The x-ray absorption spectroscopy model of solvation about sulfur in aqueous L-cysteine

Ritimukta Sarangi,^{1,a)} Patrick Frank,^{1,2,a)} Maurizio Benfatto,^{3,a)} Silvia Morante,⁴ Velia Minicozzi,⁴ Britt Hedman,¹ and Keith O. Hodgson^{1,2}

¹Stanford Synchrotron Radiation Lightsource, SLAC National Accelerator Laboratory, Menlo Park, California 94025, USA

²Department of Chemistry, Stanford University, Stanford, California 94305, USA

³Laboratori Nazionali di Frascati dell' INFN, C.P. 13, 00044 Frascati, Italy

⁴Department of Physics, University of Roma "Tor Vergata" and INFN Section of Roma "Tor Vergata," 00133 Roma, Italy

(Received 5 September 2012; accepted 30 October 2012; published online 28 November 2012)

The environment of sulfur in dissolved aqueous L-cysteine has been examined using K-edge x-ray absorption spectroscopy (XAS), extended continuum multiple scattering (ECMS) theory, and density functional theory (DFT). For the first time, bound-state and continuum transitions representing the entire XAS spectrum of L-cysteine sulfur are accurately reproduced by theory. Sulfur K-edge absorption features at 2473.3 eV and 2474.2 eV represent transitions to LUMOs that are mixtures of S–C and S–H σ^* orbitals significantly delocalized over the entire L-cysteine molecule. Continuum features at 2479, 2489, and 2530 eV were successfully reproduced using extended continuum theory. The full L-cysteine sulfur K-edge XAS spectrum could not be reproduced without addition of a water-sulfur hydrogen bond. Density functional theory analysis shows that although the Cys(H)S $\cdots$ H–OH hydrogen bond is weak (~ 2 kcal) the atomic charge on sulfur is significantly affected by this water. MXAN analysis of hydrogen-bonding structures for L-cysteine and water yielded a best fit model featuring a tandem of two water molecules, 2.9 Å and 5.8 Å from sulfur. The model included a S_{cys} $\cdots$ H–O_{w1}H hydrogen-bond of 2.19 Å and of 2.16 Å for H₂O_{w1} $\cdots$ H–O_{w2}H. One hydrogen-bonding water-sulfur interaction alone was insufficient to fully describe the continuum XAS spectrum. However, density functional theoretical results are convincing that the water-sulfur interaction is weak and should be only transient in water solution. The durable water-sulfur hydrogen bond in aqueous L-cysteine reported here therefore represents a break with theoretical studies indicating its absence. Reconciling the apparent disparity between theory and result remains the continuing challenge. © 2012 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4767350>]

I. INTRODUCTION

The organization of water around solutes is central to both the thermodynamics and the chemistry of molecular dissolution and solute-solute interactions.^{1–4} In biochemistry the solvation of amino acids is important to all aspects of protein-substrate and protein-protein binding,^{5–7} and has been studied from various experimental vantages.^{1,8–10} Electrospray mass spectrometry has been used to examine stepwise solvation of gas-phase amino acids and small peptides.^{5,11,12}

Inclusion of a third row heteroatom uniquely distinguishes L-cysteine and L-methionine among the amino acids. L-cysteine (hereinafter cysteine) is of critical importance among the cysteine metalloproteases,^{13–15} in metallothionein biochemistry,^{16,17} in regulating the *in vivo* plasma redox state,^{18,19} and in metal binding, and electron transfer.^{20–22} Nevertheless, attention has focused chiefly on solvation of polar amino acids such as arginine.

X-ray absorption spectroscopy (XAS) is used to query the electronic state of virtually the element of choice,^{23–25}

and has already been used to investigate the solvation of cysteine.^{26,27} The thiol group in cysteine can act as either a hydrogen bond acceptor or donor, e.g., HO–H $\rightarrow$ S(H)–CH₂– (mode I), or H₂O $\leftarrow$ H–S–CH₂– (mode II). Structural reviews and density functional theory (DFT) calculations have predicted that mode I sulfur hydrogen bonding in water is weak and transient.^{28,29}

The improved application of extended continuum theory to x-ray absorption spectroscopy through Minuit XANES (MXAN) analysis,^{30–32} now permits exploration of the three-dimensional structures of dissolved small molecules.^{33–36} The solution structural study reported here began when it was found that MXAN calculation of the sulfur K-edge XAS spectrum of aqueous cysteine required inclusion of at least one water molecule proximate to sulfur.

Here sulfur K-edge XAS and DFT calculations are combined to illuminate solvation at cysteine sulfur. Density functional theory was used to evaluate first-guess hydrogen-bonding models of dissolved cysteine. Using the sulfur K-edge XAS spectrum as the target,^{32,37} MXAN analysis was then used to evaluate structural models of dissolved cysteine. The structural model of cysteine with one sulfur-water hydrogen bond was energy-minimized using DFT. The orbital

^{a)}Authors to whom correspondence should be addressed. Electronic addresses: ritis@slac.stanford.edu, pfrank@ssrl.slac.stanford.edu, and Maurizio.Benfatto@lnf.infn.it.

electronic state of this sulfur was then used to explain the bound-state transitions within the sulfur K-edge XAS spectrum of dissolved cysteine, as with other sulfur-containing molecules.³⁸ Finally, the extended continuum multiple scattering (ECMS) approach of MXAN analysis is shown to reproduce accurately the continuum features of the XAS spectrum and thus complete a one-electron description of the electronic state of sulfur in aqueous cysteine. A coherent picture is then presented of the immediate solvation environment of sulfur in aqueous zwitterionic cysteine. This MXAN-DFT study focuses only on the immediate sulfur environment, and represents the first of a two-part investigation. The second part, presently in advanced preparation, will expand the empirical model using *ab initio* molecular dynamics to provide a global view of the aqueous solvation of cysteine.

II. MATERIALS AND METHODS

A. Sulfur K-edge x-ray absorption spectroscopy

Sulfur K-edge XAS spectra were measured using the SSRL 20-pole 2 T wiggler beam line 4-3, with resolution $\Delta E/E = 10^{-4} = 0.25$ eV at sulfur E_0 . The beam line was fully tuned at 2740 eV, and equipped with a Ni-coated harmonic rejection mirror and a Si(111) double-crystal monochromator. A beam size of 1 mm by 10 mm was used at the sample, which was positioned 45° with respect to the x-ray beam. Sulfur K-edge XAS were measured at room temperature as the $K\alpha$ fluorescence excitation spectrum. The fluorescence x-ray intensity was measured using a passively implanted planar silicon detector, set 90° from the x-ray beam.

The anaerobic aqueous 100 mM solution of cysteine (Aldrich chemicals, zwitterionic form) in 250 mM sodium citrate buffer, pH 5.1, was prepared in a vacuum atmospheres (VAC) nitrogen-filled glove box ($O_2 \leq 1$ ppm). The choice of pH 5.1 was guided by the pH profile of dissolved cysteine, where $\sim 100\%$ of cysteine is in the zwitterionic form (see SI, Scheme S1).^{39–42} The solution, in a teflon-spacer sample cell equipped with two polypropylene windows of $6.35\ \mu\text{m}$ thickness, was transported from the VAC glove box to the beam line in a nitrogen-filled jar and was transferred directly into a sample chamber filled with a dynamic helium flow atmosphere. The XAS spectrum was energy calibrated to the intense pre-edge transition of the sulfur K-edge spectrum of $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$, assigned to 2472.02 eV. Calibration scans were obtained before and after sample scans.

Data normalization and background subtraction were performed using the program PySpline.⁴³ Background subtraction included fitting a second order polynomial to the raw XAS over the range 2420.3–2456.3 eV. This polynomial was extrapolated to 3134.5 eV and subtracted from the raw spectrum. For normalization, the background-subtracted XAS spectrum was fit with a piecewise polynomial spline, using the following knot-points, [eV (polynomial order)]: 2490.067 (2); 2577.520 (3); 2825.075 (3); 3134.527. The background subtracted XAS spectrum was finally divided by the polynomial spline to produce the normalized XAS spectrum.

Pseudo-Voigt fits were carried out over the energy range 2470–2485 eV using the program EDG-FIT, which is part

of the EXAFSPAK suite of programs written by Professor Graham George, University of Saskatchewan. For these fits, Lorentzian-Gaussian mixing was fixed at 0.5. Twelve pseudo-Voigt linewidths were linked and refined together, and the full widths at half-height were required to be within 10% of the experimental beam line resolution of 0.64 eV at 2490 eV. Resolution was calculated as $\sigma_r = \pm\sqrt{(\Delta E/E)^2 + \gamma^2}$, where “ γ ” is the core hole lifetime width of sulfur (0.59 eV).⁴⁴ The best-fit arctangent function was found at 2476.28 eV, which approximates the ionization energy profile of cysteine sulfur.

