]

3. Results and Discussion

All amino acids have a common backbone of a carboxylic acid group and an amino group attached to a saturated carbon atom, referred to as Cα. Although more than 300 different amino acids have been found in nature, only 20 are commonly found as constituents of mammalian proteins. The 20 common amino acids have been the focus of this study. They may be subdivided into five categories, based on the nature of the side chain (R) bonded to the α-carbon atom (Cα):

1. Amino acids with simple aliphatic side chains, such as alanine (Ala), valine (Val), leucine (Leu), and isoleucine (Ile). Their NEXAFS spectra are described in section 3.2, with an emphasis on the role of o^*_{CH} bond formation.

2. Amino acids with alcohol −OH, −SH, and $-SCH_3$ group side chain, including serine (Ser), threonine (Thr), cystein (Cys), and methionine (Met). Their NEXFAS spectra are discussed in section 3.3, including discussion of the effect of hetero-substituted carbons on the C 1s spectrum.

3. Amino acids with carboxylic acid or amide group side chains, such as aspartic (Asp) and glutamic (Glu) acids, and asparagin (Asn) and glutamine (Gln).

4. Amino acids with strongly basic group side chains, such as lysine (Lys) and arginine (Arg). Their NEXAFS spectra are considered together with those of the previous category due to common trends; the splitting and energy shift observed for high-energy π^* structure due to carboxyl substitution and the spectroscopic fingerprint of C═N bond have primarily attracted our attention and are discussed in section 3.5. The result for proline (Pro) is also included with this group to distinguish the characteristic features of an imino group (cyclic amine) relative to the amino group.

5. Amino acids with aromatic side chains, such as phenylalanine (Phe), tryptophan (Trp), tyrosine (Tyr), and histidine (His) which define a spectroscopically distinct group, due to the presence of low-lying π^* structure. The analysis of experimental results versus theoretical calculations for these species is given in section 3.6.

The discussion of the first two entries in the list above are complemented with results from theoretical calculations performed for Ala, Val, Ser, and Cys.

We believe that the classification given above reflects not only the fundamental properties of amino acids, such as the types of biochemical reactions in which they participate and their influence on protein structure and reactivity, but it also highlights the common spectroscopic features within the various groups of amino acid molecules and thus helps to assign their NEXAFS features.

The discussions of the NEXAFS spectra of individual amino acids are preceded by section 3.1, which, through a discussion of the NEXAFS spectra of various preparations of glycine (Gly), describes (a) common aspects of the NEXAFS spectra for all amino acids and (b) the ion state expected from the sample preparation method employed, and its influence on the NEXAFS spectra.

3.1. C 1s NEXAFS Spectroscopy of Glycine. The evaporation rate of amino acids even close to their decomposition temperature is very low,[22] which makes it difficult to perform gas-phase NEXAFS spectroscopy in the transmission mode even for a selected set of the amino acids. The low-solubility of most amino acids also complicates NEXAFS measurements in solution even at the C K edge. Alternatively, a thin film dried from a solution or deposited from the gas phase can be an ideal sample for NEXAFS spectroscopy. However, we have observed that samples obtained by drying unbuffered aqueous solutions at room temperature tend to form textured structures, which in most cases are too dense for transmission spectroscopy in the soft X-ray region. In contrast, deposition from acidic media produces rather uniform films under the same conditions for drying, and the thickness of these films can be easily controlled by appropriate dilution. This is a major reason for our choice of trifluoro acetic acid (TFA) as a solvent, and to use naturally dried specimens as the samples for NEXAFS spectroscopy of amino acids. This choice of preparation raises some questions worth considering. What is the charge state of the amino acid samples prepared in this way? How does this state differ from the charge state of the respective amino acids in aqueous solution?

For the ground state of amino acids, gas-phase experiments have shown that the carboxyl and amine functional groups are well preserved.[23] In unbuffered aqueous solution as well as in the solid state, amino acids predominantly form a zwitterionic structure, with a protonated amino group and a deprotonated carboxyl group ($H_3N^+-CHR-COO^-$, where R is the amino acid residue). In solution there is a dynamic equilibrium between the uncharged ground state, the zwitterionic form, and possible charged species. The position of this equilibrium depends on the pH of the solution. In the case of dilute Gly aqueous solutions with a pH lower than the $pK_1 = 2.34$, most of the molecules remain in the ionized form $(H_3N-CH_2-COOH)^+$. At higher pH the Gly is predominantly in the zwitterionic form. For very high pH ($pH > pK_2 = 9.60$), the amino group loses the proton as well and forms $(H_2N-CH_2-COO)^-$ ions. Are these charged structures preserved upon drying? In an attempt to answer this question, as well as to check the influence on the NEXAFS spectra of the different charge states, we have performed NEXAFS measurements of Gly solutions with

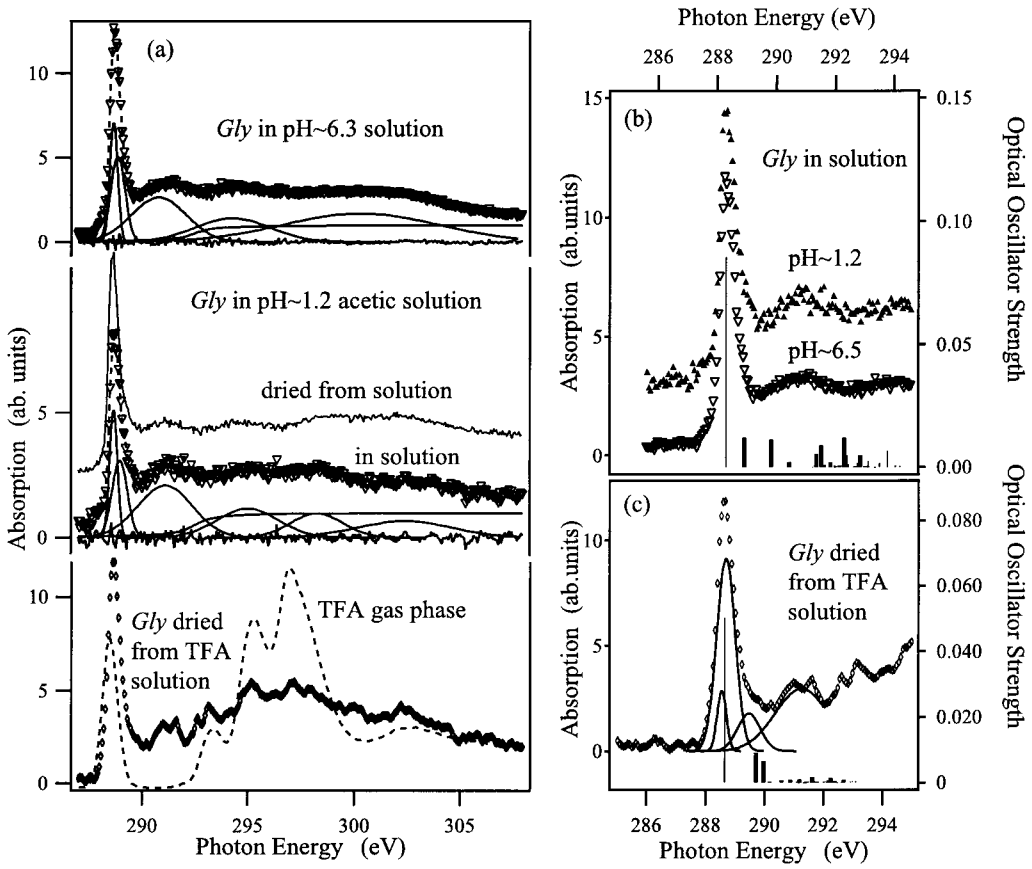

Figure 1. NEXAFS C K edge spectra of glycine. Except where noted, experimental data are shown as plotted points; individual fitted peaks and the edge step-function are shown as continuous lines; and the total fitted spectrum is shown as a dashed line. (a) Spectra of glycine in 1.3 M, pH = 6.4 aqueous solution (top), in pH = 1.2 acetic aqueous solution (middle) and dried from this solution (middle; solid line shifted upward), and dried from trifluoroacetic acid (TFA) solution (bottom) along with the gas-phase spectrum of TFA (bottom; dashed line). (b) Expansion of the NEXAFS spectra of Gly in solution (pH = 1.2 above, pH = 6.5 below) for the near-IP energy region.. The oscillator strengths (right axis) of individual transitions, as calculated in a screened STEX approximation, are shown by vertical bars (fine bars for carboxyl transitions, and thicker bars for Cα transitions). (c) Detail of the NEXAFS spectrum of Gly, dried from TFA solution. Individual transitions, as calculated using a nonpolarized core STEX modified to coincide with the experimental spectrum, are again shown by vertical bars.

TABLE 1: Absolute Energies, Term Values and Proposed Assignment for the Features in C 1s NEXAFS Spectrum of Glycine

STEX							
energy (eV)	289.7	289.9	291.				
term value[a]	4.0	3.8	4.7				
STEX origin	σ^*_{CNH}	σ^*_{CH}	$\pi^*_{C=O}$				
IP = {293.7, 295.7}							
STEX with polarized core							
energy (eV)	288.3	288.9	289.8	291.6	292.3		
term value	7.4	4.8	3.9	2.1	1.4		
STEX origin	$\pi^*_{C=O}$	σ^*_{CN}	σ^*_{CH}	σ^*/Rydberg			
IP = {293.7, 295.7}							
experiment, peak position							
IP = {292.3 (Cα), 295.1 (CO)}							
energy (eV)	**288.6**	**289.4**	**290.8**	**294.3**		**300.4**	
(acidic media)					**297.0**		**302.0**
term value[a]	6.5	3.0	1.5	−2(0)		−5.3	
(acidic media)					−1.9		−6.9
glycine (Gly) assignment	$\pi^*_{C=O}$	$\sigma^*_{CNH}/\sigma^*_{CH}$	σ^*_{CN}	σ^*_{CC}	σ^*_{CC}	σ^*_{COO-}	$\sigma^*_{C=O}$

[a] Term value = ionization potential (IP) − excitation energy, where IP values for Gly is taken from gas-phase XPS data.[24]

controlled pH. A more detailed publication will follow;[9] in this article we restrict our attention to acidic media only.

A 1.3M aqueous solution of Gly (pH = 6.3) was mixed with 4 M HCl to prepare samples with different pH. A tiny drop (∼50 μL) of amino acid solution was placed on one Si_3N_4 window. A second silicon nitride window was then gently pressed upon the lower one, and aligned and sealed with glue (Torr Seal epoxy, Vacuum Product Ltd.) to prevent water evaporation and drying. No sign of glue pollution was observed in the NEXAFS spectra. The cell was examined under an optical microscope, and Newton-type interference rings were used to estimate the thickness of the liquid layer.

The NEXAFS spectra of Gly aqueous solutions are presented in Figure 1a. The peak positions and assignments of the transitions in the NEXAFS spectrum of Gly are outlined in Table 1. Only NEXAFS spectra of the samples with normal (pH = 6.3) and lowest pH (pH = 1.2) of the dried samples are reported here; others exhibit similar trends. Fits to multiple Gaussian

lines were performed to determine the positions, widths and relative intensities of individual bands. The continuum step function centered at 292.3 eV is generated by the convolution of a square step with a Gaussian. This function was assumed to be the same for all Gly spectra. In our analysis we used a single-step function, although for Gl*y* gas phase there are two ionization potentials (IP) at 292.3 and 295.1 eV for the Cα and carboxyl groups, respectively.[24] A two-step function fit leads qualitatively to the same picture, only with intensities of the ∼291 and ∼294.3 peaks slightly renormalized. Because the qualitative picture remains the same, and because quantitative oscillator strengths are not the subject of this work, we have used only a single-step function to fit the continuum contribution and Gaussian functions for the discrete line shapes. In Figure 1 and other plots, individual fitted peaks are shown by fine lines, experimental data as markers with the sum of the fitted line drawn through, and residual functions as lines at the bottom, indicating the fitting quality. Note that the Gaussian fits used for Figure 1, and in all other spectral analyses in this paper, are free fits without fixed or constrained parameters. Fits which have different numbers of lines with quite different line parameters (widths, positions) could be envisaged which would be considered better representations of the underlying spectral physics. However, we have chosen to retain this fitting approach which has the virtues of simplicity and minimal presupposition of the final answer. Perhaps the most common limitation of these fits is the frequent occurrence of an overly large continuum step width that results in an intense low-energy tail which robs spectral intensity from some of the discrete peaks.

