# X-ray Absorption Spectroscopic Studies of Zinc in the N-Terminal Domain of HIV-2 Integrase and Model Compounds £

Martinus C. Feiters,[1]* Astrid P. A. M. Eijkelenboom,[2,#] Hans-Friedrich Nolting,[3] Bernt Krebs,[4] Fusinita M. I. van den Ent,[5,¥] Ronald H. A. Plasterk,[5,$] Robert Kaptein,[2] and Rolf Boelens[2]

1) Department of Organic Chemistry, NSR Center, University of Nijmegen, Toernooiveld, NL-6525 ED Nijmegen, The Netherlands

2) Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, NL-3584 CH Utrecht, The Netherlands

3) European Molecular Biology Laboratory Outstation at DESY, Notkestrasse 85, D-22603 Hamburg, Germany

4) Institute of Inorganic Chemistry and Institute of Biochemistry, Unversity of Münster, Wilhelm-Klemm-Strasse 8, D-49149 Münster, Germany

5) Division of Molecular Biology, The Netherlands Cancer Institute, Plesmanlaan 121, NL-1066 CX Amsterdam, The Netherlands

£ Dedicated to Prof. C. David Garner on the occasion of this 60th birthday (November 9, 2001).

# present address: Unilever R&D Vlaardingen, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

¥ present address: MRC Laboratory of Molecular Biology, Hills Road, CB2 2QH, Cambridge, UK.

$ present address: Hubrecht Laboratory, Uppsalalaan 8, NL-3584 CT Utrecht, The Netherlands

## Abstract

X-ray absorption spectroscopy (XAS), including Extended X-ray Absorption Fine Structure (EXAFS) and XANES (X-ray Absorption Near Edge Structure), has been carried out at the Zn K edge of the N-terminal part of the integrase protein of the Human Immunodeficiency Virus, type 2 (HIV-2), and of some zinc coordination compounds. In the presence of excess β-mercaptoethanol, which was present in the NMR structure elucidation of the protein (Eijkelenboom *et al.* (1997) *Current Biol.* **7**, 739-746; Eijkelenboom *et al.* (2000) *J. Biomol. NMR* **18**, 119-28), the protein spectrum was nearly identical to that recorded in its absence. Comparison of the XANES of the protein to that of model compounds and literature data permits the conclusion that the Zn ion is four-coordinated. The major shell of the EXAFS gives evidence for a mixed (N or O as well as S) coordination sphere, while the minor shells indicate imidazole coordination. Our approach to the analysis of the EXAFS, including quantification of the imidazole by multiple scattering simulations with EXCURV92, was validated on the model compounds. An important result is that with multiple scattering simulations using restraints on the parameters of the imidazole rings, the number of imidazoles as well as their orientation could be determined. The integrase spectra can be fitted with 2 sulfur ligands at 2.26 Å (Debye-Waller-type factor 0.009 $Å^2$) and 2 imidazole ligands with the nitrogens at 1.99 Å (Debye-Waller-type factor 0.005 $Å^2$). The XAS-derived geometry is fully consistent with that found in the NMR structure determination and, allowing for the volume contraction due to the temperature difference between the experiments, justifies the restraints applied in the structure calculation (Zn-S and Zn-N distances of 2.3 and 2.0 Å, respectively).

## Introduction

Integration of a DNA copy of the viral RNA into the genome of the infected cell is an important step in the infectivity cycle of the human immunodeficiency virus (HIV). This integration process is catalysed by the viral enzyme integrase (for reviews see Brown, 1997; Hansen *et al.*, 1998). Integrase (IN) can be split by partial proteolysis in three functional domains that are amenable to structure elucidation by X-ray crystallography and high-resolution 2D and 3D NMR techniques. In the last years, structures of mutants of the catalytic core (X-ray crystallography, Dyda *et al.*, 1994; Goldgur *et al.*, 1998; Maignan *et al.*, 1998; Greenwald *et al.*, 1999) and of the C-terminal DNA binding domain (NMR, Eijkelenboom et al., 1995; Lodi et al., 1995; Eijkelenboom *et al.*, 1999) of HIV-1 integrase have been reported. Recently, a crystal structure of an HIV-1 IN fragment containing both the catalytic core and the C-terminal domain has been solved (Chen *et al.*, 2000). The N-terminal (55 amino acids) part of HIV integrase contains a Zn binding HHCC ($HisX_3HisX_{23}CysX_2Cys$) motif, which does not show homology with classical

zinc-finger sequences (Klug & Rhodes, 1987; Kaptein, 1991; Berg & Shi, 1996). The NMR structures of this domain of HIV-1 integrase (Cai *et al.*, 1997) and of HIV-2 integrase (Eijkelenboom *et al.*, 1997; Eijkelenboom *et al.*, 2000) have been determined by high resolution 2D and 3D NMR techniques. Interestingly, in the course of this determination, the three-dimensional structure turned out to be a novel fold for a zinc-binding protein. Therefore, the need was felt for an independent determination of the Zn coordination; in addition, this would be an opportunity to probe the effect of the non-physiological reductant β-mercaptoethanol, which was present in the NMR samples.

Zn proteins are accessible by few spectroscopic techniques and their study by X-ray absorption spectroscopy is well-documented (Cramer, 1988; Feiters *et al.*, 1986; Garner & Feiters, 1987; Hasnain & Garner, 1987; Feiters, 1990; Clark-Baldwin *et al.*, 1998). There are various studies of zinc-finger type proteins (Diakun *et al.*, 1986; Summers *et al.*, 1992) and recent examples of Zn XAS studies include carbonic anhydrase (Bracey *et al.*, 1994; Alber *et al.*, 1999), metallothionein (Jiang *et al.*, 1994) and the related neuronal growth inhibitory factor (Bogumil *et al.*, 1998), methionine synthase (González *et al.*, 1996; Peariso *et al.*, 1998; Zhou *et al.*, 1999), other methyl transferases (Gencic *et al.*, 2001; Krüer *et al.*, 2002), and Zn-containing ferredoxin (Cosper *et al.*, 1999). The XANES (X-ray Absorption Near Edge Structure) part of the spectrum gives information on valence state and coordination geometry, while the Extended X-ray Absorption Fine Structure (EXAFS) part provides ligand distances and coordination numbers with relatively high (± 0.02 Å) and low (± 20 %) accuracy, respectively. Furthermore, the phase relationship between EXAFS and its Fourier transform together with the amplitude envelope of the EXAFS permit determination of the type of ligand Z (accuracy ± 1). As the coordination geometry derived from XANES is related to the coordination number provided (with low accuracy) by EXAFS, e.g. tetrahedral geometry with 4-coordination, and octahedral geometry with 6-coordination, the two parts of the spectrum are considered to give complimentary information.

Because of its potential to discriminate between coordination of S or low-Z (C, N, O) ligands, it has been proposed that EXAFS can be used to classify zinc sites according to ligand environment (Garner & Feiters, 1987; Hasnain & Garner, 1987; Feiters, 1990, Table 1), viz. type A sites with only S ligands (like the 'structural' site in liver alcohol dehydrogenase, Eklund *et al.*, 1976, Eklund & Brändén, 1983), type B with mixed (S as well as N or O) ligation (e.g. the 'catalytic' site of liver alcohol dehydrogenase), and type C with only N/O ligands (with carbonic anhydrase (Liljas *et al.*, 1972), thermolysin (Colman *et al.*, 1972) and superoxide dismutase (Richardson *et al.*, 1975) as early examples). Because of the association with the LADH sites it is tempting to consider type B (and C) sites as catalytically active zinc, and type A as only structural

zinc (Garner & Feiters, 1987); the lower number of S ligands when going from type A via type B to type C sites would account for increasing Lewis acidity of the zinc and hence mean more catalytic power. However, the type C site in superoxide dismutase (Richardson *et al.*, 1975; Blackburn *et al.*, 1983; Murphy *et al.*, 1996) nor the type B sites in e.g. transcription factors have catalytic activity, and in the current mechanism of purple acid phosphatase (Klabunde *et al.*, 1996), the type C zinc in the heterodinuclear Fe, Zn site in that enzyme only acts as a phosphate binding site while hydroxide anion is activated for nucleophilic attack on the nearby iron ion. Moreover, recent structural and mechanistic studies on cobalamin-independent methione synthase (González *et al.*, 1996; Peariso *et al.*, 1998; Zhou *et al.*, 1999) have revealed the power of Zn(II) ions with negatively charged ligands, including both carboxylates and thiolates, to increase the nucleophilicity of thiol. As to the potential of EXAFS to classify Zn sites, recent fundamental work (Clark-Baldwin *et al.*, 1998) has shown that care has to be taken with the possible false-positive identification of low numbers of the relatively weak low-Z (C, N, O) scatterers in the presence of larger numbers of sulfurs.

The accuracy of the distance information can be of help in the refinement of metal sites in crystal structures of metalloproteins to subatomic resolution (Hasnain & Hodgson, 1999), and this notion will undoubtedly link X-ray absorption spectroscopy of metalloproteins very strongly to X-ray diffraction. Nevertheless, it is expected that the independent determination of metal environments by EXAFS, as well as its use to complement information obtained from other than crystallographic sources, like computer predictions based on amino acid sequences and structures obtained from high-resolution 2D and/or 3D NMR studies, will continue to be important. In the past we have given an example of an EXAFS study on sheep liver sorbitol dehydrogenase (Feiters & Jeffery, 1989) to compliment a structure prediction (Eklund *et al.*, 1985) based on alignment of the amino acid sequence to the largely homologous horse liver alcohol dehydrogenase, for which the crystal structure was known (Eklund *et al.*, 1976, Eklund & Brändén, 1983). Comparison of the sorbitol dehydrogenase Zn EXAFS (Feiters & Jeffery, 1989) to that of a transcription factor (Diakun *et al.*, 1986) as well as independent simulation of the data corroborated the proposal that compared to the catalytic Zn site in liver alcohol dehydrogenase, one of the sulfur ligands to Zn is replaced by a low-Z ligand in sorbitol dehydrogenase.

The present paper deals with an EXAFS study carried out to complement an alternative structure determination, i.e. by high-resolution NMR, as well as with the development of protocols for EXAFS analysis which are checked on model compounds.

## Experimental

***Protein purification and sample preparation.*** The production and purification of the 55-residue N-terminal domain of HIV-2 integrase was performed as described previously (Eijkelenboom *et al.*, 1997). The obtained lyophilized protein was dissolved in 175 µl $H_2O$ with 1.1 molar equivalent of $ZnCl_2$ either in the presence or absence of 2 mM β-mercaptoethanol, and the pH was adjusted to 6.6. The final protein concentration was in the range 4-5 mM.

***EXAFS measurements***. EXAFS measurements at the Zn K edge were carried out in the European Molecular Biology Laboratory (EMBL) Outstation in the Hamburg Synchrotron Laboratory (HASYLAB) at the Deutsches Synchrotron (DESY) in Hamburg, Germany. During data collection, the storage ring DORIS III was operated at 4.5 GeV in the dedicated mode with ring currents between 80 and 40 mA. The EXAFS station features an order sorting monochromator, which was set at 50 % of peak intensity to suppress harmonics (Hermes *et al.*, 1984), a CANBERRA 13 element solid-state fluorescence detector, and an energy calibration device (Pettifer & Hermes, 1985). Typically 20 scans per sample were taken. The sample was kept at 20 K during the measurements and moved in between scans so that the part of the sample that was exposed to the beam was varied as much as possible. No spectroscopic differences between scans were observed. The preparation and measurements of the Zn model compounds **1-5** (cf. Chart 1 for schematic structures) were described earlier (Eggers-Borkenstein *et al.*, 1989).

***EXAFS data reduction and analysis.*** Data reduction was carried out at the EMBL and at the CAOS-CAMM Center, University of Nijmegen, with the EMBL Outstation data reduction package (Nolting & Hermes, 1992) inluding the energy calibration programmes CALIB and ROTAX, the averaging programme MEAN, and the background subtraction programme EXTRACT. Simulations of the calibrated, averaged, and background-subtracted EXAFS were carried out on the CLRC Daresbury Laboratory (Warrington, UK) dedicated EXAFS computer using the EXAFS simulation progamme EXCURV92 (Gurman *et al.*, 1984, 1986; Binsted *et al.*, 1991), including the programme MUFPOT for the ab initio calculation of phase shift and backscattering factors. Model compound crystallographic coordinates were retrieved from the Cambridge Crystallographic Data Base. Tables of distances to Zn and angles were obtained with the programme XTAFS (courtesy of R. de Gelder) for compounds **1**, **3**, **4**, and **5**, whereas data for **2** were taken directly from the literature (Klug, 1966).

## Results and Discussion

In Figure 1, the XANES (Fig. 1a), $k^3$-weighted EXAFS (Fig. 1b), and phase-corrected Fourier transform (Fig. 1c) of the integrase in presence and absence of β-mercaptoethanol are compared to a number of model compounds (Eggers-Borkenstein *et al.*, 1989; Eggers-Borkenstein, 1989) which have been shown by crystallography to have four-coordinated zinc, viz. zinc(II) ethylxanthate (**1**, Ikeda & Hagihara, 1966), zinc(II) dimethyldithiocarbamate (**2**, Klug, 1966), tetrakis(imidazole)zinc(II) diperchlorate ($[Zn(im)_4](ClO_4)_2$, **3**, Bear *et al.*, 1975), and zinc(II)(imidazole)$_2$(acetate)$_2$ ($(Zn(im)_2(acetate)_2$, **4**, Horrocks et al., 1982). The XANES (Fig. 1a) of the integrase data resembles the model compounds shown in some aspects, viz. the presence of multiple edge features and the relatively low intensity (normalized edge step < 1.5). It is different from that of other model compounds (Eggers-Borkenstein *et al.*, 1989; Eggers-Borkenstein, 1989) which have strong 'white lines' resulting in typical normalized edge steps > 1.5 and even > 2 for five- and six-coordinated zinc, respectively (not shown). The XANES spectra therefore indicate that the zinc ion in the N-terminal part of HIV-2 integrase is four-coordinated. The presence of β-mercaptoethanol did not have much influence on this part of the spectrum. For Cu(I) and Cu(II) ions, XANES features can be explained by electronic transitions, (Kau *et al.*, 1987) but for the Zn(II) ion this is not straightforward as the 3d shell is completely filled with electrons. In recent years approaches have been developed to simulate the XANES region of the X-ray absorption spectra by analysing the multiple scattering of electron waves between all atoms (intra- as well as inter-ligand scattering) around the absorber, (Della Longa et al., 1995; de Groot, 2001; Benfatto & Della Longa, 2001) but an exploration of these approaches is outside the scope of this paper.