B. Model-building

The structural models for the MXAN and DFT computational parts of the study began with the known three-dimensional coordinates of zwitterionic cysteine.⁴⁵ This structure was imported into Chem3D Pro (Cambridge-Soft, Cambridge, MA) for emplacement of solvating water molecules, but was itself used unmodified. The internal coordinates of water molecules were not changed from the standard structure supplied by Chem3D ($\text{HO}-\text{H} = 0.942\ \text{\AA}$, $\angle \text{H}-\text{O}-\text{H} = 103.7^\circ$). First-guess $\text{HO}-\text{H} \cdots \text{S}_{\text{Cys}}$ or $\text{H}_2\text{O} \cdots \text{H}-\text{S}_{\text{Cys}}$ hydrogen bond lengths were estimated from the literature.²⁸ Initial-guess structural models were derived to test potential hydrogen-bonding interactions of water molecules specifically with the sulfur atom of cysteine. These are shown as Figures S1 and S5, respectively, in the supplementary material.¹¹⁴

C. MXAN calculations

XAS data from the rising K-edge energy region to about 200 eV were analyzed, with the aim of deriving both angular and radial information around the absorbing atom.^{30,31} In general, the inverse of the scattering path operator was computed exactly, avoiding any *a priori* selection of the relevant multiple-scattering paths. The muffin-tin (MT) approximation was used for the shape of the potential,⁴⁶ and hydrogen atoms were included in the calculation.

The Coulomb and exchange parts of the total cluster potential were calculated using the total charge density, which was approximated using a superimposition of spherically symmetric self-consistent atomic charge densities. The charge densities were generated by a relativistic Dirac-Fock atomic code that automatically runs in MXAN for all atoms in the Periodic Table.^{47,48} Unless otherwise indicated, neutral electronic configurations were imposed for all atoms in the cluster surrounding the absorber while the absorber itself was treated within the so-called $Z+1$ approximation that corresponds to the fully relaxed electronic configuration. When using non-neutral atoms, atomic charges were obtained from self-consistent field (SCF) calculations wherein each atom in the cluster was assigned the Mulliken charge derived using DFT, as noted below. In some fits, the MT radii and the interstitial potentials were optimized during the fit procedure while the molecular geometry was kept fixed at the DFT energy-optimized structure, or at a previously determined MXAN structural best fit. In these fits, only the non-structural part

of the fit was allowed to change to obtain the best agreement between theory and experimental data.

A refinement of the MT radii and the interstitial potential values was used to mimic both the SCF potential and compensate for the MT approximation.³² The atomic T-matrixes depend on these two parameters via the Wronskian determinant, which was calculated to the boundary of the MT spheres for a given interstitial potential.^{49,50} The T-matrixes depend on both the charge density and the non-MT corrections. The MT radii and the interstitial potential were chosen to minimize the errors in the potential determination,⁵¹ giving better agreement between theory and experimental data.

The function minimized during the MXAN fit was R_{sq} , defined as

$$R_{sq} = n \frac{\sum_{i=1}^m w_i [(y_i^{th} - y_i^{exp}) \varepsilon_i^{-1}]^2}{\sum_{i=1}^m w_i}, \quad (1)$$

where “ n ” is the number of independent parameters, “ m ” is the number of data points, “ y_i^{th} ” and “ y_i^{exp} ” are the theoretical and experimental values of the absorption, respectively, “ ε_i ” is the error in each point of the experimental data set, and “ w_i ” is a statistical weight. When $w_i = 1$, the square residual function R_{sq} becomes the statistical χ^2 function. In this work, $w_i = 1$ was assumed and the experimental error $\varepsilon = \text{constant} = 1.0\%$ of the main experimental edge jump over the whole data set. Statistical errors were calculated using the MIGRAD routine. MXAN also introduces a systematic error of 1%–2% into the bond lengths that must be added to the MIGRAD statistical error.

For MXAN fits that excluded the K-edge bound state energy region, the weight “ w_i ” was expressed as a theta function,

$$w_i = \theta(E) = \begin{cases} 1 & \text{for } E_i - E_0 \geq 2.5 \text{ eV} \\ 0 & \text{elsewhere} \end{cases} \quad (2)$$

where E_i = energy in eV. This function introduced a vertical step at the onset of the continuum energy region of the fit to account for the rise in intensity due to ionization of the photoelectron, and constrains the assumed $w_i = 1$ to the fitted region of the XAS spectrum. The vertical step was then convolved with the Lorentzian describing the transition intensities, to produce a smooth rising intensity of unit magnitude. In general, E_0 is chosen at the onset of intensity from transitions to virtual bound states. For all the fits reported here, $E_0 = 2472.0$ eV.

Where used, the MXAN fit error $fe(E)$ is defined as

$$fe(E) = \sqrt{[y_i^{th}(E) - y_i^{exp}(E)]^2}. \quad (3)$$

where $y_i^{th}(E)$, $y_i^{exp}(E)$ is the theoretical or experimental intensity at the i th point. A greater $fe(E)$ intensity represents a poorer fit.

D. DFT computations

Gradient-corrected, (GGA) spin-unrestricted DFT calculations were carried out using the GAUSSIAN 09 (G09)⁵²

package on a 32-CPU Linux-based computer. Geometry optimizations were performed for each system. The Becke 88^{53,54} exchange functional and the correlation functional of Lee, Yang, and Parr⁵⁵ were used in combination with Vosko–Wilk–Nusair⁵⁶ local functionals as approximations to the correlation functional as implemented in the software package (B3LYP). A polarizable continuum model was used to simulate the water solvation sphere (geometry optimizations were also performed in the absence of solvation sphere for comparison, see the supplementary material¹¹⁴). In addition, calculations were performed with one, two, or three water molecules placed at H-bonding distances to the sulfur atom of the cysteine molecule. The Ahlrichs’ triple- ζ valence basis set with polarization functions (TZVP)^{57,58} was used on all atoms. Population analyses were performed by means of Weinhold’s Natural Population Analysis.^{59–61} Wave functions were visualized and orbital contour plots were generated in MOLDEN.⁶² Compositions of molecular orbitals and overlap populations between molecular fragments were calculated using the QMForge program.⁴³ Mulliken charges were obtained using the MPA keyword and were those subsequently used in MXAN fits. The zwitterionic starting structural models for geometry optimization included cysteine alone and cysteine with one water at H-bonding distance to the sulfur, cf. Figure 3. Optimization included the full cysteine-water molecular pair.

One-dimensional potential energy surfaces (PES) were calculated with the electronic structure program ORCA⁶³ on a 12-CPU Linux cluster. The B3LYP functional was chosen and the def2-TZVPP^{58,64} basis set as implemented in the ORCA package was used on all atoms. The larger basis sets were chosen for consistency with the subsequent CCSD(T) calculations. The RIJCOSX⁶⁵ approximation was used. A conductor like screening model (COSMO)⁶⁶ was chosen to mimic the water solvation. Tight convergence criterion was selected. For the Cys(H)S $\cdots$ H–OH PES the S $\cdots$ H distance was varied between 2.15 and 4.1 Å. For accurate estimation of this H-bond, coupled cluster calculations with the triples correction (CCSD(T))⁶⁷ were performed on two geometries obtained from the DFT calculations (see Figure 8). For the S_{cys}–C PES, the S–C distance was varied between 1.65 and 2.85 Å. Time dependent density functional theory (TD-DFT)⁶⁸ calculations were performed in ORCA using the protocol developed previously. Transition energy and intensity from the S 1s orbitals to the lowest unoccupied molecular orbitals were calculated.

III. RESULTS

A. Sulfur K-edge XAS

SSRL beam line 4–3 is a high-brightness 20-pole wiggler x-ray source, covering the 2.4–14 keV energy range, and dedicated to lower-Z elements. Previously, measurements of sulfur K-edge XAS at SSRL were carried out on 56-pole wiggler beam line 6–2, operating in undulator mode. Figures S2 and S3 in the supplementary material demonstrate the identity of the sulfur K-edge XAS spectrum of cysteine measured independently on each beam line.

Figure 1 shows the sulfur K-edge XAS of the experimental pH 5.1 solution of cysteine, measured on SSRL beam line

FIG. 1. (—, Violet) Sulfur K-edge XAS spectrum of cysteine in pH 5.1 citrate solution. Inset (a) (o), the rising K-edge energy region, and (—), the pseudo-Voigt fit. The dotted-dashed lines indicate the fit components including the arctangent representing the normalized ionization intensity. Inset (b) Second derivative of (o) the cysteine sulfur K-edge XAS spectrum, and (—, blue) of the pseudo-Voigt fit.

4-3. An absorption feature at 2800 eV, arising from a trace chloride impurity, limited analysis of the XAS spectrum to ≤ 2700 eV, equivalent to $k = 7 \text{ \AA}^{-1}$, and obviated EXAFS analysis.

The second derivative XAS (Figure 1, inset (b)) shows at least two distinct features. However, a pseudo-Voigt fit over the first 20 eV of the K-edge XAS of cysteine required three major lines to reproduce the rising K-edge main feature. This empirical result is consistent with a previous DFT study,²⁷ which found three major bound state transitions in the XAS of dissolved zwitterionic cysteine.