The building block principle applied to Gly predicts distinct peaks in the NEXAFS spectrum due to σ^*_{C-H}, σ^*_{C-C}, σ^*_{C-N}, σ^*_{C-O}, and $\pi^*_{C=O}$ states. The energy of C 1s→σ^* transitions can be correlated with the length of the associated bond;[25] the term values (TV = IP − *E*) for C−N(R_{C-N} = 1.46 Å), C−C(R_{C-C} = 1.52 Å), and C═O(R_{C-O} = 1.25 Å) bonds are typically 1, −1, and −5 eV, respectively.[24] For the NEXAFS spectrum of Gly (solution, pH ∼ 6.3, Figure 1a), peaks are found at 290.8, 294.3, and 300.4 eV. Assuming the ionization potentials (IPs) for Gly to be equal to 292.3 eV for the methyl and 295.1 eV for the carboxylic carbon, as found in gas-phase XPS measurements,[24] the experimental term values (TV) are 1.5, −2, −5.3 eV. We therefore identify the 290.8 eV peak as arising from C 1s(Cα)→σ^*_{C-N} transitions, the peak at 294.3 eV as C 1s(Cα)→σ^*_{C-C} transitions, and the peak at 300.4 eV as C 1s(C−O)→σ^*_{C-O} transitions. These assignments are in close agreement with those proposed for Gly adsorbed on Cu(110).[26,27] The broad peak at ∼291 eV, assigned above to C 1s→σ^*_{C-N} transitions, has a fwhm of 3 eV. This feature was not observed in the spectrum of chemisorbed Gly as an individual peak, but it may correspond to a shoulder seen in polarization resolved spectra.[26,27] The difference between the chemisorbed and solid state Gly spectra may be the result of chemisorption which modifies mostly states related to N and O, i.e., the two atoms which are closest to the Cu surface participating in the chemisorption bonding. Still, the peak seems too wide for a single bond; possibly different transitions near the IP contribute to the strength of this peak. According to the calculations (shown below) there are many low-intensity transitions associated with the carbon−carbon bonds that are in the same energy range. As a result, the relative intensity of this peak should probably increase for the larger amino acid molecules.

In highly acidic media the ∼300 eV peak in the NEXAFS spectrum of Gly is expected to spilt into two peaks. If the pH is lower than pK_1, then the carboxyl group is protonated and there is a COOH structure, as in the vapor phase. This results in two distinct C−O bonds. The bond lengths for gaseous Gly are $R_{C=O}$ = 1.21 Å, and R_{C-OH} = 1.35 Å, typical of bond lengths for protonated carboxyl groups in other organic substances. On the bases of these bond lengths and the bond length correlations,[25] the expected term values are −6 and −2 eV, respectively. The fitting analysis performed for the spectra of samples at pH below pK_1 reveals such a splitting with term values of −6.9 and −1.9 eV, calculated assuming IP_{COOH} = 295.1 eV.[24] Figure 1a supports our assignment: the same two-peak structure is observed for both dried (shown as a thin continuous line shifted up) and wet (the experimental points as triangles) specimens. The NEXAFS spectra therefore indicate that the charge state of the Gly molecule is preserved through the drying processes. This conclusion is also supported by the observation of a different texture formed by films dried from solutions with pH below and above pK_1.

The NEXAFS spectrum of dried Gly, deposited from TFA solution, is shown in the lower panel of Figure 1a and details of the low-energy region are plotted in Figure 1) (experimental points are marked as diamonds). Above the ionization edge (*E* > 292 eV), the experimental spectrum of dried Gly deposited from TFA solution shows three peaks at about 293.6, 295.7, and 297.5 eV. Because these three peaks are absent for samples from the aqueous solutions, and because the NEXAFS spectrum of gaseous TFA (drawn by a light thin continuous line in the lower panel of Figure 1a) shows three sharp σ^* resonances at the same energies,[28] we may confidently trace their origin to the TFA solvent, which apparently was not removed completely by the air-drying. The peak height of these impurity lines suggests that only a small amount of TFA is left after drying. Except for the carboxyl π^* structure, TFA does not have any intensity below 292 eV, and so the low-energy region of the Gly samples deposited from TFA solvent looks similar to the NEXAFS spectra of specimens crystallized from aqueous solutions. To be exact, the main peak is a little wider in the case of TFA-deposited Gly, and it is better fit by a combination of two separate peaks, as shown below. The low-energy intense peak is presumably coming from TFA, where it is shifted by 0.1 eV to lower energy relative to the gas phase TFA spectrum.[28]

The detailed low-energy structure for the NEXAFS spectra of Gly thin films crystallized from pH controlled solutions are shown in Figure 1b, with the spectrum of acidic solution (pH ∼ 1.2) marked as light triangles shifted up with respect to the NEXAFS spectrum of the unbuffered solution (dark triangles). The low-energy parts of the two spectra are similar: they are dominated by a sharp peak at 288.6 eV. Such a peak can be found in the NEXAFS of all carboxyl compounds. It can be assigned unambiguously to the C 1s(COOH)→π^* transition. Its position remains the same within the accuracy of our measurements (±50 meV) whether the carboxyl group is protonated. This peak, contrary to that reported for acetic acid,[29] has no visible low-energy shoulder. The fitting suggests there is an additional peak at 289.3 eV, as shown in Figure 1c. This might be considered to be an asymmetry of the main line rather than a separate peak, and could possibly be an artifact of our fitting analysis (recall that a Gaussian line shape was assumed). However, polarization measurements of the textured aqueous cast samples show that this is indeed a separate peak which is well distinguished for some crystallites and which shows a clear orientational dependence different from that of the main π^*_{COOH} peak.

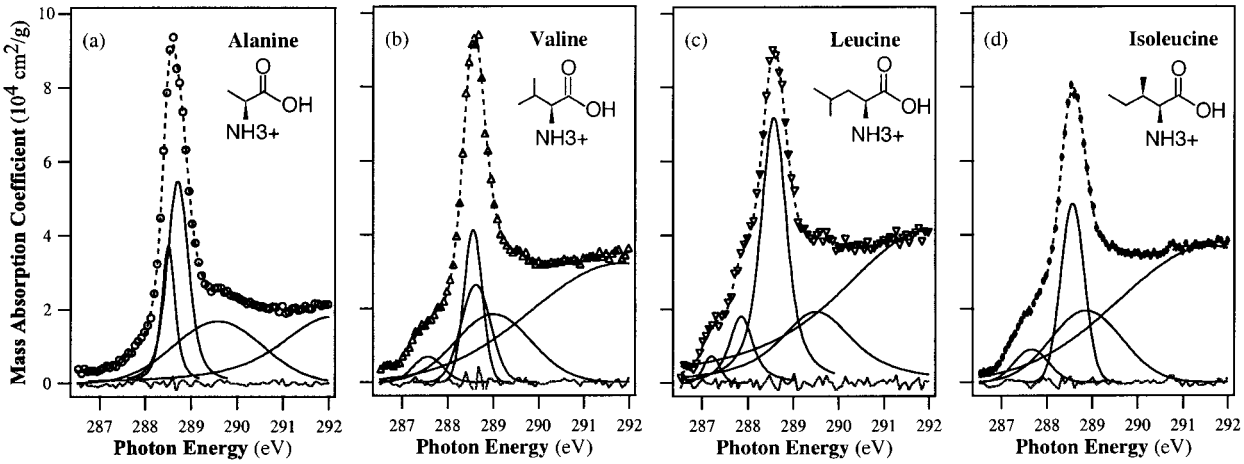

Figure 2. Detailed C K edge NEXAFS spectra of the aliphatic amino acids: alanine, valine, leucine, and isoleucine. The spectra are normalized to the mass absorption coefficients for the appropriate elemental composition[27] to give a direct measure of the molecular mass absorption coefficient. The spectra were fit to multiple Gaussian peaks (solid lines) as described in the text: the sum of these fits is shown as a dashed line, the data are shown as points near this line, and the residuals of the fits are shown at bottom.

The theoretical spectra of Gly are shown below the experimental spectra in Figure 1b,c as bar diagrams with bar heights proportional to the oscillator strength. These computed spectra are discussed later in conjunction with the results of the STEX calculations. Here we note that in both standard and polarized ion STEX approximations, the main peak at ~289.5 eV originates from the C 1s(Cα)→σ^*_{CNH} excitations. As an independent confirmation we have measured NEXAFS spectra of Gly thin films crystallized from aqueous solutions having different pH, and the specimens with pH > pK_2 (above which the amino group loses a proton) show different NEXAFS features at ~290 eV.[9] This is clearly a deviation from the simple building block approach since the NEXAFS spectrum cannot be reduced to the sum of pair bond features; instead, the whole amino group (more accurately $-CNH_3$) needs to be considered as a building block. To summarize:

- First, although the sample preparation procedure leaves the amino acid molecules in a charged state, the low-energy part of the spectra does not exhibit any additional structure with respect to the zwitterionic state and so can be directly compared with our calculations of the C K edge spectrum of the zwitterionic form.
- Second, we can identify two building blocks common for all amino acids: the carboxyl $-COOH$ group, and the $-CNH_3$ group. The carboxyl group leads to the sharp, intense line at 288.65 eV, assigned to the C $1s^{-1}\pi^*$ state. The σ^* excitations originating from the C−O bond of the carboxyl group appear on the high-energy side of the spectra and can be observed either as two separate peaks (protonated), or one broad (deprotonated) peak depending on the charge state.
- Third, the methyl group, which leads to a low-energy peak at ~288 eV in chemisorbed Gly,[26,27] cannot be considered as a good building block. Instead the $-CNH_3$ group, which can be identified by the peaks at ~290.8 eV (with mostly C $1s^{-1}\sigma_{C-N}^*$ character) and the shoulder at 289.3 eV, can be used as a building block.

3.2. Amino Acids with Simple Aliphatic Side Chains: Alanine, Valine, Leucine, Isoleucine. Aliphatic amino acids, such as alanine (Ala), valine (Val), leucine (Leu), and isoleucine (Ile), have nonpolar side chains which do not bind to a surface, which do not release protons, and which do not participate in hydrogen or ionic bonds. The side chains of these amino acids can be thought of as "oily" or lipid-like, a property that promotes hydrophobic interactions. NEXAFS spectra of this class of amino acids are shown in Figure 2 while the energies, term values, and proposed assignments of spectral features are summarized in Table 2. The spectra were normalized to atomic scattering factors[30] in order to give a direct measure of the amino acid mass absorption strength. The lower energy side of each spectrum was fit with a combination of Gaussian functions in order to determine the spectroscopic trend. The separate peak contributions are drawn in thin lines and the fitted spectrum is shown as a continuous line drawn through the experimental points shown by markers. We have restricted the analysis to the part below 292 eV in order to avoid overlap with the TFA peaks. The high-energy side is fit by a step function, broadened by a Gaussian. The number of Gaussian lines used, as well as their parameters, were determined from optimization of a free fit, i.e., without any parameter constraints. This likely leads to a situation where contributions of individual states remained unresolved in some cases and so appear as a single broad band. In particular, as in the case of Gly, both σ^*_{C-C} and σ^*_{C-N} bonds contribute to the oscillator strength near the IP. In the present analysis their contributions cannot be distinguished from near-IP excitations, so the energy position of the step function, as well as its Gaussian broadening, are left to be determined by the fitting procedure, and thus the step function represents all near-IP excitations.