In the Fourier transform of the EXAFS, taken over the full measured energy range (0-980 eV, 2 $Å^{-1}<k<16$ $Å^{-1}$) to enhance the resolution in the radial distribution function (Fig.2, Clark-Baldwin *et al.*, 1998), we observe a major peak at approx. 2.0-2.3 Å which shows clear evidence of its mixed character with N and S contributions at 2.0 and 2.3 Å, respectively, and relative intensities which are in agreement with equivalent numbers of N and S ligands when the difference in backscattering amplitude (Feiters *et al.*, 1990) is taken into account. Minor shells at 3 and 4 Å are also observed which give evidence for imidazole coordination. There is a small but significant effect of β-mercaptoethanol on the integrase Zn EXAFS (Fig.2), viz. at high k (9-16 $Å^{-1}$) but also at 5 $Å^{-1}$ in the EXAFS, and in the 'S' peak and the 3 Å imidazole C peak in the Fourier transform. These differences, however, are not found to correspond to major changes in the coordination, as the analysis (see below) shows that the data can be simulated with the same parameters, requiring only small adjustments of some of the Debye-Waller-type factors.

Analysis of the EXAFS started with phaseshift calculations by the MUFPOT subroutine of the simulation programme EXCURV92 (Binsted *et al.*, 1991). It had been shown previously (Eggers-Borkenstein *et al.*, 1989) that the major shells of the data could be satisfactorily analysed with either empirical or theoretical (McKale *et al.*, 1988) amplitude and phase functions. For the present simulations, however, we chose EXCURVE because of the possibility to simulate and iteratively refine the multiple scattering of imidazole rings with restraints (Binsted *et al.*, 1992), and because of the successful earlier application (EXCURV88, Förster *et al.*, 1996) in the simulation of Zn imidazole EXAFS spectra. The results of the present calculations were tested on the model compounds zinc(II) ethylxanthate **1** and tetrakis(imidazole)zinc(II) diperchlorate **3**, which feature Zn-sulfur (Ikeda & Hagihara, 1966) and -imidazole (Bear *et al.*, 1975) coordination, respectively. In line with the results of earlier studies on Cu-pyridine complexes (Feiters *et al.*, 1999), the iteratively refined simulations with phaseshifts calculated with the exchange potential and the ground state set to 'Xα' were preferred to those calculated with the Hedin-Lindqvist/von Bart settings, as resulting parameters gave better agreement with the crystallographic values. The contribution of the imidazole group was simulated with multiple scattering (Strange *et al.*, 1986, 1987; Pettifer *et al.*, 1986; Feiters *et al.*, 1988), and the distance, occupancy, angle and threshold energy values were iteratively refined with the interatomic distances of the imidazole ring as derived from the crystal structure of **3** (Fig. 3, Bear *et al.*, 1975) as a restraint (Binsted *et al.*, 1992). Optimum agreement for the coordination numbers for both model compounds was obtained with the amplitude reduction factor (AFAC) set to 0.75, which led to 3.9 sulfurs at 2.33 Å for **1**, and 4.4 imidazoles at 1.97 Å for **3**, which is in excellent agreement with the crystallographic values of 4 at 2.36 and 4 at 2.00 Å, respectively, considering the possibility of a slight volume contraction due to the lower temperature of the EXAFS measurements. In the final simulations for the Zn-sulfur model compounds **1** and **2**, contributions for the minor shells were also included, and $\Delta E_0$, distances, occupancies, and Debye-Waller-type factors were iteratively refined. For inclusion of each additional shell to be justified, it was required that the fit index was lowered by more than 5 % (Joyner *et al.*, 1987). As can be seen in Fig. 4 and Table 2, the agreement between EXAFS and crystallography was excellent for the major shells (agreement in occupancy <10 %, distances underestimated by only ± 0.02 Å). The results for the minor shells have to be considered as no more than a qualitative indication that certain types of atoms are present at a certain distance, because the agreement for the distances was reasonable (± 0.05 Å for **1**, ± 0.1 Å for **2**), whereas that for the occupancies was disappointing.

In the case of the imidazole compounds **3**, **4**, and di-(L-histidino)-zinc(II) dihydrate ($[Zn(his)_2](H_2O)_2$, **5**, Kretsinger *et al.*, 1963) (Fig. 5, Table 3), the multiple scattering was essential for the reproduction of the characteristic spectral features with crystallographically reasonable

parameters, in particular of the 'camel back' feature at 4-5 $Å^{-1}$ (Bordas *et al.*, 1983) which is strongest for **3**. It is of interest to inspect the final geometries of the imidazole rings after refinement, which are displayed as insets in the bottom panels of Fig. 5, and to check the final parameters against the crystallographic values in Table 3. In initial refinements, applying the 'single' (bold) restraints from Fig. 3, the Debye-Waller-type factors were allowed to float in order to obtain a reasonable estimate for their values. The values obtained for the carbon contributions at approx. 3 Å were then averaged and used in further refinements, so that they would remain the same for both carbons; the same procedure was followed for the carbon and nitrogens at just above 4 Å. In spite of the restraints applied, this approach resulted in a slight distortion of the geometry of the imidazole unit as judged from the distances to zinc (Fig. 5a, Table 3). However, the deviations in the angles were relatively small (< 2 degrees) and it can be concluded that the 'symmetric' coordination of the imidazole ring with respect to the Zn-N vector (i.e. small differences between angles with respect to Zn-N for equivalent contributions, viz. -126.03/+126.01 for the carbons at approx. 3 Å, and –161.54/+161.63 for the carbon/nitrogen at just above 4 Å) which is found in the crystal structure, is reproduced in the analysis of the EXAFS data. It is interesting to note that the symmetry of each individual imidazole unit and the similarity between the imidazole moieties are much larger in the room-temperature crystal structure of **3** (Bear *et al.*, 1975) than in another well-studied metal imidazole complexes with non–coordinating counterions, viz. $[Cu(imidazole)_4](perchlorate)_2$ (Ivarson, 1973) and $[Cu(imidazole)_4](nitrate)_2$ (MacFadden *et al.*, 1976), but that the higher order is not reflected in significant narrowing of the corresponding EXAFS signals (Strange *et al.* 1987; Feiters *et al.*, 1988); this probably means that the asymmetry/disorder present in the Cu complexes at room temperature is decreased at lower temperatures.

For both other Zn imidazole complexes the crystal structures (Horrocks *et al.*, 1982; Kretsinger *et al.*, 1963) show that significant deviations from symmetry around the Zn-N vector exist, and with the approach outlined above for the simulation of the EXAFS, the different angles are reproduced to within 3 degrees for **4,** and 7 degrees for **5** (Fig. 5b-c; Table 3). For compound **4**, the difference in the distances from Zn to the carbons at approx. 3 Å is so large (2.93 vs. 3.08 Å) that the corresponding EXAFS waves interfere destructively so as to virtually wipe out their contribution to the Fourier transform (Fig. 5b, bottom panel). The EXAFS result (Table 3) that the non-imidazole ligands in both systems coordinated at shorter (acetate oxygens in **4**) and longer (amino nitrogens in **5**) distance, respectively, is also in agreement with the respective crystal structures, although the deviations (0.03 Å for **4**, 0.05 Å for **5**) are a little larger than is typical for first shell contributions. In this context it is of interest to consider the fitting errors (given by EXCURV92 as $2\sigma$ when the fit index in the refinement approaches a minimum) although this is difficult in view of the application of restraints in the refinement. For our simulations of **3**, the

fitting error for the main shell distance is much smaller than the other errors leading to deviations from crystallographic values, like inaccuracies in the phaseshift calculations, in agreement with the analysis of [Cu(imidazole)$_4$] (nitrate)$_2$ (Binsted *et al.*, 1992). For the simulations of **4** and **5**, however, the fitting errors for the mains shell distances were much larger and are the main factor contributing to deviations from crystallographic values. The larger fitting errors for the simulations of compounds **4** and **5** as compared to **3** are due to the correlations between the two independent contributions to the main shell (imidazole nitrogens together with acetate oxygens for **4**, and imidazole nitrogens together with amine nitrogens for **5**) and to the lack of freedom for the distance parameters to move in the direction given by the decreasing fit index because of the restraints imposed by the geometry of the imidazole ring. We have confidence in the outcome of the restrained refinement in spite of the relatively large fitting errors, because of the correct analysis of both the relative distances of imidazole and non-imidazole contribution as well as the geometry of the imidazole ring, and note that even an error of 0.05 Å in a first shell distance means that the EXAFS result is still a valuable contribution toward improving the resolution of a protein structure determined by NMR or single crystal X-ray diffraction. The success in correctly analysing the relative order of the distances to Zn in spite of the fact that the contributions are not resolved is due to the use of the distance information contained in the minor shells for the imidazole, together with the multiple scattering analysis and the application of the restraints, and offers some hope for discrimination in future metalloprotein studies between N(imidazole) ligands on the one side, and oxygen and non-imidazole nitrogen ligands on the other.

Simulations of the integrase EXAFS confirmed the presence of both nitrogen (imidazole) and sulfur contributions. In order to avoid problems with possible artefacts at high k, the integrase EXAFS was analyzed over energy ranges of 2.63 - 735 eV and 2.63 – 700 eV for the spectra with and without β-mercaptoethanol, respectively. With the intra-ligand multiple scattering included it was hoped that the low energy part of the spectrum would be adequately simulated. The fact that slight discrepancies were noted (cf. the low-k region of the top panels in Fig. 6) indicates that inter-ligand scattering also contributes to the X-ray absorption spectrum in this system. No attempts were made to simulate this feature. Distances, threshold energy, Debye-Waller-type factors and angles were refined with the optimum geometry of the imidazole ring as a restraint (cf. Fig. 3) and occupancies were varied in steps between 1-4 N (imidazole) ligands and 1-4 S ligands. For the integrase EXAFS in the presence of β-mercaptoethanol, the combination of 2 nitrogen (imidazole) and 2 sulfur ligands clearly stood out as the possibility with the lowest fit index (Table 5). For the integrase EXAFS recorded in the absence of β-mercaptoethanol, the coordination numbers appeared to be less well defined and equally good fit indices were obtained with the 2N + 2S, 2N + 3S, 3N + 2S, and 3N + 3S combinations (Table 5). Of all the possibilities, however, with a total

coordination number of 4, which was indicated by the comparison of the XANES to that of model compounds (see above), the 2N + 2S combination was clearly the best (Table 5), and in the simulations with 3 imidazole ligands the Debye-Waller-type factors of the outer shells, i.e. those at 3-4 Å, refined to unrealistically high values. The final fits for the integrase spectra are presented in Fig. 6; the parameters are in Table 4. It should be noted that the problem of high fitting errors for combinations of contributions to the main shell as noted for compound **4** (imidazole nitrogen and acetate oxygen) and compound **5** (imidazole and amine nitrogen) and high correlations between them does not exist for the combination of imidazole nitrogen and sulfur, probably because the contributions are different in phase and sufficiently resolved in distance. As mentioned above there is a small but significant effect of β-mercaptoethanol on the spectrum. In the refinement of the simulations, however, the only parameter that was significantly affected by the presence of β-mercaptoethanol was the Debye-Waller-type factor for S which decreased from 0.009 to 0.007. This possibly means that a small fraction of the excess Zn is present in a complex with β-mercaptoethanol, a possibility which does not affect the validity of the NMR structure determination. In the final fits (Fig. 6, Table 4) the coordination of the imidazole residues to the Zn was totally symmetric.

In the NMR structure (Eijkelenboom *et al.*, 1997) it was noted that one of the imidazole coordinates with the ε-nitrogen, the other with the δ-nitrogen (Fig. 7). In the latter case, the methylene of the histidine side chain might be expected to be detectable by EXAFS as it would be positioned at a distance of approx. 3.5 Å to the Zn. Attempts were made to get independent evidence for this feature from analysis of the EXAFS. A shell of additional carbon when included in the simulation of both integrase EXAFS spectra refined to a distance of 3.6 Å but did not lead to a decrease in the fit index that would be sufficiently statistically significant (5 %, Joyner *et al.*, 1987) to justify its inclusion. In a comparative study of metal complexes of methyl-substituted and unsubstituted imidazoles (Feiters *et al.*, 1988), it was already noted that all imidazole EXAFS spectra show features at 3.5 Å, regardless of the presence of the methyl substituent.

The effect of additional Zn atoms at the outer limit of detectability in the radial distribution function obtained by Fourier transformation of the EXAFS (approx. 4.5 Å) was also investigated, as possible evidence for dimerization of the integrase N-terminal part, but this did not improve the quality of the simulation either; in any case, even in case of dimerization of the integrase the Zn atoms would not have been able to approach eachother to such a close distance. The question of possible false-positive identification of metal-metal distances is of wider significance in the area of zinc enzymes, in particular in view of the discrepancy that has been noted between the

crystallographic (Sträter *et al.*, 1995) and EXAFS values (Priggemeyer *et al.*, 1995; Sift *et al.*, 1999) for the Zn-Fe distance in purple acid phosphatase.