B. MXAN fitting experiments

The initial MXAN fits to the sulfur K-edge XAS spectrum of aqueous cysteine utilized coordinates obtained from the x-ray structure of the crystalline zwitterionic solid.⁴⁵ However, obtaining a good fit to the XAS spectrum proved problematic. Spectrum (a) in Figure 2 shows the best fit obtained using the x-ray structural coordinates alone ($R_{sq} = 21.8$). This fit is poor compared to the quality typically obtained using MXAN.^{32,36,37}

The poor quality of the fit is restricted to the rising K-edge energy region, and is primarily due to the constant normalization approximation used within extended continuum multiple scattering theory. This method treats XAS bound state transitions as very sharp scattering features,⁶⁹ in which the wave functions of bound and continuum states are normalized identically. The advantage is that full calculation of bound and virtual states becomes possible without needing to include an ionization energy. However, in full multiple scattering theory, bound states and continuum states are normalized differently.³⁴ Thus the extended continuum theory may not correctly reproduce bound states deep within the potential well of the scatterer. The first feature in the sulfur K-edge XAS of cysteine, at 2473.5 eV ($E-E_0 = 1.5$ eV), is well separated from the continuum part of the spectrum and derives

FIG. 2. (a) (o) Sulfur K-edge XAS spectrum of cysteine dissolved in pH 5.1 aqueous citrate, and (—, blue) the MXAN fit to the spectrum ($R_{sq} = 21.8$). (b) The sulfur K-edge XAS as in (a) offset by unity, and (—, green) the MXAN fit using the theta function to exclude the rising edge energy region from the fit ($R_{sq} = 1.46$). (Inset) Expansion of the rising edge energy region of the XAS spectrum and the two MXAN fits to the spectrum. The R_{sq} values from (a) and (b) are not comparable because the fitted energy ranges are not identical.

from bound states that are quite deep in energy. In this case, because of the normalization approximation in MXAN, the calculated intensity and linewidth of the rising K-edge bound state feature are not entirely correct.

In an attempt to improve the quality of the fit to the continuum energy region, and increase the sensitivity to geometrical structure, the bound state rising K-edge XANES feature was eliminated from the MXAN fit. To compensate for the missing XANES intensity, a theta step function was introduced at $E-E_0 = 2.5$ eV by redefining the weight in the R_{sq} error function (see Sec. II). The included theta function has the effect of re-normalizing the fitted absorption coefficient at the assigned energy. In this modification, the bound state transitions in the XANES energy region are completely neglected by the fit. The contributions from the theoretically well-defined higher energy region of the spectrum are thus emphasized.

Spectrum (b) in Figure 2 shows that removing the rising K-edge bound state feature greatly improved the fit to the continuum energy region. The improvement is especially significant near $E-E_0 = 2.5$ to 20 eV. Following this result, all the remaining MXAN fits excluded the bound state feature, and utilized the theta step function.

Beyond the rising edge energy region ($E-E_0 > 30$ eV), fits (a) and (b) in Figure 2 are approximately equivalent, as expected. Despite the improvement in the lower energy XAS region, Figure 2(b) shows that noticeable disparities between the XAS spectrum and the fit remained in the $E-E_0$ range 0–15 eV. The significant lapses were suggestive that some structural element was lacking in the model. As the fit was restricted to the structure of cysteine alone, and as the cysteine molecule itself exhausts all of the strong XAS back-scattering interactions with sulfur, the existence of an external scattering contribution was inferred.

In dissolved cysteine, proximate and localized water molecules are the only available candidates for an external

FIG. 3. The DFT geometry optimized structure for a single donor water-sulfur hydrogen bond associated with dissolved cysteine. The bulk water environment was approximated using the polarized continuum model.

photoelectron scatterer. That is, any durably localized water molecule in correlated motion with sulfur could contribute an XAS scattering resonance. The possibility of a water hydrogen-bonded to sulfur was explored using DFT. Figure 3 shows the geometry-optimized structure for a donor water-sulfur interaction, which predicted a weak Cys-(H)S $\cdots$ H-OH hydrogen bond worth ~ 2 kcal mol $^{-1}$.

It was therefore decided to include hydrogen-bonding models in the MXAN structural fits. The first test of this idea included a water molecule hydrogen bonded to sulfur (Figure 3), and second model involving a water hydrogen bond to the sulfur thiol (see Figure S1 in the supplementary material). The structural cartoon in Scheme 1 illustrates the donor-acceptor water-thiol hydrogen bonding interactions that were considered.

Single point MXAN calculations were then used to test the mode I and mode II hydrogen bonding models shown in Scheme 1. The results of these tests are shown in Figure 4 and Table I. Comparison of Figure 4 with Figure 2, spectrum (b), shows that adding a donor hydrogen bonding water molecule (mode I) improved the fit. However, adding an acceptor hydrogen bonding water molecule near the thiol hydrogen (mode II) made the fit slightly worse.

To clarify this point further, adding molecules or atoms to a MXAN fit is not equivalent to adding adjustable degrees of freedom to a numerical fit. Nearby water molecules change the local potential, and produce unique scattering resonances. If the calculated XAS resonances produced by newly added water molecules do not match the data, the MXAN fit is made worse. Therefore, merely adding a water molecule to a fit does not automatically improve R_{sq} . Rather, improvement of the fit is evidence that an added molecule has introduced photoelec-

SCHEME 1. Structural cartoons of two modes of hydrogen bonding between the thiol moiety of dissolved cysteine and solvent water molecules.

TABLE I. MXAN H-bonding metrics.

H ₂ O H-bond	S $\cdots$ O (Å) ^a	X $\cdots$ H (Å) ^a	Crystal (Å) ^b	R _{sq}
None	...	...	...	1.46
Mode I ^c	3.49	2.7 (X=S)	2.67	1.21
Mode II ^d	3.61	2.3 (X=O)	2.34	1.60

^aFixed at crystal structure distances.

^bSee Allen *et al.* (Ref. 14).

^cFigures 3 and S1.

tron scattering resonances that correspond to previously unfit intensity in the XAS spectrum.

These first experiments showed that MXAN fits were improved by addition of a single water at hydrogen bonding distance from the sulfur in aqueous cysteine. Short-lived dynamical contacts do not produce significant scattering intensity in solution XAS spectra. Any nearby water that makes a consistent contribution to the sulfur photoelectron back-scattering intensity must have a relatively long-lived and durable sulfur-water interaction.

The sensitivity of the goodness-of-fit R_{sq} value to $S_{cys}\cdots O_w$ distance was then tested by increasing this distance stepwise by 0.1 Å, starting from 2.49 Å. At each step a single-point MXAN simulation of the cysteine XAS spectrum was calculated, yielding the corresponding R_{sq} . The results of this test are shown in Figure 5, indicating a sharp minimum in R_{sq} at $O_w\cdots S_{cys} = 2.9$ Å. A second less intense minimum in R_{sq} was found at 3.4 Å, near the DFT energy-minimized structural distance ($S_{cys}\cdots O_w = 3.48$ Å, Table I). Finally, a very broad minimum in R_{sq} emerged centered around $O_w\cdots S_{cys} = 4.5$ Å.

The comparatively weak water-sulfur H-bond,^{29,70–72} implies that any water molecule durably localized near sulfur must be externally stabilized to remain in that position. Such stabilization can be envisioned as a hydrogen-bonded network of water molecules anchored to one of the charged head-groups of cysteine. Similar water molecule networks have

FIG. 4. (○), Sulfur K-edge XAS of cysteine in pH 5.1 citrate buffer. (a) (—, Red) The MXAN fit modeling one donor H-bonding water (mode I); and (b) (—, blue) the MXAN fit modeling one H-bond accepting water (mode II); see text. (Inset) Expanded display of the early continuum energy XAS region and the fits.

FIG. 5. (—○—), The change in the MXAN goodness-of-fit R_{sq} value with sulfur-oxygen distance. The test for the dependence of R_{sq} with S–O distance began with one hydrogen-bonded water molecule at 2.49 Å ($\text{HO}-\text{H} \cdots \text{S}_{\text{Cys}} = 1.7$ Å) and advanced stepwise through and beyond the DFT energy-optimized S–O distance of 3.49 Å.

been reported for other dissolved aqueous amino acids.^{73–77} Such a network would constitute an extended stable shell of water molecules about cysteine. Any durably located bridging network of hydrogen-bonded water molecules with a $\text{S}_{\text{Cys}}-\text{O}_{\text{w1}}-\text{O}_{\text{w2}}$ angle of 150° – 180° in the sulfur locale could contribute detectable back-scattering intensity.^{78–80} Therefore, one or more proximate water molecules of any such network may be sufficiently localized near sulfur to also contribute resonant intensity to the photoelectron back-scattering. The multiple R_{sq} minima in the stepwise test shown in Figure 5 are consistent with possible back-scattering intensity from additional water molecules at longer $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w}}$ distances.

The possibility of a local H-bonded network of water molecules was tested by positioning a single water molecule situated at the DFT geometry-optimized $\text{S}_{\text{Cys}}-\text{O}_{\text{w}}$ distance of 3.48 Å from sulfur. A second more distant water molecule was emplaced at 1.87 Å H-bonding distance from O_{w1} and with $\angle \text{S}_{\text{Cys}}-\text{O}_{\text{w1}}-\text{O}_{\text{w2}} = 143^\circ$. This positioning located the second water molecule at $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w2}} = 5.9$ Å. The more distant water was meant to test the presence of a localized neighboring water molecule, representing the near-sulfur terminus of a position-stabilizing bridge of water molecules. An MXAN fit was then carried out testing this two-water structural model. Both the $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w1}}$ and $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w2}}$ distances were allowed to float independently, along with the $\text{S}_{\text{Cys}}-\text{O}_{\text{w1}}-\text{O}_{\text{w2}}$ angle.

