Phase speciation by extended x-ray absorption fine structure spectroscopy

Anatoly I. Frenkel[a)]
Department of Physics, Yeshiva University, New York, New York 10016

Oded Kleifeld
Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

Stephen R. Wasserman
MediChem Life Sciences, Woodridge, Illinois 60517

Irit Sagi[b)]
Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

(Received 28 December 2001; accepted 4 March 2002)

The application of x-ray absorption spectroscopy methods to both materials and life sciences is well appreciated. However, the power of extended x-ray absorption fine structure (EXAFS) spectroscopy as a quantitative structural technique has largely been limited by its application to the microscopically *homogeneous* systems, in which the local environment around each absorbing atom in the sample is the same. The growing interest in time-resolved EXAFS studies of systems in physics, chemistry, biology, and materials science has reintroduced the requirement for an analytical tool to probe *heterogeneous* mixtures *in situ*. While long being recognized as a premiere technique for this role, EXAFS studies of mixtures have been particularly difficult due to the strong model dependence and correlations between parameters in the fit. To circumvent these drawbacks, we introduce two new techniques in EXAFS analysis: the principal component analysis and the residual phase analysis. Using a test case of a heterogeneous mixture of two organometallic Co compounds, we demonstrate that these new EXAFS modeling techniques, together with the existing one, the multiple datasets fit method are the most suitable and adequate methods for phase speciation. In addition, we discuss the application of these data analysis approaches to biological systems. [DOI: 10.1063/1.1473193]

INTRODUCTION

Time-resolved structural methods are having an increased impact on biological[1–6] and chemical[7–9] research. For example, the ability to generate detailed structural data in real time can provide invaluable mechanistic insights into protein reactions that may be used to link the static end points provided by steady-state structural methods. Most important and relevant to this work, the use of time-resolved x-ray absorption spectroscopy (TRXAS) procedures to follow structural and electronic changes at catalytic metal centers in metalloenzymes during enzyme turnover has the promise of providing valuable structural-dynamic information in solution.[5] To gain a predictive understanding of the mechanism of catalysis involving metal oxides, and their use as supports of many other catalytic materials, their *in situ* hydrogen reduction has been recently studied using a combination of structural techniques.[8] It has been demonstrated that the discrimination between the two competing kinetic models of the reduction mechanism was made possible by using the time-resolved XAS measurements.[8]

X-ray absorption spectroscopy is a unique structural technique capable of providing bond lengths, coordination numbers, and electronic structures around a target atom in millisecond-or-better time resolution.[10] This structural information can be provided with the precision of *ca.* 0.01 Å or better. XAS can be applied to the crystalline phase, low-dimensional systems, gaseous and liquid phases with equal facility.[11] Due to these remarkable structural sensitivity, XAS is largely applied in a wide range of life science fields, such as materials, environmental, and biological sciences.

The experimental and theoretical details of extended x-ray absorption fine structure (EXAFS) have been extensively described in the literature.[11] The technique has gained wide popularity recently, due to major breakthroughs both in the state of the EXAFS theory[12–14] and the data analysis methods.[15–17] With the development of the *ab initio* theories of EXAFS, which take into account multiple scatterings of electrons, the unique structural knowledge (coordination numbers, bond distances, and bond angles) can be obtained within the range of distances up to 7–8 Å in the best-case scenario of bulk homogeneous samples or compounds.[18]

Unfortunately, the analytical power of the above techniques is greatly diminished when the system under investigation is a disordered heterogeneous mixture of species. Each species that contains the absorbing element may have quite different local coordination around that element. This greatly complicates the analysis, because the number of the relevant structural parameters may be comparable to or even exceed the number of independent data points in the experimental spectra.[19] However, there are many applications in

[a)]Corresponding author. Electronic mail: afrenke2@ymail.yu.edu
[b)]Corresponding author. Electronic mail: irit.sagi@weizmann.ac.il

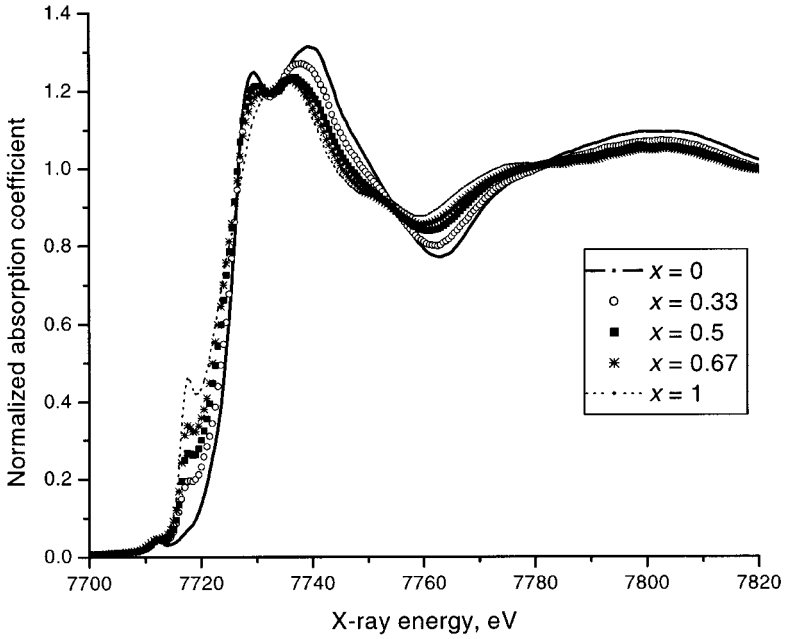

FIG. 1. Raw edge-step normalized absorption coefficients in all samples.