The fit reveals four groups of peaks present in each C K edge spectrum, as shown in Figure 2. All experimental spectra are dominated by a strong peak at 288.65 ± 0.05 eV that can be assigned unambiguously to C 1s(COO)→ π^* transitions. Except for a small shift from Gly to Ala (<0.05 eV), the peak does not change position and coincides with the π^* resonance reported previously for acetic and propionic acids.[31] Two lines separated by 0.2 eV make a better fit for the Ala+ spectrum, although for Val they are spaced more closely (Δ < 0.1 eV). Such a splitting is likely a sign of more complex internal structure of the main C 1s(COO)→π^* line. We attribute this internal structure to vibronic broadening, although an additional peak may result from TFA solvation. More precise gas or liquid phase experiments are needed to unambiguously establish the source of the apparent splitting of the $\pi^*_{C=O}$ peak in Ala and Val.

In all four molecules, the spectra of which are plotted in Figure 2, there is clear evidence for a shoulder on the high-energy side of the main line, at about ~289 eV. It is most evident in Ala. The peak position is ~0.4 eV lower than a

TABLE 2: Absolute Energies, Term Values, and Proposed Assignment for the Features in C 1s NEXAFS Spectra of Aliphatic Amino Acids

(a) Alanine (Ala)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1)		theory (STEX, no screening) IP: Cα = 292.9, C_{chain} = 291.6, COO^- = 294.3	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
CH_3 σ^*_{CH}	?287.6	3.4	287.9	3.7
			288.8	2.8
$CH-NH_3^+$ σ^*_{CNH}	**289.6(2)**	2.6	289.1	3.7
COOH π^*	288.6(1)	6.4		

(b) Valine (Val)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1), C_{branch} = 290.5(2)		theory (STEX, no screening) IP: Cα = 292.6, C_{chain} = 291.4, COO = 294.2, C_{branch} = 290.5/291.2	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
CH σ^*_{CH}	**287.5(1)**	3.0	287.6	2.9
CH_3 σ^*_{CH}		2.5	~288.5	~2.5
$CH-NH_3^+$ σ^*_{CNH}	289.0(2)	3.2	289.2	3.8
COOH π^*	**288.6(1)**	6.4		

(c) Leucine (Leu)

	exptl IP[b]: Cα = 292.2(1), COOH = 295.0(1), C_{chain}[c] = 290.7(3)	
assignment	energy (eV)	term (eV)
CH_2 σ^*_{CH}	**287.9(1)**[a]	2.8
CH, CH_3 σ^*_{CH}	**287.2(1)**	3.5
$CH-NH_3^+$ σ^*_{CNH}	**289.0(2)**	3.2
COOH π^*	**288.6(1)**	6.4

(d) Isoleucine (Ile)

	exptl IP[b]: Cα = 292.2(1), COOH = 295.0(1), C_{chain}[c] = 290.7(3)	
assignment	energy (eV)	term (eV)
CH_x σ^*_{CH}	**287.6(1)**	3.0
$CH-NH_3^+$ σ^*_{CNH}	288.8(2)	3.4
COOH π^*	**288.6(1)**	6.4

[a] The lines whose energies are in boldface are well-resolved and may be clearly seen on Figure 2. [b] Experimental XPS data for Gly gas[24] is modified for other amino acids to reflect the tendency, reported for saturated hydrocarbons.[32] [c] Estimated, based on similar data for saturated hydrocarbon.[66]

similar peak in Gly, and remains the same from Val to Ile. Its origin has already been discussed in the case of Gly. Theory suggests assignment of this peak to a σ^*_{CNH} state, with C 1s(Cα) and C 1s(side chain) excitations to σ^*_{C-H} and σ^*_{C-C} states contributing to the oscillator strength in this region.

The NEXAFS spectra of Val, Leu, and Ile also contain one or more low-energy peaks (~287.6 eV), the position and intensity of which depend on the character of the side chain. Similar low-energy side peaks have been reported for saturated hydrocarbons,[32] where they were assigned to mixed $\pi^*_{CH_3,CH_2}$/3p Rydberg transitions. Further investigations performed for oriented polyethylene films reveal a polarization dependence and suggests the presence of two bands assigned as π^* and σ^*_{C-H} bonds.[33,34] The resonances associated with σ^*/π^*_{CH} bonds become broadened upon chemisorption, forming bands instead of the individual peaks. Our calculations for Ala and Val (Figure 3a,b) support a C 1s(Cα)→C−H origin of the low-energy bands, but an assignment to C 1s(side chain)→σ^*_{CH} transitions may be more appropriate, since the π^* symmetry will not be strict for such short molecular chains consisting of branched $-CH_2-CH_3$ pairs. In addition, the line width seems to be more appropriate for σ^* bonding.

Core-excited Rydberg states have been identified in the past generally on the basis of characteristic term values, which for the lowest energy s, p, and d Rydberg states are around 4.5−3.5 eV, 3.2−2.4 eV, and 1.6 eV, respectively. Transitions to Rydberg states should form a series of sharp peaks, but these are rarely observed in core excitations on account of large natural line widths and peak overlap. As noted for gaseous[33] and solid[35] alkanes, the assignment of low-energy peaks in aliphatic systems is somewhat controversial with both Rydberg and C−H valence-type resonances assignments proposed. Outka et al.[36] have shown that low-energy C K edge peaks in oriented polyethylene and aliphatic Langmuir−Blodgett films exhibit a polarization dependence which seems inconsistent with Rydberg assignments. One may furthermore have Rydberg-valence mixing making it impossible to make an either-or assignment of the peak character. In the solid phase the extended Rydberg states form bands instead of a series of individual peaks.[37,38] As a result, the intensity of Rydberg related NEXAFS peaks is supposed to decrease. We believe that even for small amino acids, such as Gly and Ala, there is no evidence of contributions from Rydberg states so we choose to label such peaks as σ^*_{CH}.

For Val, a low-energy peak appears at 287.5 eV, while the spectrum of Leu is better fit by two closely spaced lines at 287.2 and 287.9 eV. These two lines appear to merge to form a broad line at 287.6 eV with a 1.25 eV fwhm in Ile. The C−C and C−H distances are almost the same for the different aliphatic side chains, so we do not expect a shift of the σ^* peaks due to bond length dependence. The positions of individual lines may still vary due to differences in IP or as the result of splittings due to bond-bond interactions. The calculated partial charges of the side-chain carbons[21] are as follows: $q_{C\beta}$(Ala) = −0.09 ecu; terminated atoms of Val = −0.072 ecu; Leu = −0.011 ecu, where two others have −0.074 ecu; Ile = −0.02,−0.075,−0.075, and −0.075 ecu. If we assume that the energy position of this peak is primarily determined by electrostatic energy, we should have one peak for Val. In addition to this peak, the Leu spectrum should have a shoulder/peak at high energy, and the Ile spectrum should look like that of Val, only with higher peak intensity. This is very close to the experimental observation, with only one difference: the low-energy tail for Val starts at 287 eV, while for Ala it begins at 287.3 eV. We speculate that this is evidence for electron redistribution in the C $1s^{-1}\sigma^*_{C-H}$ excited states, such that excitations on the side-chain atom next to the α-carbon are affected by the amino group.

In the experimental spectra the intensity above the $\pi^*_{C=O}$ peak grows from Gly to Ile, in proportion to the number of carbon atoms, at a more rapid rate that the intensity below the $\pi^*_{C=O}$ peak. The increase of the continuum intensity for complex molecules compared to that of a simple one (Gly) is consistent with absorption cross section calculations based on atomic scattering factors,[30] and reflects the Thomas−Reiche−Kuhn sum rule[39−41] applied to these species. However, the absence of any other distinguishable features above 290 eV may be explained in a different manner. If we assume that broad peaks around 292 eV correspond to C 1s→σ^*C−C transitions, then the low-intensity of the peaks in the 287−288 eV energy range, and the near constant intensity of the "289 eV" line in the Ala-Val-Leu-Ile sequence, suggests that a major density of the wave function of the carbon atom in the side chain is concentrated on the carbon atoms, with very little associated with C−H bonds. It thus seems that increasing delocalization

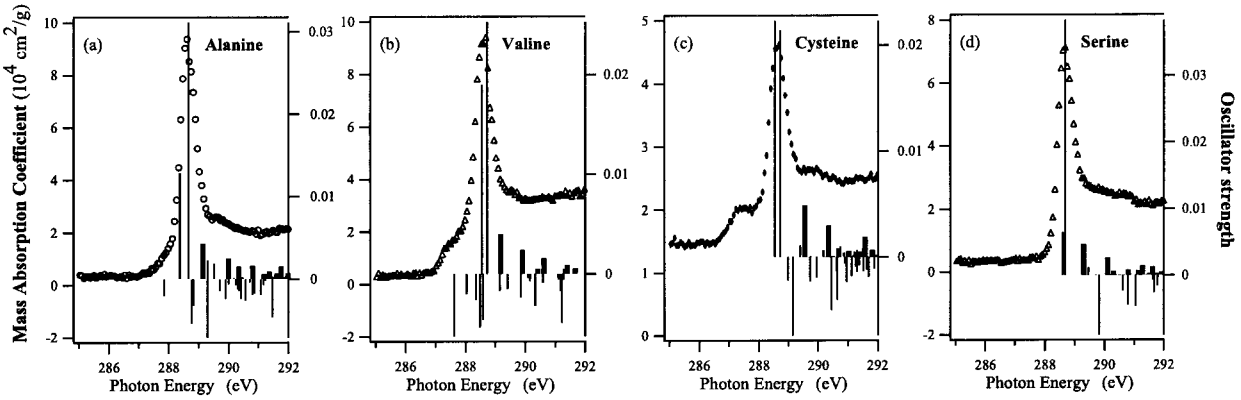

Figure 3. Calculated oscillator strengths for transitions in alanine, valine, cysteine, and serine, along with their experimentally observed spectra. For clarity, side-chain carbon contributions are shown as negative-going lines; Cα contributions are shown as wider positive-going lines, and carboxyl contributions are shown as narrower positive-going lines. The energies for oscillator transitions were calculated in the nonpolarized core STEX approximation; the theoretical energy scale had to be shifted to match the experimental scale by using IP values of 0.6 eV for alanine, 0.4 eV for valine, 1.7 eV for cysteine, and 0.9 eV for serine. The energy positions of the C 1s(COO)→$\pi^*_{C=O}$ transitions were shifted to lower energy by 2.1 eV for alanine, 2.3 eV for valine, 1.7 eV for serine, and 1.1 eV for cysteine to account for incomplete screening of π^* states in the nonpolarized STEX calculations.

in a longer chain affects excited orbitals with density concentrated along the C−C bonds more than those that are relatively localized on C−H bonds.

The computed spectra of Ala, Val, Cys, and Ser are presented in Figure 3. The bars represent the oscillator strengths of particular transitions. They are displayed in three categories: fine-type lines are associated with transitions at the carboxyl atom, while solid lines indicate C 1s and Cα excitations. The combined contributions from side-chain carbons are shown with a negative sign, so that they appear below the zero line. To compare to experiment, the sum of the theoretically calculated oscillator strengths is overlaid on the experimental spectrum (which is in turn shown by markers). As already discussed for Gly, the STEX approximation overestimates the C $1s^{-1}\pi^*$ excitation energy, while reproducing well excitations involving σ^* and Rydberg states.[3] This is of no surprise if we assume that the major unaccounted contribution in STEX is associated with a Coulomb interaction between the created core-hole and the excited electron. The perturbation of a virtual orbital by a core hole will be proportional to the weight of that orbital at the core-hole atom. As a result, the shift for π^* excitations which results from higher electron density at the excited atom is much larger relative to that for the less localized σ^* state excitations. As we see in Figure 1b, the "polarized core approximation" almost accounts for such a shift in the case of Gly, so the carboxyl π^* structure appears at an energy very close to that observed experimentally. As in the experimental spectrum, the theoretical modeling shows the C 1s→π^* transition is the lowest energy structure in Gly followed by less intense C 1s→σ^* transitions at the $-CNH_3$ and $-CH_x$ functional groups (Table 1). The resulting shift of less than 0.5 eV is also reflected in a comparison of IP values, and there is very good correspondence between experimental and theoretical data. On the other hand, the ordinary STEX underestimates the term value for C 1s→π^* transitions and predicts an extended preedge structure (Table 1). Note that the position as well as the oscillator strength associated with σ^* lines vary little, and so, in the absence of extended ab initio calculations with polarized core for the larger amino acids, we have used the following renormalization procedure to improve the ordinary STEX results in an ad hoc fashion. The "additional screening" from the core hole-excited electron Coulomb interaction may be taken into account by shifting the π^* peak *alone* to lower energy in the spectrum computed in the standard STEX approximation (by 1−2 eV as specified in each case). The resulting modified spectrum is then further shifted by the difference between the experimental and calculated IPs, which is typically 0.5−1 eV. As a result of this operation the position of the π^* peak coincides with the experimental value. The resulting total shift is 2.7 eV for Ala and Val, 2.6 eV for Ser, and 2.8 eV for Cys (Figure 3). One may argue that such an "arbitrary" shift would also affect the intensity as well as the energy position of other states. This is correct, but we expect this to be a small effect for the electronic states that originate on the side-chain carbons because of high spatial separation. Its influence on α carbon states is also minor due to the difference in polarization behavior. For Cα, the most pronounced terms originate from σ^*_{CNH} states mixed with σ^*_{CH} (∼289 eV), which have mostly an *in-plane* orientation, whereas the π^* peak has mostly *out-of-plane* character, and so they have very little overlap with each other. To prove the validity of such a semiempirical approach, we prepared the Gly spectrum according to the suggested procedure (carboxyl contribution is shifted by 2.3 eV) and compared it to the polarized core STEX calculations in Figure 1b,c. Good agreement is observed.

