ADVANCED APPLICATIONS OF X-RAY ABSORPTION SPECTROSCOPY TO THE STUDY OF PROTEIN METAL SITES

Giulia Veronesi

Advisor: Prof. Federico Boscherini   Co-advisor: Prof. Giovanni Venturoli

A thesis submitted for the degree of Doctor of Philosophy

School of Physics

ALMA MATER STUDIORUM - UNIVERSITY OF BOLOGNA

March 2010

Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

FISICA

Ciclo XXII

Settore scientifico-disciplinare di afferenza: FIS/03

Advanced applications of X-ray Absorption Spectroscopy to the study of protein metal sites.

Presentata da: Giulia Veronesi

Coordinatore Dottorato:

Prof. Fabio Ortolani

Relatore:

 Prof. Federico Boscherini

Co-relatore:

Prof. Giovanni Venturoli

Esame finale anno 2010

Contents

List of acronyms

ATR-FTIR Attenuated Total-Reflectance Fourier-Transformed InfraRed

BM Bending Magnet

COX Cytochrome Oxidase

CSD Cambridge Structural Database

DFT Density Functional Theory

DW Debye-Waller

EXAFS Extended X-ray Absorption Fine Structure

FDM Finite Difference Method

FMS Full Multiple Scattering

FTIR Fourier Transformed InfraRed

ID Insertion Device

MbCO Carboxy-Myoglobin

MD Molecular Dynamics

MDB Metalloprotein Database and Browser

MS Multiple Scattering

MT Muffin Tin

NAD Nicotinamide Adenine Dinucleotide

NADP Nicotinamide Adenine Dinucleotide Phosphate

NMR Nuclear Magnetic Resonance

PDB Protein Data Bank

PVA Polyvinyl alcohol

QE Quantum ESPRESSO

RBR Rigid Body Refinement

RC Reaction Centre

RSMS Real Space Multiple Scattering

SOD Superoxide Dismutase

SS Single Scattering

TLS Thermolysin

TH Transhydrogenase

WL White Line

WT Wild Type

XAFS X-ray Absorption Fine Structure

XANES X-ray Absorption Near-Edge Structure

XAS X-ray Absorption Spectroscopy

XRD X-Ray Diffraction

Introduction

The thesis work presented here is focused on X-ray Absorption Spectroscopy on biological systems (BioXAS), intended both as the use of XAS to investigate protein metal sites of biological relevance and as the development of methods aimed to improve the XAS analysis itself. A consistent part of this work will concern the high-resolution determination of metal binding clusters in energy transducing membrane proteins, for which X-ray Absorption Fine Structure (XAFS) spectroscopy is an ideal tool. These studies are fundamental when the protein structure has never been resolved by crystallography, because they carry unique information for the elucidation of the protein function. However, even when the metal cluster is known, the further geometrical notions retrieved by XAS can bring insight into the role of the metal, thanks to its higher resolution with respect to X-Ray Diffraction (XRD). Investigated metal centers can be either exogenously added in the protein or endogenously present in it (covering structural or catalytic functions), depending on the biophysical goal; in this work we will present one example for each case.

At present, BioXAS can be used as a valuable tool to characterize metal sites, nevertheless its analysis method is still under development. Therefore we considered the technique not simply as a tool but also as a target for this study, and tried to improve its capabilities by means of *ab initio* simulations and systematic empirical observations; this brought to the determination of an efficient analysis method, an outline of which is given in this work. For our determination of metal binding clusters in membrane proteins, we used slightly modified versions of the method, in order to adapt it to the specific biophysical problem under investigation; furthermore, the effectiveness of the method as a whole, disregarding possible semplifications due to biophysical considerations, was successfully tested for model proteins.

This work will be structured as follows:

In Chapter 1 an overview of XAFS is given, with particular attention to its applications to biological systems. The theory underlying the technique is presented without claim of exhaustivity, but with careful attention to the approximations implemented in the different analysis programs we used.

Chapter 2 includes an outline of the XAFS analysis method we developed for the unambiguous determination of protein Zn sites. The method is based on the definition of quantitative criteria related to the nature of the binding cluster; the criteria are obtained both by comparison with simulated XAFS spectra and by the research of patterns in the spectra of known compounds. A test of the method on known metal sites is presented.

In Chapter 3 we present our study on the inhibitory Zn^{2+} of Transhydrogenase (TH), a proton-translocating membrane protein. The metal binding site was characterized in native and genetically modified TH, and the evidences were interpreted also in the light of FTIR spectroscopy experiments. The results carry important information about the transmembrane domain of the protein, the crystallographic structure of which is still unavailable.

Chapter 4 is dedicated to the study of the Fe^{2+} site of the bacterial photosynthetic Reaction Center (RC), which constituted the most challenging part of this work from the experimental point of view: first of all we present a XAFS experiment on RC embedded in different matrices, then a pump-probe time-resolved XAFS study on RC in a non-interacting matrix. While the results of the static study are evident and clearly understandable in the framework of matrix effect, the time-resolved spectra show features that are barely distinguishable from the noise. However, through careful data treatment we found indications of a conformational change occurring at the Fe^{2+} site on a time scale of $\sim$500 μs after light absorption.

Chapter 5 reports of a novel way to complement XAFS analysis with *ab initio* Density Functional Theory (DFT) calculations: DFT was used to obtain information about the dynamics of the heme atoms in MbCO, in order to fix the set of dynamical parameters in XAFS analysis. It all allowed to focus the XAFS analysis on the structural parameters, and to get subtle information of biological relevance, that would have been impossible to obtain otherwise.

Chapter 1

XAFS spectroscopy in biological systems

1.1 Phenomenology of X-ray absorption

Photoelectric absorption is the main process occurring when radiation in the X-ray energy region (1-30 keV) interacts with matter: in the total photon cross section, indeed, its contribution dominates with respect to contributions due to coherent and Compton scattering (Figure 1.1)
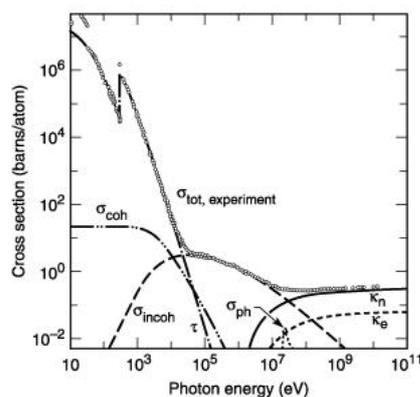
Figure 1.1: Total photon cross section (open circles) in carbon as a function of energy. The relative contributions due to photoelectric absorption (τ), coherent scattering (σ_{coh}), Compton scattering (σ_{incoh}) and other processes are reported. The figure is taken from the x-ray data booklet (http://xdb.lbl.gov/)

X-ray absorption spectroscopy is based on the measurement of the absorption coefficient $\mu(E)$ as a function of the energy of the incoming radiation on the material under study. The attenuation of a monochromatic x-ray beam of intensity

I_0 passing through an uniform layer of material of thickness x is ruled by the Lambert-Beer law:

$$I_T(E) = I_0 e^{-\mu(E)x} \tag{1.1}$$

where I_T is the transmitted beam intensity. The absorption coefficient is directly proportional to the photon cross section, and when the material is constituted by n different atomic species it can be written as a summation over the atomic densities ρ_i and the absorption cross sections σ_i of the constituent atoms:

$$\mu(E) = \sum_{i=1}^{n} \mu_i(E) = \sum_{i=1}^{n} \rho_i \sigma_i(E) \tag{1.2}$$

The experimental measurement of $\mu(E)$ can be done by direct measurement of the transmitted x-ray beam intensity as a function of the energy of the incoming photons, as suggested by the inversion of equation (1.1). The general features of the absorption coefficient for all materials are an overall decrease with increasing photon energy and the presence of sharp edges, the position of which is characteristic of each material (Figure 1.2). The edges are due to the excitation of inner
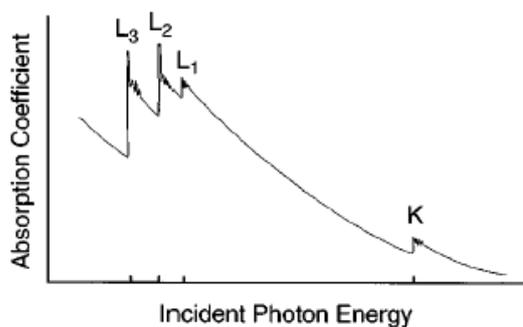
Figure 1.2: Schematic view of the x-ray absorption coefficient as a function of incident photon energy. The figure is taken from [1].