During this fit, the 3.48 Å water advanced to $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w1}} = 2.93$ Å, reproducing the R_{sq} minimum distance seen in Figure 5. This shorter $\text{S}_{\text{Cys}} \cdots \text{O}_{\text{w1}}$ distance produced an $\text{S}_{\text{Cys}} \cdots \text{H}-\text{O}_{\text{w1}}\text{H}$ hydrogen-bond of 2.19 Å. The second water molecule at 5.9 Å water concomitantly moved to 5.8 Å, while the $\text{H}_2\text{O}_{\text{w1}} \cdots \text{H}-\text{O}_{\text{w2}}\text{H}$ hydrogen-bonding distance increased to 2.16 Å. The final $\text{O}_{\text{w1}}-\text{O}_{\text{w2}}$ distance was 3.0 Å, and the final $\text{S}_{\text{Cys}}-\text{O}_{\text{w1}}-\text{O}_{\text{w2}}$ angle increased slightly to 151° .

The final MXAN fit optimized the atomic potentials of the above-described structurally optimized two-water model.

FIG. 6. (○) Sulfur K-edge XAS of cysteine in pH 5.1 citrate buffer and (—, red) the MXAN fit using the two-water model. Inset (a) (○) the early continuum energy region of the XAS spectrum and (—, red) the MXAN fit. Inset (b) Comparison of the $fe(E)$ unfit residuals of (—, green) the one water model (see Figures 3 and 4) and (—, violet) the two water model.

This yielded the final $R_{sq} = 0.48$, which is $2.5 \times$ lower than the DFT geometry-optimized R_{sq} value for the one-water model (Figure 4). The XAS fit and error metrics resulting from the two-water model are shown in Figure 6 and Table II.

Figure 6, inset (b) compares the $fe(E)$ errors of the best one water (Figure 4) and two water (Figure 6) fits. The improvement with two waters is concentrated in the continuum energy region nearest the bound state transitions, $E-E_0 = 2.5$ – 20 eV, where multiple scattering features can make a large contribution to intensity. The final $\text{S}_{\text{Cys}}-\text{O}_{\text{w1}}-\text{O}_{\text{w2}}$ angle of 151° is within the angular range that can enhance second shell scattering,^{51,81–84} which may be responsible for the improvement of the fit.

We note that if the expanded model with two water molecules was unable to reproduce the experimental back-scattering intensity, the fit would have become poorer and R_{sq} would have increased. Expanding the structural model further with additional water molecules, if correctly placed, could further reduce the positional uncertainty of each water molecule. However, even in this event the fit R_{sq} would not improve much because the correspondence between the fit and the data is already excellent.

C. The effective solvation model for cysteine sulfur

Figure 7 shows the final MXAN two-water hydrogen bonding structural model for dissolved cysteine. In this model, only one of the water molecules is within

TABLE II. Two-water MXAN fit metrics.

Water	S $\cdots$ O (Å)	Error (Å) ^a	X $\cdots$ H distance (Å)	R_{sq}
I	2.9	± 0.37	2.19 ($\text{S}_{\text{Cys}} \cdots \text{H}-\text{O}_{\text{w1}}\text{H}$)	0.48
II	5.8	± 1.6	2.16 ($\text{H}_2\text{O}_{\text{w1}} \cdots \text{H}-\text{O}_{\text{w2}}\text{H}$)	

^aStatistical uncertainty of the fitted distance.

FIG. 7. The final MXAN structural model of the cysteine zwitterion with two waters. Water “1” represents the solvent molecule durably hydrogen-bonded to sulfur.

hydrogen-bonding distance to sulfur. The MXAN fitted $S_{\text{cys}} \cdots H-O_{\text{w1}}H$ distance of 2.19 Å is significantly shorter than the consensus crystallographic hydrogen bonding distance of 2.67 Å,²⁸ or the DFT energy-minimized distance of 2.70 Å, as calculated here and elsewhere.²⁹ The second water molecule contributes detectable scattering intensity to the sulfur K-edge XAS spectrum and is consistent with the existence of an extended structure that may positionally stabilize the water molecule contributing the primary $S_{\text{cys}} \cdots H-O_{\text{w1}}H$ interaction.

This model is not offered as the true local structure, but represents an empirically effective structure that satisfies the constraints imposed by the XAS spectrum. It is the structure that, among the models tested, most closely approximates the back-scattering profile required by ECMS physical theory, as expressed in MXAN, to accurately reproduce the experimental XAS spectrum.

D. DFT calculations

The DFT geometry optimized structural parameters of cysteine itself are in reasonable agreement with the crystallographic coordinates, and the relevant bond distances and Mulliken charges are shown in Table III. The geometry-optimized DFT structural model is shown in Figure S5 in the supplementary material. Without a polarizable continuum, the zwitterionic form of cysteine was not stable in the calculated model, in that one hydrogen invariably migrated from the ammonium moiety onto one of the carboxylate oxygens. Therefore, the

TABLE III. DFT Mulliken charges for cysteine with one $S \cdots H_w$ hydrogen bond modeled without and with a polarizable continuum.

Atom	No solvation		Polarizable continuum model	
	Atomic charge	Distance (Å) ^a	Atomic charge	Distance (Å) ^a
S	−0.134	0.0	−0.175	0.0
C _β	−0.355	1.845	−0.338	1.847
C _α	−0.087	2.851	−0.133	2.841
C _{carbox}	0.220	3.324	0.208	3.437
N	−0.498	3.262	−0.269	3.192
O	−0.314	3.936	−0.494	3.772
O	−0.248	3.749	−0.523	4.130
H _{cys}	0.116	1.350	0.167	1.372
$S \cdots H_w$ (Å) ^b	...	2.541	...	2.705

^a Atom-sulfur distance.

^b Distance to the H-bonding hydrogen of the nearest water molecule.

single water H-bonded model calculated without a polarizable continuum represents neutral cysteine, rather than the zwitterion.

The effect of $S_{\text{cys}} \cdots H-OH$ hydrogen bonding on the structural parameters and on the calculated Mulliken charges of the cysteine atoms, was tested by geometry optimization of 1-water, 2-water, and 3-water models, which are shown in Figures S5–S7 in the supplementary material. The bond distances within the cysteine moiety do not change appreciably upon addition of one, two, and three H_2O molecules at H-bonding distances. The first H-bonded H_2O molecule, which potentially has the largest effect on the sulfur K-edge XAS spectrum, optimizes to a $S_{\text{cys}} \cdots H-OH$ distance of ~ 2.70 Å. This is very near the crystallographic $R(H)S \cdots H-OR$ average of 2.67 Å.²⁸

The effect of the $Cys-(H)S \cdots H-OH$ H-bond on the total energy of the molecule was tested by varying the $Cys-(H)S \cdots H-OH$ distance from 2.1–4.1 Å using a higher level of theory (see Sec. II D). The rest of the molecule was allowed to optimize and changed insignificantly throughout the one-dimensional $Cys-(H)S \cdots H-OH$ scan. The results (Figure 8) show that the different $Cys(H)S \cdots H-OH$ distances have a small effect on the total energy (1.48 Kcal/mol). A more rigorous energy calculation was performed using the CCSD(T) protocol which established the corrected H-bonding energy at 1.63 Kcal/mol. For comparison a similar test of the $S_{\text{cys}}-C$

FIG. 8. DFT calculated one-dimensional potential energy surface. (a) ± 1 Å change in the $S \cdots H-OH$ hydrogen bond. The geometric structures from the two points shown in red have been used for the CCSD(T) energy calculations. (b) ± 0.75 Å change in the $S-C$ bond.

distance results in a rapid increase in energy as the distance deviates from the optimized value.

Although the H-bonding effect is weak, the Mulliken charge on the sulfur atom increases significantly upon addition of the H₂O molecule at H-bonding distance to the cysteine sulfur, indicating that in solution the Mulliken charge on the S atom is higher than in vacuum. This is counterintuitive, because the negative charge on the S atom should be partially neutralized by the positive dipole of the H-bond. However, the H-bonding modulates the charge on all the atoms in the molecule, including the carbon and hydrogen atoms adjacent to the sulfur atom, as determined by the lowest energy electronic structure from the DFT calculations. The result is an increase in charge on sulfur.

As shown in Sec. III, the rising-edge energy region of the S K-edge XAS spectrum of cysteine in solution cannot be well simulated using the extended continuum version of multiple scattering theory. To determine the nature and spin densities of the molecular orbitals involved in the rising edge transitions, population analyses were performed on the geometry optimized structural parameters obtained from the 1-water hydrogen-bonding model of cysteine. The contour plots of the four lowest lying unoccupied molecular orbitals are shown in Figure 9.

The contour plots and the calculated Mulliken spin densities clearly indicate that all the four orbitals (LUMO to LUMO+3) are delocalized over the entire cysteine molecule (see Table S2 for spin Mulliken populations in these orbitals). Electronic transitions from the sulfur 1s orbital to these low-lying valence orbitals result in the intense features at 2473.3 eV and 2474.2 eV observed in the sulfur K-edge XAS spectrum. Therefore, the transitions cannot be solely assigned to S–C σ^* or S–H σ^* orbitals, as has been done in the past,^{26,85} but should be described as excitations into extremely delocalized orbitals with significant S–C and/or S–H σ^* character.