Although contributions from side-chain carbons to the NEXAFS spectra are obscured and mixed with transitions coming from the α-carbon atom, the calculation reveals a shift of side-chain carbon oscillator strength with respect to Cα. This is already evident in the case of the theoretical analysis of Ala, where the lower energy C−H transitions from chain carbons (shown below the zero line) are shifted by 0.7 eV relative to the lower energy Cα contribution. Experimental spectra indicate the same behavior, but suggest smaller oscillator strengths for the states associated with low-energy C−H excitations in the side chain. In the case of Val, the low-energy peak (at 287.6 eV) is dominated by C 1s(CH_3)→σ^*_{C-H} excitations. The calculated IP values (290.5 eV for Cγ_1, 291.3 eV for Cγ_2, and 291.4 eV for Cβ) are consistent with computed partial charge numbers,[21] which supports the idea that low-energy peak positions in aliphatic amino acids are primarily determined by electrostatic electron−nuclear attraction.

Another effect of R-group size on C 1s(side chain) excitations predicted by the calculation is intensity transfer from the discrete peaks to the continuum. This is consistent with the experimental trends where, as already noted, the intensity of the preedge structure (E < 289 eV) decreases relative to that of the high-energy signal (E > 289 eV) as the side-chain length increases. Because the intensity of Rydberg states at the carboxyl group

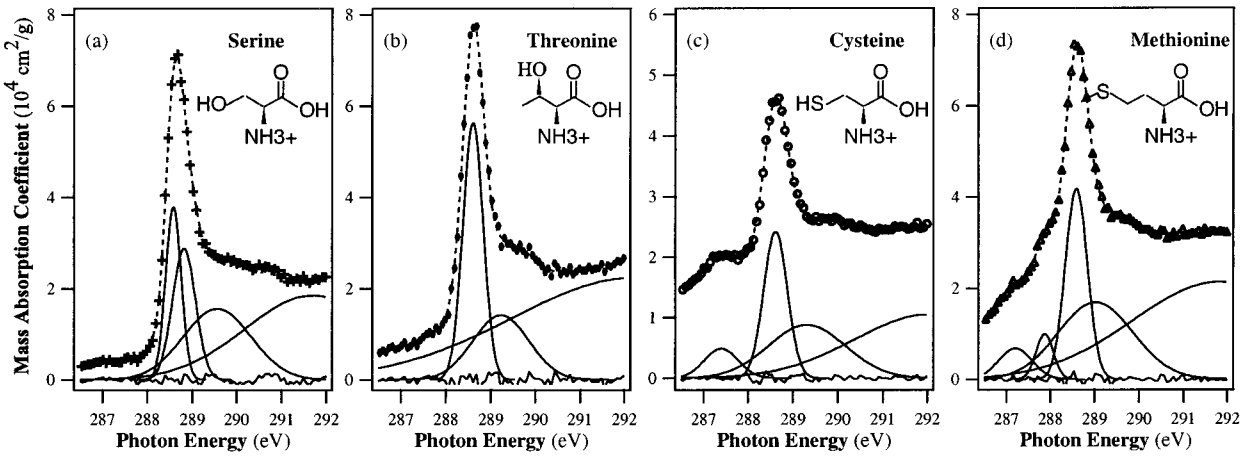

Figure 4. Detailed C K edge NEXAFS spectra of serine, threonine, cysteine, and methionine. The solid lines are the result of spectral decomposition into Gaussian components using least-squares curve fitting. Data processing and symbols as described in the caption to Figure 2.

is small or negligible, the intensity of the high-energy part of the spectrum ($E > 291$ eV) is dominated by excitations at the side-chain carbon, in particular to σ^*_{C-C} states. For instance, the contributions from the Cα atom and the carboxyl group in the NEXAFS spectrum of Ala appear to be similar to their counterparts in the Gly spectrum, but with an additional, relatively strong absorption at 290 eV originating from methyl group carbon 1s excitations to σ^*_{C-H} and σ^*_{C-C} near-degenerate levels. Further details of our interpretation of the C 1s NEXAFS spectra of the aliphatic amino acids are provided in Table 2.

3.3. Amino Acids with Side-Chain Alcohol −OH, −SH, and −SCH$_3$ Groups: Serine, Threonine, Cysteine, and Methionine. The amino acids of this group have uncharged polar side chains and a zero net charge at neutral pH. Serine (Ser), threonine (Thr), and tyrosine (Tyr, see section 3.6) each contains a polar hydroxyl group that can participate in hydrogen bond formation. This class of amino acids, as well as asparagine (Asn) and glutamine (Gln), plays a major role in the formation of the tertiary structures of proteins. Among the most important is cysteine (Cys), which contains a mercapto-group (−SH). Cys plays an active role in enzymes. In proteins the −SH group of two cysteines can become oxidized to form a dimer (cystine) which contains a covalent cross-linked disulfide bond (−S−S−). Polar hydroxyl groups (Syr and Thr) can participate and serve as an attachment site for the phosphate group and oligosaccharide chains in glycoproteins.

The NEXAFS spectra of these compounds are presented in Figure 4, while the results of theoretical calculations for Cys and Ser are shown in Figure 3c,d. Energies, term values, and proposed assignments are summarized in Table 3. In contrast to the aliphatic amino acids, the members of this group have O and S atoms in the side chains, making it appropriate to study the influence of atoms with different electronegativity (S ∼ C ∼ 2.6 and O ∼ 3.4 > C), particularly on the low-energy part of the C K edge NEXAFS spectra. As the L shell of the S atom has a substantial cross section in the investigated energy range, all S-contained amino acids (Cys and Met) are semiabsorbing even below the carbon K edge.

There is a pronounced pre-peak structure at 287.4 eV in Cys that is absent in Ala and Ser. This can be assigned unambiguously to the carbon connected to the mercapto group. Assignment of this peak to C 1s(C−S)→σ^*_{CS} transitions is supported by comparison with early studies of organo-sulfur compounds,[42−44] which showed that C 1s→σ^*_{CS} transitions are located 2−4 eV below the C 1s(C−S) IP with very little dependence on the C−S bond length. The C 1s(C−S) IP in Cys is estimated to be 290.7(3) eV, similar to that of thiolane,[42] which gives a σ^*_{CS} term value of 3.3 eV, 0.6 eV higher than that in thiolane. Although such a high term value is in agreement with bond length correlation[25] (the C−S bond length for cysteine is 1.83 Å compared to 1.53 Å for thiolane and 1.39 Å for thiophene ($\Delta = 3.2$ eV)), we believe that the peak positions for sulfites are mostly determined by the IP value, as one can see in the case of Met. After renormalization, the STEX calculation reproduces nicely the Cys experimental spectrum. As seen from Figure 3c, the intensity on the low-energy side of the $\pi^*_{C=O}$ peak is mostly of σ^*_{CS} origin. The existence of a low-energy Cα contribution is more controversial. The large shift (1.7 eV) which was needed to align the theoretical and experimental C K IP values for Cys is most likely due to the limited basis set for S used in the calculation. This basis set factor may also affect the Cα IP value and thus it could push down artificially the C 1s(Cα)→σ^*_{CNH} transition energy.

The fitting of the Met NEXAFS spectrum shows there are two peaks in the low-energy signal at 287.2 and 287.9 eV. This appears to be a contradiction to assignment of the low-energy excitation to C 1s(C−S)→σ^*_{CS} transitions, because the C−S bond lengths of −C−SH and −S−CH$_3$ groups are equivalent, and are about 0.05 Å shorter than that of cysteine, so a single peak at somewhat higher energy is expected. The calculations of the ground-state charges[21] suggest that the main difference between those two carbons arises from partial charges of −0.19 ecu and −0.12 ecu for the carbons in the −C−SH and −S−CH$_3$ groups, respectively. For aliphatic amino acids, we estimate a charge difference of 0.1 ecu can shift a peak by 1 eV, which is close to the observed 0.7 eV splitting between the two low-energy peaks in Met.

The spectrum of Ser, as that of Gly, does not exhibit structure below the main C 1s(COO)→$\pi^*_{C=O}$ transition. The integrated $\pi^*_{C=O}$ intensity is comparable to that of Ala, although the peak is more compact in Ser and so a distinct shoulder appears at the high-energy side with no intensity below the carboxyl π^* peak. The calculation of Ser seems to be in accordance with the experimental picture (Figure 3d). After renormalization of the energy scales, the C 1s(CH$_2$)→σ^*_{CH} transition appears ∼0.5 eV higher than that originating from the Cα excitation, and both peaks are predicted to occur on the high-energy side of the main π^* peak.

Based on the molecular structure, the spectrum of Thr should be related to those of Ala and Ser. It shows indeed a low-intensity preedge signal as in Ala, and an intense shoulder as in the Ser spectrum. Our fit to the C K edge spectrum of Thr

TABLE 3: Absolute Energies, Term Values, and Proposed Assignment for the Features in C 1s NEXAFS Spectra of Amino Acids with Side-Chain Alcohol −OH, −SH, and $-SCH_3$ Groups[a]

(a) Serine (Ser)

	exptl IP[b]: Cα = 292.2(3), COOH = 295.0(1), CH_2OH[d] = 292.5(3)		theory (STEX, no screening) IP: Cα = 292.8, CH_2OH = 293.4, COO^- = 294.2	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
CH_2 σ^*_{CH}			289.8	3.6
$CH-NH_3^+$ σ^*_{CNH}	289.6(2)	2.6	289.3	3.5
COOH π^*	**288.6(1)**	6.4		

(b) Threonine (Thr)

	exptl IP[b]: Cα = 292.2(3), COOH = 295.0(1), CH_3[c] = 290.7, CHOH[d] = 291.5(3)	
assignment	energy (eV)	term (eV)
CH_3 σ^*_{CH}	**287.6(1)**	3.1
COOH σ^*_{CH}		3.9
$CH-NH_3^+$ σ^*_{CNH}	**289.3(2)**	2.9
COOH π^*	**288.6(1)**	6.4

(c) Cysteine (Cys)

	exptl IP[b]: Cα = 292.2(1), COOH = 295.0(1), C^c_{chain} = 290.7(3)		theory (STEX, no screening) IP: Cα = 293.0, C_{chain} = 292.4, COO^- = 294.4	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
$HS-CH_2$ σ^*_{CHS}	**287.4(1)**	3.3	289.0	3.4
$CH-NH_3^+$ σ^*_{CNH}	289.3(2)	2.9	289.5	3.5
COOH π^*	**288.6(1)**	6.4		

(d) Methionine (Met)

	exptl IP[b]: Cα = 292.2(1), COOH = 295.0(1), C^c_{chain} = 290.7(3)	
assignment	energy (eV)	term (eV)
CH_xS^- σ^*_{CHS}	**287.2(1)**	3.5
	287.9(1)	2.8
$CH-NH_3^+$ σ^*_{CNH}	289.3(2)	2.9
COOH π^*	**288.6(1)**	6.4

[a] The lines whose energies are in boldface are well-resolved and may be clearly seen on Figure 3. [b] Experimental XPS data for Gly gas[24] is modified for other amino acids to reflect the tendency, reported for saturated hydrocarbons.[32] [c] Estimated, based on similar data for saturated hydrocarbon.[66] [d] Estimated, based on data for alcohols.[31]

suggests that the low-energy signal is from a broad edge function. However, we believe this to be an artifact of the fit, and prefer to interpret the low-energy signal as arising from C $1s(CH_3) \rightarrow \sigma^*_{CH}$ transitions as in the case of Ser.