shell electrons: when the energy of the incoming photon is equal or greater than the core-electron binding energy, the photon is absorbed and the electron, called photoelectron, is excited. When one electron is ejected from the 1s, 2s, $2p_{1/2}$ or $2p_{3/2}$ energy level the edge is called respectively K, L_1, L_2 or L_3. The photoabsorption process is depicted in Figure 1.3, and the different edges arising in the absorption coefficient are visible in Figure 1.2. The photoelectron is therefore ejected with a final energy E_f equal to the energy of the incoming photon $\hbar\omega$
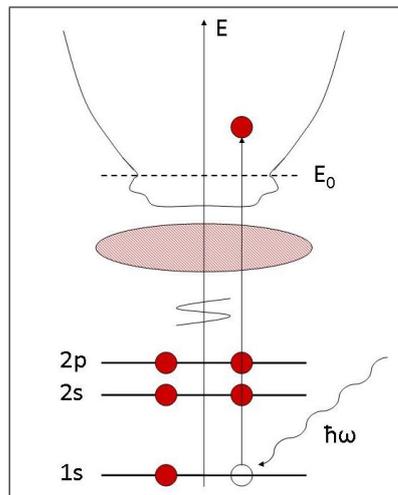
Figure 1.3: Schematic representation of the x-ray absorption process by an 1s electron. This process gives rise to the presence of a so called K-edge in the absorption coefficient.

minus its binding energy E_b, as required by energy conservation:

$$E_f = \hbar\omega - E_b \tag{1.3}$$

and it interacts with the neighboring atoms. Since the momentum of the photoelectron is $\hbar k$, its wave number k is directly derived from equation (1.3):

$$k = \frac{\sqrt{2m(\hbar\omega - E_b)}}{\hbar} \tag{1.4}$$

According to the multiple scattering description of the phenomenon, the fine structure of the absorption spectra can be interpreted by considering the photoelectron as a wave and the atoms as point scatterers: the final state of of the photoelectron is given by the interference between the photoelectron itself and its backscattered wave function. The interference gives rise to a modulation of the x-ray absorption coefficient after the edges, the so called *fine structure* which is the object of X-ray Absorption Fine Structure (XAFS) spectroscopy.

The fine structure is evident both in the near-edge region (XANES, X-ray Absorption Near Edge Structure) and in the extended region (EXAFS, Extended X-ray Absorption Fine Structure), and is caused by the same physical process in the two cases. Nevertheless, as we will see in detail in the next paragraphs, only in the EXAFS region some approximations can be done in order to make the mathematical description, and consequently the analysis strategy, much simpler. The XANES region is given by the transitions of the photoelectron to bound

states and to low-energy states in the continuum: it contains information about the 3-D arrangement of the excited atom's coordination shell and on its oxidation state. In the EXAFS region, instead, the energy of the photoelectron is such that its final state lies in the continuum and that it is weakly affected by the neighboring atoms' potential; this allows an easier mathematical formulation in terms of superposition of scattering paths, where the single scattering (SS) processes are dominant and only a limited number of multiple scattering (MS) processes need to be taken into account. The EXAFS spectrum as function of the photoelectron momentum k is:

$$\chi(k) = \frac{\mu(k) - \mu_0(k)}{\mu_0(k)} \tag{1.5}$$

where $\mu_0(k)$ is the absorption coefficient of the isolated atom, in absence of neighboring atoms. A XAFS spectrum is reported in Figure 1.4. The distinction be-
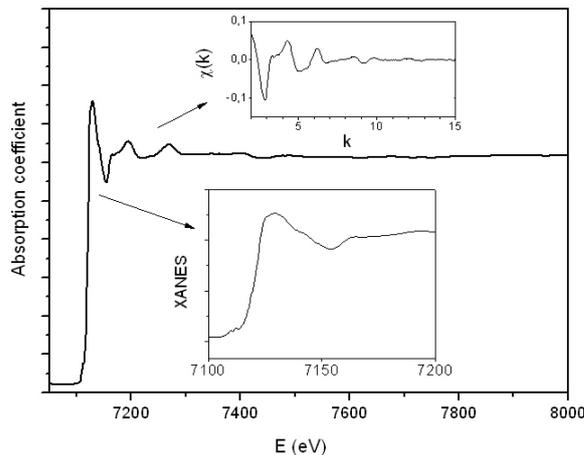
Figure 1.4: Fe K-edge X-ray absorption spectrum of Carbonmonoxy Myoglobin (MbCO). In the upper inset the EXAFS spectrum obtained after background subtraction is reported. A magnification of the XANES region is visible in the lower inset.

tween the XANES and EXAFS region is quite arbitrary and is in general located at roughly 50-70 eV after the edge. It is clear that, since a XANES study focuses on the lineshape of the absorption coefficient as a whole while EXAFS requires the extraction of its oscillatory part, the second one is more affected by noise and requires a greater experimental effort. On the other hand, due to its simpler mathematical formulation, several programs are available for a reliable quantitative fitting of EXAFS spectra [2, 3, 4], while the quantitative analysis of XANES spectra is still under development [5, 6, 7] and requires a considerable amount of computational time.

1.2 Approaches to the derivation of the absorption coefficient

In this paragraph the general lines of XAFS theory are presented. The purpose is to understand the key approximations assumed in the theoretical framework, and what kind of approaches can be chosen to treat the problem, especially in relation to the understanding of the data analysis programs used in this work. A detailed description of XAFS theory can be found in several books and reviews [8, 1, 9]. The derivation of the absorption coefficient is in principle a complicated many-body quantum mechanical problem, but a number of key approximations allow to simplify the approach. First of all the one-electron approximation must be invoked, which means that we consider only the photoelectron to be excited by the incoming radiation, while the other N-1 electrons of the irradiated atom are passive spectators. The absorption coefficient is proportional to the photoabsorption cross section (eq 1.2), the expression of which is given by the Fermi Golden Rule:

$$\mu(E) \propto \sigma(E) \propto \sum_f |M_{gf}|^2 \delta(E - E_f) \tag{1.6}$$

where M_{gf} is the matrix element providing the probability of the transition of one electron from the ground state to the final state. A second approximation, the dipole approximation, consists in neglecting the spatial variation of the electromagnetic field inducing the interaction; it allows to write the perturbing Hamiltonian as proportional to $\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r}$, where $\hat{\boldsymbol{\epsilon}}$ is the x-ray polarization vector and $\boldsymbol{r}$ the position vector. The form of the matrix element appearing in eq. 1.6 is therefore:

$$|M_{gf}|^2 = |\langle f|\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r}|i\rangle|^2. \tag{1.7}$$

The initial state $|i\rangle$ can be represented by an atomic core orbital, while the final state $|f\rangle$ can be either an unoccupied bound state or a state in the continuum. The main difficulty in the theoretical calculation of XAFS spectra resides in the computation of the final state; the problem can be addressed making use of different approaches, in particular we mention the ones that are implemented in the data analysis software employed for this work:

i) The Real Space Multiple Scattering (RSMS) approach, that avoids explicit determination of the eigenstates by describing the final state as a superposition of scattering paths.

ii) The Finite Difference Method (FDM) approach, that solves the Schrödinger

equation by discretization on a spherical volume centered on the absorbing atom and therefore directly calculates the final states.