It was also considered of interest to see if any deviation of the Zn from the plane of the imidazole ring could be detected by EXAFS, as theoretical simulations (Pettifer *et al.*, 1986) have shown that such deviations can have large effects, in particular on the intensity of the feature at just above 4 Å in the Fourier transform. In the Zn imidazole compounds included in this study, such deviations are also present, but only **5**, where the Zn is 0.44 Å out of the imidazole plane, lends itself to a more detailed investigation, as for the other compounds the deviations are smaller and the differences between the imidazole units are too large (0.02 and 0.18 Å for **3**, 0.11 and 0.23 Å for **4**). Our simulations (not shown) confirm that moving the Zn out of the imidazole plane lowers the intensity of the features at 4 Å. Work to discriminate in the approach for the refinement of the simulation between this effect and other effects that affect the intensity, like the value of the Debye-Waller-type factor, is still in progress; no further attempts were made, however, to interpret the integrase data in this respect.

**Conclusions and Outlook**

We find that the XANES and EXAFS measurements provide independent evidence for the information that was put into the NMR structure (Eijkelenboom *et al.*, 1997; Eijkelenboom *et al.*, 2000), viz. four-coordinated zinc with 2 N (imidazole) and 2 S (sulfur) at 2.0 and 2.3 Å, respectively, and that the coordination chemistry derived from the NMR structure has a solid foundation. The reagent β-mercaptoethanol, which was present during the NMR structure determination, has a small but significant effect on the EXAFS, but the only parameters changed in the analysis are some of the Debye-Waller-type factors, and the validity of the structure derived by NMR is not affected.

We have shown that the N and S contributions in a system like the Zn site in integrase can be resolved by taking the Fourier transform over a larger range of data than is typically analysed, allowing an estimate of their relative sizes. In addition, we have developed a protocol for the analysis of the EXAFS of Zn imidazole complexes, in which independent evidence for the orientation of the imidazoles with respect to the Zn-N vector is obtained, along with correct predictions for the distances of any non-imidazole ligands involved. The possibilities to analyse the effect on the EXAFS of lifting the metal ion out of the imidazole plane, and to obtain information on metal-imidazole interactions generally from XANES simulations should be explored in further work.

Provided that they can be carried out on a much smaller amount of protein than is possible nowadays, and much more readily and confidently interpreted than has been possible in the past, EXAFS studies can help identify the ligand environment in metalloproteins and provide a relatively simple experimental check for a predicted/proposed metal site. The efforts of the biological EXAFS (BIOXAS) community to continue to develop improved sources and detectors so that reliable XAS (EXAFS) measurements are possible for even smaller and more diluted samples, and to develop procedures for simulation and protocols for the analysis in order to extract information reliably deserve support at all levels. The development of protocols for the analysis and their validation on model compounds are expected to be important if the measurement and interpretation of XAS spectra of proteins are to play a role in high-throughput screening of metalloproteins in structural genomics programmes in the post-genomic era. It can be envisaged that it will be possible to identify potential metal-coordinating sites in the genetic information that is now rapidly becoming available; obviously in order for XAS to be of use on such studies, a satisfactory solution will have to be found for the problem of expression of a potential metalloprotein with incorporation of the biologically relevant metal, which, along with attachment of carbohydrate chains, modification of amino acids, incorporation of organic cofactors, and self-assembly into functional non-covalently bound protein aggregates, is an important post-transcriptional event.

**Acknowledgements.** F. v.d. E. was supported by a grant from the Netherlands AIDS foundation; A. E. by the Netherlands Foundation for Chemical Research (SON) with financial support from the Netherlands Organization for Scientific Research. B. K. thanks the BMBF (grant 05 KS1PMA/1) and the Fonds der Chemischen Industrie for support. We thank the CLRC Daresbury Laboratory (Warrington, UK) for computational facilities, including the use of EXCURV92, and the European Union for support of the work at EMBL Hamburg through the HCMP Access to Large Installations Project, Contract Number CHGE-CT93-0040. We thank Sandra Middelbeek for participating in the XAS measurements, Wim Jansen and Hilbert Bruins Slot (CAOS-CAMM, currently CMBI, Nijmegen) for help with the VAX computer and the Cambridge Crystallography Data Base, respectively, René de Gelder (Crystallography, Nijmegen) for developing and providing the XTAFS programme, and Bart Nelissen for preparing the stack plots.

## References


Abrahams, I. L., Bremner, I., Diakun, G. P., Garner, C. D., Hasnain, S. S., Ross, I. & Vasak, M. (1986). *Biochem. J.* **236**, 585-589.

Alber, B. E., Colangelo, C. M., Dong, J., Stålhandske, C., Baird, T. T., Tu, C., Fierke, C. A., Silverman, D. N., Scott, R. A. & Ferry, J. G. (1999). *Biochemistry* **38**,13119-13128.

Bear, C. A., Duggan, K. A. & Freeman, H. C. (1975). *Acta Cryst.* B**31**, 2713-2715.

Benfatto, M. & Della Longa, S. (2001) *J. Synchrotron Rad.* **8**, 1087.

Berg, J. M. & Shi, Y. (1996). *Science* **271**, 1081-1085.

Binsted, N., Campbell, J. W., Gurman, S. J. & Stephenson, P. C. (1991). EXCURV92, SERC Daresbury Laboratory.

Binsted, N., Strange, R. W. & Hasnain, S. S. (1992). *Biochemistry* **31**, 12117-12125.

Blackburn, N. J., Hasnain, S. S., Diakun, G. P., Knowles, P. F., Binsted, N. & Garner, C. D. (1983). *Biochem. J.* **213**, 765-768.

Bogumil, R., Faller, P., Binz, P.-A., Vašák, M., Charnock, J. M. & Garner, C. D. (1998). *Eur. J. Biochem.* **255**, 172-177.

Bordas, J., Dodson, G. G., Grewe, H., Koch, M. H. J., Krebs, B. & Randall, J. (1983). *Proc. R. Soc. Lond.* B **219**, 21-39.

Bracey, M. H., Christiansen, J., Tovar, P., Cramer, S. P. & Bartlett, S. G. (1994). *Biochemistry* **33**, 13126-13131.

Brown, P. O. (1997) in Retroviruses (Coffin, J. M., Hughes, S. H. & Varmus, H. E., eds), 161-203, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Cai, M., Zheng, R., Caffrey, M., Craigie, R., Clore, G. M. & Gronenborn, A. M. (1997). *Nature Struct. Biol.* **4**, 567-577.

Chen, J. C.-H., Krucinski, J., Miercke, L. J. W., Finer-Moore, J. S., Tang, A.H., Leavitt, A. D. & Stroud, R. M. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 8233–8238.

Clark-Baldwin, K., Tierney, D. L., Govindaswamy, N., Gruff, E. S., Kim, C., Berg, J., Koch, S. A. & Penner-Hahn, J. E. (1998). *J. Am. Chem. Soc.* **120**, 8401-8409.

Colman, P. M., Jansonius, J. N. & Matthews, B. W. (1972). *J. Mol. Biol.* **70**, 701-724.

Cosper, N. J., Stålhandske, C. M. V., Iwasaki, H., Oshima, T., Scott, R. A. & Iwasaki, T. (1999.) *J. Biol. Chem.* **274**, 23160-23168.

Cramer, S. P. (1988) in 'X-ray Absorption Spectroscopy' (Koningsberger, D. C. & Prins, R., editors) Wiley, New York.

de Groot, F. (2001) *Chem. Rev.* **101**, 1779.

Della Longa, S., Soldatov, A., Pompa, M. & Bianconi, A. (1995) *Comput. Mat. Sci.* **4**, 199-210.

Diakun, G. P., Fairall, L. & Klug, A. (1986). *Nature* **324**, 698-699.

Dyda, F., Hickman, A. B., Jenkins, T. M., Engelman, A., Craigie, R. & Davies, D. R. (1994). *Science* **266**, 1981-1986.

Eggers-Borkenstein, P. (1989). PhD Thesis, Univ. of Münster, Germany.

Eggers-Borkenstein, P., Priggemeyer, S., Krebs, B., Henkel, G., Simonis, U., Pettifer, R. F., Nolting, H.-F. & Hermes, C. (1989). *Eur. J. Biochem.* **186**, 667-675.

Eijkelenboom, A. P. A. M., Puras Lutzke, R. A., Boelens, R., Plasterk, R. H. A., Kaptein, R. & Hård, K. (1995). *Nature Struct. Biol.* **2**, 807-810

Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Vos, A., Doreleyers, J. F., Hård, K., Tullius, T. D., Plasterk, R. H. A., Kaptein, R. & Boelens, R. (1997). *Current Biol.* **7**, 739-746.

Eijkelenboom, A. P. A. M., Sprangers, R., Hård, K., Puras Lutzke, R. A., Plasterk, R. H. A., Boelens, R. & Kaptein, R. (1999). *Proteins* **36**, 556-564.

Eijkelenboom, A. P. A. M., van den Ent, F. M. I., Wechselberger, R., Plasterk, R. H. A., Kaptein, R. & Boelens, R. (2000). *J. Biomol. NMR* **18**, 119-128.

Eklund, H. & Brändén, C.-I. (1983) in Metal Ions in Biology, Vol. V: Zinc Enzymes (Spiro, T. G. series and volume editor), Wiley-Interscience, New York, pp. 123-152.

Eklund, H., Horjales, E., Jörnvall, H., Brändén, C.-I. & Jeffery, J. (1985). *Biochemistry* **24**, 8005-8012.

Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., Söderberg, B.-O., Tapia, O., Brändén, C.-I. & Åkeson, Å. (1976). *J. Mol. Biol.* **102**, 27-59.

Feiters, M. C. (1990). *Comments Inorg. Chem.* **11**, 131-174.

Feiters, M. C., Boelens, H., Veldink, G. A., Vliegenthart, J. F. G., Navaratnam, S., Allen, J. C., Nolting, H.-F. & Hermes, C. (1990). *Recl. Trav. Chim. Pays-Bas* **109**, 133-146.

Feiters, M. C. Klein Gebbink, R. J. M., Solé, V. A., Nolting, H.-F., Karlin, K. D. & Nolte, R. J. M. (1999). *Inorg. Chem.* **38**, 6171-6180.

Feiters, M. C. & Jeffery, J. (1989). *Biochemistry* **28**, 7257-7262.

Feiters, M. C., Little, C. & Waley, S. G. (1986). *J. de Phys.* **12-47-C8**, 1169-1172.

Feiters, M. C., Navaratnam, S., Al-Hakim, A., Allen, J. C., Spek, A. L., Veldink, G. A. & Vliegenthart, J. F. G. (1988). *J. Am. Chem. Soc.* **110**, 7746-7750.

Förster, M., Brasack, I., Duhme, A.-K., Nolting, H.-F. & Vahrenkamp, H. (1996). *Chem. Ber.* **129**, 347-353.

Garner, C. D. & Feiters, M. C., in Springer Series in Biophysics, Vol.2: Biophysics and Synchrotron Radiation (Bianconi, A. & Congiu Castellano, A., Editors), Springer-Verlag, Berlin, Heidelberg (1987) 136-146.

Gencic, S., LeClerc, G. M., Gorlatova, N., Peariso, K., Penner-Hahn, J. E. & Grahame, D. A. (2001) *Biochemistry* **40**, 13068-13078.

Goldgur, Y., Dyda, F., Hickman, A. B., Jenkins, T. M., Craigie, R. & Davies, D. R. (1998). *Proc. Natl. Acad. Sci. USA* **95**, 9150–9154.

González, J. C., Peariso, K., Penner-Hahn, J. E. & Matthews, R. G. (1996). *Biochemistry* **35**, 12228-12234.

Greenwald, J., Le, V., Butler, S. L., Bushman, F. D. & Choe, S. (1999). *Biochemistry* **38**, 8892–8898.

Gurman, S. J., Binsted, N. & Ross, I. (1984). *J. Phys. C., Solid State Phys.* **17**, 143-151.

Gurman, S. J., Binsted, N. & Ross, I. (1986). *J. Phys. C., Solid State Phys.* **19**, 1845-1861.

Hansen, M. S. T., Carteau, S., Hoffmann, C., Li, L. & Bushman, F. (1998). Retroviral cDNA integration: mechanism, applications and inhibition. In Setlow, J.K. (ed.), Genetic Engineering. Principles and Methods, Vol. 20. Plenum Press, New York and London, 41–62.

Hasnain, S. S. & Garner, C. D. (1987). *Prog. Biophys. Molec. Biol.* **50**, 47-65.

Hasnain, S. S. & Hodgson, K. O. (1999). *J. Synchrotron Rad.* **6**, 852-864.

Hermes, C., Gilberg, E. & Koch, M. H. J. (1984). *Nucl. Instrum. & Methods* **222**, 207-214.

Horrocks Jr., W. D., Ishley, J. N. & Whittle, R. R. (1982). *Inorg. Chem.* **21**, 3265-3269.

Ikeda, T. & Hagihara, H. (1966). *Acta Cryst.* **21**, 919-927.

Ivarson, G. (1973). *Acta Chem. Scand.* **27**, 3523-3530.

Jiang, D. T., Heald, S. M., Sham, T. K. & Stillman, M. J. (1994). *J. Am. Chem. Soc.* **116**, 11004-11013.

Joyner, R. W., Martin, K. J. & Meehan, P. (1987). *J. Phys. C., Solid State Phys.* **20**, 4005-4012.

Kaptein, R. (1991). *Curr. Opin. Struct. Biol.* **1**, 63-70.

Kau, L.-S., Spira-Solomon, D. J., Penner-Hahn, J. E., Hodgson, K. O. & Solomon, E. I. (1987) *J. Am. Chem. Soc.* **109**, 6433–6442.

Klabunde, T., Sträter, N., Fröhlich, R., Witzel, H. & Krebs, B. (1996). *J. Mol. Biol.* **259**, 737-748.

Klug, H. P. (1966). *Acta Cryst.* **21**, 536-546.

Klug, A. & Rhodes, D. (1987). *Trends Biochem. Sci.* **12**, 464-469.

Kretsinger, R. H., Cotton, F. A. & Bryan, R. F. (1963). *Acta Cryst.* **16**, 651-657.

Krüer, M.; Haumann, M.; Meyer-Klaucke, W. M., Thauer, R. K. & Dau, H. (2002) *Eur. J. Biochem.* **269**, 2117-2123.