3.4. Amino Acids with Side Chains Containing Carboxylic Acid or Amine Groups: Aspartic Acid, Glutamic Acid, Asparagine, and Glutamine. Most biological reactions take place in aqueous solution, where the reactivity is crucially influenced by the pH of the solution. Being proton donors, amino acids may control the ionic content of solution. At neutral pH the side chains of amino acids that contain carboxylic acid groups (Asp and Glu) are fully ionized, containing a negatively charged carboxylate group $(-COO)^-$. The side chains of some basic amino acids, such as lysine (Lys) and arginine (Arg) (section 3.5), accept protons and are positively charged. Apart from these effects, protonation of the amino group in amino acids results in the creation of zwitterions. In extremely acidic media (such as in TFA solution), all these side groups as well as the backbone amino and carboxyl groups are protonated, and as we saw in the case of Gly, the drying preserves the charged state that existed in solution. As indirect evidence of protonation due to the presence of TFA, we note that the peaks at 294−298 eV (not shown) that are associated with residual TFA were about twice as intense as those for Arg and Lys as for the other amino acids. Twice as much TFA is needed to neutralize the doubly charged Arg and Lys ions deposited from TFA solution. With such a distinction in mind, we anticipate to observe: (a) an influence of substituted carbonyls (−COX, X = NH_2 or OH) on the location of the C $1s^{-1}\pi^*_{CO}$ structure, (b) spectroscopic features of C═N double bonding, and (c) a possible distinction between imino and amino groups in regards to their appearance at the carbon K edge.

The C K edge spectra of this group are presented in Figure 5 while energies, term values, and assignments are listed in Table 4. The spectra of the amino acids with two carboxyl groups (Glu and Asp) have an intense line centered at 288.60(5) eV with a fwhm of 0.6 eV, which is narrower than the $\pi^*_{C=O}$ peak of Gly. The fitting for both Asp and Glu suggests that the main $\pi^*_{C=O}$ peak consists of two lines separated by 0.2 eV. High-resolution gas-phase experiments could help to confirm this suggestion. The major difference from Gly is the presence of a high-energy shoulder on the 289.5−292 eV region. This feature is similar to that found in the aliphatic amino acids. In contrast, Gly has a distinct peak at 291.2 eV.

Asparagine (Asn) and glutamine (Gln) have structures similar to those of Asp and Glu, respectively, except for the substitution of the −OH group by $-NH_2$. Both have a main $\pi^*_{C=O}$ peak with maximum at 288.5(1) eV, which is wider (FWHM = 0.8 eV) than that in Asp or Glu (Figure 5). The fit indicates that the additional width is associated with a second low-energy peak with a separation of 0.3 eV (Asn) or 0.4 eV (Gln). The effect of substitutients (X) with different electronegativity on the energy and oscillator strength of the $\pi^*_{C=O}$ orbital has already been reported in a study of formamide, formic acid, and formyl fluoride.[45] It was found that the smaller electronegativity of X = $-NH_2$ relative to −OH decreases the term value by 1.2 eV, although the energy position changes only a little (−0.1 eV) because the IP value is lowered by the same amount. The oscillator strength of the major C $1s \rightarrow \pi^*$ peak also decreases from −COOH to $-CONH_2$. Comparison of the Asp (Glu) and the Asn (Gln) spectra shows exactly the same trend. Following the same arguments as in ref 45 we conclude that the $\pi^*_{C=O}$ orbital is less tightly bound when the hydroxyl group is substituted by $-NH_2$. There is clearly an effect of hyperconjugation of the $\pi^*_{C=O}$ bond, which results in shortening of the C−N distance (R_{C-N} = 1.33 Å for Asn).

Comparison of the spectra of this group with those of the aliphatic amino acids (section 3.2) clearly shows that the oscillator strength related with transitions at side-chain atoms and amino groups is seriously suppressed for a side-chain terminated by a C═O group. Even for Glu with a relatively long chain the low-energy region of the NEXAFS spectrum is closer to Ala than to Val, and shoulders assignable to C $1s-(CH_x) \rightarrow \sigma^*_{CH}$ transitions are not observed. The results also suggest a shift of the $\pi^*_{C=O}$ peak to higher energy for this group of amino acids, or the presence of an additional peak on the high-energy side of the main carboxyl π^* structure. In the case of $-NH_2$ substitution it may be an additional C $1s^{-1}\sigma^*_{CNH}$ transition, which in the case of formamide was found to lead to a separate peak at 290.7 eV.[45] The origin of the high-energy signal around 290 eV for Asp and Glu is unclear.

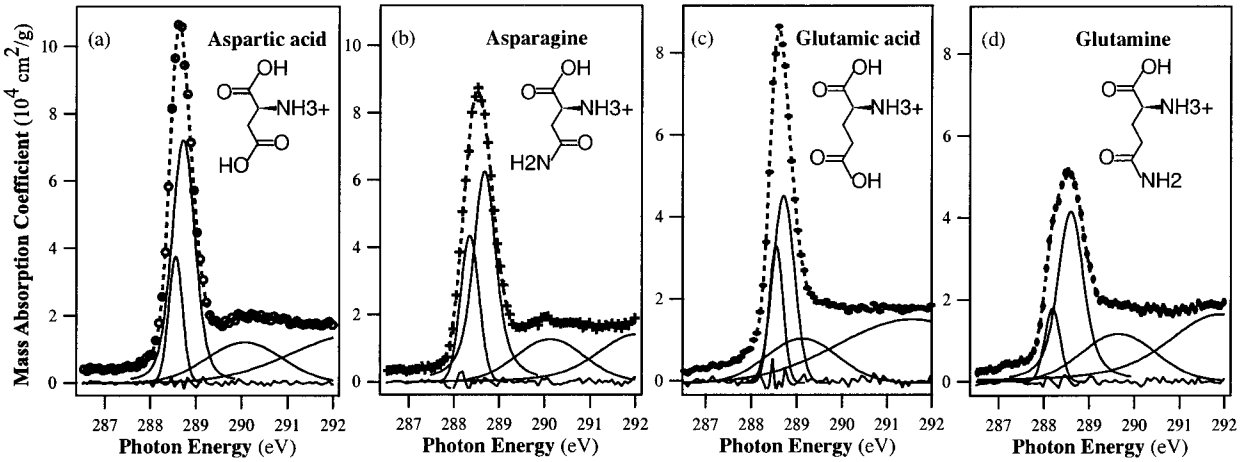

Figure 5. Detailed C K edge NEXAFS spectra of aspartic acid, asparagine, glutamic acid, and glutamine. Data processing and symbols as described in the caption to Figure 2.

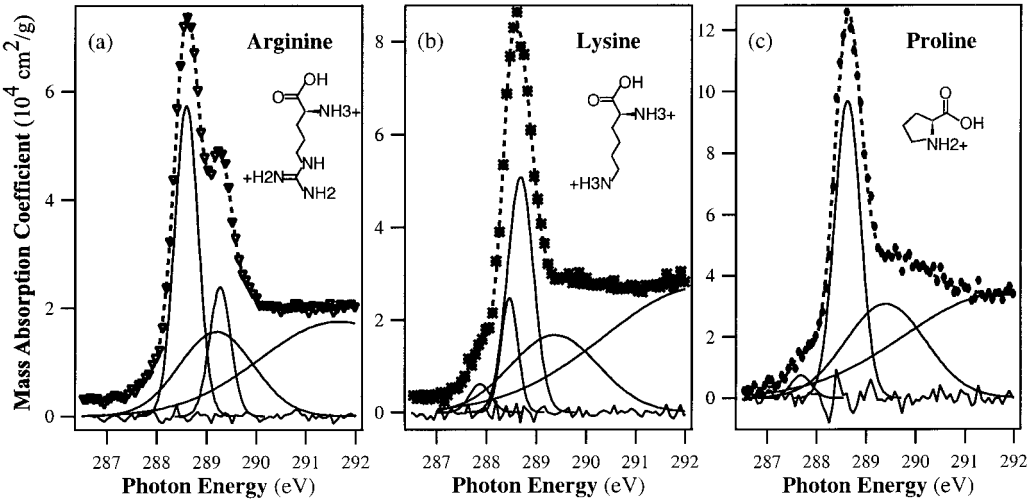

Figure 6. Detailed C K edge NEXAFS spectra of arginine, lysine, and proline. Note that the chemical formula appearing in these figures reflects the expected amino acid charge state. Data processing and symbols as outlined in caption to Figure 2.

TABLE 4: Absolute Energies and Proposed Assignment for the Features in C 1s NEXAFS Spectra of Amino Acids with Side Chains Containing Carboxylic Acid or Amide Groups, Strong Basic Group, and Cyclic (Secondary) Amino Acids

	aspartic acid (Asp)			asparagine (Asn)			glutamic acid (Glu)			glutamine (Gln)		
energy (eV)	288.50(5)	288.70(5)	290.1(1)	288.30(5)	288.60(5)	289.8(1)	288.50(5)	288.70(5)	289.1(1)	288.20(5)	288.60(5)	289.7(1)
assignment	π^*_{COOH}	π^*_{COOH}	σ^*_{CNH}	π^*_{CONH2}	π^*_{COOH}	σ^*_{CNH}	π^*_{COOH}	π^*_{COOH}	σ^*_{CNH}	π^*_{CONH2}	π^*_{COOH}	σ^*_{CNH}

	arginine (Arg)			lysine (Lys)				proline (Pro)		
energy (eV)	288.60(5)	289.30(5)	289.2(1)	287.8(1)	288.50(5)	288.70(5)	289.5(1)	287.6(1)	288.65(5)	289.3(1)
assignment	π^*_{COOH}	$\pi^*_{-HNCNH_2-NH_2}$	σ^*_{CNH}	σ^*_{CH}	π^*_{TFA}	π^*_{COOH}	σ^*_{CNH}	σ^*_{CH}	π^*_{COOH}	σ^*_{CNH}

3.5. Amino Acids with Side Chains Containing Strongly Basic or Cyclic Groups: Lysine, Arginine, and Proline. The spectra of the amino acids with basic groups arginine (Arg), lysine (Lys), and proline (Pro) are presented in Figure 6 while the energies, term values, and proposed assignments are listed in Table 4. In arginine (Arg) the discrete structure consists of two strong peaks. The lower energy peak at 288.6 eV is the $\pi^*_{C=O}$ signal based on comparison to all the other amino acids. The higher energy peak at 289.24(5) eV is thus attributed to C 1s(C═N)→$\pi^*_{C=N}$ transitions. At first sight the high energy of the $\pi^*_{C=N}$ peak is puzzling since the higher electronegativity of O relative to N would suggest the C 1s(C═N)→$\pi^*_{C=N}$ transition might occur below the C 1s(COOH)→$\pi^*_{C=O}$ transition. However, the C═N carbon is actually bonded to three N atoms, which shifts the C 1s(C═N) IP and thus the π^* peak above that of a C 1s(COO^-) site. Recent systematic studies of the positions of $\pi^*_{C=O}$ transitions as a function of the electronegativity of adjacent groups[46,47] support this interpretation.