The RSMS approach has been developed starting from the early stages of the EXAFS technique [10] and extended later on to the interpretation of the XANES spectra [11]; an interesting summary of its historical evolution is given in reference [9]. According to a modern formulation of the RSMS theory [2], the golden rule 1.6 can be written in the spectral representation, in terms of the one-electron Green's function G:

$$\mu(E) \propto \frac{2}{\pi} Im\langle i|\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r}' G(\boldsymbol{r}', \boldsymbol{r}, E)\hat{\boldsymbol{\epsilon}} \cdot \boldsymbol{r}|i\rangle \tag{1.8}$$

This formula shows how the problem of the computation of the final states for the photoelectron is replaced by the computation of the propagator $G(\boldsymbol{r}', \boldsymbol{r}, E)$. Once the potential is chosen, radial wave functions and partial phase shifts can be obtained by solving the Dirac equation and used to construct the Green's function. The choice of the potential is crucial, and the approximations employed at this stage dramatically influence the agreement between theoretical calculations and experimental data, in particular in the XANES region; near the absorption edge, indeed, the photoelectron has a low kinetic energy and is therefore strongly sensitive to the details of the scattering potential. The RSMS approach is always associated to the Muffin Tin (MT) potential, a description of which will be given in paragraph 1.2.1. The propagator $G(\boldsymbol{r}', \boldsymbol{r}, E)$ is given by the sum of a central atom G^C and a scattering contribution G^{SC}. The latter is the sum over the quantum numbers $L = (l, m)$ and $L' = (l', m')$ of a series of contributions $G^{SC}_{Lc,L'c}$ (where the subscript c indicates the central atom) that can be expressed formally by the Multiple Scattering (MS) expansion:

$$G^{SC}_{Lc,L'c}(E) = \sum_i G^0_{Lc,L_1i}t_{l_1i}G^0_{L_1i,L'c} + \sum_{i_1,i_2} G^0_{Lc,L_1i_1}t_{l_1i_1}G^0_{L_1i_1,L_2i_2}t_{l_2i_2}G^0_{L_2i_2,L'c} + ...$$
$$\tag{1.9}$$

where $G^0_{L,L'}$ are the matrix elements of the free propagator and t_l the elements of the scattering matrix T, which is diagonal ($T_{l,l'} = t_l\delta_{l,l'}$) when the potential is spherically symmetric, as for MT potential; the indices i run over the different atomic centers in the structure and summation over angular momentum indices is implicit. Equation (1.9) is one of the possible formulations of the MS series: it tells us that the x-ray absorption coefficient is given in principle by the summation of an infinite number of contributions, corresponding to all of the possible

scattering paths experienced by the photoelectron in the material. A graphical representation of some scattering paths is given in figure 1.5, where the central atom is Fe.
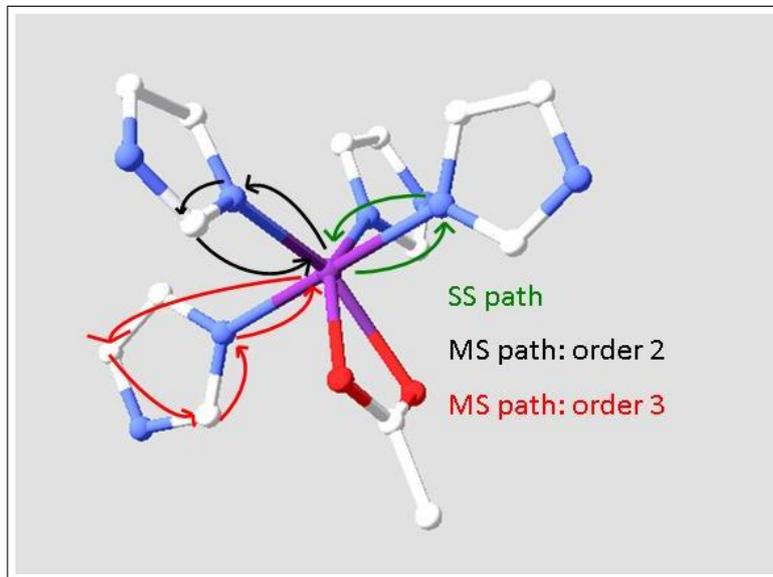
Figure 1.5: Visual representation of single and multiple scattering paths corresponding to the terms of sum 1.9. The depicted cluster is the Fe site of photosynthetic Reaction Center.

The definitions of EXAFS and XANES regions stem from this formula, being the EXAFS region of the absorption spectrum the one where the series 1.9 converges and a few terms approximate very well the modulations of the absorption coefficient (see paragraph 1.2.2), to the contrary of the XANES region. To obtain a reliable prediction of the XANES spectrum for a given material it is necessary to sum MS contributions over all orders; it can be done in the so called Full Multiple Scattering (FMS) treatment, by inverting the matrix in:

$$G^{SC} = (1 - G^0 t)^{-1} G^0 \tag{1.10}$$

where the angular momentum and atomic site indices have been omitted for brevity. The RSMS approach described so far provides a clear explanation of the physical phenomena involved in XAS; it is implemented in the code feff8.2 [2], which is able to calculate both EXAFS and XANES spectra making use of the approximations described above, together with the MT shape for the scattering potential.

1.2.1 Choice of the potential

The potential experienced by the photoelectron in the the atomic cores regions is dominated by the atomic charge densities, in the outer regions of the atoms and in the interstitial regions by the distribution of charge determined by the bonding properties of the material. In the latter case, the potential is anisotropic and it depends on the nature of atoms and bonds present in the material. The total potential appearing in the Schrödinger equation is given by the sum of this Coulomb contribution V_{coul} and a self-energy term $\Sigma(E)$:

$$V(E) = V_{coul} + \Sigma(E). \tag{1.11}$$

Photoelectrons with energies greater than $\sim$30 eV above the absorption edge are strongly scattered by the core regions and negligibly influenced by the details of the anisotropic interstitial potential. In such a condition, i.e. in the EXAFS energy region, the Muffin Tin (MT) approximation for the shape of the potential is particularly suitable: the potential is spherically averaged inside MT spheres centered at each atomic core, while a constant value is assigned to the interstitial region. A pictorial representation is given in figure 1.6(a). The approximation can be extended also to the near-edge energy region, provided that nonspherical corrections are included; *feff* includes these corrections by allowing MT spheres to overlap, furthermore it couples the use of the MT shape with Self Consistent Field (SCF) calculations of the potential [2]; finally it adds the Hedin-Lundqvist form of the self-energy term of equation (1.11) and uses this potential to compute the propagators appearing in equation (1.9).

The tools mentioned above allow to extend the MT approximation of the potential to the calculation of the XANES spectra for a wide set of structures; nevertheless, independent tests of the method performed on metalloproteins [12, 13] showed that the approximation holds as long as the cluster of interest is close to spherical symmetry (for example in 6-coordinated octahedral metal sites), but it fails when the cluster is radically different from this geometry (for example in 4-coordinated tetrahedral metal sites).

An alternative method to deal with theoretical calculations of XANES spectra is the Finite Difference Method implemented in the FDMNES program [5, 14]. The FDM is a general way to solve differential equations by discretizing them over a grid of points in the volume of interest, and in this case it is applied to the Schrödinger equation over a cluster centered on the absorbing atom. In this approach the volume is divided in three regions, corresponding to three shapes

of the potential:

1. atomic-core regions (up to 0.65 Å around each atomic center), where the potential is spherically symmetric

2. interatomic region, where no approximation is made on its shape

3. outer-sphere region, where it is assumed to be constant.

The three zones are represented in figure 1.6(b). In region 2), the Local Density Approximation (LDA) is used to calculate the potential, as described in [5]: at first the electronic density is computed starting from superposition of atomic densities, then the Poisson equation is solved to get the Coulomb term of the potential of equation (1.11), finally the exchange-correlation term is added following the Hedin-Lundqvist approach. The program solves the Schrödinger equation on
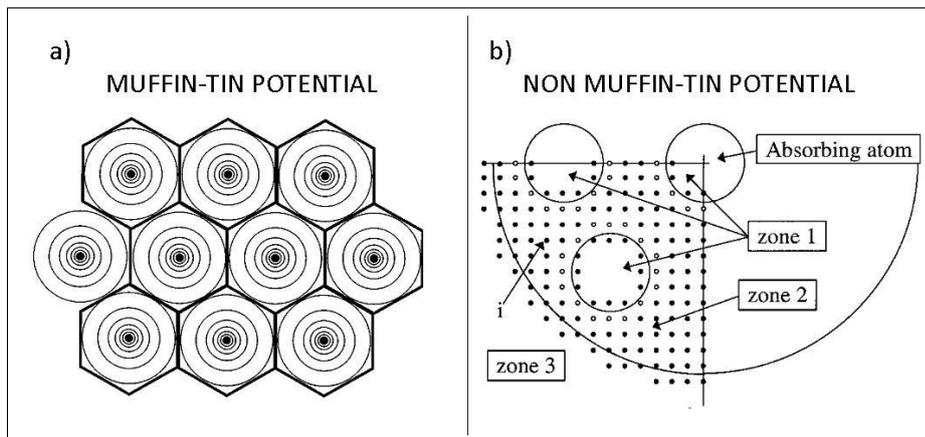
Figure 1.6: Schematic two-dimensional representation of (a) Muffin Tin potential and (b) the three regions considered by FDMNES in the non-MT approach. Panel (a) is taken from reference [1], panel (b) from reference [5].

the mesh points imposing the continuity of the wave function and of its derivative between the different regions. This procedure corresponds to the numerical calculation of the final states $|f\rangle$ of equation (1.7).

