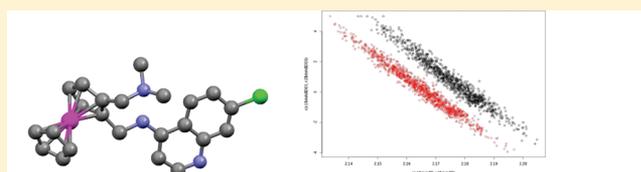


Statistical Methodology for the Detection of Small Changes in Distances by EXAFS: Application to the Antimalarial Ruthenoquine

Emmanuel Curis,^{*,†} Faustine Dubar,[‡] Ioannis Nicolis,[†] Simone Bénazeth,[†] and Christophe Biot[¶][†]Laboratoire de biomathématiques, Plateau iB², Département Santé publique et biostatistiques, faculté de Pharmacie, Université Paris Descartes, Paris, France, Université Lille Nord de France[‡]Université Lille 1, UCCS, UMR CNRS 8181, 59652 Villeneuve d'Ascq Cedex, Lille, France[¶]Université Lille Nord de France, Université Lille 1, UGSF, CNRS UMR 8576, IFR 147, 59650 Villeneuve d'Ascq Cedex, Lille, France

ABSTRACT: Antimalarial compounds ruthenoquine and methylruthenoquine were studied by X-ray absorption spectroscopy both in solid state and in solution, in normal (aqueous or CH₂Cl₂ solutions) and oxidative (aqueous solution with H₂O₂, either equimolar or in large excess) conditions, to detect small changes in the coordination sphere of the ruthenium atom. Since changes in the EXAFS spectra of these compounds are quite subtle, a complete procedure was developed to assess the different sources of uncertainties in



Monte-Carlo simulations and accounting for correlations between ΔE_0 and R allows for the detection of subtle changes in distances by XAS (right). This method shows that ruthenoquine (left) may be partially oxidized in solution by H₂O₂.

fitted structural parameters, including the use of multivariate statistical methods for simultaneous comparison of edge energy correction ΔE_0 and distances, which can take into account the very strong correlation between these two parameters. Factors limiting the precision of distance determination depend on the recording mode. In transmission mode, the main source of uncertainty is the data reduction process, whereas in fluorescence mode, experimental noise is the main source of variability in the fitted parameters. However, it was shown that the effects of data reduction are systematic and almost identical for all compounds; hence, they can be ignored when comparing distances. Consequently, for both fluorescence and transmission recorded spectra, experimental noise is the limiting factor for distance comparisons, which leads to the use of statistical methods for comparing distances. Univariate methods, focusing on the distance only, are shown to be less powerful in detecting changes in distances than bivariate methods making a simultaneous comparison of ΔE_0 and distances. This bivariate comparison can be done either by using the Hotelling's T^2 test or by using a graphical comparison of Monte Carlo simulation results. We have shown that using these methods allows for the detection of very subtle changes in distances. When applied to ruthenoquine compounds, it suggests that the implication of the nonbinding doublet of the aminoquine nitrogen in either protonation or methylation enhances the tilt of the two cyclopentadienyls. It also suggests that ruthenoquine and methylruthenoquine are, at least partially, oxidized in the presence of H₂O₂, with a small decrease in the Ru–C bond length and increase in the edge energy.

INTRODUCTION

X-ray absorption spectroscopy is well suited to study the local environment of the metal atom in bioinorganic compounds, both in solid state and in solution. It can give insights on the degree of oxidation of the metal atom through the edge energy and on the distance between this atom and the atoms of its first coordination shell. However, although good precision is generally assumed for distances, there is, to our knowledge, no standardized procedure to determine if distances or edge energies obtained for two samples are, indeed, different. In fact, the question is complex for two reasons: First, these values are obtained through a fitting procedure involving also Debye–Waller factors, and a straight comparison using standard errors or confidence intervals of these parameters does not take into account the correlation between them. Second, uncertainties in these values not only come from experimental noise, but also from various sources such as the choice of the initial values in the fit process, the EXAFS oscillation extraction procedure, etc.

Some hints for a correct estimation of distances and their comparisons are presented in the reports of the Standards and Criteria Committee.^{1,2} Usually, precision is considered to be around 0.01 Å for distance determination, and 0.001 Å for distances comparisons. For mineral crystals such as oxides, methods were developed to allow a very precise determination and comparison of distances, with an accuracy ranging from 2×10^{-4} Å (20 fm)³ to 10^{-5} Å (1 fm).⁴ Such a precision was achieved by using third-generation synchrotron sources with a high beam stability, especially in energy; some of the approaches use differential EXAFS.^{4,5} Analysis involves taking into account thermal expansion of the crystal; hence, studies at different temperatures and introducing special parameters known as cumulants.^{3,6}

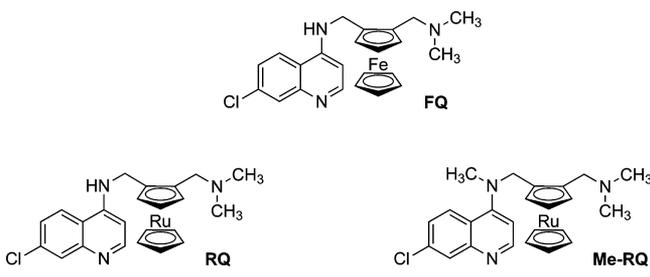
Received: February 23, 2012

Revised: May 16, 2012

Published: May 18, 2012

The aim of this paper is to present a potential procedure for handling such comparisons for bioinorganic compounds in solution or in solid state. For such compounds, the thermal expansion approach cannot be used, since they are not in crystalline form. Our procedure is based on the use of suited statistical tools and Monte Carlo simulations. This procedure will be illustrated on samples of ruthenoquine (RQ) and its methylated derivative, methylruthenoquine (Me-RQ) (Chart 1)⁷

Chart 1. Chemical Structures of Ferroquine (FQ), Ruthenoquine (RQ), and Methylruthenoquine (Me-RQ)



in solid state and solutions. Since the cyclopentadienyl ligands are tightly bound to the ruthenium atom and are quite rigid, only very small changes in Ru–C distances are expected, even if the oxidation state of the ruthenium changes, thereby giving a higher challenge to the procedure.

We selected ruthenoquine, the cold tracer of the antimalarial ferroquine,⁸ as a case study because of the need to gain insights into its structure. Recently, we performed a feasibility study on single erythrocytes infected by the HB3 strain of *Plasmodium falciparum*.⁹ We showed that the drugs can be detected and mapped by X-ray fluorescence in a parasite exposed to concentrations that are comparable to dosages of ferroquine in clinical assays.¹⁰ Moreover, the present study enables us to compile a library of reference spectra that will be used as a basis in further X-ray imaging experiments in which micro-XANES measurements will give insights into the localization and chemical state of RQ in the parasite. We also included methylruthenoquine (Me-RQ, Chart 1) in this study.