To summarize, the following tendency for the C $1s^{-1}\pi^*$ peak is obtained. The carboxyl peak position −288.6 eV is lowered due to the $-NH_2$ substitution by −0.1 eV as a result of conjugation of $\pi^*_{C=O}$. For Arg the second peak appears shifted by 0.6 eV making it a fingerprint of $H_2NHN{=}C{-}NH{-}$ group.

The influence of a side-chain amino group is evident by comparison of the NEXAFS spectra of Lys and Pro (Figure 6b,c) with those of the aliphatic amino acids (Figure 2). Although the peak positions are very similar (consistent with expectations since these molecules contain many of the same "building blocks"), the relative intensity of the low-energy step (∼287.6 eV) and high-energy shoulder (∼290 eV) differ. In the case of Pro and Lys, the low-energy peak is less pronounced, whereas the high-energy shoulder gains intensity relative to the

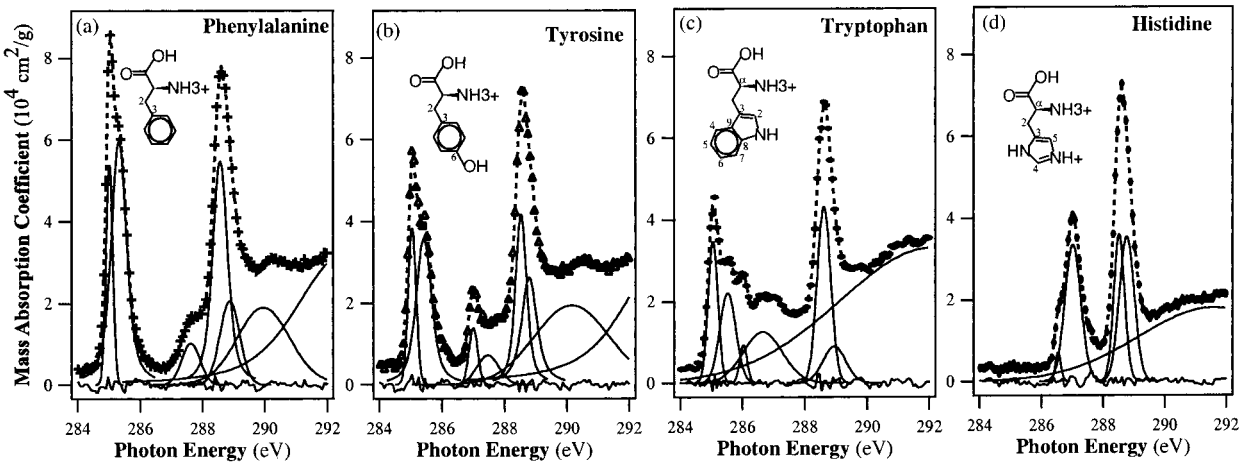

Figure 7. Detailed C K edge NEXAFS spectra of phenylalanine, tyrosine, tryptophan, and histidine. Some carbon atoms are numbered in the displayed chemical formulas to clarify their spectroscopic assignment. Data processing and symbols as outlined in caption to Figure 2.

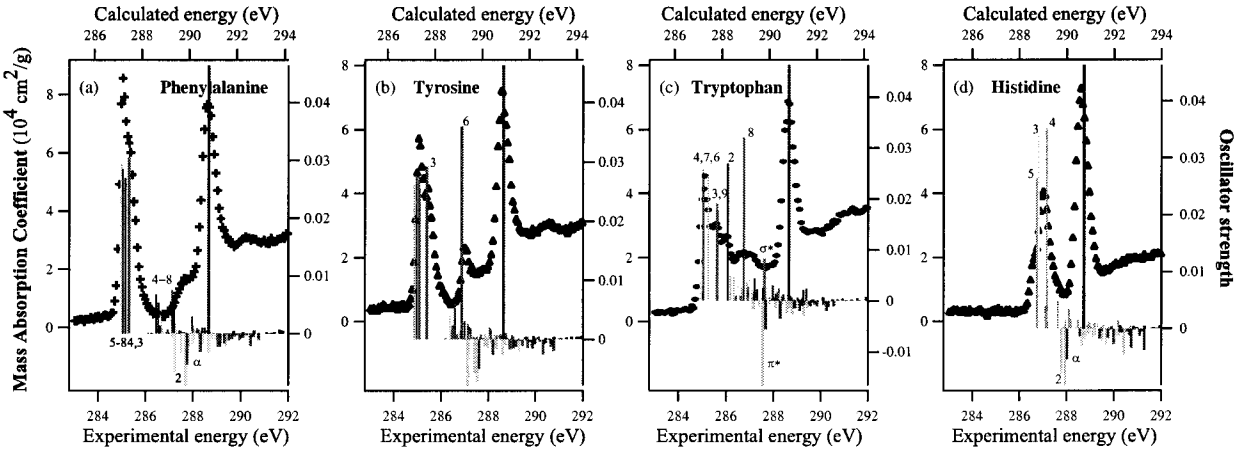

Figure 8. Computed NEXAFS spectra of phenylalanine, tyrosine, tryptophan, and histidine. The energy positions and intensities of optical transitions are calculated in the nonpolarized STEX approximation. The experimental mass absorption coefficients are shown, along with bars that indicate the energy and oscillator strengths of various transitions. Side-chain carbon and Cα contributions are shown below the zero line (negative intensities); aromatic ring and carboxyl contributions are shown above the zero line (positive intensities). In addition, transitions originated from different carbon atoms are shown by different line style. Some lines are numbered and correspond to transitions at the numbered atoms in Figure 7. The computed carboxyl contribution has been shifted to lower energy by 1.3 eV for phenylalanine, 0.4 eV for tyrosine, 1.3 eV for tryptophan, and by 1.4 eV for histidine to reflect the core ion relaxation effect, as described in the text.

$\pi^*_{C=O}$ peak. As we noted earlier, most of the C 1s(side chain) oscillator strength contributes to the near IP featureless step function (290−292 eV), and, to a smaller extension, to low-energy σ^*_{CH} excitations. The same tendency exists in the amino acids with side groups containing electronegative hydroxyl groups (Ser and Thr), where no low-energy side shoulder is seen. We conclude that there is a general tendency whereby substitution of carbon in the side chain by more electronegative atoms (N and O) primarily affects the intensity of the low-lying σ^*_{CH} states, although their location (∼287.5 eV) does not change. There is a small shift of the high-energy shoulder at ∼289−290 eV to higher energy in Lys and Pro relative to the other species. The C K edge spectrum of pyrrolidine[48] exhibits a pronounced peak at 290.7 eV, which is assigned to σ^*_{C-N} states. This comparison suggests that C 1s(C−N)→σ^*_{CN} transitions may also contribute to the oscillator strength around 290 eV. This actually may explain a similar shift observed for Asn and Gln.

3.6. Amino Acids with Aromatic Side Chains: Phenylalanine, Tryptophan, Tyrosine, and Histidine. Amino acids with aromatic side chains are readily distinguished from other amino acids since their NEXAFS spectra show pronounced structures around 285 eV, associated with the $\pi^*_{C=C}$ states of the aromatic ring. The fitted experimental spectra of the four common aromatic amino acids phenylalanine (Phe), tyrosine (Tyr), tryptophan (Trp), and histidine (His) are presented in Figure 7, while the corresponding spectra computed with the STEX method are presented in Figure 8. The energy scale of the calculated spectra was shifted by 2.1 eV for Phe, by 2.2 eV for Tyr, by 2.1 eV for Trp, and by 2.0 eV for His. In addition, the carboxyl peak alone was moved toward lower energy by 1.3 eV for Phe, 0.4 eV for Tyr, 1.3 eV for Trp, and by 1.4 eV for His. Table 5 lists peak energies, term values, and proposed assignments for these species.

The "building block" principle predicts that the spectrum of Phe should resemble the sum of the spectra of benzene and Ala. The spectrum of gas-phase benzene consists of three low-energy lines.[49−51] The most intense line with a maximum at 285.1 eV is identified as a C 1s→$1\pi^*(e_{1u})$ transition. The next peak is a low-intensity line at 287.2 eV attributed to σ^*_{CH}/Rydberg excitations. The third line at 288.9 eV is attributed to a C 1s→$2\pi^*(b_{2g})$ transition, although there is some controversy over that assignment.[52] In addition to these low-lying features, there are strong, broad continuum resonances in benzene of σ^*_{C-C} character at 293.5 and 299.8 eV.[49,50] The aliphatic acids have a strong $\pi^*_{C=O}$ peak at 288.6 eV, with a low-energy tail

TABLE 5: Absolute Energies, Term Values, and Proposed Assignment for the Features in C 1s NEXAFS Spectra of Aromatic Amino Acids[a]

(a) Phenylalanine (Phe)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1), C_{ring}[c] = 290.3(1)		theory (STEX, no screening) IP: Cα = 293.0, C_{chain} = 292.8, COO^- = 295.7, C_{ring} = 291.3	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
$C_{ring}\ \pi^*_1$	**285.0(1)**	5.3	287.2	4.1
$C_3\ \pi^*_2$	285.3(1)	5.0	287.5	3.8
$CH_{ring}\ \sigma^*_{CH}/\pi^*$	**287.6(1)**	~2.5	288.6	4.2
COOH π^*	**288.6(1)**	6.4		
$CH-NH_3^+\ \sigma^*_{CNH}$	289.0(2)	3.3	289.9	3.1
$C_{ring}\ \sigma^*$/IP	**290.3(2)**	~0		

(b) Tyrosine (Tyr)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1), C_{ring}[c] = 290.2(1), COH[d] = 292.0(2)		theory (STEX, no screening) IP: Cα = 293.0, C_{chain} = 291.8, COO^- = 294.8, C_{ring} = 291.3, COH = 293.0	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
$C_{ring}\ \pi^*_1$	**285.0(1)**	5.3	287.2	4.1
$C_3\ \pi^*_2$	**285.4(1)**	4.9	287.6	3.8
$C_6\ \pi^*_2$	**287.0(1)**	3.6	289.1	3.9
$CH_{ring}\ \sigma^*_{CH}/\pi^*$	**287.5(1)**	~3.5	289.6	2.2
COOH π^*	**288.5(1)**	6.5		
$CH-NH_3^+\ \sigma^*_{CNH}$	288.8(2)	3.4	289.9	3.1
$C_{ring}\ \sigma^*$/IP	290.3(2)	~0		

(c) Tryptophan (Trp)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1), C_{ring}[c] = 290.3(2), CNH[e] = 291.8		theory (STEX, no screening) IP: Cα = 292.8, C_{chain} = 291.8, COO^- = 295.5, C_{ring} ~ 291.0, CNH = 291.8	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
$C_{ring}\ \pi^*_1$	**285.0(1)**	5.3	287.2	3.8
$C_{3/9}\ \pi^*_2$	**285.5(1)**	4.7	287.6	3.2
$C_2\ \pi^*_2$	**286.0(1)**	5.0	289.1	3.6
$C_8\ \pi^*_2{}^+\ CH_{ring}\ \sigma^*_{CH}/\pi^*$	**286.6(1)**	~5.2	288.9	2.9
			289.8	2.0
COOH π^*	**288.6(1)**	6.4		
$CH-NH_3^+\ \sigma^*_{CNH}$	288.9(2)	3.3	289.9	3.1

(d) Histidine (His)

	exptl IP[b]: Cα = 292.2(1), C_{chain} = 291.0(2), COOH = 295.0(1), C_4[e] = 291.8, C_5[e] = 290.5, C_3[e] = 291.1		theory (STEX, no screening) IP: Cα = 293.2, C_{chain} = 292.3, COOH = 296.0(1), C_4 = 293.1, C_5 = 291.8, C_3 = 292.6	
assignment	energy (eV)	term (eV)	energy (eV)	term (eV)
$C_3\ \pi^*$	286.5(1)	4.6	288.7($C_{3/5}\ \pi^*$)	3.4
$C_{5/4}\ \pi^*$	**287.0(1)**	~5	289.2 ($C_4\ \pi^*$)	~3.3
?TFA	288.5			
COOH π^*	**288.7(1)**	6.3		
			289.8 (σ^*_{CH})	2.5
			290.0 (σ^*_{CNH})	3.2

[a] Carbon atoms are assigned as shown in Figure 7; the energy positions of their peaks are in boldface and are clearly resolved and primed by dash. [b] Estimated, based on XPS data for gaseous benzene.[53] [c] Estimated, based on XPS data for gaseous phenol.[67] [d] Estimated, based on XPS data for gaseous imidazole.[65]

and a low-intensity shoulder on the high-energy side. At first sight, there appears to be a good match between the sum of the spectra of benzene and Ala with the spectrum of Phe (Figure 7a). Quantitative comparison reveals that although the overall picture is correct, the line at 287.65 eV and the high-energy shoulder at 290 eV are both more intense than would be expected from a simple sum of constituted parts.