Liljas, A., Kannan, K. K., Bergstén, P. C., Waara, I., Fridborg, B., Strandberg, B., Carlbom, U., Järup, L., Lövgren, S. & Petef, M. (1972). *Nature* **235**, 131-137.

Lodi, P. J., Ernst, J. A., Kuszewski, J., Hickman, A. B., Engelman, A., Carigie, R., Clore, G. M. & Gronenborn, A. M. (1995). *Biochemistry* **34**, 9826-9833.

MacFadden, D. L., McPhail, A. T., Garner, C. D. & Mabbs, F. E. (1976). *J. Chem. Soc., Dalton Trans.,* 47-52.

McKale, A. G., Veal, B. W., Paulikas, A. P., Chan, S.-K. & Knapp, G. S. (1988). *J. Am. Chem. Soc.* **110**, 3763-3768.

Maignan, S., Guilloteau, J. P., Zhou-Liu, Q., Clement-Mella, C. & Mikol, V. (1998). *J. Mol. Biol.* **282**, 359–368.

Murphy, L. M., Strange, R. W. & Hasnain, S. S. (1996). *Structure* **5**, 371-379.

Myers, L. C., Terranova, M. P., Nash, H. M., Markus, M. A. & Verdine, G. L. (1993). *Science* **261**, 1164-1167.

Nolting, H.-F. & Hermes, C. (1992). EXPROG Exafs data reduction package, EMBL Outstation Hamburg, Germany.

Peariso, K., Goulding, C. W., Huang, S., Matthews, R. G. & Penner-Hahn, J. E. (1998). *J. Am. Chem. Soc.* **120**, 8410-8416.

Pettifer, R. F., Foulis, D. F. & Hermes, C. (1986). *J. Phys.* **12-47-C8**, 545-550.

Pettifer, R. F. & Hermes, C. (1985). *J. Appl. Crystallogr.* **18**, 404-412.

Priggemeyer, S., P. Eggers-Borkenstein, P., Ahlers, F., Henkel, G., Körner, M., Witzel, H., Nolting, H.-F., Hermes, C. & Krebs, B. (1995). *Inorg. Chem.* **34**, 1445-1454.

Rees, D. C., Lewis, M., Konzatko, R. B., Lipscomb, W. N. & Hardman, K. D. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 3408-3412.

Richardson, J. S., Thomas, K. A., Rubin, B. H. & Richardson, D. C. (1975). *Proc. Natl. Acad. Sci. USA* **72**, 1349-1353.

Sift, B. H., Durmus, A., Meyer-Klaucke, W. & Krebs, B. (1999). *J. Synchrotron Rad.* **6**, 421-422.

Strange, R. W., Hasnain, S. S., Blackburn, N. J. & Knowles, P. F. (1986). *J. de Phys.* **12-47-C8**, 593-596.

Strange, R. W., Blackburn, N. J., Knowles, P. F. & Hasnain, S. S. (1987). *J. Am. Chem. Soc.* **109**, 7157-7162.

Sträter, N., Klabunde, T., Tucker, P., Witzel, H. & Krebs, B. (1996). *Science* **268**, 1489-1492.

Summers, M. F., Henderson, L. E., Chance, M. R., Bess Jr., J. W., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder III, R. C., Hare, D. R. & Arthur, L. O. (1992). *Protein Sci.* **1**, 563-574.

Yachandra, V. K., Powers, L. & Spiro, T. G. (1983). *J. Am. Chem. Soc.* **105**, 6596-6604.

Zeppezauer, M., Haas, C., Maret, W., Hermes, C. & Pettifer, R. F. (1986), *J. Phys.* **12-47-C8**, 1165-1168.

Zhou, Z. S., Peariso, K., Penner-Hahn, J. E. & Matthews, R. G. (1999). *Biochemistry* **38**, 15915-15926.

Table 1. Classification of Zn-sites in proteins by EXAFS and/or crystallography, adapted from (Garner & Feiters, 1987; Hasnain & Garner, 1987; Feiters, 1990).

| **Type** | **Characteristic** | **Examples** |
|---|---|---|
| A | S ligands only | LADH "structural" (Eklund *et al.*, 1976; Zeppezauer *et al.*, 1986)<br>Metallothionein (Abrahams *et al.*, 1986)<br>Ada protein (Myers *et al.*, 1993)<br>Neuronal growth inhibiting factor (Bogumil *et al.*, 1998) |
| B | N/O and S ligands | LADH "catalytic" (Eklund *et al.*, 1976)<br>Sorbitol dehydrogenase<br>(Eklund *et al.*, 1985; Feiters & Jeffery, 1989)<br>Transcription factor (Diakun *et al.*, 1986)<br>Nucleocapsid zinc finger (Summers *et al.*, 1992)<br>Methionine synthase<br>(González *et al.*, 1996; Peariso *et al.*, 1998; Zhou *et al.*, 1999)<br>Methyl transferase (Gencic *et al.*, 2001; Krüer *et al.*, 2002)<br>Spinach carbonic anhydrase (Bracey *et al.*, 1994) |
| C | N/O ligands only | Carbonic anhydrase (Liljas *et al.*, 1972; Yachandra *et al.*, 1983)<br>*Methanosarcina* carbonic anhydrase (Alber *et al.*, 1999)<br>Thermolysin (Colman *et al.*, 1972)<br>Superoxide dismutase<br>(Richardson *et al.*, 1975; Blackburn *et al.*, 1983; Murphy *et al.*, 1996)<br>Carboxypeptidase (Rees *et al.*, 1981)<br>Zn-containing ferredoxin (Cosper *et al.*, 1999)<br>Fe(III)-Zn(II) purple acid phosphatase<br>(Priggemeyer *et al.*, 1995; Sträter *et al.*, 1995; Klabunde *et al.*, 1996) |

Table 2. Parameters for the final simulations of the EXAFS (Fig. 4) and comparison with crystallographic values for zinc(II) ethylxanthate **1** and zinc(II) dimethyldithiocarbamate **2**. Amplitude reduction factor, 0.75; imaginary potential, -1 V; energy range, 25 - 730 eV.

| | Zn(II) ethylxanthate **1** (Fig. 4a) | | Zn(II) dimethyldithiocarbamate **2** (Fig. 4b). | |
|---|---|---|---|---|
| | EXAFS | crystal[a] | EXAFS | crystal[b] |
| Type | occupancy, distance | occupancy, distance | occupancy, distance | occupancy, distance |
| S | 3.6 at 2.341 Å (0.005)[c] | 4 at 2.36 Å | 4.2 at 2.346 Å (0.008)[c] | 4 at 2.36 Å |
| C | 1.7 at 3.171 Å (0.004)[c] | 4 at 3.21 Å | - | - |
| S | 1.0 at 3.417 Å (0.005)[c] | 4 at 3.46 Å | 1.2 at 2.960 Å (0.030)[c] | 1 at 3.04 Å |
| S | - | - | 1.8 at 4.197 Å (0.014)[c] | 1 at 4.27 Å |
| $\Delta E_0$ | 9.48 eV | | 10.33 eV | |
| Fit index | 0.00050 | | 0.00036 | |

[a]) crystal structure of **1**from Ikeda & Hagihara, 1966.

[b]) crystal structure of **2** from Klug, 1966.

[c]) Debye-Waller-type factors, quoted as $2\sigma^2$ ($Å^2$), in parentheses.

Table 3. Parameters for the simulations of the EXAFS (Fig. 5) and comparison with crystallographic values for Zn imidazole compounds. Amplitude reduction factor, 0.75; imaginary potential, -1 V.

| | $[Zn(im)_4](ClO_4)_2$ (**3**, Fig. 5a) | | $Zn(im)_2(acetate)_2$ (**4**, Fig. 5b) | | $[Zn(his)_2](H_2O)_2$ (**5**, Fig. 5c) | |
|---|---|---|---|---|---|---|
| Type | EXAFS | Xtal[a] | EXAFS | Xtal[b] | EXAFS | Xtal[c] |
| No of Im | 4 | 4 | 2 | 2 | 2 | 2 |
| N(im) | | | | | | |
| R[d], a[e] | 1.972, 0.000 | 2.000 | 1.995, 0.004 | 1.999 | 1.990, 0.008 | 2.039 |
| C(im) | | | | | | |
| R[d], a[e] | 2.928, 0.005 | 3.002 | 2.930, 0.007 | 2.940 | 3.007, 0.014 | 3.029 |
| angle[f] | - 124.30 | - 126.03 | - 121.62 | - 123.35 | - 128.36 | - 130.50 |
| C(im) | | | | | | |
| R[d], a[e] | 3.019, 0.005 | 3.037 | 3.080, 0.007 | 3.074 | 2.929, 0.014 | 3.003 |
| angle[f] | 126.59 | 126.01 | 132.91 | 130.38 | 120.21 | 120.30 |
| N(im) | | | | | | |
| R[d], a[e] | 4.060, 0.006 | 4.149 | 4.070, 0.009 | 4.098 | 4.144, 0.023 | 4.098 |
| angle[f] | - 163.20 | - 161.54 | - 158.88 | - 157.71 | - 165.97 | - 159.36 |
| C(im) | | | | | | |
| R[d], a[e] | 4.220, 0.006 | 4.182 | 4.237, 0.009 | 4.176 | 4.133, 0.023 | 4.108 |
| angle[f] | 160.88 | 161.63 | 165.60 | 164.50 | 157.46 | 153.41 |
| | | | 2 O(acetate) | | 2 N(amino) | |
| R[d], a[e] | | | 1.940, 0.004 | 1.973 | 2.054, 0.008 | 2.061 |
| Energy range | 2.4 – 755 eV | | 2.4 – 985 eV | | 2.4 – 755 eV | |
| $\Delta E_0$ | 12.84 eV | | 13.16 eV | | 13.58 eV | |
| Fit index | 0.00034 (0.00131)[g] | | 0.00021 (0.00084)[g] | | 0.00039 (0.00151)[g] | |

[a]) Crystallographic values for **3** from Bear *et al.*, 1975, cf. Fig. 3.

[b]) Crystallographic values for **4** from Horrocks *et al.*, 1982.

[c]) Crystallographic values for **5** from Kretsinger *et al.*, 1963.

[d]) R, distance to absorber Zn atom in Å.

[e]) a, Debye-Waller-type factor, quoted as $2\sigma^2$ ($Å^2$).

[f]) angle with respect to Zn-N, positive values correspond to upper halves of imidazoles in Fig. 3.

[g]) values without restraints in parentheses.

Table 4. Parameters for the final simulations of the EXAFS (Fig. 6) of Zn in the N-terminal part of HIV integrase-2 N. Amplitude reduction factor, 0.75; imaginary potential, -1 V.

| | with β-mercaptoethanol (Fig.6a) energy range, 2.6 - 700 eV | | without β-mercaptoethanol (Fig.6b) energy range, 2.6 - 735 eV | |
|---|---|---|---|---|
| Type | occupancy, distance | angle with respect to Zn-N[a] | occupancy, distance | angle with respect to Zn-N[a] |
| S | 2 at 2.263 Å (0.007)[b] | | 2 at 2.264 Å (0.009)[b] | |
| N (im) | 2 at 1.984 Å (0.005)[b] | | 2 at 1.989 Å (0.005)[b] | |
| C (im) | 2 at 2.979 Å (0.022)[b] | - 125.63[a] | 2 at 2.797 Å (0.020)[b] | - 125.59[a] |
| C (im) | 2 at 3.025 Å (0.022)[b] | + 126.38[a] | 2 at 3.013 Å (0.020)[b] | + 125.45[a] |
| N (im) | 2 at 4.135 Å (0.033)[b] | - 161.60[a] | 2 at 4.130 Å (0.033)[b] | - 162.03[a] |
| C (im) | 2 at 4.169 Å (0.033)[b] | + 162.49[a] | 2 at 4.163 Å (0.033)[b] | + 161.85[a] |
| $\Delta E_0$ | 14.84 eV | | 14.68 eV | |
| Fit index | 0.00022 (0.00090)[c] | | 0.00018 (0.00073)[c] | |

[a]) positive values correspond to upper halves of imidazoles in Fig. 3.

[b]) Debye-Waller-type factors, quoted as $2\sigma^2$, in parentheses.

[c]) values without restraints in parentheses.

Table 5. Dependence of fit index (* $10^3$, corrected for offset due to restraints) of EXAFS simulations of integrase on S and N(im) occupancies.

| Total Coordination Number: | 2 | | 3 | | 4 | | 5 | | 6 | |
|---|---|---|---|---|---|---|---|---|---|---|
| | with β-mercaptoethanol | | | | | | | | | |
| 0 N | | | | | 4 S | 214 | | | | |
| 1 N | + 1 S | 177 [a,b] | + 2 S | 118 [a] | + 3 S | 112 | + 4 S | 128 | | |
| 2 N | | | + 1 S | 119 [b] | + 2 S | 89 | + 3 S | 100 | + 4 S | 119 |
| 3 N | | | | | + 1 S | 109 | + 2 S | 93 | + 3 S | 104 |
| 4 N | | | | | + 0 S | 383 | + 1 S | 124 | + 2 S | 112 |
| | without β-mercaptoethanol | | | | | | | | | |
| 0 N | | | | | 4 S | 212 | | | | |
| 1 N | + 1 S | 187 [a] | + 2 S | 114 [a] | + 3 S | 101 | + 4 S | 111 | | |
| 2 N | | | + 1 S | 113 | + 2 S | 73 | + 3 S | 79 | + 4 S | 94 |
| 3 N | | | | | + 1 S | 97 | + 2 S | 75 | + 3 S | 80 |
| 4 N | | | | | + 0 S | 340 | + 1 S | 111 | + 2 S | 92 |

[a]) Debye-Waller type factor for coordinating nitrogen refines to negative value.

[b]) Debye-Waller type factor for sulfur refines to negative value.

**Figure Legends**

Figure 1. XANES (top panel, $E_0$ = 9660 eV), $k^3$-weighted EXAFS (middle panel), and modulus of phase-corrected Fourier transform (bottom panel) of 4 model compounds and 2 protein samples, viz. (top to bottom) Zn ethylxanthate **1**; Zn dimethyldithiocarbamate **2**; integrase with β-mercaptoethanol; integrase; $[Zn(imidazole)_4](ClO_4)_2$ **3**; $Zn(imidazole)_2(acetate)_2$ **4**.