which the difference within the series of experimental spectra is due to the small number of externally controlled parameters; nevertheless, all the individual spectra can be theoretically described by the same or similar models. The most recent examples of the XAS studies of the multicomponent mixtures are (a) time-resolved measurements, in which the mixing fractions of species change with time;[8] and (b) the studies of metal complexes in solutions with different pH[20] or different molar concentrations of ligands.[21]

In these cases, if all the experimental spectra are analyzed *concurrently* and physically reasonable constraints are introduced, the increase in the total number of the data points is smaller than the increase in the number of structural parameters. Such an increase of the degrees of freedom in the fit can be executed in the framework of the multiple dataset (MDS) if implemented, for example, in the FEFFIT program of the UWXAFS data analysis package.[22] However, the MDS fit method has a typical drawback: By involving additional (and sometimes rather subjective) constraints among the structural variables, the undesirable model dependence of the results is increased.

In this work, we develop two alternative approaches for the EXAFS data analysis of mixtures: principal component analysis (PCA) and residual phase analysis (RPA). With the first approach, we can obtain the number of chemically distinct species in the heterogeneous samples model independently, by analyzing the samples' EXAFS spectra. In some cases, when the suitable experimental or theoretical standards are available, the PCA allows us to identify the species and obtain the mixing fractions of all the species in the samples. The second approach, the RPA, is a superior technique if the identification of the samples is difficult when using the MDS fit or the PCA. If both the number of the species as well as the identities of *some* of the species in the samples are known in advance, the RPA method allows us to identify the remaining species in the samples and obtain their mixing fractions.

To compare the reliability and limitations of the MDS fit, PCA, and RPA, we used all three techniques to study a test case of a heterogeneous mixture of two organometallic compounds: Cobalt acetylacetonate (Co-ACAC) and cobalt tetraphenyl-prophine (Co-TPP).

MATERIALS AND METHODS

Sample preparation. Cobalt acetylacetonate (Co-ACAC) and cobalt tetraphenyl-prophine (Co-TPP) in an analytical grade were purchased from Sigma. Five mixtures of $[\text{Co-ACAC}]_{1-x}[\text{Co-TPP}]_x$ were prepared by mixing, with $x=0$, 0.33, 0.5, 0.67, and 1.0. The mixtures were weighed by analytical balance, ground with mortar and pestle, and sieved, to ensure that the particles size is small compared to the inverse absorption edge jump at the Co K edge. The separate mixtures were brushed onto Scotch tape and mounted in copper sample holders. The accuracy of the mixing was ~5%.

Data collection and analysis. X-ray absorption data were measured using the Agere/UIUC beamline X16C at the National Synchrotron Light Source, located at the Brookhaven National Laboratory in Upton, New York. X-ray absorption coefficients in the samples were measured in fluorescence mode by using the ion chamber for the incident and the Stern–Heald detector for fluorescent radiation intensities, respectively. For the beam energy calibration, the reference Co metal foil was measured simultaneously with the sample in the transmission mode. The energy was varied from 200 eV below to 1000 eV above the Co K edge. The x-ray absorption coefficients were aligned in the absolute energy and normalized by the absorption edge jumps.

RESULTS

The local structural environments in the Co-ACAC and Co-TPP pure compounds are dramatically different. In Co-ACAC [Fig. 1(a)], the Co(III) ion is coordinated by six oxygen atoms, with average bond distances of 1.88 Å, forming a

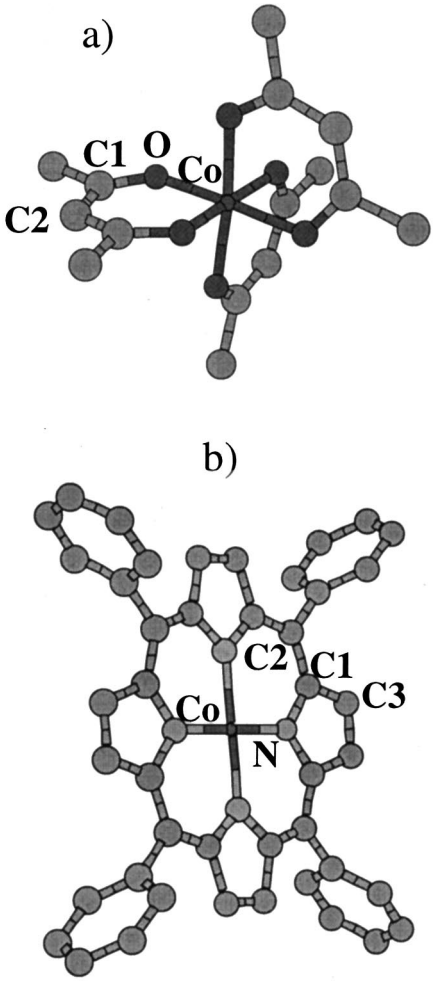

FIG. 2. Schematic of the Co-ACAC and Co-TPP structures.