Last, there are still some questions about the behavior of ruthenoquine in an oxidative environment, as may be found in the digestive vacuole of the parasite. Electrochemical oxidation of ruthenocene leads to complex species that may involve dimerization^{11,12} and in which the oxidation state of the ruthenium core is not clear; the same difficulties are then expected for RQ and related compounds. Fine studies of the X-ray absorption edge and of the ruthenium–carbon distances may give insights on this question. For this reason, experiments were made with RQ and Me-RQ in aqueous solutions with either H₂O₂ or hemine, the hemoglobin degradation product that is present in the digestive vacuole, as the oxidative agent, to be close to biologic conditions.

MATERIALS AND METHODS

Sample Preparation. RQ = ruthenoquine, RQ·2HCl = dihydrochloride salt of ruthenoquine (by protonation of the amine groups), Me-RQ = methylruthenoquine, Me-RQ·2HCl = dihydrochloride salt of methylruthenoquine.

Synthesis. RQ and methylruthenoquine Me-RQ were synthesized as previously reported.^{7,13} Boron nitride and ruthenocene were purchased at Sigma Aldrich.

Synthesis of RQ and Me-RQ Dichlorhydrate Salts. RQ or Me-RQ (0.500 g, 1 mmol) was dissolved in anhydrous dichloromethane (5 mL). Gaseous HCl (obtained by reaction of H₂SO₄ on NaCl) was bubbled into this solution until precipitation of the dichlorhydrate salt. The precipitate was filtered (RQ, 0.552 g, 1 mmol; Me-RQ, 0.574 g, 1 mmol).

Sample Preparation for XAS Experiments. *Preparation of Solid Samples.* Using a mortar, RQ (or Me-RQ, RQ·2HCl, Me-RQ·2HCl) (10 mg, 20 μmol) or ruthenocene (10 mg, 40 μmol) and boron nitride (10 mg, 400 μmol) were mixed and ground to a fine powder. The powder was then transferred in Evacuatable Pellet Dies. The pellet was made with a load of 2 tons and a pellet diameter of 6 mm.

Preparation of Solution Samples. For dichloromethane solutions, RQ (2.4 mg, 5 μmol) or ruthenocene (1.2 mg, 5 μmol) was dissolved in dichloromethane (1 mL) for a final concentration of 5 mM. The analysis was performed on a fresh solution.

For aqueous solutions, RQ·2HCl, or Me-RQ·2HCl (2.8 mg, 5 μmol) was dissolved in distilled water (1 mL) for a final concentration of 5 mM. The analysis was performed on a fresh solution.

Equimolar aqueous solutions of RQ or Me-RQ, and H₂O₂ were prepared from 20 μL of 30% aqueous H₂O₂ dissolved in 19.98 mL of distilled water. In 1 mL of this solution, 6.6 mg (11 μmol) of RQ or Me-RQ was dissolved for a final concentration of 11 mM.

Solutions of RQ or Me-RQ with an excess of H₂O₂ were prepared by dissolving 6.6 mg of RQ or Me-RQ in 1 mL of a solution obtained by mixing 150 μL of 30% aqueous H₂O₂ in 19.85 mL of distilled water.

The RQ solution in HEPES was prepared by mixing 4 mL of distilled water, 5 mL of a solution of 5.5 mg (11 μmol) of RQ in 10 mL DMSO (solution A), and 1 mL of a solution of 25 mg (100 μmol) HEPES (solution B) in 5 mL of distilled water for a final RQ concentration of 5.5 mM.

The hemine solution in HEPES (solution 4) was prepared by mixing 4 mL of distilled water, 1 mL of solution B, and 5 mL of a solution of 8 mg (12 μmol) of hemine in 10 mL of DMSO.

The equimolar solution of RQ and hemine was prepared by mixing 1 mL of the RQ solution in HEPES and 1 mL of the hemine solution in HEPES for a final RQ concentration of 2.75 mM.

The solution of RQ with an excess of hemine was prepared by mixing 1 mL of the RQ solution in HEPES and 2 mL of the hemine solution in HEPES for a final RQ concentration of 1.83 mM.

XAS Experiments. Data Recording. Spectra were recorded on the SAMBA beamline (SOLEIL synchrotron, Saint-Aubin, France), at the K-edge of the ruthenium (22.1 keV). A ruthenium foil spectrum was recorded for energy calibration. To avoid saturation of the ionization chambers (in transmission mode) and of the fluorescence detector (in fluorescence mode), an aluminum filter had to be added; its thickness was adapted to have a beam intensity as high as possible, maintaining the linearity of the detectors.

Solid state samples (RQ, Me-RQ, RQ·2HCl, Me-RQ·2HCl, and ruthenocene) spectra were recorded on pellets in transmission mode with two ionization chambers. The edge region was recorded with a constant energy scale step (pre-edge, 5 eV step; acquisition time, 1 s/point; edge, 0.5 eV, 1 s/point; postedge, 1 eV, 1 s/point), whereas the EXAFS region was recorded with a constant *k*-scale step (assuming an edge

energy of 22 117 eV) with the acquisition time being increased from 2 to 4 s/point to get a constant signal/noise ratio after k^2 weighting. All spectra were recorded at least three times.

Solution samples (RQ in CH_2Cl_2 , RQ in $\text{H}_2\text{O} + \text{HCl}$, Me-RQ in $\text{H}_2\text{O} + \text{HCl}$, RQ in $\text{H}_2\text{O} + \text{H}_2\text{O}_2$, Me-RQ in $\text{H}_2\text{O} + \text{H}_2\text{O}_2$, and RQ and hemine in HEPES buffer) spectra were recorded in fluorescence mode using a seven-element germanium detector. Potential spectrum distortion induced by self-absorption was estimated using the approach previously described¹⁴ using the LASE software, which was written by one of us,^{15,16} based on the solid state sample spectra, and it was found to be negligible for such diluted solutions. Spectra were recorded using the same energy grid as above, with an acquisition time of 4 s in the edge region and from 4 to 8 s in the k -scale region; each sample was recorded at least three times.

Data Analysis. All EXAFS analyses were made using the LASE software. Background absorption was removed using either a Victoreen model (transmission) or a straight line (fluorescence); edge energy was taken at 22 126 eV (inflection point of the ruthenocene edge) for all spectra to ensure a comparable k -scale. EXAFS oscillations were extracted using a polynomial smoothing in E and k -spaces, controlled to minimize the very short distance peak of the Fourier transform. For all analyses, spectra were averaged after background removal and error bars propagated throughout EXAFS oscillation extraction and Fourier transform using previously published results.¹⁷ These unfiltered EXAFS oscillations were then fitted to obtain structural parameters values. All fits were made both on averaged spectra and on individual spectra, on unfiltered EXAFS oscillations, to investigate the effect of the extraction itself on parameters values.