1.2.2 The EXAFS function

Sufficiently far from the absorption edge, the series 1.9 converges; this allows a relatively straightforward mathematical description of the oscillatory part of $\mu(k)$ in terms of the superposition of photoelectron waves backscatterd by the surrounding shells of atoms. The standard EXAFS formula, the components of

which will be explained in this paragraph, reads:

$$\chi(k) = S_0^2 \sum_i \frac{N_i |f_{eff}(k, R_i)|}{k R_i^2} sin(2kR_i + 2\delta_c + \Phi) e^{-2R_i/\lambda(k)} e^{-2\sigma_i^2 k^2} \qquad (1.12)$$

where the index i runs over the shells of atoms, each one charachterized by a distance R_i from the absorber, and composed by N_i atoms. S_0^2 is the amplitude reduction factor, a parameter that accounts for the corrections to the one-electron approximation by considering the relaxation of the N-1 passive electrons. It is defined as the superposition of the many-body wave function of the N-1 passive electrons before ($|\Psi^i\rangle$) and after ($|\Psi^f\rangle$) ejection of the photoelectron:

$$S_0^2 = |\langle \Psi_{N-1}^i | \Psi_{N-1}^f \rangle|^2. \qquad (1.13)$$

The amplitude reduction factor is in general empirically determined and its value ranges from 0.7 to 1.0, depending on the experimental conditions.

$|f_{eff}(k, R_i)|$ is the modulus of the effective backscattering amplitude relative to an atom in shell i, and Φ is its phase factor: $f_{eff}(k, R) = |f_{eff}(k, R)| e^{i\Phi(k)}$; for a correct prediction of the EXAFS function, the calculation of f_{eff} must take into account the curved-wave nature of the photoelectron [15]. δ_c is the phase shift due to the central atom and $\lambda(k)$ the energy-dependent mean free path of the photoelectron. The exponential term $e^{-2R/\lambda}$ is therefore responsible for the decay of the EXAFS signal due to the finite lifetime of the photoelectron, which determines the short-range nature of this technique; EXAFS allows indeed to probe a cluster of a few tens of Å in crystals and ~ 5 Å in biomolecules, where no long range order is present.

The last element of the formula is the Debye-Waller (DW) factor σ_i^2, which represents the mean square variation in the distance between the absorber and the atoms of shell i. The Radial Distribution Function is described in the Gaussian approximation (as shown by the term $e^{-2k^2\sigma^2}$ in equation (1.12)), for which σ_i^2 is the second order cumulant. The DW factors are therefore defined as the following configurational average:

$$\sigma_i^2 = \langle [\hat{\boldsymbol{r}}_{ci} \cdot (\boldsymbol{u}_c(t) - \boldsymbol{u}_i(t))]^2 \rangle \qquad (1.14)$$

where $\hat{\boldsymbol{r}}_{ci}$ is the unit vector relative to the segment connecting the absorber and atom i, $\boldsymbol{u}_c$ and $\boldsymbol{u}_i$ are the instantaneous displacements of the two atoms with respect to their equilibrium positions. The presence of the DW factors results in the damping of each contribution to the EXAFS oscillations, since they appear

in equation (1.12) as exponential terms; the reasons for the damping are both disorder and thermal vibrations in the material, generally mentioned as statical and dynamical component of the disorder respectively.

DW factors are crucial for the correct interpretation of EXAFS spectra, to the contrary they can often be neglected in XANES calculations, since the exponential decay depends on k^2. In recent years, several studies based on Density Functional Theory (DFT) [16, 17, 18, 19, 20] have been developed, aimed to perform *ab initio* calculations of the DW factors in aperiodic systems. In particular Dimakis and Bunker, in their works focused on Zn sites [17, 18], provide a parametrization of DW factors relative to Zn-amino acid bondings as a function of temperature that can be immediately applied in any other study relative to Zn-metalloproteins; in Chapter 3 an application of this parametrization will be shown.

In the approach of Vila et al. [19], the damping due to the DWs is expressed in terms of cumulants expansion up to third order, which represents a correction to the Gaussian approximation (equation (1.14) is the second order term in the expansion), and they are extracted from the dynamical matrix calculated with DFT methods; in Chapter 5 an original test of this approach on Fe-MbCO will be presented.

1.3 Experimental techniques

XAFS techniques require first of all continuous x-ray over a wide energy range, which can be provided only by a synchrotron radiation source. Nevertheless, the development of BioXAS occurred much later than XAFS studies on crystalline solids, as a consequence of its more demanding experimental requirements that have been fully satisfied only with the advent of third-generation synchrotron light sources, in the early 1990s. Metalloproteins belong indeed to the category of diluted samples (absorber concentration $\leq 1\%$), they require therefore a high-intensity and focussed x-ray beam to give rise to a detectable XAFS signal; because of dilution their absorpion coefficient must most often be measured in the fluorescence mode (see Paragraph 1.3.1).

Two major types of beamlines (and relative experimental stations) are available in synchrotron radiation laboratories:

- Bending magnet (BM) beamlines: a magnet placed on the storage ring curves the electron beam and causes electrons to emit synchrotron radiation tangentially to the ring. The X-ray radiation obtained in this way is characterized by moderately high flux (10^9-10^{10} photons/s).

13

- Insertion device (ID) beamlines: an array of magnets is inserted into a straight section of the storage ring, producing high-intensity synchrotron light. The flux of the X-ray radiation in this case is higher (10^{12}-10^{13} photons/s).

Two kinds of insertion devices exist: undulators and wigglers. In undulators, the period of the array of magnets is tuned to the frequency of the radiation produced by the electrons. Hence, when electrons traverse the structure, they undergo small amplitude oscillations that interfere with each other, producing very intense radiation in a narrow energy band. In wigglers instead, the applied magnetic fields are higher and they are not tuned to the radiation produced by the electrons; each electron radiates therefore independently, giving rise to a broader spectrum.

When XAFS spectra are recorded on a BM beamline, practicle problems could be encountered if the sample is very diluted, because in such case the flux could be insufficient to record good signal-to-noise (S/N) spectra; however, the flux of BM beamlines is suitable for the great majority of BioXAS experiments. In ID beamlines, to the contrary, good S/N spectra are more easily achieved, but attention must be payed to the preservation of the sample during the measurements; radiation damage could indeed occur [21], as well as a partial degradation of the protein structure upon irradiation. Besides the x-ray dose, also temperature and exposure time concur to activate radiation damage, therefore it is common practice to irradiate the protein at low temperature and minimize the exposure time; however, since the effects of radiation damage are clearly visible in the XANES region, it is always important to check if any changes in the lineshape occur between following spectra acquired on the same spot of the sample.