Figure 2. Comparison of N-terminal part of HIV integrase-2 with (dashed line) and without (solid line) β-mercaptoethanol; top panel, extended range (k 2-16 $Å^{-1}$) of $k^2$-weighted EXAFS; bottom panel, corresponding modulus of the phase-corrected Fourier transform.

Figure 3. Averaged geometry of the two types of imidazole coordinating to zinc(II) in $[Zn(imidazole)_4](ClO_4)_2$ **3** (Bear *et al.*, 1975), showing the restraints used in the refinement of the EXAFS data. Bold, so-called 'single' restraints used in all refinements; italic, optional 'double' restraints; in parentheses, distance values that were refined in the iteration.

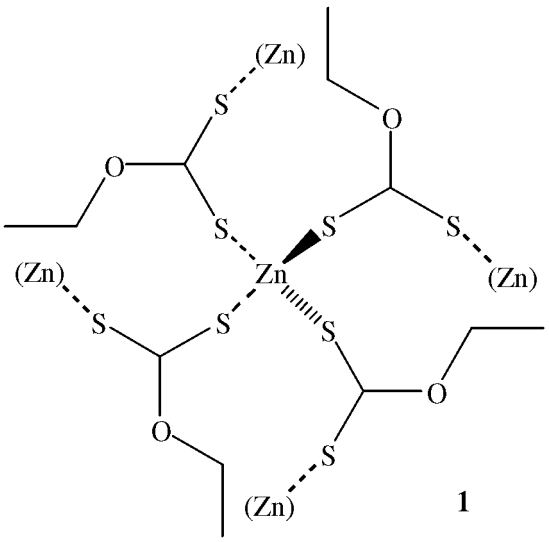
Figure 4. $k^3$-weighted EXAFS (top panels) and modulus of phase-corrected Fourier transform (bottom panels) of experimental (solid lines) and simulations (dashed lines, parameters in Table 2) of (left) zinc(II) ethylxanthate **1** and (right) Zn(II) dimethyldithiocarbamate **2**.

Figure 5. $k^3$-weighted EXAFS (top panels) and modulus of phase-corrected Fourier transform (bottom panels) of experimental (solid lines) and simulations (dashed lines, parameters in Table 3) of $[Zn(imidazole)_4](ClO_4)_2$ **3**, $Zn(imidazole)_2(acetate)_2$ **4**, and di-(L-histidino)-zinc(II) dihydrate **5**. Insets of bottom panels: final geometry of imidazole ring after refinement, with Zn as the central atom, nitrogen atoms (1) and (4), and carbon atoms (2), (3), and (5).

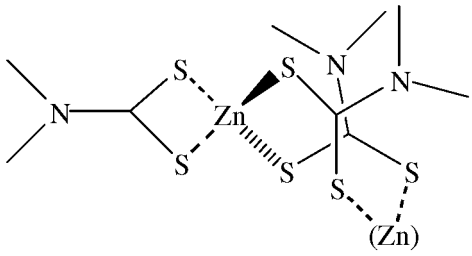
Figure 6. $k^3$-weighted EXAFS (top panels) and modulus of phase-corrected Fourier transform (bottom panels) of experimental (solid lines) and simulations (dashed lines, parameters in Table 4) of the N-terminal part of HIV-II integrase-2 with (left) and without (right) β-mercaptoethanol. Insets of bottom panels: final geometry of imidazole ring after refinement, with Zn as the central atom, nitrogen atoms (1) and (4), and carbon atoms (2), (3), and (5).

Figure 7. Coordination topology of the zinc center in the N-terminal domain of HIV-2 integrase, proposed in the NMR study (Eijkelenboom *et al.*, 1997) and confirmed by the present EXAFS results.

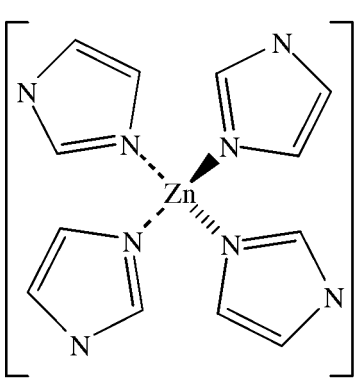
Chart 1



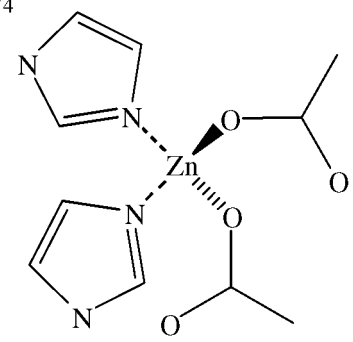

**1**



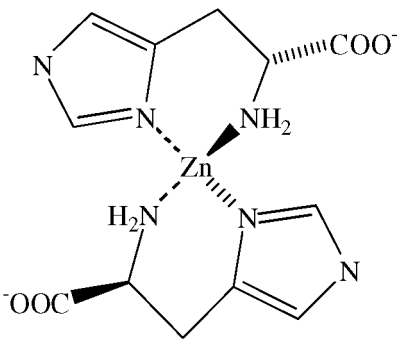

**2**




**3**




**4**




**5**

Energy (eV, $E_0$ = 9660)

Figure 1a

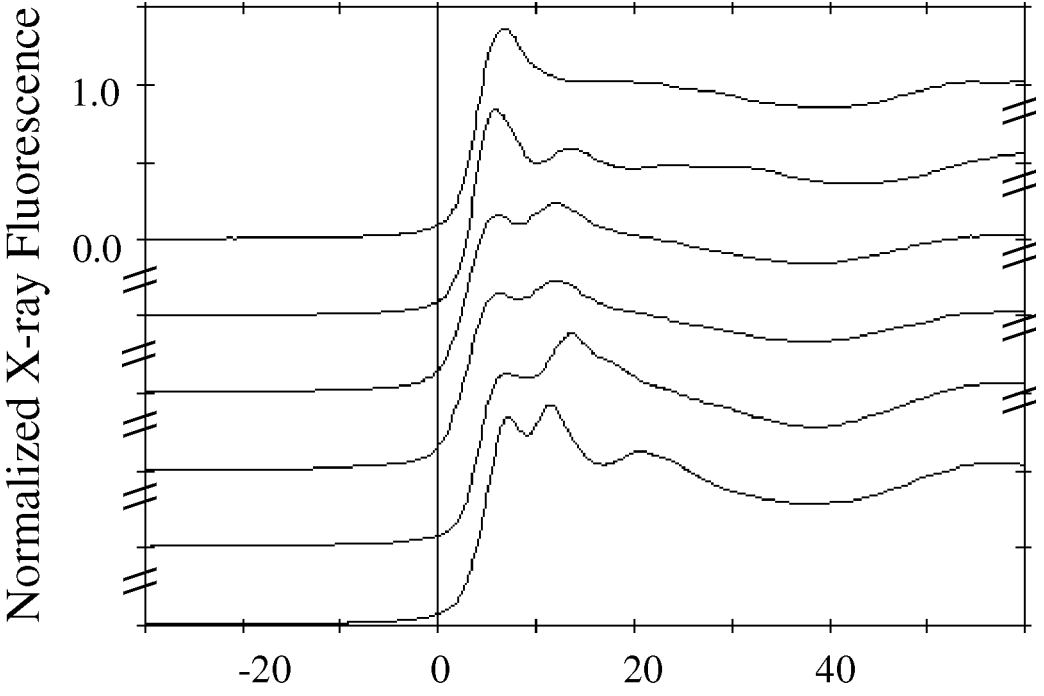


Figure 1b

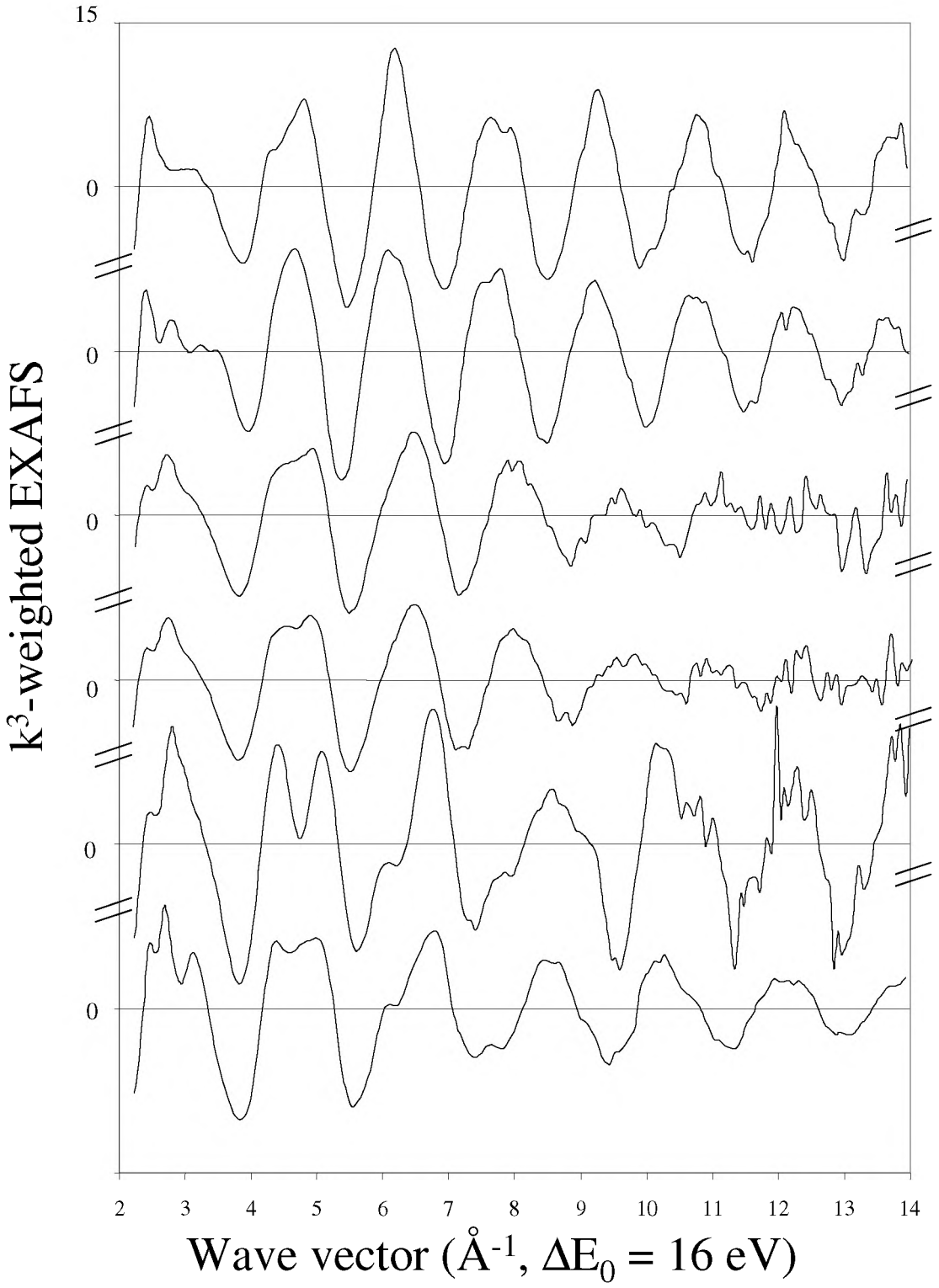
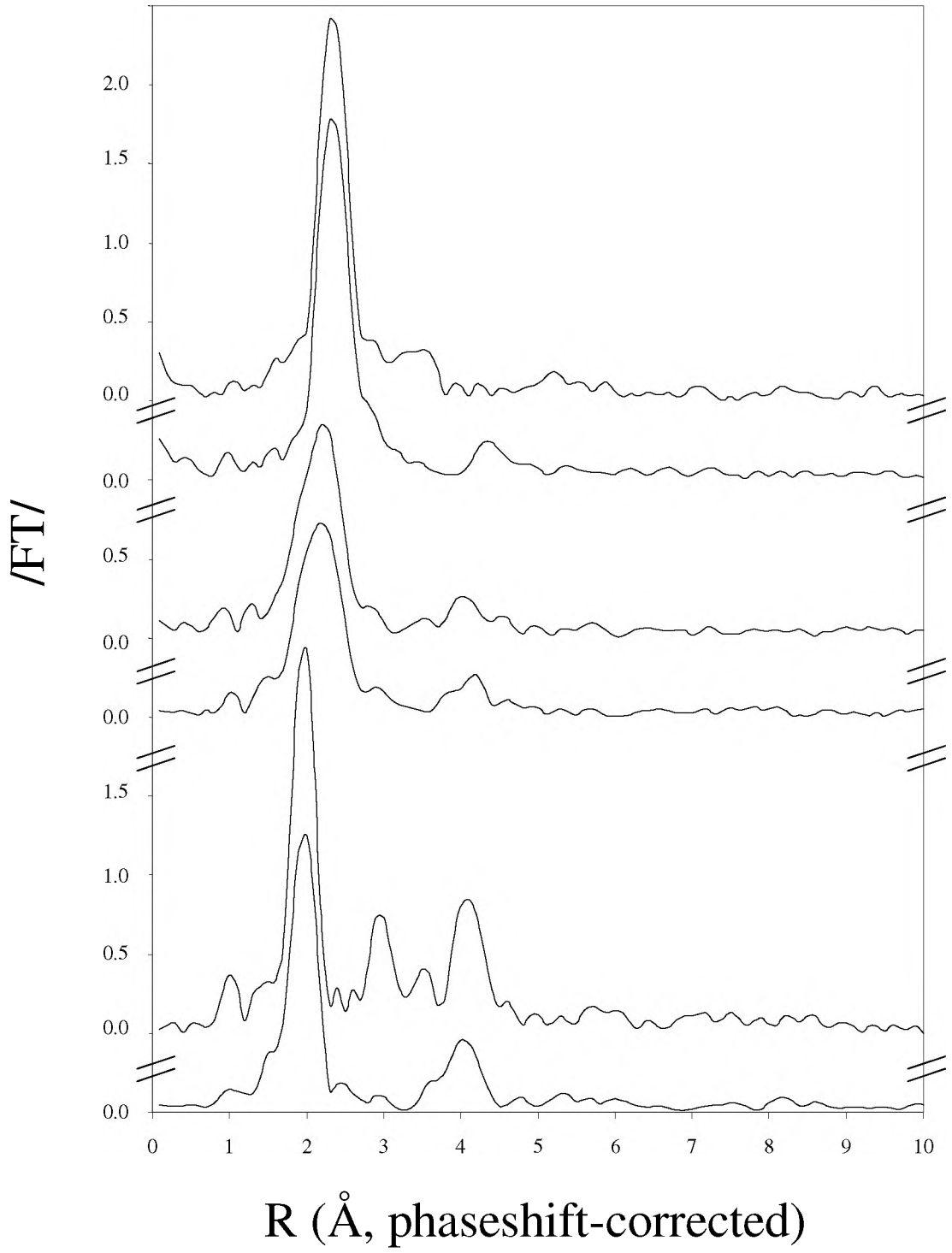