Unfiltered k^2 -weighted EXAFS oscillations were analyzed without weighting by the experimental errors (since acquisition conditions ensure constant error with k^2 weighting) between $k = 4$ and 14 \AA^{-1} . Fitted parameter uncertainties and correlations were estimated through the Monte Carlo procedure previously described:¹⁸ $n = 1000$ random unfiltered spectra were generated based on a normal distribution with expectation equal to the average of the experimental, unfiltered EXAFS oscillations, and a diagonal covariance matrix, with diagonal values taken as the experimental squared standard deviation on each experimental point, as obtained when averaging the experimental spectra. As shown before,¹⁷ unfiltered spectra have uncorrelated data points, but this is not true for spectra after Fourier-transform filtering (either used for noise removal or for peak selection). Each of these random n spectra was fitted, leading to a set of n fitted parameter values, then studied using standard statistical methods. Fits were performed using either reference, experimental phase and amplitude, or theoretical phases and amplitudes.

Reference, experimental phase, and amplitude were extracted from the ruthenocene pellet spectrum by filtering the Fourier transform peak corresponding to the first coordination sphere, the back Fourier transform leading to raw phase and amplitude of the corresponding EXAFS spectrum. There, raw phase and amplitude were distance-corrected using the average Ru–C distance (2.188 \AA) found in the several available ruthenocene crystal structures,^{19–21} then used to fit experimental EXAFS spectra of other compounds.

To generate theoretical phases and amplitudes, including multiple scattering paths, a structural model was built from the most recent ruthenocene crystal structure at ambient temper-

ature²⁰ and the ruthenoquinone crystal structure²² and used in FEFF 8.2.^{23,24} When using the crystal structure of ruthenocene, FEFF generates numerous scattering paths because of slight differences in distances in the crystal; such paths were averaged in LASE before fitting. The resulting model was fitted to both filtered and unfiltered data, and path selection was made to have the best fit with the simplest model.

Since only single scattering from the first coordination shell leads to a significant contribution for the EXAFS oscillations after this path selection step, the following parameters were finally fitted, using reference phase and amplitudes: R_C , the average distance of the 10 carbon atoms from this first coordination sphere; σ_C^2 , the associated Debye–Waller factor; ΔE_0 , the correction of the edge energy and S_0^2 , the amplitude correction factor to account for multielectronic transitions. The fitted equation was then

$$\chi(k) = S_0^2 \frac{N_C A(k')}{R_C k} e^{-2k^2 \sigma_C^2} \sin(2kR_C + \Phi(k'))$$

where $N_C = 10$ is the number of carbon atoms in the first coordination shell of the ruthenium atom; $A(k')$ and $\Phi(k')$ are the electronic phases and amplitudes, including the mean free path, either computed by FEFF or extracted from the reference (RQ pellet) spectrum, and corrected for the different edge energy, with $k' = (k^2 + 2(m_e/\hbar)\Delta E_0)^{1/2}$ (m_e being the electron mass and \hbar the reduced Planck constant, expressed in suitable units). Least-square estimates are used, with a minimization procedure based on stepwise gradient optimization, with analytical computation of the derivatives and no constraint. Initial values for these parameters were $\Delta E_0 = 0 \text{ eV}$, $R_C = 2.1880 \text{ \AA}$, and $\sigma_C^2 = 0 \text{ \AA}^2$, corresponding to the values used to extract experimental phases and amplitudes. Influence of the initial values was checked by randomly selecting $n' = 1000$ random alternative initial values uniformly in the 3-dimensional box $[\Delta E_{0_{\min}}, \Delta E_{0_{\max}}] \times [R_{C_{\min}}, R_{C_{\max}}] \times [\sigma_{C_{\min}}^2, \sigma_{C_{\max}}^2]$ and comparing the fit results.

Statistical Procedure. All statistical analyses were made using the R software, version 2.14,²⁵ and a set of scripts written by one of us. After the fit procedure presented above, several sets of parameter values were available for each compound, including the distances to be compared. A two-step procedure was used to make these comparisons.

First, the various sources of parameter estimate variability were explored. To evaluate the influence of experimental noise (“random errors”), the same model was fitted on the average experimental spectrum, on each individual spectrum, and on the filtered spectrum. To evaluate the influence of initial values in the fit process, several initial values were used, and the fit results were compared. To evaluate the influence of the data reduction steps, data reduction was made several times, and the results were compared. All these comparisons relied on both graphical comparisons of the Monte Carlo simulation results and numerical inspection of the results. For clarity, details on these method are presented below, together with the results. Second, suited statistical comparison procedures were applied. These procedures were either univariate procedures, with direct comparisons of the distances only, or multivariate procedures, comparing the whole set of parameter estimates.

Univariate procedures were of two kinds. First, the usual analysis of variance (ANOVA) and subsequent multiple comparison procedures were applied when distance estimation distributions could be considered as approximately normal

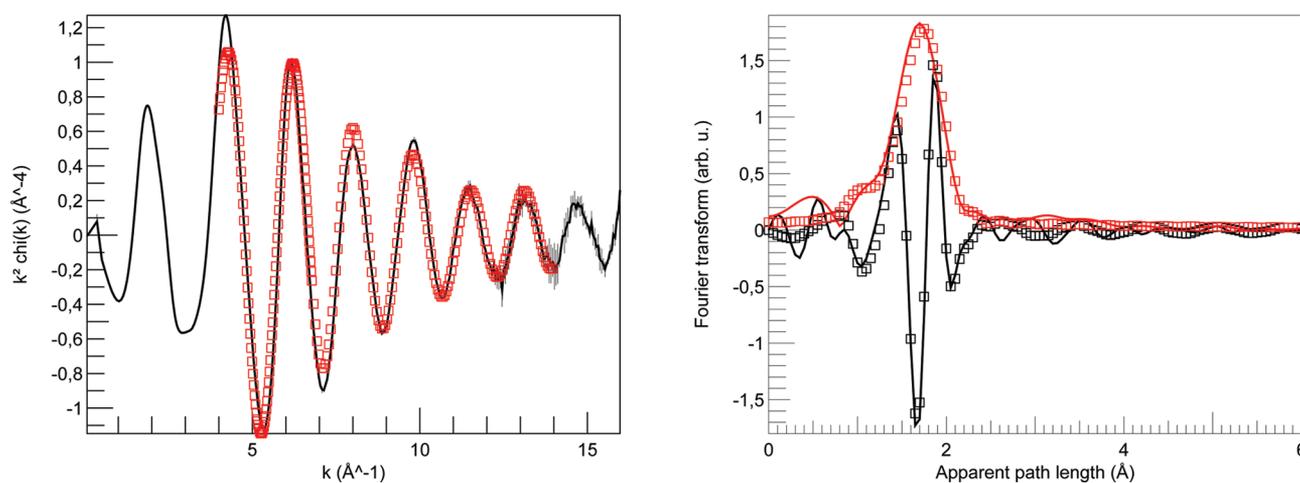


Figure 1. EXAFS (left) and its Fourier transform (right; black is the imaginary part, red is the modulus) for solid ruthenoquine. Continuous line, experiment; squares, model. EXAFS oscillations are k^2 -weighted and shown with associated experimental error bars. Fourier transform was made with a Gaussian apodization function; its error bars are shown on the imaginary part.