1.3.1 Detection modes

The most straightforward way to measure the x-ray absorption coefficient is the transmission mode, which stems directly from the inversion of equation (1.1). It requires the direct measurement of the flux of the incoming beam (I_0) and of the transmitted beam (I_T), in order to evaluate the quantity:

$$\mu(E)x = ln\frac{I_0(E)}{I_T(E)}. \tag{1.15}$$

Fluxes are in general measured by means of ionization chambers, choosing the gas and its pressure depending on the energy range of interest. It has been demonstrated [10] that the value $\mu x \approx 2$ is the one that maximizes the signal

S/N in transmission measurements. Samples must be thin, typically of the order of 10 μm, and horizontally homogeneous; spectra measured in the transmission mode usually have a high S/N, resulting in high quality data when it can be applied.
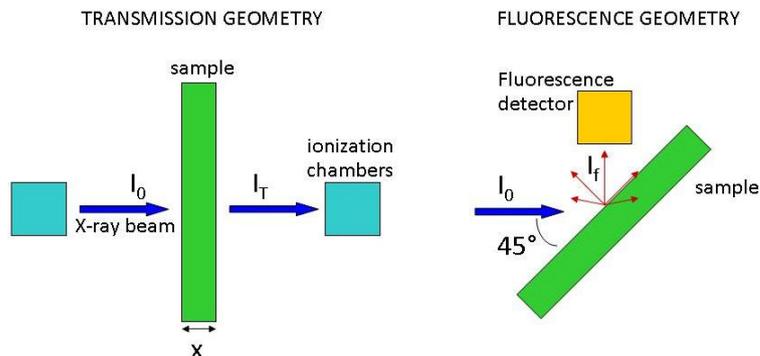
Figure 1.7: Schematic representation of typical detection geometries for transmission (left panel) and fluorescence (right panel) XAFS experiments.

Unfortunately this is almost never the case for metalloproteins, because in diluted samples the variations in the absorption coefficient due to the absorbing metals are not distinguishable from the background. In such cases the fluorescence detection technique is most often employed: it relies on the physical principle of radiative de-excitation of the absorber upon photoemission and allows therefore to collect only the signal due to a specific element; this property makes the fluorescence detection mode particularly suitable for diluted samples. The standard setup for a fluorescence experiment is shown in figure 1.7: the detector is placed at an angle of 90° with respect to the incoming beam, i.e. in the same direction of the polarization vector of the beam, since the elastic scattering is minimum in this direction. In this condition, and for uniform samples, the intensity of fluorescent radiation revealed by the detector is [10]:

$$I_f(E) = I_0 \epsilon_f \frac{\Omega}{4\pi} \cdot \frac{\mu_A(E)}{\mu_T(E) + \mu_T(E_f)} \tag{1.16}$$

where ϵ_f is the fluorescence yield, $\frac{\Omega}{4\pi}$ is the solid angle acceptance of the detector, μ_A is the contribution of the absorber A to the total absorption coefficient μ_T, and E_f is the characteristic fluorescence energy. If $\mu_A \ll \mu_T$, the intensity of the fluorescence radiation is proportional to μ_A, which explains why this technique is suitable for diluted samples. The fluorescence energies are characteristic of each metal, and they are conventionally named depending on the energy level of the electron that undergoes the transition, as illustrated in Figure 1.8. To exploit
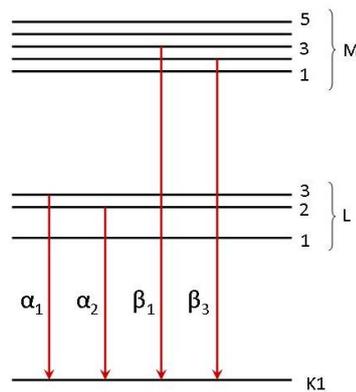
Figure 1.8: Transitions that give rise to the fluorescence emission.

the selectivity of this technique, the fluorescence detector must be able to isolate the energy region corresponding to the emission line of the absorber, in such a way that background signal due to Compton and elastic scattering and to Bragg peaks do not contribute to the signal.

1.3.2 Standard XAFS beamline setup

In a beamline designed for XAFS experiments, the broad band X-ray radiation produced by the electron beam accelerated in the storage ring is first of all conveyed through a series of optical elements in such a way to produce a monochromatic focused X-ray beam; the beam is then conveyed to the experimental hutch, where an experimental chamber and detection elements for transmission and fluorescence experiments are present. A schematic view of the beamline layout is given in Figure 1.9.
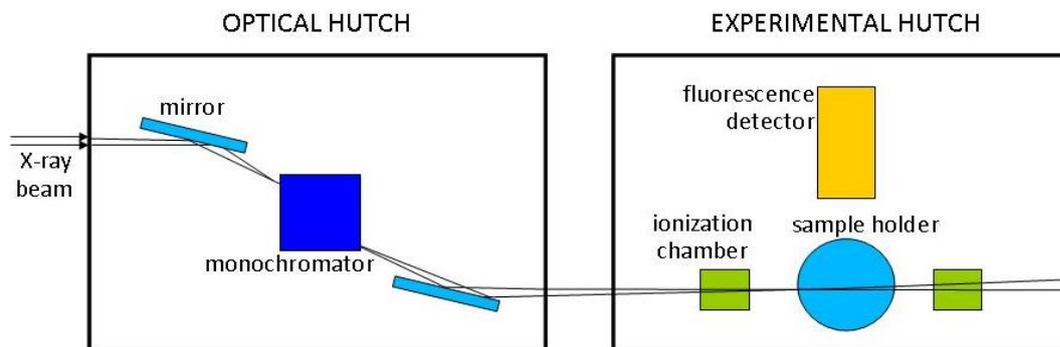
Figure 1.9: Pictorial side view of the basic components of a XAFS dedicated beamline.

The central element in the optical hutch is the monochromator, which selects the energy $E = \hbar\omega$, within a width ΔE, from the incoming white beam.

Monochromators are based on the principles of perfect crystals, the outcoming wavelength λ is the one allowed by Bragg's law:

$$2d_{hkl}sin\theta = n\lambda \qquad (1.17)$$

where d_{hkl} is the distance between crystallographic plans, θ the incidence angle and n an integer; clearly, not only the fundamental wavelength is reflected but also the harmonics for which the form factor F_{hkl} is not zero; the harmonics are a source of distortion for XAFS. The energy tuning is achieved by varying θ, i.e. rotating the crystal around an axis parallel to the Bragg plans and perpendicular to the beam direction. An important parameter for monochromators is the energy resolution $\Delta E/E$, obtained by differentiating equation (1.17):

$$\frac{\Delta\lambda}{\lambda} = \frac{\Delta E}{E} = \Delta\theta \cdot cotg\theta \qquad (1.18)$$

it depends therefore on the angular spread $\Delta\theta$, determined by the angular spread of the incident beam and the intrinsic reflection width of the monochromator; typical values for $\Delta E/E$ are in the range $10^{-4} \div 10^{-5}$. The exact theory of monochromator physics for design and realization is derived in the dynamical diffraction framework [22]. In general the monochromator is associated to the presence of mirrors that reject the harmonics and focus the beam. Harmonic rejection is based on the principle of total reflection: since the critical angle for total reflection depends on the wavelength, it is possible to choose the incidence angle in such a way to reflect only the fundamental wavelength; moreover, the mirror surface is slightly bent to focus the beam in the vertical plane.

The sample is mounted on the sample holder in the experimental hutch. Even if measurements at room temperature can in principle be carried out also in air, a vacuum system is highly desirable when fluorescence emission has to be detected, and in particular for low count rates, because it helps to maximize the fluorescence yield and minimize the air scattering. Moreover, when proteins are embedded in solid matrices, the vacuum environment reduces the amount of residual water in the sample and consequently the mobility of the protein, which would the EXAFS signal to damp more quickly. A temperature regulation system is always integrated with experimental chambers, with a cooling system based either on liquid N (77 K) or liquid He (4 K) flow; low temperature is the ideal condition to measure protein samples, since it lowers the thermal disorder and minimizes the radiation damage. Several types of fluoresce detectors exist: inorganic and organic scintillators, solid-state detectors, multi-element energy-

resolved detectors. The latter are multi-element semiconductor (generally Ge) detectors, and constitute at present the standard detection system in BioXAS experiments.

The great majority of the experiments presented in this work has been carried out at the beamline GILDA BM08 of ESRF. For this reason, GILDA is taken here as an example of XAS-dedicated beamline and its specifications are recalled.