FIG. 9. Contour plots of LUMO (33), LUMO+1 (34), LUMO+2 (35), and LUMO+3 (36). These orbitals have significant sulfur character and are expected to contribute to the sulfur K-edge XAS spectrum.

FIG. 10. Comparison of (—), the experimental S K-edge XAS spectrum of cysteine, pH 5.1; (---, blue dashed) the TD-DFT simulated spectra of cysteine, broadened to the core-hole lifetime width of 0.6 eV FWHM,⁴⁴ showing the transitions that dominate the spectrum; and (—, red) the fully simulated XAS spectrum, broadened to the experimental resolution of 1 eV FWHM. (Top) 0-water model, (bottom) 1-water model.

To test the accuracy of the DFT calculated ground state and valence level Mulliken spin densities, TD-DFT calculations were performed to compare with the experimental S K-edge XAS spectrum. A comparison of the calculated and experimental spectra is shown in Figure 10.

The calculated spectrum reproduces the experimental spectrum reasonably well, and indicates that there are two main transitions under the intense low-lying feature. The first and second peaks are due to a combination of the sulfur 1s core-to-valence transitions ($S1s \rightarrow 33$ plus $S1s \rightarrow 34$), and ($S1s \rightarrow 35$ plus $S1s \rightarrow 36$), respectively (see Figure 9 for the contour plots of MO 33, 34, 35, and 36). As can be seen from Figure 10, DFT fails to address spectral features above ~ 2475 eV, especially the broad peak at ~ 2480 eV. This supports the MXAN extended continuum calculations, which successfully simulated this broad peak as a continuum multiple scattering feature. Sulfur K-edge XAS TD-DFT calculations were also performed on the 1-water model (Figure 10, bottom). The addition of one-water at H-bonding distance to the S of cysteine leads to an increase in intensity of the third transition in the simulated spectra and a small decrease in intensity of the second transition. The energies of transitions 1 and 2 are perturbed less than 0.1 eV, while transition 3 is shifted up by ~ 0.2 eV (see Figure S8 in the supplementary material for a direct comparison of the two simulated spectra). This indicates that the inclusion of the H-bonding water molecule has a small effect on the electronic structure,

FIG. 11. (○) Sulfur K-edge XAS spectrum of dissolved cysteine, pH 5.1; (—, red) TD-DFT simulation of the rising edge energy region, broadened by 1 eV FWHM (see Figure 10); (—, green) individual transitions as calculated using TD-DFT; and (—, blue) the MXAN fit to the higher energy and continuum regions. The gray rectangle marks the energy region where the TD-DFT and MXAN simulations overlap.

especially at the sulfur. Since the H-bonding is to the sulfur atom, the effect on other distant atoms of cysteine is likely even smaller. Thus, H-bonding of S_{cys} to a water molecule does not significantly perturb the electronic structure of cysteine itself.

Finally, the MXAN and TD-DFT simulations were combined to reproduce the full sulfur K-edge XAS spectrum of dissolved cysteine. This result is shown in Figure 11.

Both the rising K-edge and continuum energy regions are successfully explained by combining the two formulations of atomic theory. The gray rectangle in Figure 11 shows the energy region where TD-DFT and extended continuum theory are each independently applicable. This is a bound-state energy region residing just below the continuum energy. The MXAN simulation in this energy region thus includes elements of the bound state intensity of transition three. This result is an explicit demonstration that extended continuum theory is successfully applicable to higher energy virtual bound states.

IV. DISCUSSION

In this study, the solvation about sulfur in aqueous cysteine has been investigated by applying a combination of MXAN and DFT to the room temperature S K-edge x-ray absorption spectrum of the dissolved molecule. Until this study, attempts to reproduce complete XAS spectra have either employed multiple scattering theory, best represented by MXAN and FEFF, which does not fully describe bound states,^{34,86} or have used DFT, represented by transition potential DFT (TP-DFT, such as implemented in StoBe), which only poorly reproduce continuum energy features,^{26,27,87,88} and cannot accurately predict extended solvation effects on the post-edge structure.²⁹ The combined approach developed here has produced a complete solution of the sulfur K-edge XAS spectrum of cysteine. Accurate bound-state transitions and ener-

gies have resulted from DFT, while ECMS theory has reproduced the higher energy bound state and the full continuum energy regions. In each case, the theory has been advanced over what has previously been used. Large basis sets have been introduced with the DFT approach, and the specific transitions have been identified using molecular orbital theory. The effect of charge has been introduced into ECMS theory, resulting in a more accurate simulation. The combination of improved DFT and ECMS theories has successfully reproduced the sulfur K-edge XAS spectrum of L-cysteine over its entire energy range.

DFT analysis of zwitterionic cysteine reveals that the H-bond between acceptor cysteine-sulfur and a donor water molecule is weak. The contour plots shown in Figure 9 illustrate that the valence orbitals contributing to the rising edge white line in the S K-edge XAS spectrum of cysteine are heavily mixed and are best described as delocalized mixtures of S-C σ^* and S-H σ^* orbitals. TD-DFT calculations of the S K-edge XAS spectrum of cysteine show that the two main features in the intense rising edge represent transitions from the sulfur 1s to these lowest energy delocalized orbitals. However, the Coulombic attraction between the positive charge in the core hole and the transition electron heavily localizes the excited bound-state electron on sulfur. H-bonding and other perturbations in distant parts of the cysteine molecule will not materially impact the transition energies of the core sulfur electron despite the delocalization of the excited state orbitals. That is, the probability distribution of the excited state electron within the delocalized orbitals attenuates rapidly with distance from sulfur. Therefore, solvent interactions with the charged carboxylate or ammonium groups of the cysteine molecule have little effect on the sulfur K-edge XAS transition energy or intensity.

Previous studies have assigned the first rising K-edge transition to a S-H σ^* based orbital and the second lower intensity transition to a S-C σ^* orbital.^{42,85} The assignment to a S-H σ^* transition was consistent with the decrease in intensity of the first peak upon deprotonation and hence widely accepted. Nevertheless, as demonstrated here and elsewhere,²⁷ the transition final states are somewhat more complicated and the relevant orbitals are delocalized over the entire cysteine framework. TD-DFT calculations also reveal that the 1-water model fits additional intensity near 2475.3 eV (Figure 10), and improves the correspondence between the calculated and experimental XAS spectra in the higher bound-state energy region. Clearly, the weak H-bonding interaction is modulating the valence orbital occupation and hence the transition probability into the corresponding anti-bonding orbitals. Thus, the TD-DFT calculation independently corroborates the MXAN result, in that it supports the idea of a durable hydrogen bonding interaction between a water molecule and sulfur in dissolved cysteine.

Figure 10 also shows, however, that the first two transitions remain essentially unperturbed with the addition of the water molecule. This is not surprising, because the main transitions are to orbitals that are not expected to change significantly upon H-bonding. This would also be true if the H-bonding was of the $S_{cys}-H \cdots OH_2$ type, in which case the water molecule would perturb the molecular framework valence

orbitals even less. A network of H-bonding water molecules (as expected in solution cysteine) may perturb the spectrum further, but any additional effect should be even weaker than the first H-bond to the sulfur. Note that since the effect of the inclusion of the water molecule on the S K-edge spectra is small, TD-DFT simulations alone cannot be used to establish the presence or absence of H-bonding to the sulfur atom.

A striking observation from the TD-DFT calculations is the lack of any transition intensity beyond ~ 2475.3 eV, even when molecular orbitals up to 15 eV above the LUMO were considered. This is a clear demonstration of the limitation of DFT in calculating transitions to higher valence orbitals. Unoccupied virtual orbitals become increasingly diffuse at greater energy and hence cannot be modeled well with DFT. For the XAS higher energy region, the best agreement is obtained with multiple scattering theory as deployed in MXAN^{32,37} and FEFF.^{86,89} A popular technique for fitting the post-edge region is by employing TP-DFT methods.⁹⁰ This method utilizes a diffuse basis set and employs a large number of basis functions to calculate transitions into the higher energy continuum states. Applications of TP-DFT to cysteine and other biologically relevant thiols and thioethers,²⁷ show that although TP-DFT works well in the rising edge region, where the near-valence levels contribute, the method performs poorly in the post edge region (~ 2475 eV and higher), where the intensity pattern is not well reproduced. This is because, beyond the edge, these methods give reasonable simulations of the experimental data only after special normalization techniques are applied. Some TP-DFT protocols still use the linearized spectral shape with huge broadening of the lines after the edge. In principle, this is an inaccurate approximation as the error grows rapidly beyond the edge.

In some cases, multi-electron excitations and autoionization states, which cannot be simulated by the inherently single-electron DFT calculations, can result in features near and beyond the rising edge. Such results indicate that DFT needs further improvements before both the near-edge and the post-edge energy regions can be simulated with reasonable accuracy.

Figure 12 below compares the sulfur K-edge XAS spectrum of dissolved cysteine with the theoretical XAS spectra, calculated using extended continuum multiple scattering theory,^{32,37} of a 3-atom-CSH fragment and the entire cysteine molecule.

The spectrum of the 3-atom fragment shows that virtually all the main cysteine XAS features represent the electronic state of sulfur and the nearest neighbor atoms alone. Thus all the important features of the XAS spectrum are reproduced using the CSH model, although better correlations with the energies and intensities of the features are obtained upon inclusion of the entire cysteine molecule.