Table 1. Structural Parameters for Each Compound, After Fitting the Single-Shell, Single Scattering Model with Experimental Phases and Amplitudes Extracted from the Reference Compound EXAFS Spectrum on the Unfiltered EXAFS Oscillations^a

compound	fitted structural parameters						
	solid state samples (pellets); transmission mode			solution samples; fluorescence mode			
	ΔE_0 (eV)	R_C (Å)	$10^4 \sigma_C^2$ (Å ²)	solution	ΔE_0 (eV)	R_C (Å)	$10^4 \sigma_C^2$ (Å ²)
ruthocene	-0.35(10)	2.1896(7)	1.3(4)	no solution studied			
RQ	0.31(1)	2.1889(1)	-6.4(2)	H ₂ O	-0.33(111)	2.1857(82)	-0.6(5)
RQ·2HCl ^d	-0.48(10)	2.1869(8)	0.8(3)	H ₂ O ₂ 1:1 ^b	0.36(150)	2.1641(110)	6.9(60)
				H ₂ O ₂ 1:15	0.18(190)	2.1684(138)	7.0(80)
				CH ₂ Cl ₂ ^b	0.79(149)	2.1753(103)	-6.1(72)
				HEPES	0.79(232)	2.1654(162)	-3.8(83)
				HEPES + hemine 1:1	-0.78(371)	2.1796(271)	-4.1(150)
				HEPES + hemine 2:1	-6.30(419)	2.2159(321)	-5.6(145)
Me-RQ	-0.06(5)	2.1872(4)	-3.2(4)	H ₂ O	-0.34(64)	2.1857(45)	-0.6(30)
Me-RQ·2HCl ^d	-0.18(8)	2.1866(7)	-4.1(3)	H ₂ O ₂ ^c 1:1	-0.69(125)	2.1771(92)	3.4(52)
				H ₂ O ₂ ^c 1:15	-0.77(166)	2.1764(123)	8.0(70)

^aNumber in parentheses gives the standard deviation, after Monte Carlo simulation. ^bVisual comparison in Figure 3, left. ^cVisual comparison in Figure 3, right. ^dVisual comparison in Figure 4.

(parametric method). The principles of the analysis of variance, which can be seen as a generalization of the Student's T test, can be found in numerous statistics courses and will not be given here; see for instance, refs 26–28. Second, and especially when the parametric analysis of variance assumptions were not met, nonparametric analysis of the Monte Carlo results was proposed. This analysis is based on the properties of nonparametric prediction intervals.²⁹ Let us assume, as a null hypothesis, that the two sets of parameters come from the same distribution (have the same theoretical value). Then, the first sample of n values obtained by the Monte Carlo simulation can be used as a $1 - (2/(n + 1))$ prediction interval for the second set, that is, an interval that has probability $1 - (2/(n + 1))$ to contain a single value of the second set. Since values of the second set are independent, the probability that all the n' values of the second set are outside this interval (that is, the probability of no overlap between the two sets of values) is then $(2/(n + 1))^{n'}$. With $n = n' = 1000$, this is a probability of about 10^{-2699} , completely negligible. Hence, the hypothesis that parameter values for two samples are the same can be ruled out if the two Monte Carlo set of values for these two samples do not overlap. This concept can be adapted to estimate a p -value

in the case of partial overlap: assuming that the k smaller and the k larger values of the first set are omitted, the probability for the second set to completely lay outside this reduced first set is $(2(k + 1)/(n + 1))^{n'}$.

Multivariate procedures applied were the equivalents of the univariate ones, using multivariate analysis of variance (MANOVA) and Hotelling's test when data were multidimensionally normally distributed, and nonparametric approaches when otherwise, using prediction regions instead of prediction intervals. Details about multivariate procedures can be found in refs 30 and 31, for instance.

In addition, when several samples are to be compared, multiplicity correction may be required; for instance, here, with 15 samples to be compared, one should use $\alpha = (0.05/(15 \times 14/2)) \approx 4.7 \times 10^{-4}$ instead of 0.05 to perform the test (Bonferroni correction). Note that even with this correction, the complete nonoverlapping of two clouds of points still leads to reject the null hypothesis of similar parameter values.

RESULTS AND DISCUSSION

EXAFS Data Analysis. As expected, all recorded spectra lead to very similar spectra, in agreement with the common

ruthenocenic core of all compounds. The Fourier transform shows a single peak, around 1.7 Å (uncorrected for phase shift; Figure 1). Fitting the structural model using scattering paths generated by FEFF 8.2, only single scattering on the first coordination shell (10 carbon atoms at around 2.21 Å) was necessary to reproduce this first peak of the Fourier transform and the EXAFS oscillations to a good extent (Figure 1). To reproduce secondary peaks of the Fourier transform, models including multiple scattering paths were built. Such models only slightly improve the overall fit; however, they lead to an inconsistent set of structural parameters for these shells, such as abnormally high ($>0.5 \text{ \AA}^2$) or low ($<10^{-4} \text{ \AA}^2$) Debye–Waller factors or inconsistent distances. This probably results from complex interferences between various multiple scattering paths, as already observed.¹⁸ Since adding these paths makes the fit process unstable without changing the first shell parameters values, to allow comparisons between the results, all subsequent fits were made using reference phases and amplitudes to avoid the biases due to the theoretical approximations made.

This single-scattering, single-shell model was fitted to all experimental data sets. The results are given in Table 1, and these values, which differ slightly from one compound to another, are the values for which the question is raised “How to compare them?”.

Since the fluorescence mode operated on dilute solution samples leads to higher noise than the transmission mode, statistical uncertainties are, as expected, much higher for fluorescence spectra than for transmission spectra. For fluorescence spectra, a quick glance at the fitted values and at the statistical uncertainties suggests that the results are not very different; for transmission spectra, however, differences in the distances are much higher than the statistical uncertainties. The aim of the procedure presented below is to obtain more reliable answers than would be obtained through this simple comparison applied to this family of compounds but in a way that can easily be generalized to any family of closely related compounds.

Assessing the Source of Variability. There are two main sources of uncertainties in the fitted parameters: systematic errors and statistical errors. Systematic errors, or biases, are associated with any unverified assumption during the experiment recording, the data analysis procedure, or the modeling of the data. Statistical errors are basically due to the noise. They have different consequences for distances determination and comparison: systematic errors preclude obtaining a correct distance value. However, since they are deterministic, they do not change from one spectrum to another if all assumptions made are the same; that is, experiments and analyses are made under the same conditions, and the model is the same for similar enough compounds. Consequently, they tend to vanish when taking the difference between two results and do not preclude, to a certain extent, making comparisons. Conversely, statistical errors affect not the exactness of the results, but its precision; hence, they are the key factor to limiting comparisons of the distances, assuming systematic errors vanish, as said before.