15
0
0
0
0
0
0
k3-weighted EXAFS
2
3
4
5
6
7
8
9
10
11
12
13
14
Wave vector (Å-1, ΔE0 = 16 eV)

Figure 1c


/FT/
2.0
1.5
1.0
0.5
0.0
0.0
0.5
0.0
0.0
1.5
1.0
0.5
0.0
0.0
0
1
2
3
4
5
6
7
8
9
10
R (Å, phaseshift-corrected)

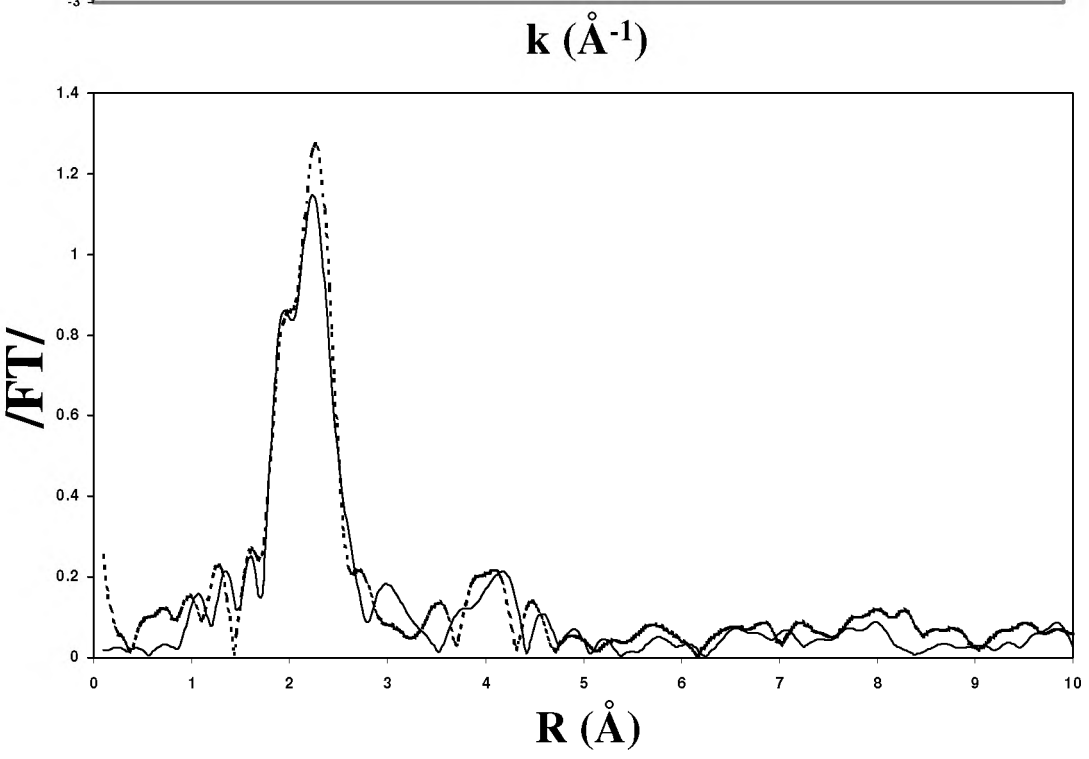

k (Å-1)
/FT/
R (Å)
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
1
2
3
4
5
6
7
8
9
10

Figure 2

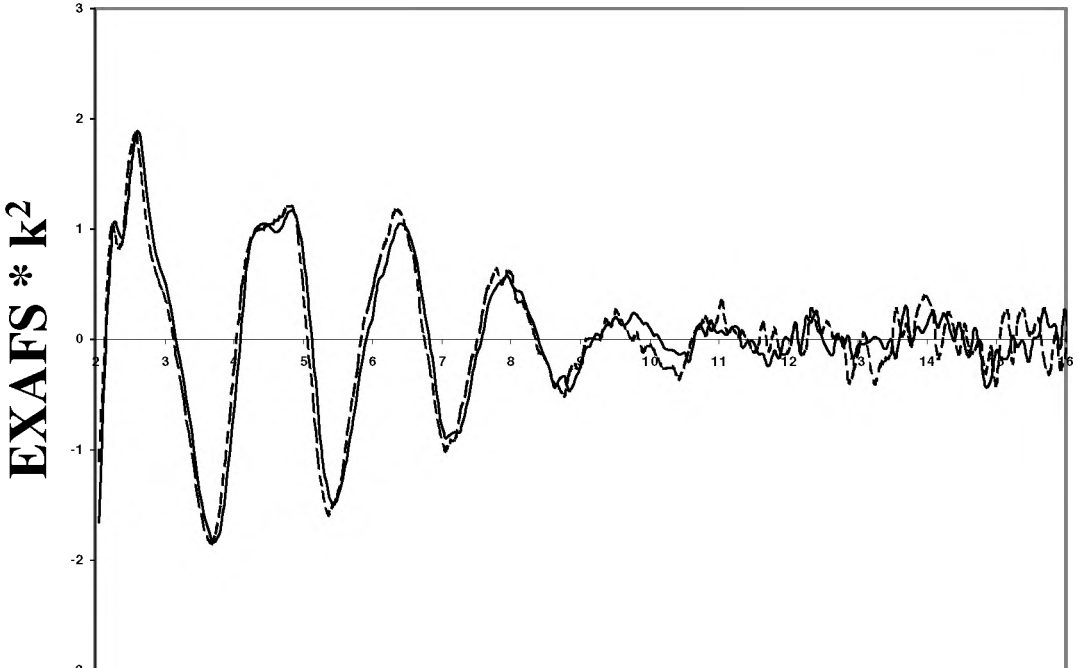
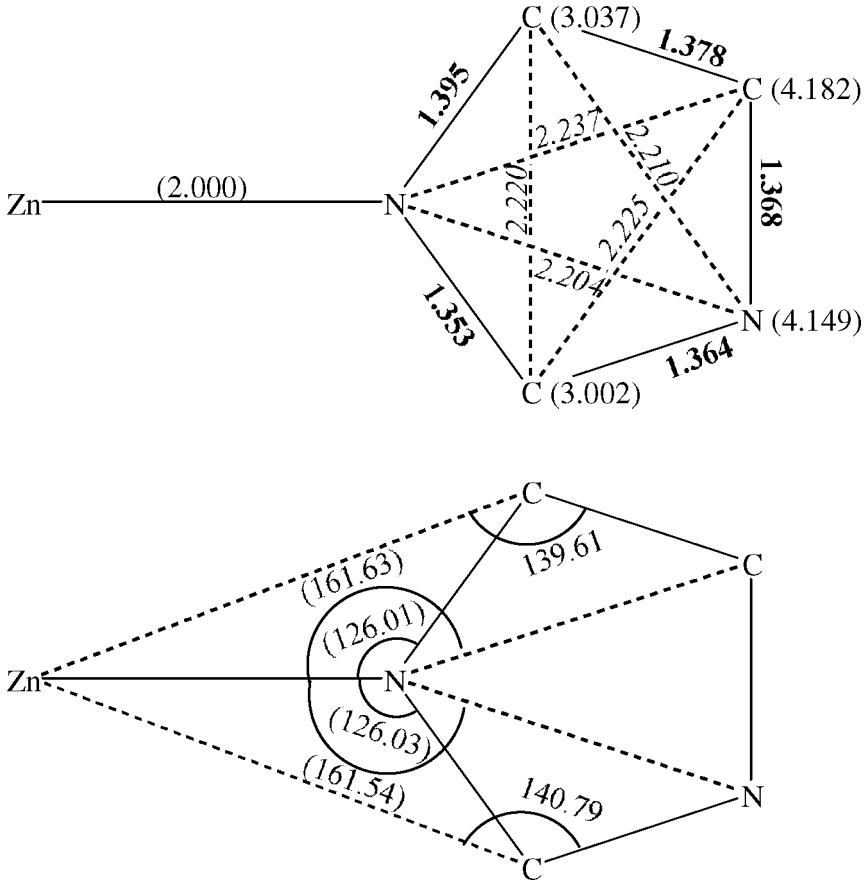


Figure 3


C (3.037)
1.378
C (4.182)
1.395
2.237
2.210
1.368
Zn
(2.000)
N
2.220
2.225
2.204
1.353
N (4.149)
1.364
C (3.002)
C
(161.63)
139.61
C
(126.01)
Zn
N
(126.03)
(161.54)
140.79
N
C

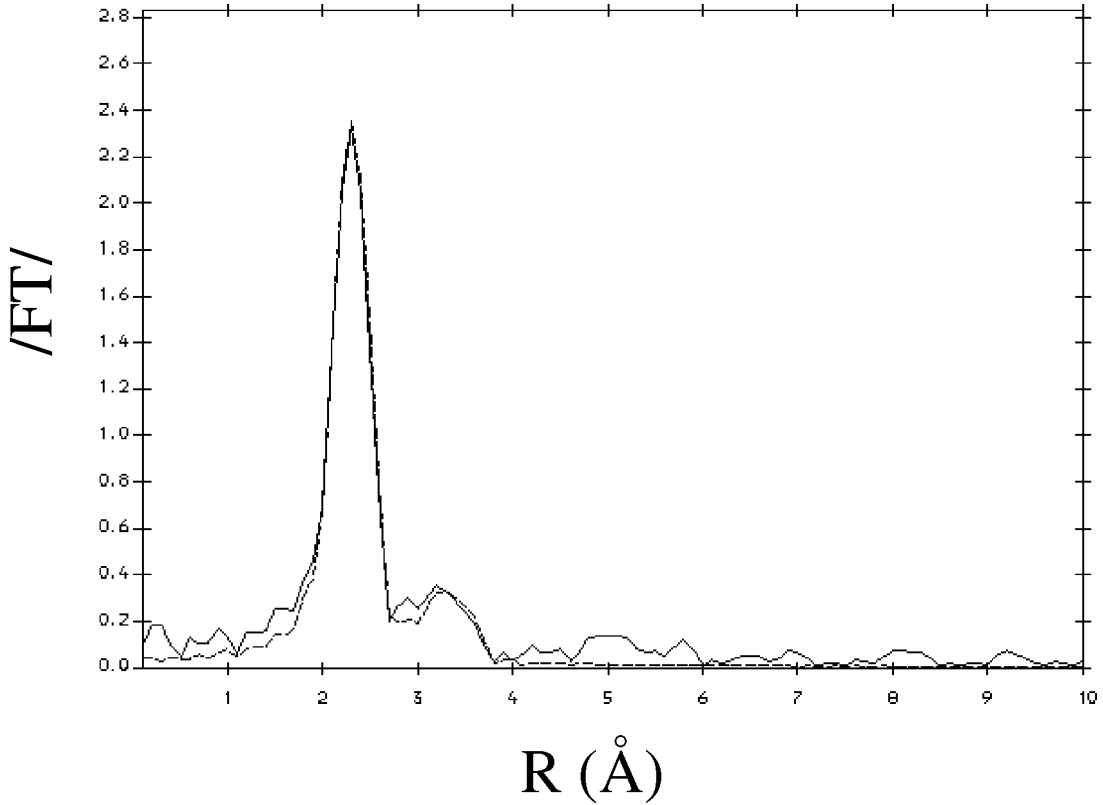
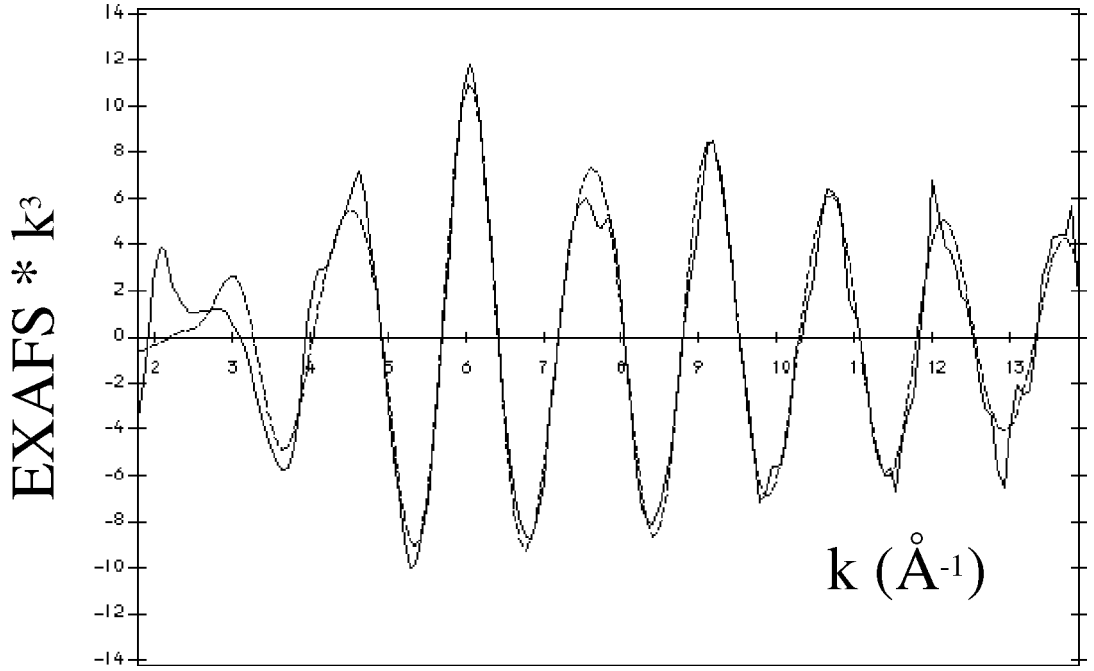