GILDA BM08 beamline of ESRF. The General purpose Italian beamLine for Diffraction and Absorption (GILDA) [23] exploits a 0.8 T Bending Magnet on the European Synchrotron Radiation Facility (ESRF) operating at 6 GeV with typical currents of 100-200 mA. The beamline is characterized by an energy resolution $\Delta E/E \sim 10^{-4}$ and a maximum flux of 10^{11} ph/s, and operates in the energy range 5-50 keV. Energy tuning, in the energy range 5-30 keV, is performed by a double crystal monochromator made of two Si(111) or Si(311) single crystals (the Si(111) configuration provides higher flux but lower energy resolution than the Si(311)) employing dynamical sagittal focusing [24]; the horizontal focusing is achieved by allowing the second Si crystal to bend. Harmonic rejection and vertical focusing are performed by means of two mirrors, the surface of which is divided into two parts, one coated with Pd, the other with Pt; the two different coatings allow to reject the harmonics in the whole 5-30 keV range. The provided fluorescence detector is a 13-element hyper-pure Ge detector equipped with fast digital electronics and a peaking time of 1 μs [25].

1.4   Current status of BioXAS

Protein metal sites raise great interest in the scientific community, since they are responsible for the life-sustaining processes operated by protein complexes. Nowadays, crystallographic structures are provided for a large number of proteins, however, a high-resolution structure of the metal site is not always attainable, neither with XRD nor with NMR; the knowledge of the details of the metal ion binding, of its electronic structure and oxidation state are most often necessary to account for the protein function and to face the subject of structure-to-function relation underlying structural biology. XAFS allows to resolve the geometry of a cluster of ~ 5 Å around the absorber, with a precision of the order of ~ 0.01 Å, therefore, since it was first applied to biomolecules, it has been considered as a valuable tool to complement crystallographic information. A modern application of the role of XAFS as a technique complementary to XRD consists in the inte-
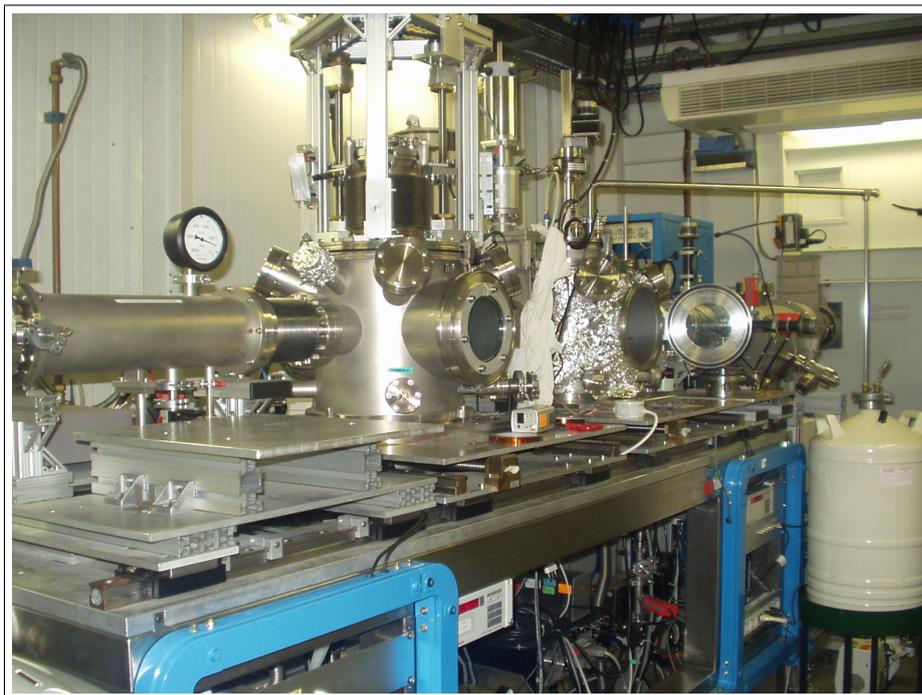
Figure 1.10: Experimental hutch of the beamline GILDA-BM08 of ESRF.

gration of both techniques in one experimental setup, in order to allow *in situ* observation of the metal site during diffraction measurements and track eventual changes in its geometry and/or its oxidation state [26]; in this regard, the combined approach has also been used for the determination of the best experimental conditions for x-ray crystallography, in order to avoid radiation damage [27]. Besides crystallography, several other techniques have been matched to XAFS to bring insight into protein structure (reviewed in [28]), like Raman spectroscopy, anomalous dispersion, and computational chemistry.

Even though its short range contributed to develop the consideration of XAFS as a complementary technique, in recent years much effort has been put in making XAS a self-sufficient tool: data analysis softwares and strategies that serve this purpose are now available or under development, as we will explain in detail in Chapter 2. However, some limitations are intrinsically included in this technique and they cannot be overcome with the improvement of data analysis: when several binding sites for the same metal are present, the recorded XAFS spectrum is given by their averaged signal, from which the individual contributions cannot be singled out. This aspect can at present only be treated empirically; in the case where the metal is exogenously added, as for inhibitory metals, the best solution is to incubate the sample with sub-stoichiometric amounts of metal, in order to maximize the occupancy of the highest affinity site. The case of double

metal sites, instead, can only be treated in the analysis procedure and brings the disadvantage of a significant increasing of the number of putative models to be considered.

The most challenging subjects faced at present with XAS as a standalone technique are probably the time-resolved studies. Time-resolved XAS was initially conceived in association to dispersive geometry [29], therefore measurements could only be performed in the transmission mode, which made it impossible to extend the technique to proteins (i.e. diluted samples). A way to perform pump-probe time-resolved XAS studies in the fluorescence mode was adopted by Haumann et al. [30]: it consists in the irradiation of the protein sample at a fixed energy value, corresponding to a significant feature of the absorption spectrum of the target metal. In this way, the kinetics of the reactions involving the metal can be tracked, and transient reactions intermediates can be detected, as it was done for the Mn site of Photosystem II (for a review see [31]). Subsequently, Kleifeld et al. proposed to trap the protein sample in its transient state by rapid freeze-quench, in order to increase the time available for fluorescence measurements and collect full XAFS spectra [32].

Chapter 2

XAFS analysis methods for metalloproteins

XAFS allows high resolution ($\sim$0.01 Å) determination of the local structures of metal sites in proteins, nevertheless the analysis procedure is at present feasible only for expert users and it is in general complicated by the presence of multiple solutions and by correlation between variables. Moreover, a disadvantage of XAFS is that, to bring relevant contribution to the knowledge of a protein system, it often requires a priori information on the protein itself, for example its whole structure resolved by X-Ray Diffraction (XRD) or Nuclear Magnetic Resonance (NMR); for these reasons XAFS has often been considered as a complementary technique.

Only very recently the first "black-box" approach for BioXAS analysis has been developed [33], the aim of which is to make the technique available also to non expert users. However, in spite of the usefulness of this approach in spreading EXAFS through the protein research community, we believe that also new users might prefer a non-blind approach to EXAFS analysis. Therefore we have developed an original analysis method based on a combination of XANES simulations, observations of recurring spectral features, and parametrization of the Debye Waller factors based on DFT calculations for mononuclear Zn sites. This collection of evidences based on physical phenomena helps to reduce the number of putative clusters to fit to the experimental data and makes the analysis procedure easier to handle in an insightful way.

In this chapter we introduce the basic steps of a XAFS analysis; at the same time we present our collection of phenomenological evidences and their validation on known Zn sites. It all proves that our method can be generalized in order to lead to the unambiguous interpretation of BioXAS data, with the final objec-

tive of making this technique free from the need of a priori information. The results were published as: Giachini et al. *Synergic approach to XAFS analysis for the identification of most probable binding motifs for mononuclear zinc sites in metalloproteins* J. Synchrotron Rad., 2010, 17, 41-52 (reference [34]).

2.1 XANES analysis

The near-edge region of the absorption spectrum is very sensitive to the oxidation state of the absorbing atom and to the 3-dimensional arrangement of first shell ligands; the measurement of this region of $\mu(E)$ is convenient for the user, because it does not require the atomic background subtraction and it usually has a high signal-to-noise ratio. On the other hand, the importance of multiple scattering effects in this region makes the mathematical formulation so complicated that only a few programs are able to fit the XANES spectra with a limited number of structural models (MXAN [35], FDMNES [5] that implement both simulation and fitting tools, and FitIt [7] that fits given calculated spectra using multidimensional interpolation approximation), and they all require a significant amount of computational time. XANES analysis is at present mainly qualitative, in particular one monitors some features of the spectra that are considered as directly related to physical properties of the metal site, for example:

- The edge position changes when the oxidation state of the absorber changes [36, 37]; for example a Fe(II) $\rightarrow$ Fe(III) transition in a non-heme Fe compound causes an energy shift of the order of 2-3 eV [38]. The edge position is therefore monitored also when the oxidation state of the absorber is not the direct target of the experiment, in order to check if the sample has undergone radiation damage.