distorted octahedron; in Co-TPP [Fig. 1(b)], the Co(II) ion is surrounded by four nitrogen atoms at an average bond distance of 1.95 Å in quasiplanar geometry. The differences in the coordination number between the two compounds are reflected in the pre-edge transitions of the raw XAS data (Fig. 2). Specifically, Co-TPP contains the $1s$-$4p$ peak intensity at 7715 eV, which is typical of square planar geometry.[23] This transition is absent in Co-ACAC, due to its octahedral geometry around the Co ion. The extent of mixing is reflected in the gradual reduction of the $1s$-$4p$ transition and in the change of the other edge features. The coordination numbers and distances of the next neighboring shells were determined by EXAFS data analysis (Table I). The results of the pure compounds by EXAFS fitting are in good agreement with those reported by x-ray crystallography. These results outline the feasibility of differentiating between the two compounds by EXAFS analysis.

TABLE I. Best fit values for the nearest neighbor distances (in Å) to Co in Co-ACAC and Co-TPP species obtained by the non-linear least square fits of EXAFS theory (FEFF7) to the data (Fig. 3). The notations of Co neighbors in Co-ACAC and in Co-TPP are the same as in Fig. 2. The subscript next to the neighboring atom symbol indicates the coordination number of the neighbor. Uncertainties in the distances obtained from EXAFS are given in parentheses.

	Co-ACAC			Co-TPP		
Neighbor	O_6	Cl_6	$C2_3$	N_4	Cl_8	$C2_4$
XRD	1.88	2.78	3.13	1.95	2.99	3.31
EXAFS	1.89(1)	2.82(2)	3.13(4)	1.96(1)	2.99(1)	3.31(1)

Multiple dataset fits. The theoretical EXAFS signals Co-ACAC and Co-TPP were calculated using FEFF7[12] The atomic coordinates for FEFF7 for the pure compounds were taken from the available crystallography data.[24] As a result of FEFF7 calculations, the partial Co–O, Co–N, and Co–C theoretical contributions (corresponding to the most significant, single-scattering nearest neighbor interactions in Co-ACAC and Co-TPP, respectively) were constructed. We verified the reliability of the analysis by fitting the FEFF7 theory to the experimental data for both Co-ACAC and Co-TPP (pure). Both the fit quality (Fig. 3), and the good agreement of the best-fit values of the bond lengths between Co and its nearest neighbors with the crystallography data (Table I) show that our procedure is reliable.

To fit the mixed system $[\text{Co-ACAC}]_{1-x}[\text{Co-TPP}]_x$, EXAFS data for $x=0.33$, 0.5, and 0.67 (the theoretical EXAFS signals for the pure Co-ACAC and Co-TPP phases) were weighted with the mixing factors $1-x$ and x, respectively, multiplied by k^2, and then Fourier transformed into r space, where the nonlinear least squares fits were performed, concurrently, on the three datasets, using the program

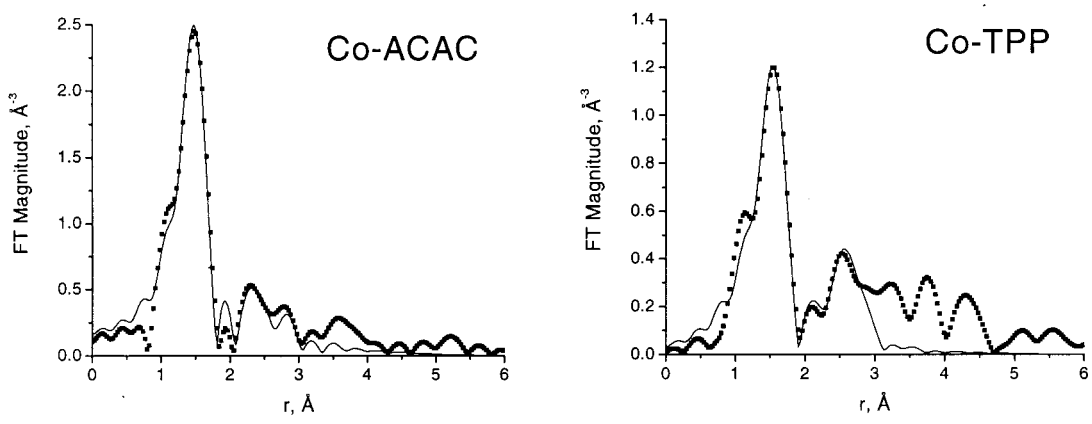

FIG. 3. Fourier transform magnitudes of the k^2-weighted EXAFS data (symbols) and FEFF fit (solid) for pure Co-ACAC and Co-TPP compounds.