Bias (Systematic Errors) Introduced by the Model Itself. The incorrectness of the fitted model itself is one of the major systematic errors in EXAFS analysis. To assess its influence, the fit of the ruthenocene pellet spectrum using experimental phases and amplitudes from ruthenocene itself was compared between the filtered and unfiltered EXAFS

oscillations. Obviously, the fit on the filtered spectrum should be almost perfect (to the precision of the computer); the difference between the two fits will then give the contribution of all model imperfections (distortions due to Fourier filtering, mainly) to the parameter value precision. This comparison leads to $\Delta(\Delta E_0) = -0.37 \text{ eV}$, $\Delta R_C = 0.00162 \text{ \AA}$ and $\Delta(10^4 \sigma_C^2) = 1.5 \text{ \AA}^2$. These results give a lower bound for the precision of the parameters; fitting using theoretical electronic parameters would lead to even higher lower bounds, since all theoretical assumptions will also limit the precision of fitted values. These bounds are well below the uncertainties for fluorescence spectra; however, they are higher than the Monte Carlo estimated statistical uncertainties for transmission spectra (Table 1). This observation is in agreement with the well-known fact that noise is not the limiting factor in transmission mode experiments^{3,4}

To note, fitted distances are shifted toward higher values ($\Delta R_C > 0$), which is coherent with the lack of the high-R contributions to the spectrum when filtering the first peak of the Fourier transform; in that case, one can expect a similar bias for any fit using the same phases and amplitudes on structurally close compounds, as with the ruthenoquinone family compounds here. To check this point, for each of the solid state samples, phases and amplitudes of the first peak were isolated from the average transmission spectra using the same procedure and were used to fit the corresponding unfiltered EXAFS oscillations. Results of these fits are given in Table 2. As can

Table 2. Fitted R Values on Unfiltered EXAFS Oscillations, Using Phase and Amplitude Extracted from the First Peak of the Corresponding Fourier Transform, with $R_C = 2.1880 \text{ \AA}$ ^a

compound	ruthenocene	RQ	RQ-2HCl	Me-RQ	Me-RQ-2HCl
fitted R_C (Å)	2.1896	2.1913	2.1905	2.1899	2.1910

^aResults are for solid state samples only, recorded in transmission mode.

be seen here, the bias is similar, and always positive, for all spectra and is around $0.0025 \pm 0.0007 \text{ \AA}$ (mean and standard deviation). This result suggests that, for estimating absolute distances, this bias is the factor limiting the precision. However, when comparing distances, the difference between the two estimated distances will almost cancel out this bias, and the precision is then $\sim 0.0007 \text{ \AA}$, rounded to 0.001 \AA , comparable to statistical uncertainties.

Bias Induced by the Extraction Process. A second source of uncertainty is the extraction process used to obtain the (unfiltered) EXAFS oscillations to be fitted. To assess its influence, for each compound, EXAFS oscillations were obtained either from the average spectrum or from each of the three recorded spectra; each of these EXAFS oscillations was fitted, and the results were compared. Typical results are shown in Figure 2. As can be seen, both for transmission and fluorescence mode, variations between individual spectra (including extraction effects) are comparable with the variability estimated by the Monte Carlo procedure. In addition, performing data reduction several times, slightly changing the data reduction parameters, led to very similar results. For this reason, they are not discussed further.

Influence of the Initial Values in the Fit Procedure. Since the EXAFS equation is strongly nonlinear relative to its parameters, the choice of the initial values may have a

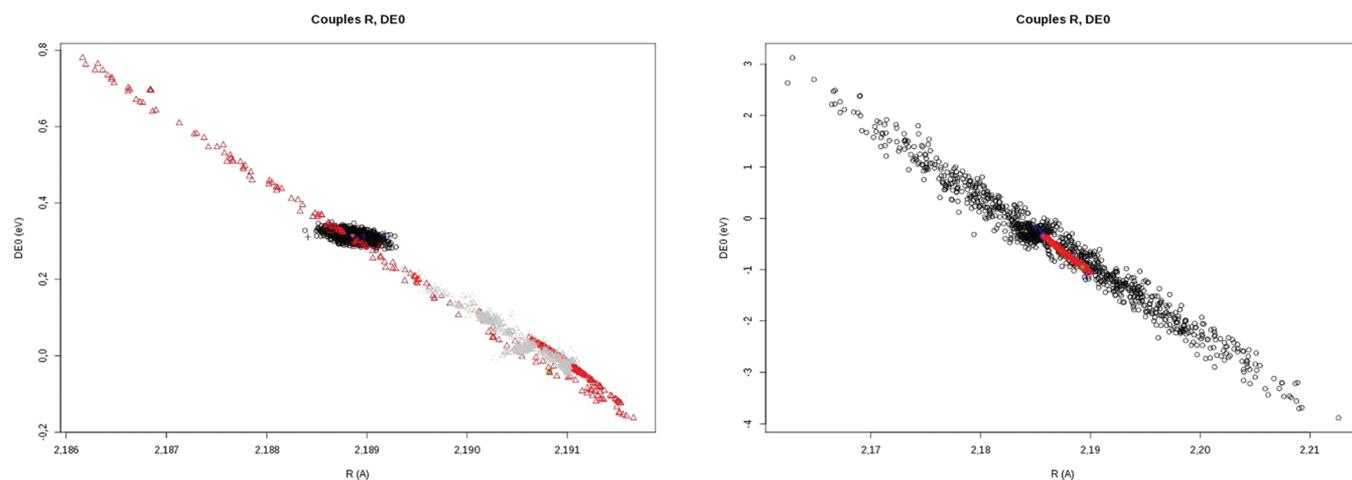


Figure 2. Fitted ΔE_0 and R_C values obtained from various procedures. Left: fit for RQ in pellet (transmission mode). Right: fit for RQ in aqueous solution (fluorescence mode). Black circles: Monte Carlo simulations for estimating statistical uncertainties. Red triangles: changing the initial values. Blue crosses: fit on individual spectra, instead of average. See text for comments.