- Some pre-peak features are related to transitions that are forbidden by the selection rule $\Delta l = \pm 1$, derived in the dipole approximation [39]. Since the selection rule holds strictly only for atomic orbitals with spherical geometry, the further the coordination sphere of the absorber is from sphericity, the higher will be the intensity of the pre-peak.

- The correlation between the white-line (WL) intensity and the coordination number is very well documented for metalloproteins for a wide variety of absorbing atoms [40, 12, 41, 42, 43]. In particular, it has been observed that the WL intensity increases with the coordination number. The existence of

such correlation can be understood qualitatively in the framework of molecular orbital theory. In fact, the density of final states due to unoccupied molecular orbitals is expected to increase, for similar ligands with similar bond lengths, with the numbers of neighbors; since the absorption cross section is directly proportional to the density of final states, the observed correlation is reasonable.

These features are in general compared to the corresponding ones in reference compounds; this approach is very effective, but since the model compounds are chosen on purpose to be compared with the particular system under study, it often lacks generality. In order to obtain more general information about the XANES features of Zn protein sites, we performed *ab initio* simulations of the spectra relative to a selected set of Zn binding motives; the approach consists in identifying the commonest binding motives occurring in data bases and systematically simulating their XANES, in such a way to have a spectra data base besides the structural one. Some information from the simulated spectra can be extracted and applied whenever an unknown binding site is investigated.

2.1.1 Database search

Before tackling the XANES simulations, a data bank search is needed to determine which starting clusters should be considered; this can be achieved by means of the Metalloprotein Database and Browser (MDB, [44]) statistical analysis feature and of the MESPEUS database [45], both available on the web with open access. For each chosen metal and coordination number, the MDB search gives an output histogram where the commonest sites and their frequencies are represented. We chose only mononuclear Zn sites where no ligands other than amino acids from the protein are involved; for each coordination number, the binding motives encountered in more than 15 PDB entries, reported in Table 2.1, were singled out to form the initial set of 16 clusters to be simulated. The set included only Zn sites with a coordination number of 3, 4 and 5, it was therefore completed by adding the motif *His His H_2O H_2O Asp Asp*, which is the commonest one for coordination 6, with 6 PDB entries.

We notice that no Cys residues appear in 5- and 6- coordinated clusters identified so far. To extract information about the presence of Cys in these sites, a further database search can be done: searching Zn-Cys bonds in high resolution (≤ 2.0 Å) crystal structures where Zn is 5-coordinated, by means of MESPEUS, we encountered 49 protein structures presenting this pattern, 7 of which have 2

Coordination number	Binding motifs	Counts in MDB
	His His His	39
	Asp His His	19
3	Cys Cys His	17
	His H_2O H_2O	17
	Cys Cys Cys Cys	484
	Asp His His His	187
	Cys Cys Cys His	183
	Cys Cys His His	157
	His His His H_2O	74
4	Cys Cys His H_2O	56
	Glu His His H_2O	27
	Cys Cys Cys H_2O	22
	Asp Asp His Ser	16
	Asp His H_2O H_2O	20
	Asp Asp His His	16
5	Glu Glu His His H_2O	35
	Asp Asp His His His	19

Table 2.1: Ligand patterns for zinc metalloproteins with more than 15 PDB entries according to the MDB. For the coordination number 6 all the ligand patterns have less than 15 PDB entries in the MDB

Cys in the Zn cluster while in the remaining 42 have only one Cys out of five ligands. The same search for 6-coordinated Zn yields no results. We can therefore neglect the presence of more than two Cys in 5-coordinated Zn clusters and of Cys at all in 6-coordinated clusters.

Finally, since Zn clusters with a low number of S and a high number of N/O ligands did not appear among the commonest clusters retrieved through the MDB search, and since the presence of Cys is known to have a crucial influence on XAFS spectra [46], we performed a further set of simulations on a representative set of clusters. For each coordination number N, the first binding motif with one Cys out of the N ligands indicated by the MDB statistical analysis tool was selected. The clusters were: *Cys His His* for N=3 (8 PDB entries), *Cys His His His* for N=4 (12 entries), *Cys His Asp H_2O H_2O* for N=5 (4 entries).

2.1.2 XANES spectra data base

Ab initio XANES simulations of protein metal sites, when performed using programs based on the muffin-tin approximation, often provide poor agreement with experimental data, especially in the case of tetrahedral sites [12, 13]. This can be ascribed to the remarkable difference between the potential experienced by

the photoelectron in the tetrahedral site and the spherical potential represented by the *muffin-tin* approximation; this difference is less pronounced in octahedral sites.

We chose therefore to make use of the Finite Difference Method (FDM), free from muffin-tin approximation, implemented in the FDMNES program [5] to perform the XANES simulations.

In order to build input clusters for simulations, we need to know first of all the binding geometry for each coordination number; in the case of Zn, the ideal stereochemistry is trigonal pyramidal for N=3, tetrahedral for N=4, trigonal bipyramidal or square-based pyramidal for N=5, octahedral for N=6 [47]. It is important to consider that metal coordination shells in proteins are in general quite far from the ideal geometry determined by coordination chemistry, nevertheless there is no evidence to help us deciding what kind of deformation the cluster should present; so for a systematic approach it is reasonable to use the regular models. In addition we need the internal structural parameters from amino acids; we adopted the bond length and angles provided by Engh and Huber [48], derived from a statistical survey of X-ray structures of small compounds from the Cambridge Structural Database (CSD). Finally, as metal-ligands distances we chose the target distances reported in Table 2.2, provided by M.M.Harding and based on the combined analysis of CSD and Protein Data Bank (PDB) determined at or near atomic resolution [49, 50, 51, 52, 53].

	O_{H_2O}	$O_{Asp/Glu}$ monodentate	N_{His}	S_{Cys}
	(Å)	(Å)	(Å)	(Å)
Fe	2.09	2.04	2.16	2.30
Cu	2.13	1.99	2.02	2.15
Zn	2.09	1.99	2.03	2.31

Table 2.2: Metal-ligand distances provided by M.M.Harding [53]