/FT/
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
R (Å)

Figure 4a



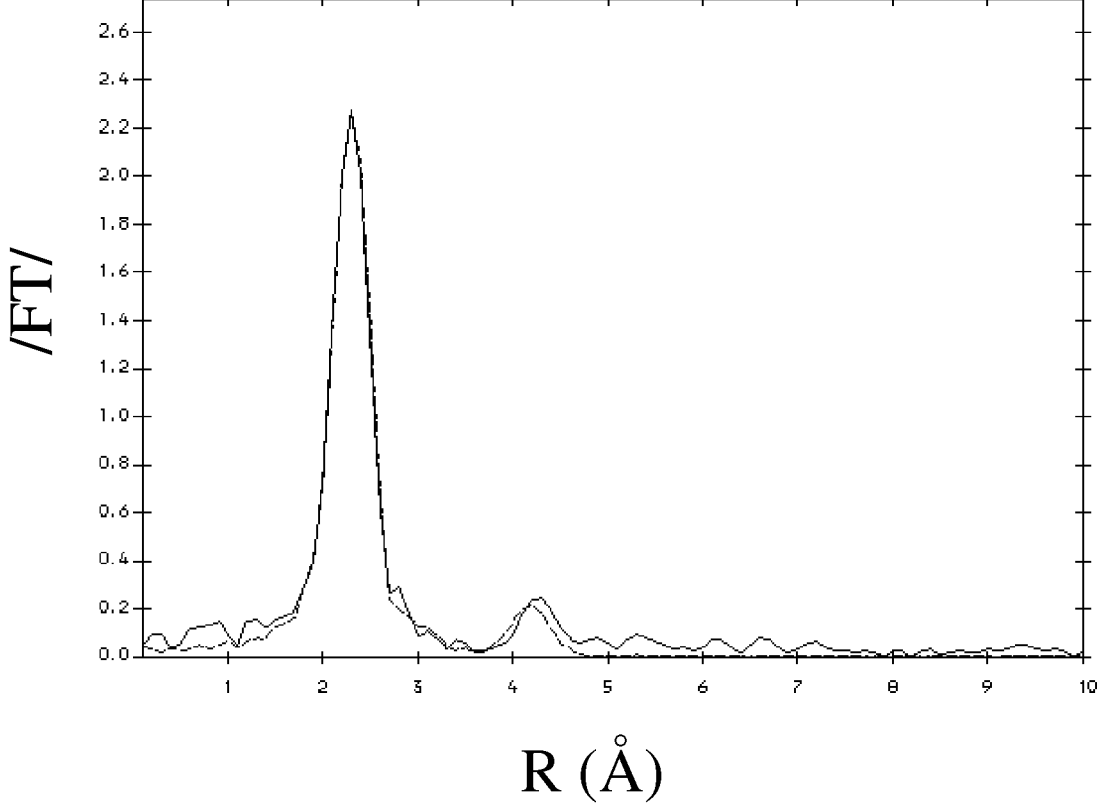
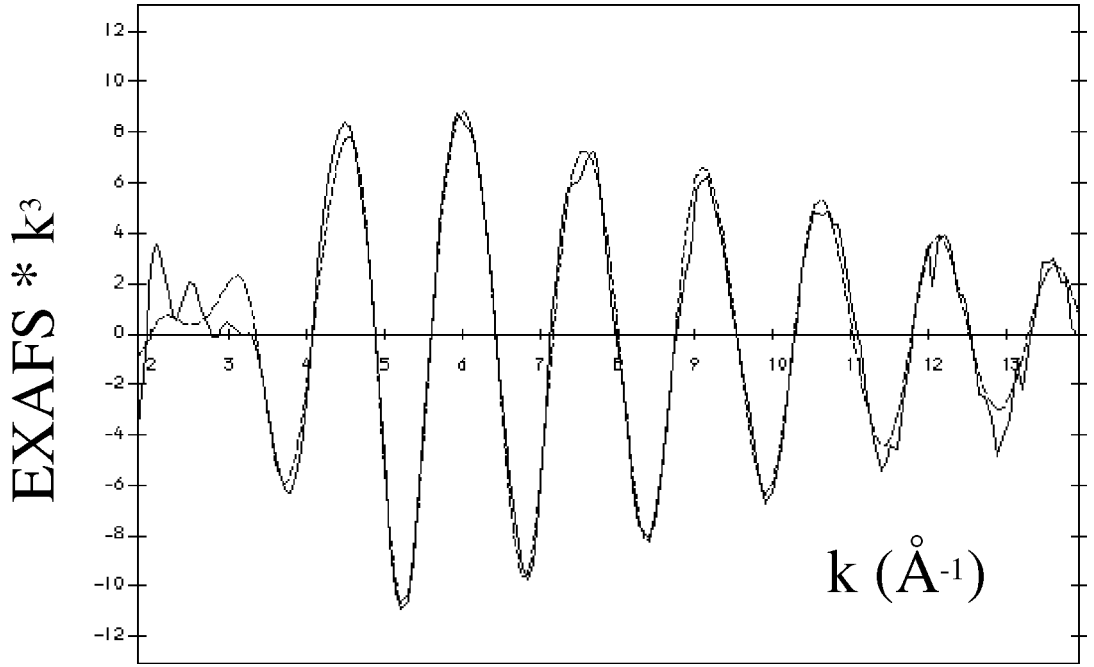

/FT/
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
9
10
R (Å)

Figure 4b



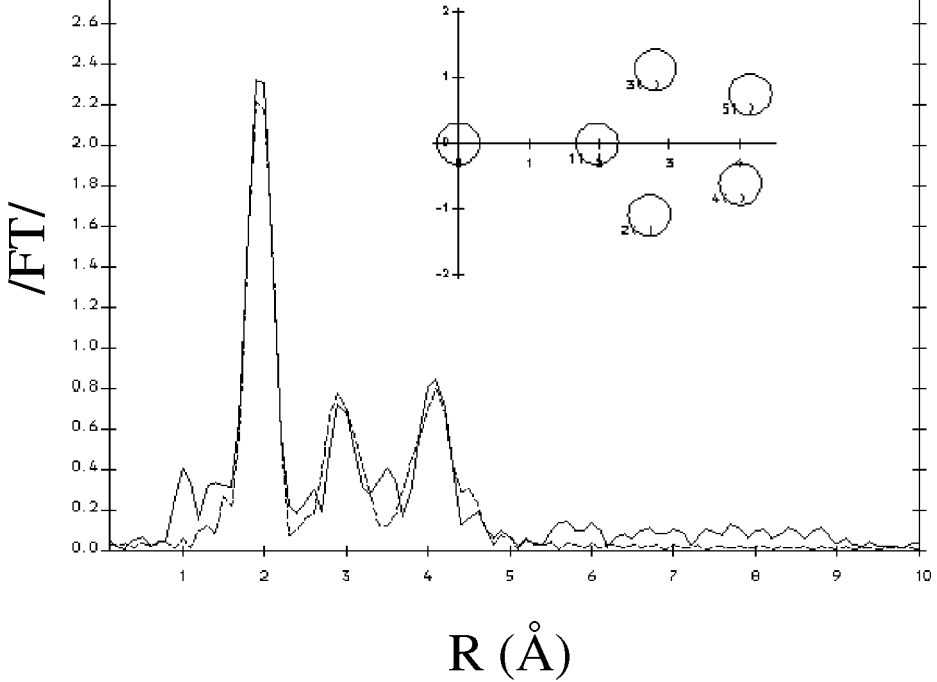
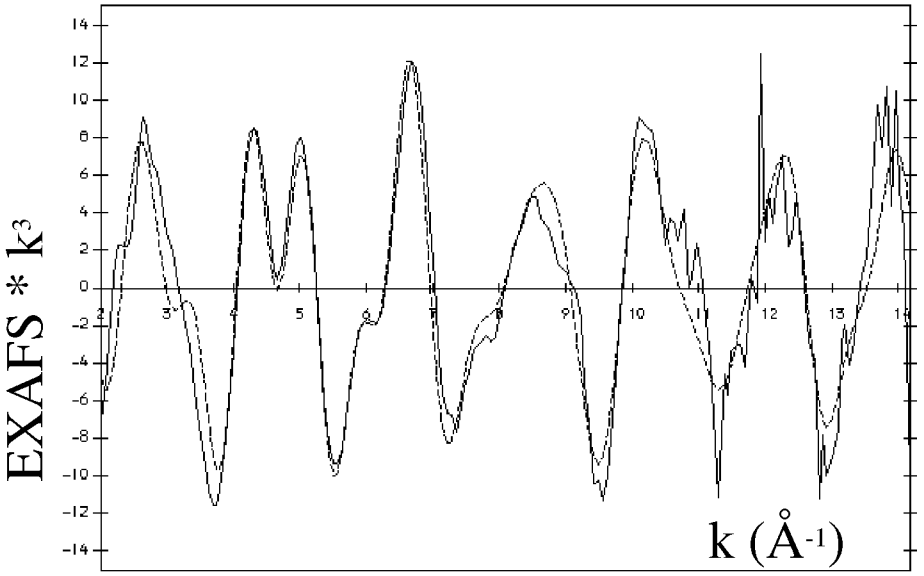

/FT/
R (Å)

Figure 5a



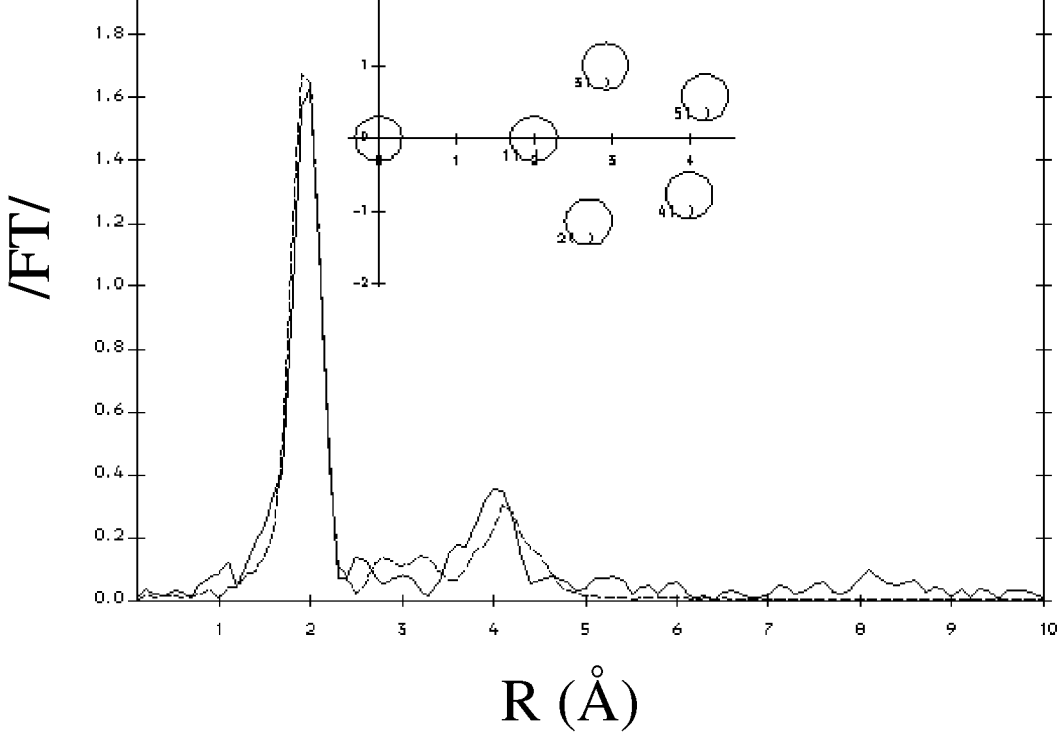
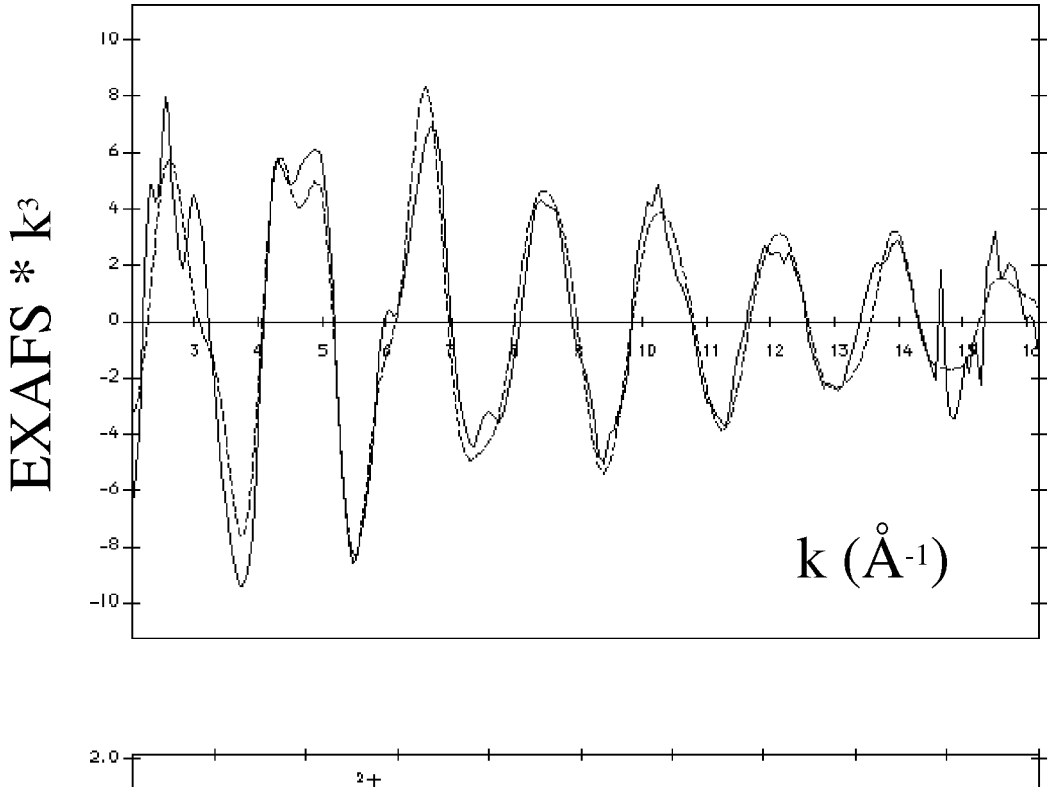