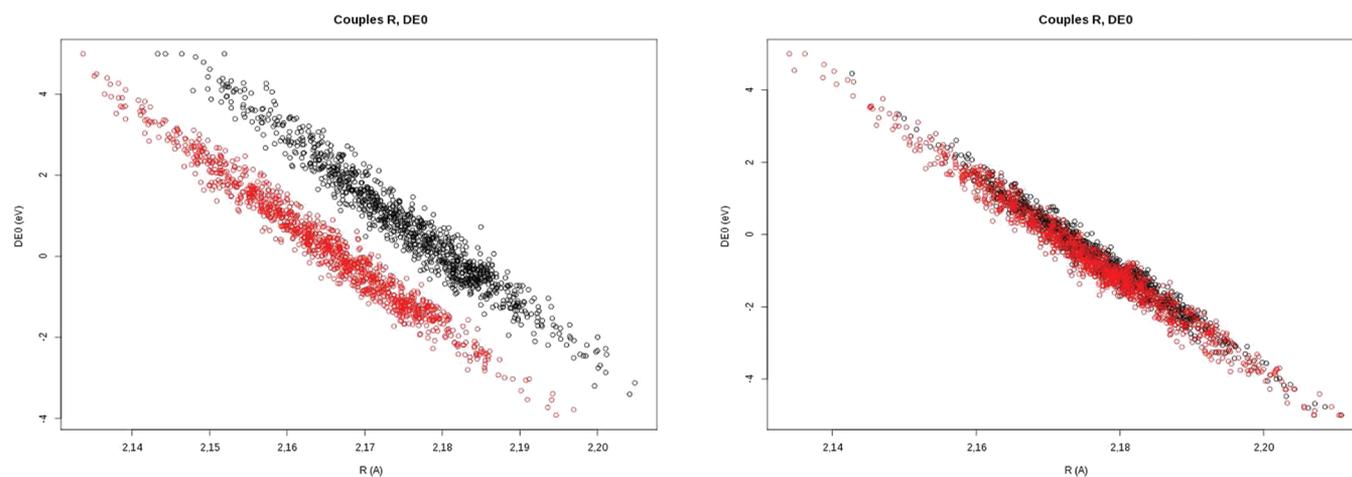


Figure 3. Comparisons of two-dimensional variability of the $(R_C, \Delta E_0)$ parameter couple obtained from Monte Carlo simulation. Left: results for RQ in CH_2Cl_2 (black) and in equimolar H_2O_2 (red). Right: results for Me-RQ in equimolar (black) and 1:15 (red) H_2O_2 .

significant influence on the fit results. To assess the importance of this choice, after each fit and using the initial values given in the method (expected to be the “best” initial values, owing to the obtention mode of phases and amplitudes), 1000 sets of alternative initial values were randomly selected in the parameter space, and the fit was performed again with these new initial values. Some of these fits were clearly worse than the reference one, with a residual sum of square higher than 10 times the reference residual sum of square; others led to obviously unrealistic distances, close to 2.5 Å. These fits (corresponding to 0–12.1% of the fits) were removed before analysis of the results.

Typical results of changing initial values, considering only valid fits, are presented in Figure 2. Here again, for fluorescence mode spectra, the effect is much lower than the statistical uncertainties estimated by the Monte Carlo procedure; for transmission mode spectra, however, results are more difficult to interpret.

It seems that the very strong correlation between ΔE_0 and R_C leads to a valley in the surface of the minimized function, $\xi^2 = \sum_{i=1}^n k_i^4 (\chi_{\text{exp}}(k_i) - \chi(k_i))^2$, where n is the number of experimental points and $\chi_{\text{exp}}(k_i)$ is the experimental spectrum. According to the choice of the initial values, the fit converges

somewhere inside the valley; the noise in the spectrum creates small bumps in the valley, preventing the algorithm from finding the “best” minimum. Values of the residual sum of squares ξ^2 are very close. To select a “correct” minimum, the behavior of the fitted parameters distribution can be a hint: for a real minimum, the fitted parameters should be at the center of the cloud of the Monte Carlo simulations, and ideally, this cloud should look like an ellipsis. This is the case here for the fitted values obtained starting from the ideal values. Making Monte Carlo simulations using initial values leading to one of these other local minima (gray circles on the figure, generated for the green fitted values) shows that, although the residual sum of squares is not statistically different, this secondary minimum is not the real one, since it does not have this ellipsoidal shape and is not even centered on the fitted values!

Statistical Comparison of the Distances. Comparison is straightforward for fluorescence mode spectra, since, as shown above, statistical uncertainties are the major source of variability in fitted values. Hence, comparisons can be made directly on the Monte Carlo simulation results. Furthermore, the cloud of Monte Carlo simulated values is very close to an ellipsis, coherent with the hypothesis of a multivariate Gaussian (normal) distribution for fitted parameters.

For transmission spectra, comparisons using statistical uncertainties should be taken with more care because of other factors cited above.

Direct Comparison. Direct comparison of distances is then straightforward using Student's *T* test for comparing two distances, or one-way analysis of variance with subsequent all pairwise comparisons when more distances are to be compared. Using this approach, the only significant differences in distances are among the solid state samples, leading to two sets of groups of compounds: (1) ruthenocene and RQ and (2) RQ·2HCl, Me-RQ, and Me-RQ·2HCl.

Solution spectra cannot be distinguished by this approach; however, the equimolar RQ-H₂O₂ spectrum is significantly different from the second group of compounds and from RQ.

These differences are consistent with small changes in the orientation of the cyclopentadienyl rings binding the ruthenium atom, the two cycles being almost parallel in ruthenocene and RQ, but being slightly tilted in RQ·2HCl, Me-RQ, and Me-RQ·2HCl. Such a structural change was observed in the crystal structure of the equivalent iron compounds, ferroquine³² and protonated ferroquine.⁷

Comparison Also Using ΔE_0 . It is well-known that the ΔE_0 and R_C values obtained by fitting are strongly correlated, and Figure 2 confirms these strong correlations. Consequently, simply comparing the distances without taking into account the changes in fitted ΔE_0 values will artificially increase the standard error on distances.

For this reason, comparisons of R_C values taking into account the ΔE_0 values were made. The importance of this correction is illustrated in Figure 3: when clouds representing each parameter variability are superimposed (as on the right), there is no difference between the fitted parameter values; when they do not superimpose (as on the left), parameters do differ, but in both cases, the projection on the two axes masks this difference. This effect is a well-known consequence of correlation in multivariate studies. To correct for this, we propose two methods.

First, the graphical comparison of the fitted parameters obtained after the Monte Carlo simulation is an easy way to distinguish between clouds of values clearly overlapping or not, as exemplified in Figure 3. However, interpretation is more difficult in the case of partial overlapping, since determining which prediction region must be used to estimate *p*-values is not straightforward.

Second, one can use a Hotelling's T^2 test to simultaneously compare the fitted parameters values for two compounds, corresponding to a two-sample MANOVA. This test is a multidimensional version of the Student's *T* test and presents the same limitations, that is, assuming normality and a similar variance-covariance matrix between the two samples. Such conditions are met for the compounds given in Figure 3: the two clouds look like ellipses (normality) with parallel axes (same correlation) and similar extension (same variances), but this is not always the case. This test should be performed on the fitted parameters obtained by fitting the model on individual spectra (testing on Monte Carlo simulation results would artificially increase the sample size; hence, the test power).

Despite limitations of the multiplicity correction, when applied to our experimental data and fit results, the Hotelling T^2 test finds the following differences:

- Me-RQ and Me-RQ·2HCl in pellets are different from Me-RQ in H₂O₂ (both 1:1 and 1:15; $p < 0.0007$), with a slightly longer distance for the latter.
- RQ with hemine in excess is different from all solid-state compounds (ruthenocene, RQ, RQ·2HCl, Me-RQ, and Me-RQ·2HCl; $p < 0.0003$).
- RQ with equimolar hemine is different from RQ in pellets ($p < 0.0002$).
- RQ in pellets is different from Me-RQ and Me-RQ·2HCl pellets and Me-RQ in H₂O₂ (both 1:1 and 1:15; $p < 0.0003$).