As a general rule in the extended continuum approach, the calculated resonance broadens with the depth of the bound state within the potential energy well. Figure 12 shows that although the XAS features of the bound and the continuum electronic state are represented, extended continuum theory does not adequately reproduce the intensity or energy position of the principal cysteine bound-state feature. This result

FIG. 12. Sulfur K-edge XAS spectrum of (•••) cysteine in pH 5.1 citrate buffer; (—, red) the extended continuum theory MXAN fit to the full experimental XAS spectrum; and (—, blue) the three atom fragment also calculated using extended continuum theory. (Inset) Expansion of the XANES and near-edge continuum energy regions, also showing the structure of the minimal three-atom fragment.

underscores the importance of treating bound and continuum states with the appropriate forms of theory.

The present study demonstrates that the methodological combination of DFT and multiple scattering theory can be used to simulate and explain an entire XAS spectrum (Figure 11). The near-edge spectrum is independently simulated and explained using DFT and the post-edge spectrum is likewise explained using multiple scattering theory. The results from the two different analyses combine to yield a global one-electron picture of the geometric and electronic structure of the system under investigation.

The complete description of the sulfur K-edge XAS spectrum requires inclusion of nearby water molecules. Within the sulfur K-edge XAS, solvent-shell back-scattering intensity can arise only if one or more water molecules are durably localized near sulfur.

The single-scattering intensity of an absorber-scatterer interaction can be approximated as

$$I \approx \exp(-2k\sigma_j^2) \sin[2kR_j + \Phi_{ij}(R_j, k)], \quad (4)$$

where k is the photoelectron wave vector ($\text{\AA}^{-1}$), σ_j^2 is the thermal mean square displacement in absorber-scatterer distance, R_j is the distance between the absorber and the j th backscattering atom, and $\Phi_{ij}(R_j, k)$ is the atomic phase-shift function of the back-scattered photoelectron.⁹¹ The dependence of back-scattering intensity on R and σ^2 means that randomly organized water shells cannot be the source of extra-molecular x-ray absorption back-scattering intensity. The self-diffusion coefficient of a water molecule in pure water is $D = 2.27 \times 10^{-9} \text{ m}^2 \text{ s}^{-1}$ at 298 K.⁹² The rate of displacement across 4 $\text{\AA}$ for any given water molecule in pure water is given by $d^2/D \sim 10^{10} \text{ s}^{-1}$. In a dynamic and fluxional environment, any non-stabilized water solvation shell will undergo approximately 10^9 rearrangements per 0.1 s, which is the typical time constant for an XAS data point. Thus, any photoelectron back-scattering from individual water molecules in bulk thermal motion around cysteine sulfur should average

to near zero intensity in 0.1 s. Therefore, the presence of features in the sulfur K-edge XAS spectrum of cysteine that cannot be explained by reference to the amino acid structure itself, provides evidence for an external source of photoelectron back-scattering, implying the durable presence of solvent water localized near sulfur. MXAN simulations that included a two-water structural model of the hypothesized H-bonding network greatly improved the fit over that of cysteine alone, strengthening the hypothesis of a durable water network near the cysteine sulfur.

In this regard, Raman data revealed a mode II $S_{\text{cys}}\text{-H}\cdots\text{OH}_2$ donor-acceptor H-bonded structure in water solution,⁹³ whereas the present study found no evidence of this interaction. Nevertheless, the finding from XAS does not necessarily refute the Raman result. If mode II $(\text{Cys})\text{S-H}\cdots\text{OH}_2$ H-bonds indeed exist in water solution, the lack of XAS intensity most likely means that these structures are dynamic on the XAS measurement time-scale (~ 0.1 s). As discussed above, fluctuating dynamical processes can randomize photoelectron back-scattering, leading to undetectable XAS intensity. Only nearby scattering atoms in correlated motion with the absorber can contribute intensity to an XAS spectrum. It is likely that mode II H-bonding if present, corresponds to a minor fraction of the overall $S_{\text{cys}}\text{-H}\cdots\text{OH}_2$ interaction and does not have a significant signature in the XAS spectrum. Therefore, rejection by MXAN does not necessitate absence of mode II H-bonded structures. Strictly, the rejection means mode II H-bonding, if present, produces no sulfur K-edge XAS intensity under the present experimental conditions. The XAS result does mean, however, that only mode I H-bonding produces a consistently localized water molecule.

However, the number and positions of the sulfur-proximate water molecules cannot presently be fixed with accuracy. The weak dependence of R_{sq} on water-sulfur distance and the large statistical positional error in the final fit imply that the specific distance is much less important than the positive presence of a sulfur-water backscattering contribution to the XAS.

The weak interaction between cysteine sulfur and nearby water molecules makes it likely that any water molecule durably proximate to sulfur must be stabilized in that position, possibly by participation in an array of hydrogen-bonded water molecules anchored to the charged ammonium or carboxylate groups.^{76,77} If present, the waters within this array will exchange with waters in the bulk solvent, but at some rate much slower than the bulk exchange rate.⁹⁴

If this picture is accurate, the solvation region around sulfur will include slow-exchange water molecules that can contribute some back-scattering intensity to the sulfur K-edge spectrum. In using a structural model containing only one or two nearby water molecules, the MXAN algorithm minimized fit error by moving these waters so that they best reproduced the net photoelectron scattering of what is likely to be several proximate waters. This possibility rationalizes the weak dependence of the goodness-of-fit R_{sq} on the position of only one or two nearby water molecules. Further studies using *ab initio* molecular dynamics are underway to test this idea.

The structure of cysteine in water solution has been the subject of a number of other studies.^{26,27,29,95-98} The sulfur K-edge XAS spectrum of the cysteine zwitterion reported here is identical to those reported previously,^{27,42} apart from calibration differences. However, this work presents the first evidence that the sulfur of aqueous cysteine maintains a durable solvation shell consisting of at least one hydrogen-bonded water molecule.

Although a complete explanation of the sulfur K-edge XAS spectrum of cysteine requires inclusion of the scattering from proximate water, this result presents a conundrum. DFT calculations, both as reported earlier by others,²⁹ and as described here, indicate the $\text{Cys}(\text{H})\text{S}\cdots\text{H-OH}$ hydrogen bond is intrinsically weak and should be transient in water solution. Four of the many solvated zwitterionic cysteine structures investigated by Bachrach *et al.*,²⁹ included a water molecule H-bonded to the thiol sulfur (their Z5o, Z6vv, Z6ggg, and Z6nnn models), however none of them represented the lowest-energy structure. There is no reason to distrust these predictions. The central question that emerges, then, is how to reconcile the empirical finding that a hydrogen-bonding water molecule is durably proximate to sulfur with the convincing theoretical prediction of its absence, albeit with a limited solvation model.

We suggest that a major element toward this reconciliation must include the full hydrogen-bonding environment, which comprises all the solvent-solute interactions including any durable first solvation shell. The configurational thermodynamics and ground-state energy of the dissolved cysteine zwitterion will be governed in part by these extended interactions. The structure of water remains incompletely understood,^{99,100} and although advances have been made,¹⁰¹⁻¹⁰⁴ a complete description of molecular solvation is not available. However, the present work demonstrates that sulfur K-edge XAS provides an empirical target to assess the accuracy of theoretical models, namely, a water molecule that is persistently localized near sulfur. Agreement between theory and experiment should emerge from an accurate solvation model.

Ab initio molecular dynamics calculations of the Car-Parrinello type (CP-MD) are well suited to study the bonding dynamics in molecular systems,¹⁰⁵⁻¹⁰⁷ including hydrogen bonding dynamics among water molecules and between solute and solvent as a function of temperature.¹⁰⁸⁻¹¹²

Following discovery of water molecules durably located and proximate to sulfur, we began extensive CP-MD simulations to investigate the organization of water molecules around dissolved cysteine.¹¹³ Preliminary low-temperature *in silico* experiments have produced H-bonding networks that include stabilized water molecules positionally compatible with the MXAN fit. If this structural phenomenon persists up to room temperature, a first principles explanation will be available for the empirical structure discussed here. The structural model may then be refined in the context of the *ab initio* configurations.

Work along these lines is currently in progress in an attempt to reconcile the experimental discovery of water durably hydrogen-bonded to cysteine sulfur with the DFT theoretical prediction of its absence.

ACKNOWLEDGMENTS

M.B. acknowledges financial support from the SSRL Structural Molecular Biology program during a stay as a visiting scientist. This paper was made possible by Award No. P41 RR001209 (K.O.H.) from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). XAS data were measured at the Stanford Synchrotron Radiation Lightsource (SSRL), a Directorate of SLAC National Accelerator Laboratory, and an Office of Science User Facility operated for the U.S. Department of Energy Office of Science by Stanford University. The SSRL Structural Molecular Biology Program is supported by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and by the Department of Energy, Office of Biological and Environmental Research (BER). The paper was in part supported by Grant No. 5 P41 RR001209 from the NCRR, a component of the NIH and its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH.