/FT/
R (Å)
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
1
2
3
4
5
6
7
8
9
10

Figure 5b



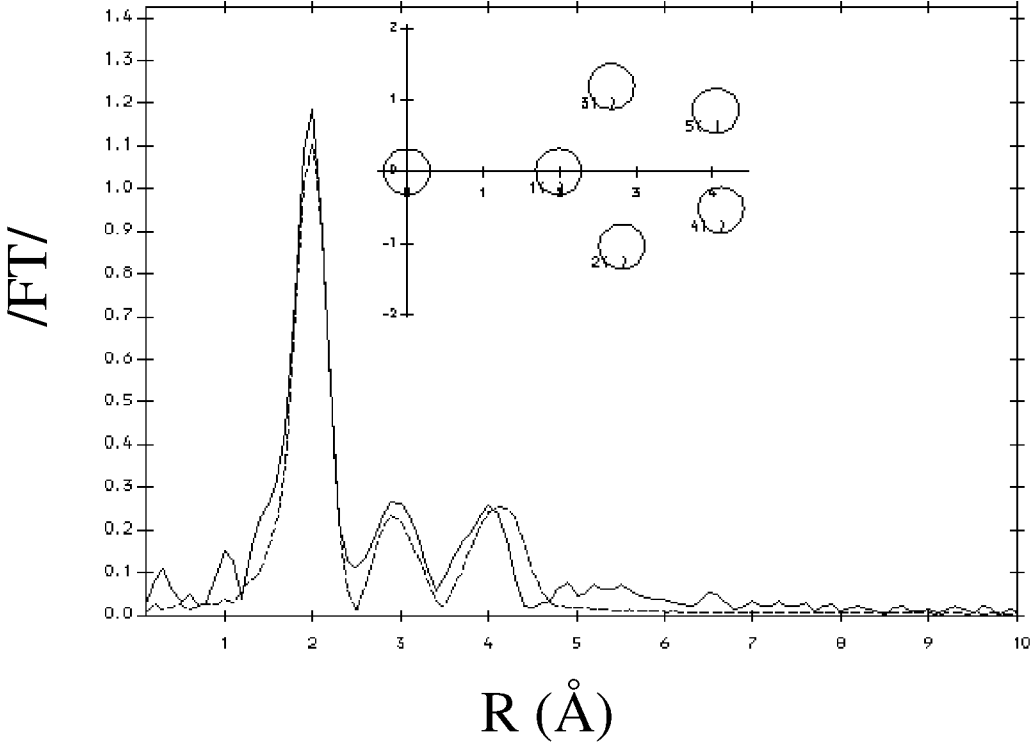
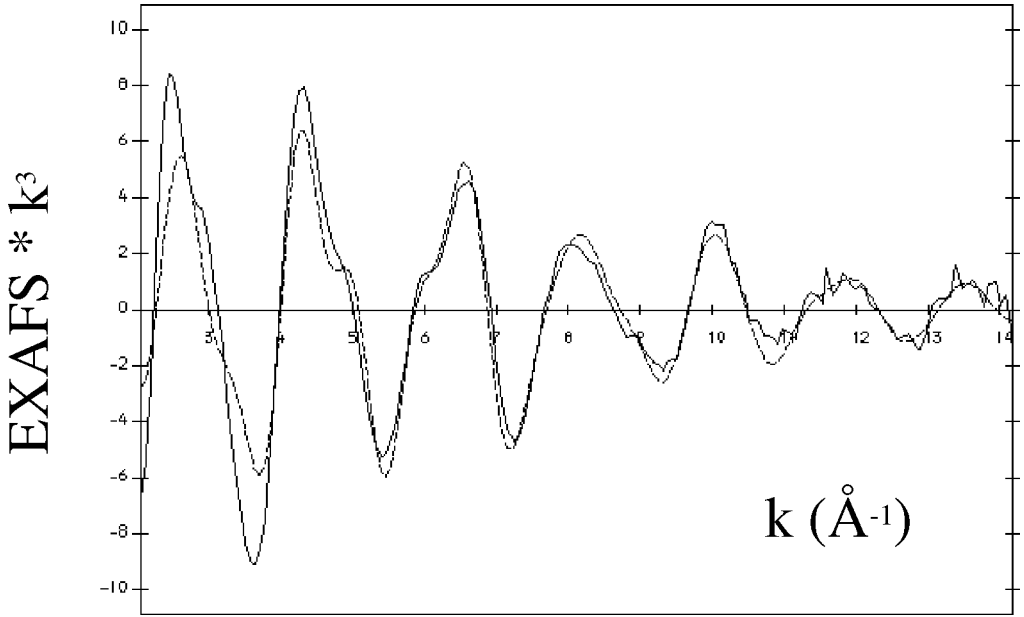

/FT/
R (Å)

Figure 5c



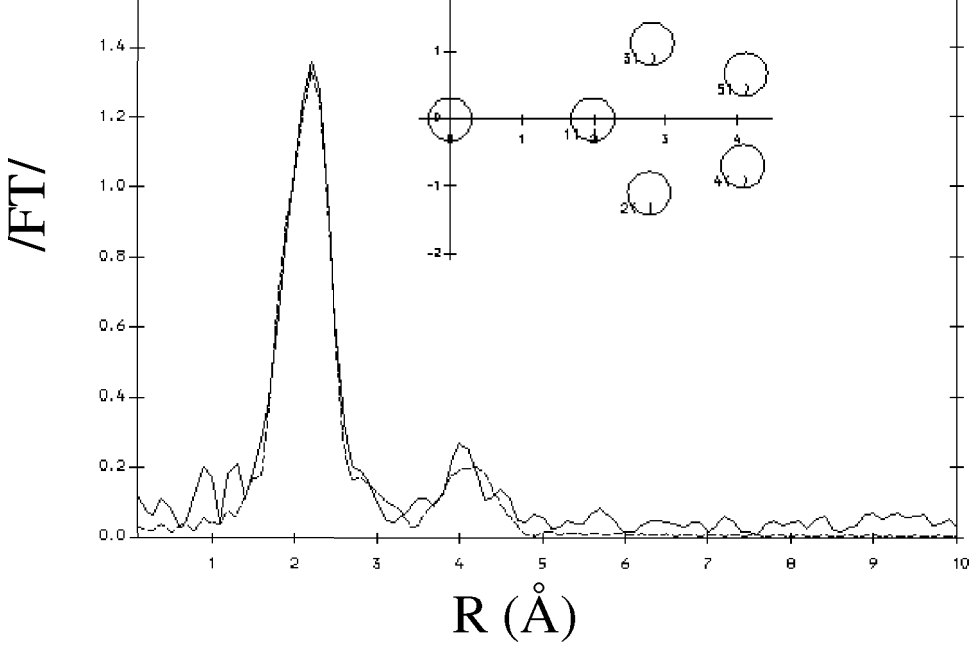
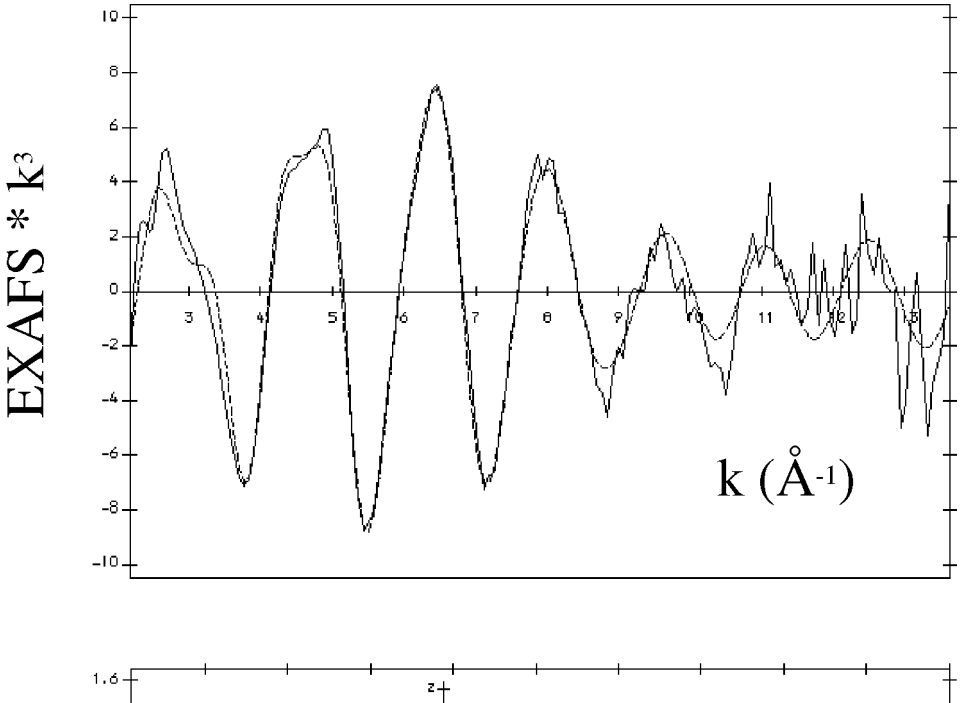

/FT/
R (Å)

Figure 6a



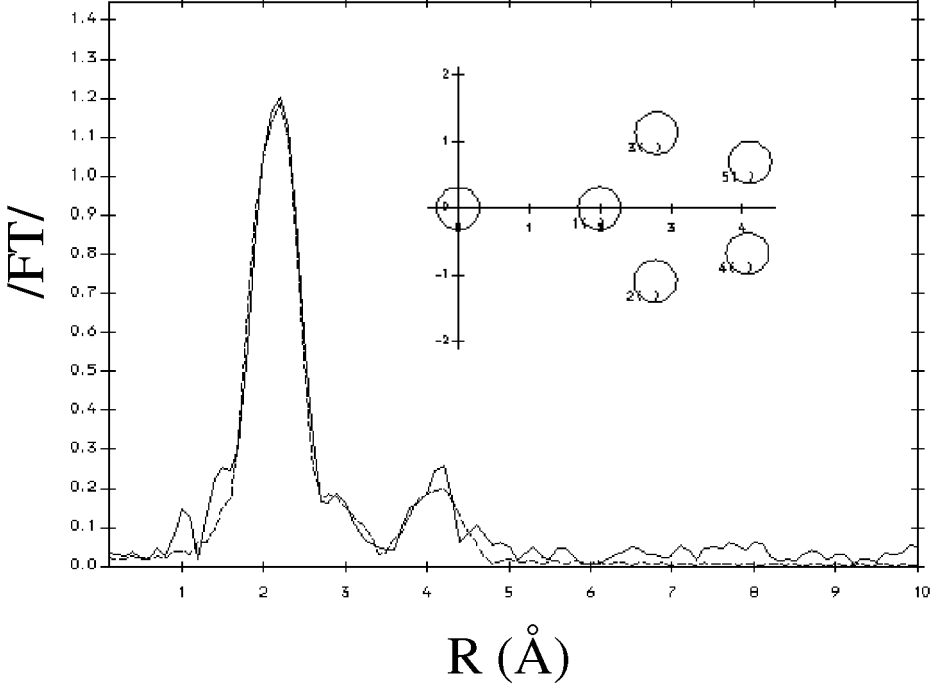

/FT/
R (Å)

Figure 6b

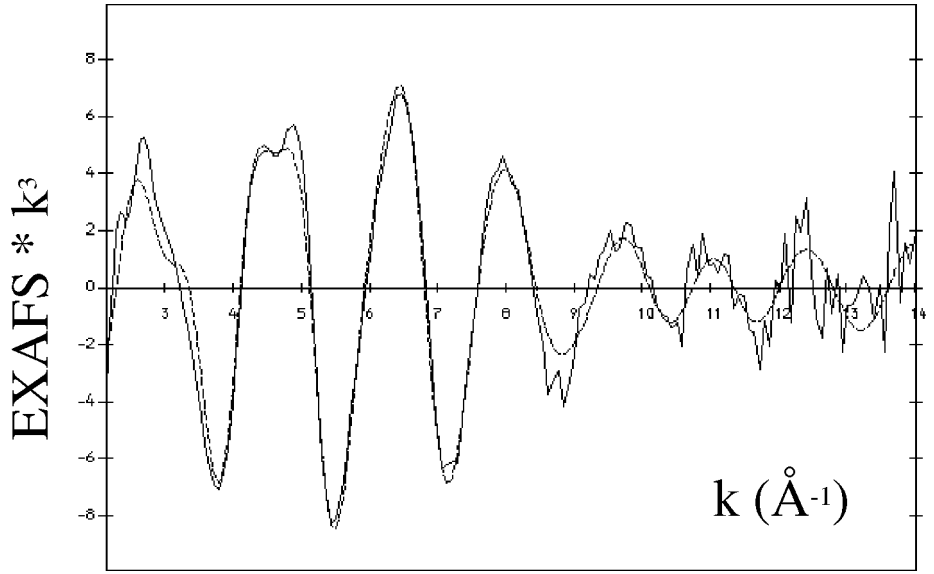


Figure 7