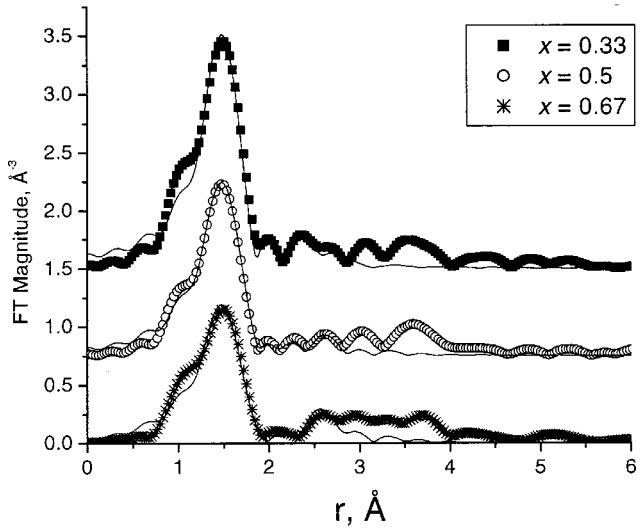

FIG. 4. Multiple dataset fit results. Shown are the Fourier transform magnitudes of the k^2-weighted EXAFS data (symbols) and FEFF fit (solid) for the $[\text{Co-ACAC}]_{1-x}$:$[\text{Co-TPP}]_x$ mixture at $x=0.33$, 0.5, and 0.67.

FEFFIT[22] (Fig. 4). The *k* ranges and *r* ranges in the fits were from 3 to 12 Å^{-1} and from 1.1 to 3 Å, respectively. The variables in the fits were the following.

(1) The corrections to the model distances for the three Co–O, Co-C1, Co-C2 nearest neighbors in the Co-ACAC phase and for the three Co–N, Co-C1, Co-C2 nearest neighbors in the Co-TPP phase (Fig. 2).

(2) The mean square deviations (EXAFS Debye–Waller factors).

(3) Three mixing fractions, x_{fit}.

(4) The corrections to the photoelectron energy origin (E_0).

(5) The passive electrons amplitude reduction factors[11] for Co.

The total number of variables (17) was, therefore, much smaller than the total number of relevant independent data points in the five EXAFS spectra (38). The best fit results for the mixing fractions are reported in Table II.

Principal component analysis. The standard PCA scheme represents each experimental spectrum as a vector $\mathbf{x}_i$ $(i=1,...,M)$ in the *N*-dimensional space, where *N* is the number of data points in each spectrum and *M* is the number of spectra. The data matrix **D**, of the dimension $M\times N$, is constructed from all the datasets. By finding the *M* eigenvectors and eigenvalues of **D**, and by arranging the eigenvectors in the descending order of eigenvalues, one can construct an ordered orthogonal basis. Each original spectrum can be represented as a linear combination of *M* basic vectors, or *components*. By selecting the eigenvectors having the *largest* eigenvalues and neglecting those with the *smallest* ones, one can represent all the datasets by using a linear combination of just a few (M_c) principal components (eigenvectors). Because $M_c<M<N$ (in most practical cases, $M_c\ll N$), the PCA provides a convenient way to reduce the dimension of the representation.

For the phase speciation purpose, the translation of the standard PCA scheme into the EXAFS language is straightforward. The EXAFS oscillations, normalized and background subtracted, now play a role in the PCA vectors $\mathbf{x}_i$. The dimension *N* is the number of energy points in the spectrum. The M_c principal components necessary to reconstruct the original data within the experimental noise now correspond to M_c distinct species in the original spectra.[25] Thus, it is possible to discover how many different species are in the sample without any *a priori* knowledge of the identity of these species. At the subsequent stages of the analysis, the species are identified by a comparison of the suitable experimental (or theoretical) standards with the linear combination of the obtained components. The mixing fractions of different species in all samples are then obtained by a linear least-square fitting.

In this work, we used the MATHEMATICA™-based program,[26] implementing the PCA method to analyze the $k^2\chi(k)$ data in the *k*-range from 2 to 11.5 Å^{-1}. For five samples with $x=0$, 0.33, 0.5, 0.67, and 1, we obtained the set of five eigenvalues and eigenvectors (components). By examining the decay of the eigenvalues with the component number, it is possible to obtain the least number of components (species in the sample) using the ''scree test.'' This is a graphic method for determining the number of principal components. The eigenvalues are plotted in the sequence of their decrease, and the number of principal components is chosen, where the curve levels off to a linear decline. Figure 5(a) suggests a two-component mixture, because the eigenvalues level off, beginning with the third component. By a *linear*[27] fit to the reference-pure compounds, two of them, Co-ACAC and Co-TPP, provided excellent fit, as expected. This proves that the two pure compounds indeed serve as good standards for this problem.

The reproduction of the data, for all *x* using the two principal components, is shown in Fig. 5(b). After rotating the matrix of the components onto the matrix of the standard compounds data, the mixing fractions x_{PCA} were obtained (Table II). It should be noted that the end values of the fractional coefficients (0 and 1), as obtained by the PCA, are always correct, and the comparison should be made only with the three intermediate concentrations.

Residual phase analysis (RPA). The RPA approach utilizes one of the known components, in this case the pure compound, as a ''starting phase.'' The ''starting phase'' is then fractionated and iteratively subtracted from the total XAS signal to produce corresponding residual spectra. The individual residual spectra are analyzed to obtain the best fit.