- ¹T. Head-Gordon, J. M. Sorenson, A. Pertsemliadis, and R. M. Glaeser, *Bio-phys. J.* **73**, 2106–2115 (1997).
- ²R. Ludwig, *Angew. Chem., Int. Ed.* **40**, 1808–1827 (2001).
- ³B. Hribar, N. T. Southall, V. Vlachy, and K. A. Dill, *J. Am. Chem. Soc.* **124**, 12302–12311 (2002).
- ⁴A. Cooper, *Biophys. Chem.* **115**, 89–97 (2004).
- ⁵S. Xu, J. M. Nilles, and K. H. Bowen, *J. Chem. Phys.* **119**, 10696–10701 (2003).
- ⁶Y. Deng and B. Roux, *J. Phys. Chem. B* **108**, 16567–16576 (2004).
- ⁷T. Wytenbach and M. T. Bowers, *Chem. Phys. Lett.* **480**, 1–16 (2009).
- ⁸A. Pertsemliadis, A. M. Saxena, A. K. Soper, T. Head-Gordon, and R. M. Glaeser, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10769–10774 (1998).
- ⁹M. N. Blom, I. Compagnon, N. C. Polfer, G. von Helden, G. Meijer, S. Suhai, B. Paizs, and J. Oomens, *J. Phys. Chem. A* **111**, 7309–7316 (2007).
- ¹⁰B. Hernandez, F. Pfluger, M. Nsangou, and M. Ghomi, *J. Phys. Chem. B* **113**, 3169–3178 (2009).
- ¹¹T. Wytenbach, D. Liu, and M. T. Bowers, *Int. J. Mass. Spectrom.* **240**, 221–232 (2005).
- ¹²B. Gao, T. Wytenbach, and M. T. Bowers, *J. Am. Chem. Soc.* **131**, 4695–4701 (2009).
- ¹³H. E. Van Wart and H. Birkedal-Hansen, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5578–5582 (1990).
- ¹⁴H. A. Chapman, R. J. Riese, and G.-P. Shi, *Annu. Rev. Physiol.* **59**, 63–88 (1997).
- ¹⁵V. Turk, B. Turk, and D. Turk, *EMBO J.* **20**, 4629–4633 (2001).
- ¹⁶M. Capdevila, R. Bofill, Ö. Palacios, and S. Atrian, *Coord. Chem. Rev.* **256**, 46–62 (2012).
- ¹⁷M. Namdarghanbari, W. Wobig, S. Krezoski, N. M. Tabatabai, and D. H. Petering, *J. Biol. Inorg. Chem.* **16**, 1087–1101 (2011).
- ¹⁸Y. S. Nkabyo, Y.-M. Go, T. R. Ziegler, and D. P. Jones, *Am. J. Physiol. Gastrointest. Liver Physiol.* **289**, G70–G78 (2005).
- ¹⁹B. Stanic, M. Katsuyama, and F. J. Miller, Jr., *Arterioscler., Thromb., Vasc. Biol.* **30**, 2234–2241 (2010).
- ²⁰M. L. Kennedy and B. R. Gibney, *Curr. Opin. Struct. Biol.* **11**, 485–490 (2001).
- ²¹N. M. Giles, A. B. Watts, G. I. Giles, F. H. Fry, J. A. Littlechild, and C. Jacob, *Chem. Biol.* **10**, 677–693 (2003).
- ²²D. C. Johnson, D. R. Dean, A. D. Smith, and M. K. Johnson, *Annu. Rev. Biochem.* **74**, 247–281 (2005).
- ²³J. G. Parsons, M. V. Aldrich, and J. L. Gardea-Torresdey, *Appl. Spectrosc. Rev.* **37**, 187–222 (2002).
- ²⁴J. Yano and V. Yachandra, *Photosynth. Res.* **102**, 241–254 (2009).
- ²⁵E. I. Solomon, B. Hedman, K. O. Hodgson, A. Dey, and R. K. Szilagy, *Coord. Chem. Rev.* **249**, 97–129 (2005).
- ²⁶Y. Mochizuki, H. Agren, L. G. M. Pettersson, and V. Carravetta, *Chem. Phys. Lett.* **309**, 241–248 (1999).
- ²⁷E. D. Risberg, F. Jalilehvand, B. O. Leung, L. G. M. Pettersson, and M. Sandström, *Dalton Trans.* 3542–3558 (2009).
- ²⁸F. H. Allen, C. M. Bird, R. S. Rowland, and P. R. Raithby, *Acta Crystallogr. B* **53**, 696–701 (1997).
- ²⁹S. M. Bachrach, T. T. Nguyen, and D. W. Demoin, *J. Phys. Chem. A* **113**, 6172–6181 (2009).
- ³⁰M. Benfatto and S. Della Longa, *J. Synchrotron Radiat.* **8**, 1087–1094 (2001).
- ³¹M. Benfatto, S. Della Longa, and C. R. Natoli, *J. Synchrotron Radiat.* **10**, 51–57 (2003).
- ³²M. Benfatto and S. Della Longa, *J. Phys.: Conf. Ser.* **190**, 1–5 (2009).
- ³³P. D'Angelo, M. Benfatto, S. Della Longa, and N. V. Pavel, *Phys. Rev. B* **66**, 064209–064217 (2002).
- ³⁴S. Della-Longa, L. X. Chen, P. Frank, K. Hayakawa, K. Hatada, and M. Benfatto, *Inorg. Chem.* **48**, 3934–3942 (2009).
- ³⁵P. Frank, M. Benfatto, R. K. Szilagy, P. D'Angelo, S. Della Longa, and K. O. Hodgson, *Inorg. Chem.* **44**, 1922–1933 (2005).
- ³⁶P. Frank, M. Benfatto, B. Hedman, and K. O. Hodgson, *Inorg. Chem.* **47**, 4126–4139 (2008).
- ³⁷M. Benfatto, S. Della Longa, and P. D'Angelo, *Phys. Scr.* **T115**, 28–30 (2005).
- ³⁸R. Sarangi, P. Frank, K. O. Hodgson, and B. Hedman, *Inorg. Chim. Acta* **361**, 956–964 (2008).
- ³⁹I. Puigdomenech, The Royal Institute of Technology, Stockholm, 2004.
- ⁴⁰A. Albert, *Biochem. J.* **50**, 690–697 (1952).
- ⁴¹R. E. Benesch and R. Benesch, *J. Am. Chem. Soc.* **77**, 5877–5881 (1955).
- ⁴²I. J. Pickering, R. C. Prince, T. Divers, and G. N. George, *FEBS Lett.* **441**, 11–14 (1998).
- ⁴³A. Tenderholt, B. Hedman, and K. O. Hodgson, in *X-ray Absorption Fine Structure - XAFS13*, edited by B. Hedman and P. Pianetta (American Institute of Physics, Stanford University, 2007), Vol. 882, pp. 105–107.
- ⁴⁴M. O. Krause and H. H. Oliver, *J. Phys. Chem. Ref. Data* **8**, 329–338 (1979).
- ⁴⁵S. A. Moggach, S. J. Clark, and S. Parsons, *Acta Crystallogr.* **61**, 2739–2742 (2005).
- ⁴⁶I. P. Grant and M. A. Whitehead, *Mol. Phys.* **32**, 1181–1185 (1976).
- ⁴⁷J. P. Desclaux, *Comput. Phys. Commun.* **9**, 31–45 (1975).
- ⁴⁸J. J. Rehr, J. Mustre de Leon, S. I. Zabinsky, and R. C. Albers, *J. Am. Chem. Soc.* **113**, 5135–5140 (1991).
- ⁴⁹C. R. Natoli, M. Benfatto, and S. Doniach, *Phys. Rev. A* **34**, 4682–4694 (1986).
- ⁵⁰S. C. Lovatt, *J. Phys. A* **27**, 3787–3794 (1994).
- ⁵¹R. Sarangi, M. Benfatto, K. Hayakawa, L. Bubacco, E. I. Solomon, K. O. Hodgson, and B. Hedman, *Inorg. Chem.* **44**, 9652–9660 (2005).
- ⁵²M. J. Frisch, G. W. Trucks, H. B. Schlegel *et al.*, GAUSSIAN 09, Revision A.02, Gaussian, Inc., Wallingford, CT, 2009.
- ⁵³A. D. Becke, *Phys. Rev. A* **38**, 3098–3100 (1988).
- ⁵⁴A. D. Becke, *J. Chem. Phys.* **98**, 5648–5652 (1993).
- ⁵⁵C. T. Lee, W. T. Yang, and R. G. Parr, *Phys. Rev. B* **37**, 785–789 (1988).
- ⁵⁶S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.* **58**, 1200–1211 (1980).
- ⁵⁷A. Schaefer, H. Horn, and R. Ahlrichs, *J. Chem. Phys.* **97**, 2571–2577 (1992).
- ⁵⁸A. Schaefer, C. Huber, and R. Ahlrichs, *J. Chem. Phys.* **100**, 5829–5835 (1994).
- ⁵⁹J. E. Carpenter and F. Weinhold, *THEOCHEM* **46**, 41–62 (1988).
- ⁶⁰J. P. Foster and F. Weinhold, *J. Am. Chem. Soc.* **102**, 7211–7218 (1980).
- ⁶¹A. E. Reed, L. A. Curtiss, and F. Weinhold, *Chem. Rev.* **88**, 899–926 (1988).
- ⁶²G. Schaftenaar and J. H. Noordik, *J. Comput.-Aided Mol. Des.* **14**, 123–134 (2000).
- ⁶³F. Neese, ORCA: An *ab initio*, DFT and semiempirical SCF-MO package, Version 2.8.20, 2010.
- ⁶⁴F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297–3305 (2005).
- ⁶⁵F. Neese, F. Wennmohs, and A. Hansen, *J. Chem. Phys.* **130**, 114108 (2009).
- ⁶⁶S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, and F. Neese, *J. Phys. Chem. A* **110**, 2235–2245 (2006).
- ⁶⁷F. Neese, A. Hansen, F. Wennmohs, and S. Grimme, *Acc. Chem. Res.* **42**, 641–648 (2009).
- ⁶⁸S. DeBeer George, T. Petrenko, and F. Neese, *Inorg. Chim. Acta* **361**, 965–972 (2008).
- ⁶⁹T. A. Tyson, K. O. Hodgson, C. R. Natoli, and M. Benfatto, *Phys. Rev. B* **46**, 5997–6019 (1992).