The graphical procedure is less conservative and more powerful, so it gives a better classification of the compounds. From this approach, it appears that

- Ruthenocene is different from all other compounds, except maybe the Me-RQ pellet, for which there is a slight overlap of the two clouds. Ruthenocene seems to have slightly larger distances.
- RQ (pellet) is also different from all other compounds, with slightly larger distances and higher ΔE_0 .
- RQ·2HCl (pellet), however, is comparable to RQ in CH₂Cl₂ and RQ and Me-RQ H₂O solution, but presents higher distances than other solutions and shorter distances than Me-RQ and Me-RQ·2HCl pellets.
- RQ in CH₂Cl₂ and in H₂O are almost identical; both are barely distinguishable from RQ with hemine (both 1:1 and 1:2). They have longer distances than RQ and Me-RQ in H₂O₂ (both 1:1 and 1:15).
- RQ and Me-RQ in H₂O₂ (both 1:1 and 1:15) cannot be distinguished and present shorter distances than Me-RQ and RQ in H₂O.
- Me-RQ (pellet) has longer distances than Me-RQ·2HCl (pellet).
- Comparison of RQ in water, RQ in H₂O₂, and RQ in HEPES shows that RQ in HEPES is similar to RQ in H₂O₂ and different from RQ in water, but distance comparison is difficult because of the high noise level of RQ in HEPES spectra. Unexpectedly, RQ in HEPES leads to results very similar to RQ in H₂O₂ with or without hemine.

Taken together, these results suggest that, in the presence of H₂O₂, something occurs in solution that tends to shorten the Ru–C distance. This would be coherent with oxidation (partial or complete) of the ruthenium atom. However, there is no net change in the Fourier transform of the RQ or Me-RQ in the presence of H₂O₂, as would have been expected in the case of the formation of a Ru–Ru bond, since Ru is a heavy atom that would lead to an important EXAFS signal. This suggests that this dimerization either does not occur with H₂O₂ or that only a small amount of RQ is oxidized in such conditions.

Whereas spectra of RQ in HEPES with hemine would suggest an oxidation of RQ by the hemine, the unexpected result that RQ would be also oxidized in HEPES alone precludes this interpretation. In fact, a few papers mention that HEPES is able to generate H₂O₂ when exposed to light,^{33,34} but that this effect would be mediated via a coupling with riboflavin present in cell culture media.³⁵ There is no riboflavin in our solutions; however, the irradiation with the X-ray beam is much more intense than a usual source light. Moreover, ruthenoquine has several aromatic rings, especially the aminoquinolone part, which is known to absorb and fluoresce in visible light. Hence, we assume that the results observed for RQ in HEPES are the

consequence of H_2O_2 generation by HEPES under X-ray beam, maybe by coupling with the aminoquinolone part of the ruthenoquinone itself.

In addition, quite interestingly, RQ-2HCl, Me-RQ, and Me-RQ-2HCl exhibit a clear bimodal pattern for the $(R_C, \Delta E_0)$ distribution (Figure 4) that was not found for any other fit but

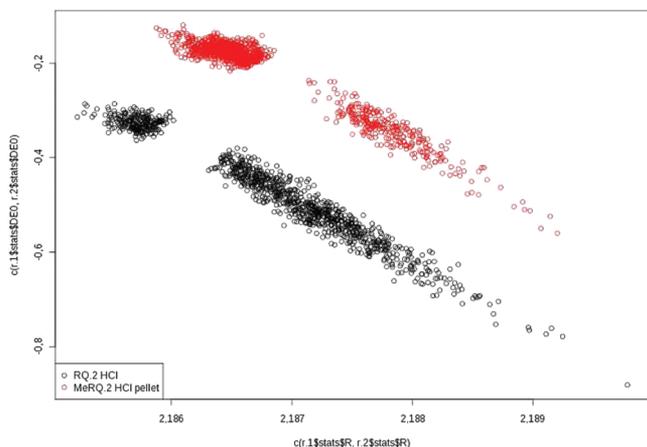


Figure 4. Fit results and Monte Carlo simulations for RQ-2HCl (black) and Me-RQ-2HCl (red). A bimodal pattern clearly exists. Despite this, these distributions clearly do not overlap, suggesting a longer distance and smaller ΔE_0 for Me-RQ-2HCl.

has the same shape for these three compounds. Previous work on the interpretation of such bimodal patterns in Monte Carlo simulations suggests that there may be a combination of two distances in these compounds.¹⁸ Hence, this bimodal pattern might also be a consequence of the tilt of the two cyclopentadienyl ligands, leading to two sets of distances, too close to be distinguished by using two different scattering paths, but different enough to let the fit converge toward one or the other according to the noise pattern.

CONCLUSION

Precise determination and comparisons of distances between a metal atom and its first coordination sphere are often some of the aims of X-ray absorption studies, with a generally admitted rule of thumb that the precision is around 0.01 Å. In this work, we developed an approach to estimate the influence of several factors on the precision of these distance determinations and comparisons.

In fluorescence mode, the main factor limiting the precision of distance determination is the noise level of the data. Consequently, the most suited methods for estimating the precision and analyzing comparisons in such experimental conditions are statistical methods, including analysis of Monte Carlo simulation results.

In transmission mode, however, noise level of the data is much less important, leading to statistical estimates of distance precision almost ten times smaller than the 0.01 Å rule of thumb. However, as was shown, under these conditions, the choice of the initial value in the fit procedure and the limitations of the model used to fit the data lead to important bias in the results. The influence of initial values may depend on the specific choice of the algorithm used to minimize the data. As was exemplified, analysis of the repartition of Monte Carlo simulation results around the minimum may help to distinguish the true minimum from spurious ones. Model

limitations are, indeed, the limiting factor for distance determination, as with our example, a limit of 0.003 Å. However, for distance comparisons, this effect does not play the major role. Consequently, for distance comparisons, the same statistical methods as above are also adapted.

The main contribution of this work is the introduction of simultaneous comparison of distances and edge energy correction using either Hotelling's T^2 test or graphical comparison of Monte Carlo results. As we have shown, using such methods strongly enhances the ability to detect changes in distances as small as 0.001 Å as a consequence of taking into account the strong correlation between ΔE_0 and the distance. Tools for applying these methods are included in our LASE software¹⁵ for the Monte Carlo simulations and as a template R script, available from the author, for further analyses.

Applications of the methodological developments to ruthenoquinone and related compounds suggest that the nonbinding electron doublet of the nitrogen atom linking the aminoquinolone to the ruthenoquinium may be involved in the orientation of the two cyclopentadienyl groups. Indeed, either its protonation or its methylation lead to changes in distances that may be related to an increase of the tilt between the plane of the two rings. In addition, results suggest a partial oxidation of the ruthenium atom in aqueous solution in presence of H_2O_2 , either directly introduced or generated in situ in HEPES solutions.