For simplicity, we can examine the case of the two-phase mixture, where the total experimental EXAFS data can be written as

$$\chi_D^{\text{ex}}(k)=x\chi_s^{\text{ex}}(k)+(1-x)\chi_R(k), \tag{1}$$

and where the first term on the right side of the Eq. (1) describes EXAFS from the starting (known) phase, which will be subtracted from the total signal. The second term denotes the EXAFS, originating from the residual phase only. The weighting coefficient *x* in Eq. (1) is the actual composition of species that is to be determined by this method. We can assume that, in a heterogeneous mixture, the local environment around the absorbers in each species is

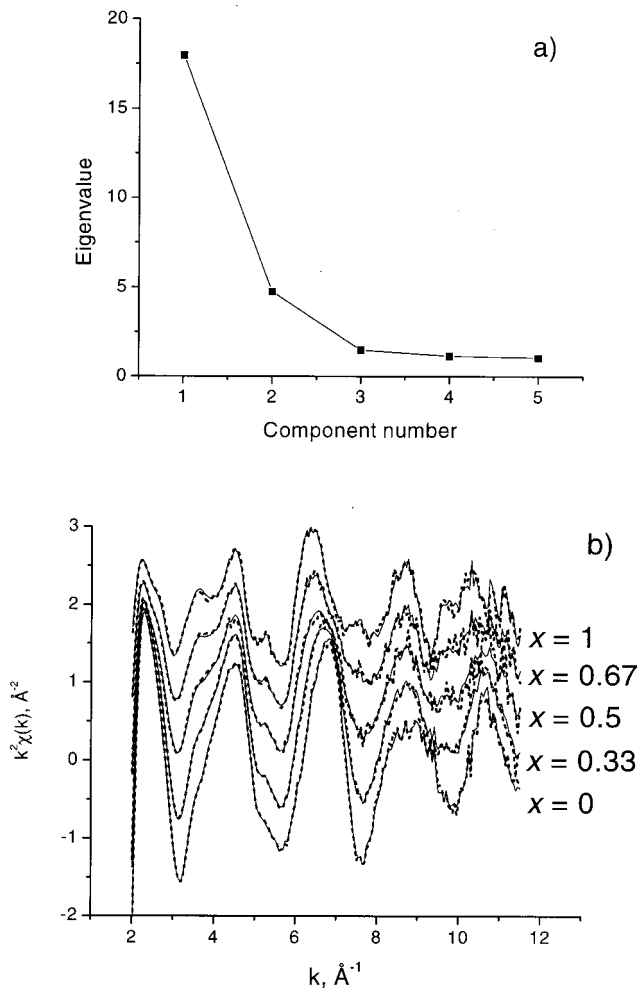

FIG. 5. Principal component analysis results. (a) The ''scree test'' demonstrates that using only two principal components with the highest eigenvalues must be sufficient to reproduce all the data. (b) Data reproduction using the two principal components.

unaffected by the presence of other species (this assumption is validated by the well-established transferability of amplitudes and phases in EXAFS). Under this assumption, Eq. (1) is exact.

If the appropriate models of the coordination environment of the absorber in the residual phase can be constructed, we can obtain the approximation for x as well, even though we know neither the actual value of x nor the identity of the residual phase. By introducing an adjustable mixing fraction y, we can approximate the experimental EXAFS originating from the residual phase and normalized per one absorbing atom by

$$\tilde{\chi}_R(k)=\frac{\chi_D^{\mathrm{ex}}(k)-y\chi_S^{\mathrm{ex}}(k)}{1-y}. \tag{2}$$

Note that $\tilde{\chi}_R(k)$ is not expected to be equal to the unknown residual phase signal $\chi_R(k)$ unless $y=x$. We can construct the theoretical EXAFS signal, $\chi_R^{\mathrm{th}}(k)$, corresponding to the residual phase, $\chi_R(k)$. Then, by varying the structural parameters in the theory, while fitting the χ_R^{th} to $\tilde{\chi}_R$,[28] we can obtain the best fit values for these parameters for each value of the external parameter y, at the same time minimizing the statistical chi square χ^2:

$$\chi^2=\frac{P}{N\epsilon^2}\sum_{i=1}^{N}(\tilde{\chi}_R-\chi_R^{\mathrm{th}})^2, \tag{3}$$

where P is the number of independent parameters in the fit, N is the number of data points in the fit, and ϵ^2 is the measurement uncertainty. By substituting the $\tilde{\chi}_R$ [Eq. (2)] and χ_D^{ex} [Eq. (1)] into Eq. (3), and by assuming that $\chi_R^{\mathrm{th}}(k)$ is a good model for $\chi_R(k)$, we obtain

$$\chi^2=\frac{(x-y)^2}{(1-y)^2}\frac{P}{N\epsilon^2}\sum_{i=1}^{N}(\chi_S^{\mathrm{ex}}-\chi_R^{\mathrm{th}})^2. \tag{4}$$

Therefore, the statistical χ^2 will change quadratically with $x-y$ in the vicinity of $y=x$ and will exhibit a *minimum* at the actual value of the mixing fraction x, which solves the problem of the residual phase analysis. Equation (4) confirms the intuitive prediction that the quality of the fit will be the best at the actual value of x, because the systematic errors associated with the subtraction will be minimized when x is correct.