- ⁷⁰A. Vila and R. A. Mosquera, *Chem. Phys.* **291**, 73–80 (2003).
- ⁷¹M. Nocker, S. Handschuh, C. Tautermann, and K. R. Liedl, *J. Chem. Inf. Model.* **49**, 2067–2076 (2009).
- ⁷²P. Zhou, F. Tian, F. Lv, and Z. Shang, *Proteins* **76**, 151–163 (2009).
- ⁷³V. Lalitha, E. Subramanian, and J. Bordner, *Int. J. Pept. Protein Res.* **24**, 437–441 (1984).
- ⁷⁴D. S. Eggleston and D. J. Hodgson, *Int. J. Pept. Protein Res.* **26**, 509–517 (1985).
- ⁷⁵T. Ebata, T. Hashimoto, T. Ito, Y. Inokuchi, F. Altunsoy, B. Brutschy, and P. Tarakeshwar, *Phys. Chem. Chem. Phys.* **8**, 4783–4791 (2006).
- ⁷⁶S.-P. Ju, W.-J. Lee, C.-I. Huang, W.-Z. Cheng, and Y.-T. Chung, *J. Chem. Phys.* **126**, 224901 (2007).
- ⁷⁷J. Chang, A. M. Lenhoff, and S. I. Sandler, *J. Phys. Chem. B* **111**, 2098–2106 (2007).
- ⁷⁸B.-K. Teo, *J. Am. Chem. Soc.* **103**, 3990–4001 (1981).
- ⁷⁹C. Brouder, *J. Phys. Condens. Matter.* **2**, 701–738 (1990).
- ⁸⁰T. E. Westre, A. Di Cicco, A. Filipponi, C. R. Natoli, B. Hedman, E. I. Solomon, and K. O. Hodgson, *J. Am. Chem. Soc.* **117**, 1566–1583 (1995).
- ⁸¹T. E. Westre, A. Di Cicco, A. Filipponi, C. R. Natoli, B. Hedman, E. I. Solomon, and K. O. Hodgson, *J. Am. Chem. Soc.* **116**, 6757–6768 (1994).
- ⁸²M. J. Scott, H. H. Zhang, S. C. Lee, B. Hedman, K. O. Hodgson, and R. H. Holm, *J. Am. Chem. Soc.* **117**, 568–569 (1995).
- ⁸³A. M. Rich, P. J. Ellis, L. Tennant, P. E. Wright, R. S. Armstrong, and P. A. Lay, *Biochemistry* **38**, 16491–16499 (1999).
- ⁸⁴R. Sarangi, R. K. Hocking, M. L. Neidig, M. Benfatto, T. R. Holman, E. I. Solomon, K. O. Hodgson, and B. Hedman, *Inorg. Chem.* **47**, 11543–11550 (2008).
- ⁸⁵C. Dezaud, M. Tronc, and A. P. Hitchcock, *Chem. Phys.* **142**, 455–462 (1990).
- ⁸⁶J. J. Rehr, J. J. Kas, F. D. Vila, M. P. Prange, and K. Jorissen, *Phys. Chem. Chem. Phys.* **12**, 5503–5513 (2010).
- ⁸⁷A. Mijovilovich, L. G. M. Pettersson, F. M. F. de Groot, and B. M. Weckhuysen, *J. Phys. Chem. A* **114**, 9523–9528 (2010).
- ⁸⁸S. Behyan, H. Yongfeng, and S. G. Urquhart, *J. Chem. Phys.* **134**, 244304 (2011).
- ⁸⁹J. J. Rehr and R. C. Albers, *Rev. Mod. Phys.* **72**, 621–654 (2000).
- ⁹⁰L. Triguero, L. G. M. Pettersson, and H. Ågren, *Phys. Rev. B* **58**, 8097 (1998).
- ⁹¹P. Behrens, in *Molecular Sieves - Science and Technology*, edited by H. G. Karge and J. Weitkamp (Springer, Berlin/Heidelberg, 2004), Vol. 4, p. 2733.
- ⁹²K. Tanaka, *J. Chem. Soc., Faraday Trans. 1* **74**, 1879–1881 (1978).
- ⁹³D. Garfinkel and J. T. Edsall, *J. Am. Chem. Soc.* **80**, 3823–3826 (1958).
- ⁹⁴M. Ide, Y. Maeda, and H. Kitano, *J. Phys. Chem. B* **101**, 7022–7026 (1997).
- ⁹⁵A. Fernández-Ramos, E. Cabaleiro-Lago, J. M. Hermida-Ramón, E. Martínez-Núñez, and A. Peña-Gallego, *J. Mol. Struct.: THEOCHEM* **498**, 191–200 (2000).
- ⁹⁶M. Canle Lopez, D. R. Ramos, and J. A. Santaballa, *Chem. Phys. Lett.* **417**, 28–33 (2006).
- ⁹⁷J. C. Dobrowolski, J. E. Rode, and J. Sadlej, *J. Mol. Struct.: THEOCHEM* **810**, 129–134 (2007).
- ⁹⁸Z. Tian, A. Pawlow, J. C. Poutsma, and S. R. Kass, *J. Am. Chem. Soc.* **129**, 5403–5407 (2007).
- ⁹⁹F. Paesani and G. A. Voth, *J. Phys. Chem. B* **113**, 5702–5719 (2009).
- ¹⁰⁰A. K. Soper, *Pure Appl. Chem.* **82**, 1855–1867 (2010).
- ¹⁰¹T. Head-Gordon and G. Hura, *Chem. Rev.* **102**, 2651–2670 (2002).
- ¹⁰²C. Huang, K. T. Wikfeldt, T. Tokushima, D. Nordlund, Y. Harada, U. Bergmann, M. Niebuhr, T. M. Weiss, Y. Horikawa, M. Leetmaa, M. P. Ljungberg, O. Takahashi, A. Lenz, L. Ojamae, A. P. Lyubartsev, S. Shin, L. G. M. Pettersson, and A. Nilsson, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 15214–15218 (2009).
- ¹⁰³G. N. I. Clark, C. D. Cappa, J. D. Smith, R. J. Saykally, and T. Head-Gordon, *Mol. Phys.* **108**, 1415–1433 (2010).
- ¹⁰⁴S. Merchant, J. K. Shah, and D. Asthagiri, *J. Chem. Phys.* **134**, 124514 (2011).
- ¹⁰⁵S. Furlan and G. La Penna, *Phys. Chem. Chem. Phys.* **11**, 6468–6481 (2009).
- ¹⁰⁶F. Guerrieri, V. Minicozzi, S. Morante, G. C. Rossi, S. Furlan, and G. La Penna, *J. Biol. Inorg. Chem.* **14**, 361–374 (2009).
- ¹⁰⁷P. Giannozzi, K. Jansen, G. La Penna, V. Minicozzi, S. Morante, G. Rossi, and F. Stellato, *Metallomics* **4**, 156–165 (2012).
- ¹⁰⁸C. Molteni and M. Parrinello, *J. Am. Chem. Soc.* **120**, 2168–2171 (1998).
- ¹⁰⁹S. Pratihar and A. Chandra, *J. Chem. Phys.* **134**, 024519 (2011).
- ¹¹⁰J. Stare, J. Mavri, J. Grdadolnik, J. Zidar, Z. B. Maksic, and R. Vianello, *J. Phys. Chem. B* **115**, 5999–6010 (2011).
- ¹¹¹J. T. Titantah and M. Karttunen, *J. Am. Chem. Soc.* **134**, 9362–9368 (2012).
- ¹¹²E. M. Tuckerman, K. Laasonen, M. Sprik, and M. Parrinello, *J. Phys. Chem.* **99**, 5749–5752 (1995).
- ¹¹³S. Morante, V. Minicozzi, R. Sarangi, M. Benfatto, P. Frank, B. Hedman, and K. O. Hodgson, “The solvation structure of aqueous cysteine: A QMMD, ECMS, and DFT theoretical study” (unpublished).
- ¹¹⁴See supplementary material at <http://dx.doi.org/10.1063/1.4767350> for starting models of solvated cysteine for the MXAN calculations, a comparison of the S K-edge spectra of L-cysteine from two different beam lines at SSRL, pH profile for dissolved L-cysteine, MXAN calculations with and without water molecules, DFT calculated models and xyz coordinates for the 1, 2, and 3 water models, tables showing Mulliken charges on atoms in the three different solvation models and Mulliken population analysis of L-cysteine and a comparison of TD-DFT computed S K-near-edge region of 0 water and 1 water models.