AUTHOR INFORMATION

Corresponding Author

*E-mail: emmanuel.curis@parisdescartes.fr.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

These experiments were carried out thanks to the acceptance of two proposals at SOLEIL: 20090133 and 20090602. We thank the team of the SAMBA beamline at SOLEIL, especially Stéphanie Belin and Valérie Briois, for their help in these experiments.

REFERENCES

- (1) Sayers, D. E. *Report of the International XAFS Society Standards and Criteria Committee*. International XAFS Society, 2000; http://ixs.iit.edu/subcommittee_reports/sc/.
- (2) IXS Standards and Criteria Committee. *Report of the International XAFS Society Standards and Criteria Committee*; 2000; http://ixs.iit.edu/subcommittee_reports/sc/.
- (3) Purans, J.; Dalba, G.; Fornasini, P.; Kuzmin, A.; Panfilis, S. D.; Rocca, F. *AIP Conf. Proc.* **2007**, 422–424.
- (4) Pettifer, R. F.; Mathon, O.; Pascarelli, S.; Cooke, M. D.; Gibbs, M. R. H. *Nature* **2005**, 435.
- (5) Ruffoni, M. P.; Pettifer, R. F.; Pascarelli, S.; Mathon, O. *J. Synchrotron Radiat.* **2007**, 14, 169–172.
- (6) Dalba, G.; Fornasini, P.; Grisenti, R.; Purans, J. *Phys. Rev. Lett.* **1999**, 82, 4240–4243.
- (7) Dubar, F.; Egan, T.; Pradines, B.; Kuter, D.; Ncokazi, K.; Forge, D.; Paul, J.; Pierrot, C.; Kalamou, H.; Khalife, J.; Buisine, E.; Rogier, C.; Vezin, H.; Forfar, I.; Slomianny, C.; Trivelli, X.; Kapishnikov, S.; Leiserowitz, L.; Dive, D.; Biot, C. *ACS Chem. Biol.* **2011**, 6, 275–287.
- (8) Biot, C.; Glorian, G.; Maciejewski, L. A.; Brocard, J. S.; Domarle, O.; Blampain, G.; Millet, P.; Georges, A. J.; Abessolo, H.; Dive, D.; Lebibi, J. *J. Med. Chem.* **1997**, 40, 3715–3718; PMID: 9371235.
- (9) Dubar, F.; Bohic, S.; Slomianny, C.; Morin, J.; Thomas, P.; Kalamou, H.; Guéardel, Y.; Cloetens, P.; Khalife, J.; Biot, C. *Chem.*

Commun. (Camb.) **2012**, *48*, 910–912 ; Epub 2011 Dec 5. PMID: 22143053.

(10) Supan, C.; Mombo-Ngoma, G.; Dal-Bianco, M.; Ospina Salazar, C.; Issifou, S.; Mazuir, F.; Filali-Ansary, A.; Biot, C.; Ter-Minassian, D.; Ramharther, M. *Antimicrob. Agents Chemother.* **2012**, *56* (6), 3165–3173.

(11) Trupia, S.; Nafady, A.; Geiger, W. E. *Inorg. Chem.* **2003**, *42*, 5480–5482.

(12) Watanabe, M.; Sano, H. *Chem. Lett.* **1991**, 555–558.

(13) Beagley, P.; Blackie, M. A. L.; Chibale, K.; Clarkson, C.; Moss, J. R.; Smith, P. J. *Dalton Trans.* **2002**, *23*, 4426–4433.

(14) Curis, E.; Osán, J.; Falkenberg, G.; Bénazeth, S.; Török, S. *Spectrochim. Acta B* **2005**, *60*, 814–849.

(15) <http://xlase.online.fr>, 2011.

(16) Curis, E. Développement d'outils informatiques et statistiques pour l'analyse des spectres EXAFS — Application aux systèmes bioinorganiques. Ph.D. Thesis, Université Paris-Sud, 2000.

(17) Curis, E.; Bénazeth, S. *J. Synchrotron Radiat.* **2000**, *7*, 262–266.

(18) Curis, E.; Bénazeth, S. *J. Synchrotron Radiat.* **2005**, *12*, 361–373.

(19) Hardgrove, G. L.; Templeton, D. H. *Acta Crystallogr.* **1959**, *12*, 28–32.

(20) Seiler, P.; Dunitz, J. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 2946.

(21) Borissova, A.; Antipin, M. Y.; Perekalin, D. S.; Lyssenko, K. A. *CrystEngComm* **2008**, *10*, 827.

(22) Beagley, P.; Blackie, M. A. L.; Chibale, K.; Clarkson, C.; Moss, J. R.; Smith, P. J. *J. Chem. Soc. Dalton Trans.* **2002**, 4426–4433.

(23) Rehr, J. J.; de Leon, J. M.; Zabinsky, S.; Albers, R. *J. Am. Chem. Soc.* **1991**, *113*, 5135–5140.

(24) http://leonardo.phys.washington.edu/feff/html/feff_team.html, 2009.

(25) R Development Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2011; ISBN 3-900051-07-0.

(26) Scheffé, H. *The Analysis of Variance*; Wiley Interscience, 1999.

(27) Sokal, R. R.; Rohlf, F. J. *Biometry*, 3rd ed.; W. H. Freeman and Company: New York, 1995; Chapter 8, p 179.

(28) Mickey, R. M.; Dunn, O. J.; Clark, V. A. *Applied Statistics: Analysis of Variance and Regression*, 3rd ed.; Wiley Series in Probability and Statistics; Wiley Interscience: New York, 2004.

(29) Guttman, I. *Statistical Tolerance Regions: Classical and Bayesian*; Griffin Statistical Monographs and Courses 26; Griffin: London, 1970.

(30) Rencher, A. C. *Methods of Multivariate Analysis*, 2nd ed.; Wiley Series in Probability and Statistics; Wiley: New York, 2002.

(31) Mardia, K. V.; Kent, J. T.; Bilby, J. M. *Multivariate Analysis; Probability and Mathematical Statistics: A Series of Monographs and Textbooks*; Academic Press: New York, 2003.

(32) Beagley, P.; Blackie, M. A. L.; Chibale, K.; Clarkson, C.; Meijboom, R.; Moss, J. R.; Smith, P. J.; Su, H. *Dalton Trans.* **2003**, *15*, 3046–3051.

(33) Lepe-Zuniga, J. L.; Zigler, J., J. S.; Gery, I. *J. Immunol. Methods* **1987**, *103*, 145.

(34) Zigler, J., J. S.; Lepe-Zuniga, J. L.; Vistica, B.; Gery, I. *In Vitro Cell. Dev. Biol.* **1985**, *21*, 282–287.

(35) Mahns, A.; Melchheier, I.; Suschek, C. V.; Sies, H.; Klotz, L.-O. *Free Radical Res.* **2003**, *37*, 391–397.