We used the pure Co-TPP EXAFS data as the known phase $\chi_S^{\mathrm{ex}}(k)$ [Eq. (1)] for the subtraction purpose. The residual phase data, $\tilde{\chi}_R(k)$ [Eq. (2)], were constructed for all y between 0 and 1, with the increment of 0.1 for the mixtures' data with $x=0.33$, 0.5, and 0.67.

The residual phase theoretical EXAFS signal $\chi_R^{\mathrm{th}}(k)$ was constructed with FEFF in accordance with the Co-TPP structure. The fitting variables included the corrections to the Co–N, Co-C1, and Co-C2 distances [Fig. 2(b)] and their mean square disorders. The correction ΔE_0 to the photoelectron energy origin was varied as well. The passive electron reduction factor, S_0^2, was fixed at 0.77 in the fits. This value is the average between the $S_0^2=0.73$ and 0.80 obtained in the FEFF7 fits to the experimental EXAFS data of pure Co-ACAC and Co-TPP compounds, respectively. The total number of fitting variables, 7, was smaller than the total number of the independent data points (10).

To run the RPA automatically, a suite of UNIX scripts was developed to allow for both the rapid change of the theoretical model parameters used to calculate $\chi_R^{\mathrm{th}}(k)$ and the sequential fits of the experimental data while incrementing y from 0 to 1. The statistical chi-square values obtained in each fit were then analyzed as the function of y, and their minima were obtained (Fig. 6). Then, the corresponding mixing fractions were defined as the positions of the minima (Table II), in accordance with Eq. (4). It is important to mention that, despite the statistical chi square being in qualitative agreement with Eq. (4), the minima in Fig. 6 are greater than 0. The reason is that Eq. (4) was derived under assumption that the theoretical EXAFS function provides the perfect fit to the experimental data. However, due to both systematic and statistical errors in the theory and in the experiment, the theory always deviates from the experiment, explaining the residual positive chi-square values at all the minima in Fig. 6.

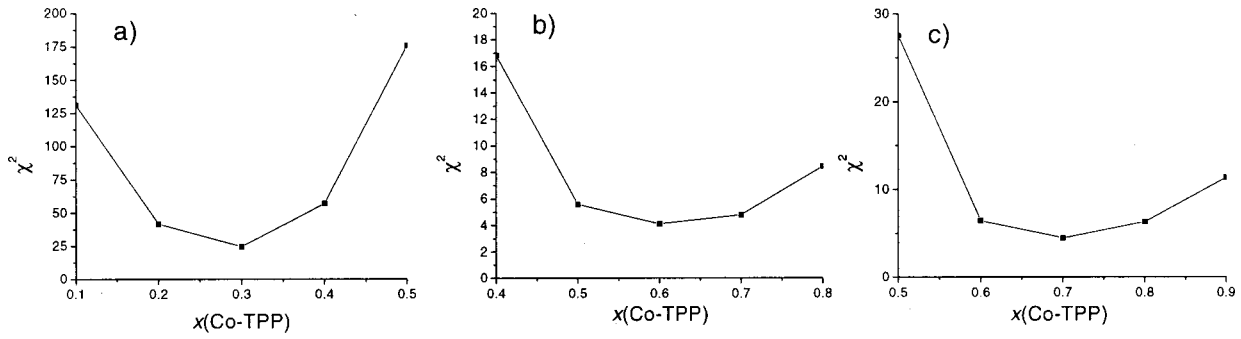

FIG. 6. Reduced statistical χ^2 obtained for the compositions $x=0.33$, 0.5, and 0.67, using the residual phase analysis.

DISCUSSION

Table II outlines the results of the quantitative comparisons of the three analytical approaches appropriate for the chemical phase speciation, using EXAFS spectroscopy measurements of the heterogeneous mixtures. Each analysis method resulted in the determination of the mixing fraction x of the phases, Co-ACAC and Co-TPP in the test mixture: $[\text{Co-ACAC}]_{1-x}[\text{Co-TPP}]_x$. For all three methods (the multiple dataset fit, principal component analysis, and residual phase analysis), the resultant mixing fractions x were in good agreement with the known concentrations of the Co-TPP species. However, these three methods have their unique areas of application. The MDS fit relies heavily on theoretical modeling of the EXAFS data in *each* species. It is the most powerful tool when the number of variables is much smaller than the number of the experimental data points.

If the number of variables approaches the number of data points, it usually means that either the data quality is poor or the theoretical model relies heavily on the large number of adjustable parameters. In both cases, confidence in the results is diminished, and it becomes dangerously easy to overinterpret the data. Such a limitation, predicted by the information theory,[29] is usually called the information bottleneck of EXAFS data analysis. The fitting procedure becomes even more ambiguous if the number of species is unknown. In that case, it is virtually impossible to analyze the local structure of the mixed species reliably. A typical example of such a complication is the analysis of catalytic intermediate states during biocatalysis by, e.g., metalloenzymes. In most cases, the enzymatic catalysis is rapid and involves a chain of intimate chemical changes within the catalytic site of the protein. Often, these changes can be related to structural changes that may be directly correlated with the reaction mechanism. However, the determination of the number and the identity of the intermediate states during biocatalysis is still a very complicated task, due to the lack of sufficient experimental and computational tools

TABLE II. Concentration x of the Co-TPP phase in the $[\text{Co-ACAC}]_{1-x}[\text{Co-TPP}]_x$ samples obtained using the three different methods:(1) by fitting the FEFF7 theory to the EXAFS data, (2) by using the principal component analysis and (3) by the residual phase analysis.