Theory and experiment agree that the peak at 287.65 eV is C 1s→σ^*_{CH} in character but the excitation site is not clear. The C 1s IP of benzene is 290.3 eV, whereas that of saturated hydrocarbons is ~290.7 eV in the gas phase.[53] Consequently, σ^*_{CH} lines from the aromatic ring carbons should appear 0.7 eV lower than those associated with the CH_2 side-chain carbon. This is clear from the calculated spectra, where two groups of the peaks (4−8; both marked as σ^*_{CHring}) are located 1.8 eV above the $\pi^*_{C=O}$ peak. The side-chain contribution, shown with reversed sign below the zero line, is shifted by 0.7 eV to higher energy. As the location of the C 1s(side chain) peaks coincides well with the experimentally observed peak, we assign them to $\sigma^*_{CHchain}$. As no other peaks are seen in the spectrum, we believe that there is an electron density redistribution relative to the "parent" compounds such that transitions to σ^*_{CHring} are suppressed, while for the chain carbons the intensities of the $\sigma^*_{CHchain}$ transitions are enhanced.

A broad intense peak leading to a "bump" is apparent in the spectra of Phe at 290.3 eV and Tyr at 290.5 eV. Such a peak is

absent in the NEXAFS spectra of both aliphatic amino acids and of the benzene molecule. We assign this peak to σ^*_{CC} states with possible admixture from the $2\pi^*(b_{2g})$ states.[52] This assignment requires a significant shift of the σ^*_{CC} line relative to its counterpart in benzene, which appears at 293.5 eV. This can be rationalized by the reduction of the 6-fold symmetry of the phenyl ring in Phe. As a result, the C−C bond lengths, which are 1.40 Å for benzene, 1.42 Å for C_4-C_5, C_7-C_8 pairs, and 1.36 Å in the apex (C_3-C_4, C_5-C_6, C_6-C_7, and C_8-C_3) (carbon atoms are numbered as shown in the Figure 8) of Phe. For Tyr, due to an additional −OH group, the bond lengths are even more distributed with values of 1.38 Å for atoms close to the side-chain carbon, 1.36 Å for carbons next to C−O bond, and 1.44 Å between the middle atoms in the ring. The $1/R_{C-C}$ empirical relationship has a slope of −5 eV/0.1 Å.[25] When this is applied to the σ^*_{CC} peak position in the phenyl rings of Phe and Tyr, it predicts an average −2 eV shift of the σ^* peak in Phe compared to benzene. Although such a "predicted" position is still about 1 eV higher than that we found in the experiment, we mark this peak as $\sigma^*_{C-Cring}$. We note that the application of the bond length correlation to benzene is a special case[54] with considerable controversy.[55]

Strictly speaking all six C atoms combined in the benzene ring are not chemically equivalent in Phe. However, as was already observed in previous investigations of substituted benzene molecules,[56−58] the unique C 1s(C−X)→1π^* line (X = the C atom directly connected to the chain) shows a remarkable shift (Figure 8a). All other π^*_{ring} states have almost the same value with maximum at 285.05(5) eV (Figure 7a). The fit to the Phe spectrum indicates that the 285 eV line consists of two peaks separated by 300 meV. Although this is very close to that expected from theory (200 meV between the peaks marked as 5−8 and 3, Figure 8a), such a result must be interpreted cautiously. High-resolution spectra of benzene[59] and other systems with phenyl substituents such as polystyrene[60] clearly show vibrational fine structure with a similar line shape as seen in Phe (and Tyr): a sharp "$\nu = 0$" spike with a broader shoulder about 400 meV at higher energy. At the present energy resolution (FWHM ∼ 150 meV) we cannot differentiate this expected vibrational structure from the expected shifts of the 1s→π^* states associated with the different carbon sites around the ring. However, we clearly see a difference between Phe and Tyr since the second shoulder in Tyr is shifted by 400 meV, which is in agreement with the theoretical prediction (Figure 8b).

The role of chemical substitution is emphasized in the spectrum of the Tyr (Figure 7b). Tyr can be considered as a derivative of Phe in which an −OH group is substituted in the para position. Calculations (Figure 8) show that the contributions of the carbon atoms not directly involved in the −OH bond formation are almost unchanged. For the carbon atom bonded to the hydroxyl group, the strong electronegativity of the O atom induces a large positive shift which leads to a separate narrow peak at 287.0(1) eV. The splitting of the C 1s(C−OH)→1π^* and C 1s(C−H)→1π^* transitions in Tyr is 1.9 eV, in agreement with calculation. This value is similar to what was found previously for phenol,[50,61] so one can unambiguously identify the peak at 287.0 eV as the C 1s(C−OH)→1π^* transition. The intensity redistributions caused by hydroxyl group substitution produce an apparent reduction in intensity of the lines near 285.1 eV relative to the intensity of the carboxyl π^* peak. This is primarily due to a splitting of the oscillator strength and transfer of intensity to the C 1s(C−OH)→1π^* peak. The fit suggests that the −OH substitution also increases the intensity of the C 1s→σ^*_{C-C} transitions around 290 eV.

Due to the complexity of the tryptophan (Trp) spectrum (Figure 7c), we limit our discussion to the π^* peaks coming from the aromatic ring. The influence of aromatic states on position and intensity of electronic transitions, originating from the "aliphatic amino acid building block", is difficult to identify. The fit leads to an unrealistically broad step feature in the spectrum. The aromatic component of Trp may be considered as a fusion of two aromatic molecules: benzene and pyrrole.[6] It is interesting to consider if a building block approach to describing its NEXAFS spectra has any validity. As in the case of benzene derivatives such as aniline, phenol, and monofluorobenzene,[62] the break up of the six- fold symmetry splits the 1π^* peak. To account for such symmetry reduction we compare the Trp spectrum to those of aniline[63] and pyrrole[58] rather than benzene. Pyrrole has two intense lines centered at 286.8 eV with a peak separation of 0.6 eV which correspond to $C_1\rightarrow\pi_4^*$ and $C_2\rightarrow\pi_4^*$ transitions, where the C_2 is the atom adjacent to N. In addition to π^* structure similar to that of benzene (285.4 eV), aniline shows a low-intensity peak (π_2^* (b_1)) symmetry, which is shifted by 1.2 eV.[63,64] The experimental spectrum of Tyr shows a five peak structure with the maxima at 285.1, 285.5, 286.0, 286.6, and 287.0 eV. It is clear that the substantial conjugation in the fused ring structure of Trp leads to a different electronic structure, one which cannot be derived from the "building block" principle. The STEX calculations[3] suggest the following assignment. The group of peaks centered at 285.1 eV is due to C 1s→1π^* transitions at the four carbon atoms of the benzene ring distant from the benzene−pyrrole double bond. The energy positions of two π^* peaks, one originated from the carbon atom connected to the chain, and another from its neighbor carbon participating in the benzene-to-pyrrole double bond, coincide at 287.8 eV. Accounting for an expected ∼2.2 eV normalization shift, these π^* peaks should correspond to the experimentally observed peak at 285.5 eV. The contribution of the carbon placed between N and the side-chain carbon is shifted down by 1.45 eV, so the 286.0 eV peak may be assigned to $\pi^*_{C_2}$. The carbon atom which participates in the benzene-pyrrole double bond formation, next to the N atom, has the largest shift, ∼1.7 eV compared to the "bare" benzene line. In the experimental spectrum this peak appears 0.2 eV lower in energy, possibly due to overlap with σ^*_{CH} states. The unique feature suggested by the calculation is due to π^* bond formation and to conjugation of the aromatic rings and the carboxyl group. The experiment does not show any substantial oscillator strength in that energy region. We speculate that if this line is shifted 0.5 eV to lower energy, it will be mixed with $\pi^*_{C_8}$ and so be part of the broad peak at 286.6 eV (Figure 7c).

In histidine (His), the imidazole group takes place of the phenyl ring of Phe, and thus it is convenient to analyze the spectrum of His by comparison with the spectrum of imidazole[65] which has two peaks in the energy range below 290 eV. By analogy, the intense peak at 286.7 eV in His (Figure 7d) has been assigned to the C 1s→1π^* state. However, its fwhm = 1.5 eV suggests that this peak arises from the overlap of a number of unresolved C 1s→1π^* transitions associated with different C 1s chemical shifts for the three chemically inequivalent carbon atoms. It is followed by a less pronounced peak at 288.5 eV, assigned to the C 1s→2π^* transition. His (Figure 7d) shows a peak at 287.0(1) eV, which apparently is not a single state. There is a pronounced low-energy shoulder at 286.5 eV and the high-energy side may contain a separate peak as well. Theory suggests the following interpretation: due to the larger electronegative shift, the contribution from the carbon

atom that is placed between the two nitrogen atoms shifts to higher energy by 0.35 eV, whereas the C 1s(C_3)→1π* and C 1s(C_5)→1π* transitions essentially coincide. An alternative assignment is the following. The relaxation shift, which we discussed in the case of Phe, could lead to the low-energy shoulder of the main C 1s→π* peak. This shift could reverse the sign, if carbons in the ring become more electronegative relative to the carbon connected to the chain. This would lead to the shoulder at the low-energy side of the main $\pi^*_{C=O}$ peak (∼287 eV), as found in the fit. This suggests that the 286.5 eV peak could be the C 1s(C_3)→1π* transition, while σ^*_{CH} states contribute to the high-energy side (287.6 eV) of the same peak. As we already reported in the case of acidic amino acids, more electronegative substitution in the chain leads to a suppression of oscillator strength associated with transitions to the σ^*_{CH} and σ^*_{CC}. The comparison of His and Phe spectra follow the same path.

4. Summary

The C K edge NEXAFS of all 20 amino acids commonly occurring in nature have been reported. From our spectral analysis we conclude that the common backbone structure of a carboxylic acid group and an amino group attached to a saturated carbon atom is reflected in the C edge NEXAFS spectra. In particular there is a sharp $\pi^*_{C=O}$ peak (always within 0.1 eV of 288.6 eV). In addition there is often a peak or prominent shoulder at higher energy around 290 eV, associated with σ^*_{CNH} states. These features may be used as fingerprints of amino acids. The first peak does not change its position for any of the amino acids and does not depend on the charge state, be it a zwitterion or a positively charged ion as in the case of acidic media. A similar feature has been observed in the C 1s spectrum of all other acids,[29,31] and can be firmly identified as a carboxyl/carbonyl C $1s^{-1}\rightarrow\pi^*$ transition. Although the FWHM (∼0.7 eV) does not change substantially among the small amino acid molecules, there are indications of peak structure, possibly due to vibrational excitations, or in some cases overlapping transitions as suggested in detailed discussion above. Asparagine (Asp), glutamine (Glu), and particular arginine (Arg) show a splitting or a shift of the 288.6 eV line. An additional structure, identified as being due to a C 1s($H_2N-C=O$)→π* transition in the case of Asp (Glu), is shifted by 0.1 eV to a lower energy. In addition, the $(H_xN)_2-C=NH_2$ functional group in Arg has a C 1s(C═N)→$\pi^*_{C=N}$ peak at 289.3 eV. The amino group signals are less pronounced, since the Cα contribution overlaps other transitions. Still we can identify the ∼289 eV peak as a C 1s-($-CNH_3^+$)→π* transition on the bases of its pH dependence and on the theoretical analysis.