	Composition x of Co-TPP		
As prepared	0.33	0.5	0.67
MDS FIT	0.3(1)	0.62(15)	0.8(2)
PCA	0.36	0.57	0.79
RPA	0.3	0.6	0.7

If the number of the species in the sample is unknown, the principal component analysis is the ideal method for obtaining this number, model independently. In addition, as discussed in greater detail elsewhere, the PCA allows the detection of the changes in the data that would normally be considered below the sensitivity level of traditional EXAFS analysis methods.[30] We would like to mention that PCA has recently been recognized as an important tool for phase speciation using XANES data.[25] The important prerequisite for the success of PCA analysis of XANES spectra, however, is the presence of distinctly different features in the spectra of each individual species. It is very common for most biologically or environmentally important systems, for example, to contain XAS-active metals (Cu, Mn, Cr, S, etc.) in several oxidation states, each of which generates quite different fingerprint features in their XANES spectra. Therefore, PCA is a very powerful tool for chemical phase speciation in these cases. If, however, the target elements are present in the same oxidation states and have similar coordination, or if the XANES regions are relatively featureless, as in the case of heavy elements with significant lifetime broadening effects, the power of PCA to analyze XANES data is greatly diminished.

In the situations where both the MDS fit of EXAFS data and the PCA analysis of XANES are not plausible because of the above-mentioned reasons, the phase speciation can be achieved by applying PCA to EXAFS data measured in the mixtures. The only available example[31] of the successful use of PCA to analyze EXAFS data has dealt with *homogeneous* systems, i.e., where all absorbing atoms are equivalent with respect to their environment throughout the sample. Our goal has been a different one: to characterize the feasibility of PCA as a *phase speciation* technique where there is *macroscopic* segregation of different species throughout the sample. We have demonstrated that the PCA provides correct mixing fractions of different species in the system, within 10–15% uncertainty.

Even though PCA provides a model-independent determination of the number of species in the mixtures, establishing their identities and, therefore, the mixing fractions depends on the availability of the appropriate experimental or theoretical standards. The problem becomes complicated if

there is no preferred theoretical or experimental standard for all the species in the data, e.g., the structure of catalytic intermediate states in enzymatic catalysis. In these cases, as proposed by this work, the phase speciation problem may be then solved by the residual phase analysis. The deconvolution of the species may be simplified by subtracting the contribution of the starting phase from the total experimental data, as described above using the RPA, if there is a high degree of confidence about the local structure of the starting phase in the time-resolved measurements and, in addition, if these species are present in the sample at all times of the reaction. The analysis of species remaining after subtracting the known species can be best performed by fitting the theoretical model to the EXAFS data. This procedure reduces the number of variables in the analysis and increases confidence in the results. Therefore, RPA may be most suitable for the XAS analysis of complex and dynamic systems such as metalloenzymes. The availability of three-dimensional protein structures and accessibility to the protein data bank (PDB) provide a new means for structural dynamic investigation by XAS. Specifically, the development of time-resolved XAS has the potential to provide detailed structural data in real time and, therefore, can provide invaluable mechanistic insight into enzyme reactions. Using RPA for phase speciation in time-resolved XAS data analysis, in conjunction with available PDB coordinates as structural models, may result in resolving catalytic intermediate states that evolve during enzymatic catalysis.

CONCLUSIONS

In this work, we compared different strategies appropriate for the chemical speciation of heterogeneous mixtures using EXAFS spectroscopy, which include multiple datasets fit, and two new EXAFS modeling techniques: principal component analysis and residual phase analysis. The MDS fit method is most appropriate for mixed systems with a small number of species that are relatively well ordered and characterized. In that case, both the mixing fractions of the species and the local structures around the absorbing element in each species can be reliably obtained.

PCA, the other technique commonly used in statistics, can also be utilized to solve the phase speciation problem of multicomponent mixtures, using their EXAFS data. The unique advantage of this method is its robust, model-independent determination of the number of unique species in the samples. If good experimental standards exist to represent each species, this method can also reliably obtain both the identities and the mixing fractions of all the species in the sample. The main result, however, is that the true number of different species, as given by the number of principal components, remains valid, regardless of whether the identities of the species and their mixing fractions were obtained.

RPA is the last method, which was developed in the course of this work. RPA is the superior technique if both the EXAFS and PCA fail to establish the identities and mixing fractions of the species. This method is particularly powerful if the identities of one or more species in the mixture are known in advance (e.g., the starting phase in the time-resolved EXAFS measurements). In that case, the known phase EXAFS data is subtracted from the experimental data at each time step. The residual phase is analyzed with the fewer number of the adjustable variables providing the best-fit values of the mixing fractions of the residual phases, as well as their local structural characteristics. In addition, this procedure has the promise of application in the analysis of time-resolved XAS of biological systems.

ACKNOWLEDGMENTS

One of the authors (A.F.) acknowledges the support by the Yeshiva University research funds. I.S. acknowledges the support of the BSF Project No. 6602. The beamline X16C at the NSLS is supported in part by the U.S. Department of Energy through the Seitz Materials Research Laboratory (DEFG02-91-ER45439). The NSLS is supported by the Divisions of Materials and Chemical Sciences of the Department of Energy.

[1] E. A. Galburt and B. L. Stoddard, Phys. Today **54**, 33 (2001), and references therein.
[2] C. Walsh, Nature (London) **409**, 226 (2001).
[3] M. R. Chance, Biochem. Biophys. Res. Commun. **287**, 614 (2001).
[4] C. Y. Ralson, B. Sclavi, M. Sullivan, M. L. Deras, S. A. Woodson, M. R. Chance, and M. Brenowitz, Methods Enzymol. **317**, 353 (2000).
[5] O. Klefeld, A. Frenkel, and I. Sagi, J. Synchrotron Radiat. **8**, 978 (2001).
[6] M. R. Chance, L. Miller, R. F. Fischetti, E. Scheuring, W. X. Huang, B. Sclavi, Y. Hai, and M. Sullivan, Biochemistry **35**, 9014 (1996).
[7] P. Norby and J. Hanson, Catal. Today **39**, 301 (1998), and references therein.
[8] J. Rodriguez, J. Hanson, A. I. Frenkel, J.-Y. Kim, and M. Perez, J. Am. Chem. Soc. **124**, 346 (2002).
[9] M. S. Nashner, A. I. Frenkel, D. Somerville, C. W. Hills, J. R. Shapley, and R. G. Nuzzo, J. Am. Chem. Soc. **120**, 8093 (1998).
[10] L. X. Chen, W. J. H. Jäger, G. Jennings, D. J. Gosztola, A. Munkholm, and J. P. Hessler, Science **292**, 262 (2001).
[11] E. A. Stern and S. M. Heald, in *Handbook on Synchrotron Radiation*, edited by E. E. Koch (North-Holland, Amsterdam, 1983), Vol. 1, Chap. 10.
[12] S. I. Zabinsky, J. J. Rehr, A. L. Ankudinov, R. C. Albers, and M. J. Eller, Phys. Rev. B **52**, 2995 (1995).
[13] A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson, Phys. Rev. B **58**, 7565 (1998).
[14] For a review, see J. J. Rehr and R. C. Albers, Rev. Mod. Phys. **72**, 621 (2000).
[15] G. Bunker, Nucl. Instrum. Methods **207**, 437 (1983).
[16] E. A. Stern, Y. Ma, O. Hanske-Petitpierre, and C. E. Bouldin, Phys. Rev. B **46**, 687 (1992); A. I. Frenkel, E. A. Stern, A. Voronel, A. Rubshtein, Y. Ben-Ezra, and V. Fleurov, *ibid.* **54**, 884 (1996).
[17] J. Mustre, Y. Yacoby, E. A. Stern, and J. J. Rehr, Phys. Rev. B **42**, 10843 (1990); M. Vaarkamp, I. Dring, R. J. Oldman, E. A. Stern, and D. C. Koningsberger, *ibid.* **50**, 7872 (1994).
[18] A. I. Frenkel, E. A. Stern, A. Voronel, M. Qian, and M. Newville, Phys. Rev. B **49**, 11662 (1994).
[19] E. A. Stern, Phys. Rev. B **48**, 9825 (1993).
[20] A. I. Frenkel and G. V. Korshin, in *Humic Substances: Versatile Components of Plants, Soil and Water*, edited by E. A. Ghabbour and G. Davies (Royal Society of Chemistry, Cambridge, 2000), p. 227.
[21] A. I. Frenkel, M. Vairavamurthy, and M. Newville, J. Synchrotron Radiat. **8**, 669 (2001).
[22] E. A. Stern, M. Newville, B. Ravel, Y. Yacoby, and D. Haskel, Physica B **208&209**, 117 (1995).
[23] D. M. Wirt, I. Sagi, and M. R. Chance, Biophys. J. **63**, 412 (1992).
[24] G. J. Kruger and E. C. Reynhardt, Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. **30**, 822–824 (1974); S. G. DiMango, A. K. Wertsching, and C. R. Ross, J. Am. Chem. Soc. **117**, 8279 (1995).
[25] S. R. Wasserman, J. Phys. IV **7**, C2-203 (1997).
[26] The PCA software was written by and available from S. R. Wasserman.

[27] In the PCA method, the fit of M_c principal components to the M_c experimental references is a *linear* least-square fit that requires M_c-1 variables (only one variable was used in this work, where two principal components were used). This method of fitting is much less ambiguous than the *nonlinear* least square fit used in the theoretical methods of EXAFS analysis.
[28] We omit the argument of the residual phase EXAFS functions χ_R^{th} and $\tilde{\chi}_R$ because fits may be performed in either k space or in r space.
[29] L. Brillouin, *Science and Information Theory* (Academic, New York, 1962).
[30] S. R. Wasserman, M. Denecke, J. Rothe, and K. Dardenne (unpublished).
[31] S. R. Wasserman, P. G. Allen, D. K. Shuh, J. J. Bucher, N. M. Edelstein, J. Synchrotron Radiat. **6**, 284 (1999).