Although the NEXAFS spectrum of each amino acid is unique, only five of them have sufficiently distinct features which can be used readily for chemical identification. Arg has the C 1s(C═N)→$\pi^*_{C=N}$ peak at 289.3 eV. Aromatic amino acids (Phe, Tyr, Trp, and His) have a characteristic NEXAFS π^*_{ring} signal around 285 eV, which uniquely identify them. Hydroxyl substitution in Tyr leads to a large shift of C 1s(C−OH)→1π* and a unique identifying signal at 287 eV. The pyrrole side group in Trp can be identified by a set of narrow peaks at 285.0, 285.5, 286.0 eV followed by broad band at 286.6 eV. The imidazole group in His is characterized by an asymmetric peak at 287.0 eV with a low-energy shoulder.

The detailed C K edge spectral analysis for each amino acid group helps to identify and assign spectroscopic features of other functional units. For aliphatic amino acids, the growing of the side chain can be seen as an addition of low-intensity peaks in the energy range 287−288 eV. The origin for this spectral weight involves CH C $1s^{-1}\rightarrow\pi^*$ transitions. It seems that the exact position may depend on the partial charge at a particular carbon in the chain rather than due to chain geometry/ branching. The intensity is controlled by the presence and position of hetero substituted atoms. For instance, in a sequence of similar molecules such as Ala, Cys, and Ser, the presence of a sulfur atom in Cys leads to a prominent peak at 287.4 eV. For Ala (no substitution), there is only a tail for 288 eV, and this signal is completely absent in Ser, where the side chain is terminated by an −OH group. As the side chain grows, a ∼287 band increases relative to the carboxyl peak, but the high-energy signal (290−292 eV) grows more rapidly.

A detailed spectrum above the IP (E > 285 eV) was measured only in the case of glycine (Gly), but for a different charging state of the molecule. We expect that this part of the spectrum has a character common to all amino acids, as the main peaks which contributes to a high-energy side originated from C−C and C−O C 1s→π* transitions. The experimental data clearly shows the difference between carboxyl (−COOH) and carbonyl ($-COO^-$) functional groups and may so be used to determine the charging state of the amino acid molecules. Two different bonding lengths of CO pairs in carboxyl lead to two separate broad peaks (at 297 and 302 eV), whereas carbonyl has only one peak at 300.4 eV.

The limitations of early STEX calculations[2] were overpassed by accounting for core hole polarization.[3] This newly developed calculational scheme showed good correspondence with experimental data for glycine and led to an empirical correction scheme that has allowed a more satisfactory correlation of theory and experiment than available previously.

In summary, based on comparison among related molecules and ab initio calculations, we have assigned all major peaks in the carbon K edge NEXAFS spectra of the commonly occurring 20 amino acids. The fingerprinting capability of NEXAFS has been demonstrated. The ability to make plausible, internally consistent assignments of all spectral features suggests that NEXAFS spectroscopy may be useful at differentiating peptide sequences, at least for species where averaging over a large number of residues does not blur the underlying differentiation based on the spectral features of the individual residues. A more challenging application would be to use differences in the NEXAFS spectra for mapping different proteins. This is probably only possible for smaller proteins that have substantial differences in composition (e.g., aliphatic-rich versus sulfide rich, etc). Still, the present results provide a firm basis for further experimental and theoretical work aimed at exploring the limits of applicability of NEXAS spectroscopy as an analytical tool for peptides and proteins. In our opinion such studies are highly warranted.

Acknowledgment. K.K. gratefully acknowledges guidance by O. Gerasimov (Brookhaven National Laboratory, Chemistry Department) in sample preparation. All spectra were recorded with the X-1A STXM, with support by the Office of Biological and Environmental Research, U. S. DoE under Contract DE-FG02-89ER60858, and the NSF under Grant DBI-9605045. We especially thank S. Wirick (Stony Brook) for her excellent assistance in microscope and beamline operation. The work of Hitchcock is supported by NSERC and the Canada Research Chair program.

References and Notes

(1) Boese, J.; et al. *J. Electron Spectrosc. Relat. Phenom.* **1997**, *85*, 9−15.

(2) Carravetta, V.; Plashkevych, O.; Ågren, H. *J. Chem. Phys.* **1998**, *109* (4), 1456−1464.

(3) Carravetta, V.; Plashkevych, O.; Ågren, H. *Chem. Phys.* **2001**, *263*, 231.
(4) Mochizuki, Y.; et al. *Chem. Phys. Lett.* **1999**, *309*, 241.
(5) Hitchcock, A. P.; Mancini, D. C. *J. Electron Spectrosc. Relat. Phenom.* **1994**, *67* (1), 0.
(6) Stöhr, J. *NEXAFS Spectroscopy*; Springer-Verlag: Berlin, 1992.
(7) Francis, J. T.; Hitchcock, A. P. *J. Phys. Chem.* **1992**, *96*, 6598.
(8) Urquhart, S. G.; et al. *J. Phys. Chem. B* **1997**, *101*, 2267−2276.
(9) Kaznacheyev, K.; Osanna, A.; Jacobsen, C. 2001. In preparation.
(10) Pawlak, J.; Cheng, P. C.; Shinozaki, D. M. A Simple Procedure for the Fabrication of Si_3N_4 Windows. In *X-ray Microscopy: Instrumentation and Biological Applications*; Cheng, P. C., Jan, G. J., Eds.; Springer-Verlag: Berlin, 1987; pp 336−345.
(11) Kang, J. F.; et al. *Langmuir* **2000**, *16*, 3791.
(12) Jacobsen, C.; et al. *Opt. Commun.* **1991**, *86*, 351−364.
(13) Zhang, X.; et al. *Nucl. Instrum. Methods Phys. Res. A* **1994**, *347*, 431−435.
(14) Winn, B.; et al. *J. Synchrotron Radiat.* **2000**, *7*, 395−404.
(15) Williams, S.; et al. *J. Microsc.* **1993**, *170*, 155−165.
(16) Ma, Y.; et al. *Phys. Rev. A* **1991**, *44* (3), 1848−1858.
(17) Smith, A. P.; Coffey, T.; Ade, H. Characterization of the XI-STXM Spectroscopy Acquisition Mode Utilizing Carbon Dioxide. In *X-ray Microscopy and Spectromicroscopy*; Thieme, J., Schmahl, G., Umbach, E., Rudolph, D., Eds.; Springer-Verlag: Berlin, 1998.
(18) Hunt, W. J.; Goddard, W. A., III *Chem. Phys. Lett.* **1969**, *3*, 414.
(19) Ågren, H.; et al. *Chem. Phys.* **1994**, *222*, 75.
(20) Kendall, R. A.; Dunning, J.; T. H., Harrison, R. J. *J. Chem. Phys.* **1992**, *96*, 6796.
(21) Momany, F. A.; et al. *J. Phys. Chem.* **1975**. *79*, 2361.
(22) Gross, D.; Grodsky, G. *Chem. Phys.* **1955**, *77*, 1678.
(23) Larsen, D. W.; Friberg, S. E.; Christenson, H. *J. Am. Chem.Soc.* **1980**, *102*, 6566.
(24) Slaughter, A. R.; Banana, M. S. *J. Phys. Chem.* **1988**, *92*, 2165.
(25) Sette, F.; Stöhr, J.; Hitchcock, A. P. *Chem. Phys.* **1984**, *81*, 4906.
(26) Hasselstrom, J.; et al. *Surf. Sci.* **1998**, *407*, 221.
(27) Nyberg, M.; et al. *J. Chem. Phys.* **2000**, *112*, 5420.
(28) Robin, M. B.; et al. *J. Electron Spectrosc. Relat. Phenom.* **1988**, *47*, 53.
(29) Ishii, I.; Hitchcock, A. P. *J. Chem. Phys.* **1987**, *87*, 830.
(30) Henke, B. L.; Gullikson, E. M.; Davis, J. C. *At. Data Nucl. Data Tables* **1993**, *54*, 181−342.
(31) Ishii, I.; Hitchcock, A. P. *J. Electron Spectrosc. Relat. Phenom.* **1987**, *46*, 55.
(32) Hitchcock, A. P.; Ishii, I. *J. Electron Spectrosc. Relat. Phenom.* **1987**, *42*, 11.
(33) Stöhr, J.; et al. *Phys. Rev. B* **1987**, *36*, 2976.
(34) Ohta, T.; et al. *Phys. Scr.* **1990**, *41*, 150.
(35) Weiss, K.; Bagus, P. S.; Woll, C. *J. Chem. Phys*, **1999**, *111*, 6834.
(36) Outka, D. A.; et al. *J. Chem. Phys.* **1988**, *88*, 4076.
(37) Robin, M. B., Ed. *Higher Excited States of Polyatomic Molecules*; Academic Press: New York, 1978; Vol. 3.
(38) Robin, M. B., Ed. *Higher Excited States of Polyatomic Molecules*; Academic Press: New York, 1974; Vol. 2.
(39) Thomas, W. *Naturwissenschaft* **1925**, *13*, 627.
(40) Reiche, F.; Thomas, W. *Z. Phys.* **1925**. *34*, 510.
(41) Kühn, W. *Z. Phys.* **1925**, *33*, 408.
(42) Hitchcock, A.P.; Horsley, J. A.; Stöhr, J. *J. Chem. Phys.* **1986**, *85*, 4835.
(43) Hitchcock, A. P.; Tronc, M.; Modelli, A. *J. Phys. Chem.* **1989**, *93*, 3068.
(44) Hitchcock, A. P. *Phys. Scr.* **1990**, *T 31*, 159.
(45) Hitchcock, A. P.; Ishii, I. *J. Electron Spectrosc. Relat. Phenom.* **1987**, *42*, 1126.
(46) Urquhart, S. G.; et al. 2001. In preparation.
(47) Lessard, R. J.; Gordon, M.; Hitchcock, A. P. 2001. In preparation.
(48) Newbury, D. C.; Ishii, I.; Hitchcock, A. P. *Can. J. Chem.* **1986**, *64*, 1145.
(49) Horsley, J. A.; et al. *J. Chem. Phys.* **1985**, *83*, 6099.
(50) Solomon, J. L.; Madix, R. J.; Stöhr, J. *Surf. Sci.* **1991**, *255*, 12.
(51) Ma, Y.; et al. *Phys. Scr.* **1990**, *41*, 833.
(52) Schwarcz, W. H. E.; et al. *Chem. Phys.* **1987**, *117*, 73.
(53) Bakke, A. A.; Chen, H. W.; Jolly, W. L. *J. Electron Spectrosc. Relat. Phenom.* **1980**, *20*, 333.
(54) Hitchcock, A. P.; Stöhr, J. *J. Chem. Phys.* **1987**, *87*, 3523.
(55) Rennie, E. E.; et al. *Phys. Rev. B* **2000**, *133*, 7362.
(56) Carravetta, V.; et al. *Chem. Phys. Lett.* **1998**, *288*, 37.
(57) Yang, L.; et al. *J. Phys. IV* **1997**, *C2*, 227.
(58) Petterson, L. G. M.; et al. *Int. J. Quantum Chem.* **1997**, *63*, 749.
(59) Ma, Y.; et al. *Phys. Rev. Lett.* **1989**, *63*, 2044.
(60) Urquhart, S. G.; et al. *Chem. Phys. Lett.* **2000**, *322*, 412.
(61) Francis, J. T.; Hitchcock, A. P. *J. Phys. Chem.* **1994**, *98*, 3650.
(62) Plashkevych, O.; et al. *Chem. Phy.* **1997**, *222*, 125.
(63) Turci, C. C.; Urquhart, S. G.; Hitchcock, A. P. *Can. J. Chem.* **1996**, *74*, 851.
(64) Luo, Y.; et al. *Phys. Rev. A* **1995**, *52*, 3730.
(65) Apen, E.; Hitchcock, A. P.; Gland, J. L. *J. Phys. Chem.* **1993**, *97*, p 6859.
(66) Perry, W. B.; Jolly, W. L. *Inorg. Chem.* **1974**, *13*, 1211.
(67) Ohta, T.; Fujikawa, T.; Kuroda, K. *Bull. Chem. Soc. Jpn*. **1975**, *48*, 2017.