For each selected pattern mentioned in paragraph 2.1.1, the corresponding target cluster was built according to these criteria and the XANES spectrum was simulated; in order to be compared, all of the simulated spectra were normalized at their value corresponding to the energy of 80 eV after the edge. The resulting set of simulated spectra, the majority of which is shown in Figure 2.1, can be used to identify recurring spectral features and can help to extract information about unknown Zn sites just observing its XANES spectrum.
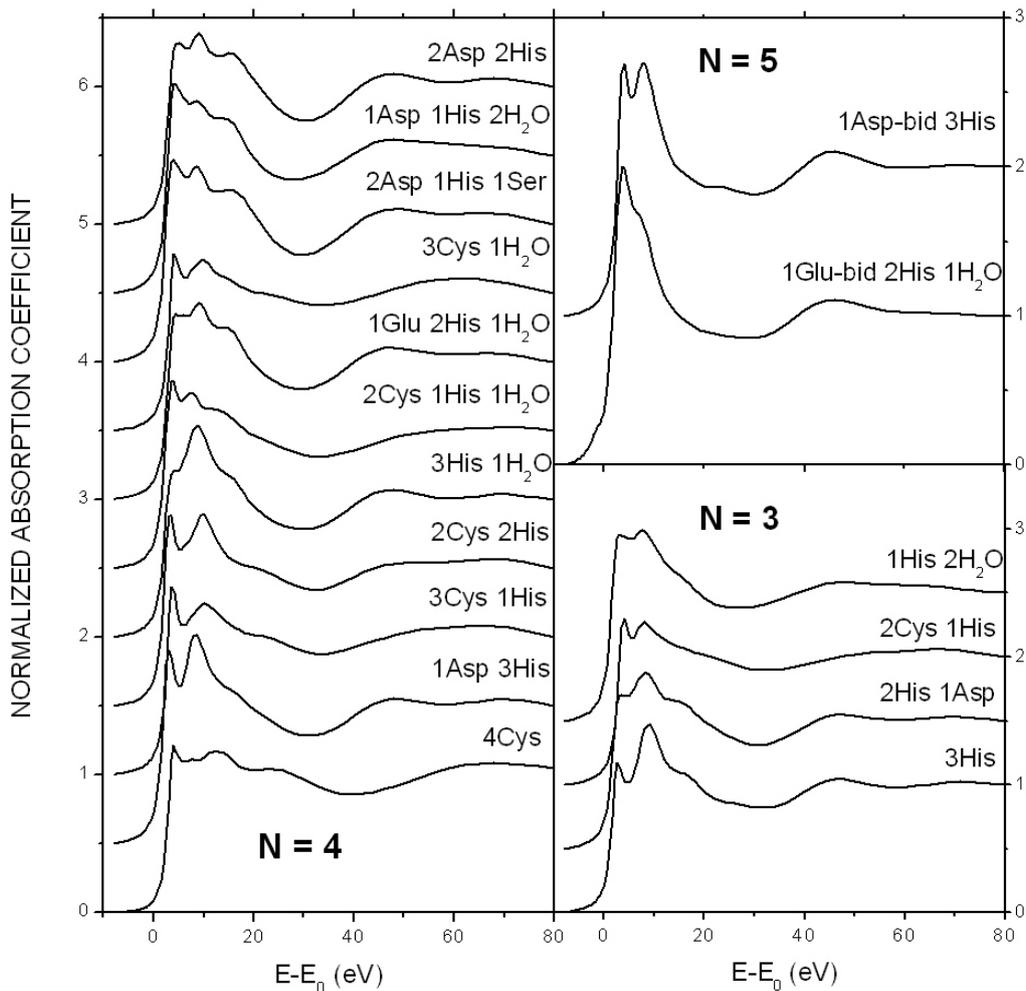
Figure 2.1: XANES simulations performed with the Finite Difference Method implemented in FDMNES for the 17 commonest Zn binding motives retrieved through a MDB search. Each spectrum was normalized to its value corresponding to the energy of 80 eV after the edge.

2.1.3 Correlation between spectral features and ligation patterns

The correlation between WL intensity and coordination number encountered in several experiments is systematically reproduced by simulations. In particular, observing the spectra in Figure 2.1, we notice that the maximum of the normalized simulated absorption coefficient is never higher than 1.5 when Zn occupies a tetrahedral site, in agreement with Feiters et al [40] on experimental XANES spectra of model compounds, and the upper limit of 1.5 is reached by clusters where no Cys residues are involved. To the contrary, for coordination numbers of 5 and 6 (see Figure 2.2(a)), we never observed in normalized spectra a WL intensity lower than 1.6.

26
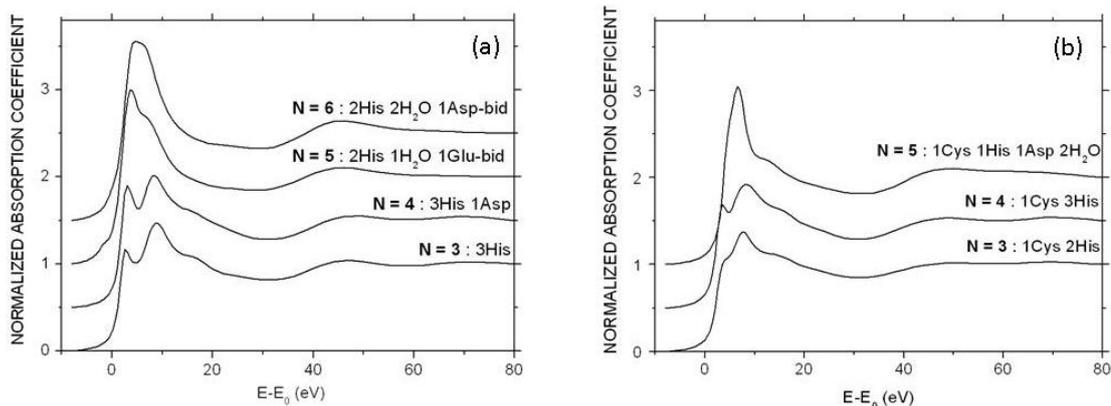
Figure 2.2: Simulated Zn K-edge XANES spectra for binding motives involving the same ligands but with different coordination numbers. Clusters reported in panel (a) present Zn first shells characterized only by ligands with low Z donors (His, carboxylic acids, water molecules); in clusters reported in panel (b) Zn has a mixed first shell, binding one S from a Cys residue and N-1 low Z ligands.

In Figure 2.2(a) we show the simulated XANES spectra for a set of clusters containing the same ligands (His, carboxylic acids and water molecules) but characterized by different coordination numbers and coordinating geometries. This clearly shows that when similar ligands are involved, the WL intensity increases progressively with the coordination number. The clusters giving rise to the spectra reported in Figure 2.2(b) are characterized by the presence of one Cys out of N ligands: we see that the progressive increase of the WL intensity with the coordination number holds also in this case; on the other hand we notice that the WL of the spectrum relative to N=4 with one Cys is lower than the one of the spectrum relative to N=3 with no Cys.

The presence of Cys residues has a strong influence on the site projected density of states, and consequently on the XANES spectrum. This is reasonable, since the donor atom for Cys is S, the atomic number of which (Z=16) differs radically from the one of the other usual donors (N, Z=7 and O, Z=8). Considering the coordination number N=4, for which a wide number of simulations are available (Figure 2.1, left panel), we notice that in the presence of Cys the normalized absorption coefficient extends over a narrower range of values than in the absence of Cys: the difference between its maximum value (reached at or in the vicinity of the white line) and its minimum value (corresponding to the post-edge trough) is indeed lower. Moreover, increasing the number of Cys in the cluster, the post-edge minimum appears at progressively higher energy values (up to a maximum of 40 eV after the edge for the 4Cys cluster). These differences are not marked enough to establish an absolute criterion, nevertheless they can

be taken into account any time the XANES of an unknown Zn site is compared to reference spectra.

2.2 EXAFS analysis in the Rigid Body Refinement scheme

An important step for BioXAS as a technique for structural investigation of metal sites in proteins was reached with the introduction of the Rigid Body Refinement (RBR) scheme for data analysis [4]. The RBR consists in considering the amino acid residues as structural units, allowing them to move rigidly with respect to the absorber. In a quantitative EXAFS analysis it corresponds to the definition of geometrical constrains in the refinement of each atomic position, in such a way to define at most two degrees of freedom for each residue: rigid translation and rotation with respect to the absorber (see Figure 2.3). The analysis package EXCURV98 [54] is currently the only one that implements RBR by including a data base of constrains for each amino acid residue, while in other packages the constrains must be introduced by hand.
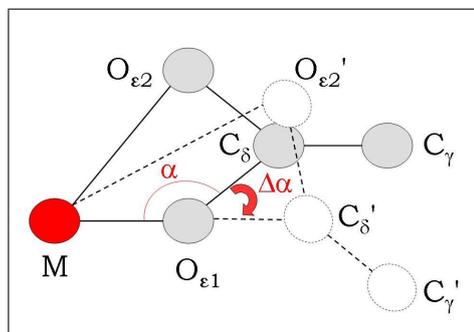
Figure 2.3: Rotation of a Glu residue taken into account in the Rigid Body Refinement EXAFS analysis. The atom marked as M is the metal, the initial positions of the Glu atoms are drawn with continuous lines, the final positions after rotation with dashed lines.

In this paragraph an outline of the EXAFS analysis procedure is given; the EXAFS data presented in the next chapters have been treated with this approach, introducing only slight variations whenever required by the specific scientific problem. Although a great number of EXAFS analysis packages exists, special attention will be given to the *IFEFFIT* package [55] and its graphic interfaces Athena and Artemis [56], since they have been used for the data analysis relative to this work.

Raw data treatment. Once the experimental data have been collected, the first step to undertake an EXAFS analysis is to extract the EXAFS function of equation (1.5) by subtraction of the atomic background. Athena implements the program AUTOBK [57], that automatically performs background subtraction on the raw absorption coefficient $\mu(E)$. The EXAFS function can be extracted using the relation:

$$\chi(E) = \frac{\mu(E) - \mu_0(E)}{\Delta \mu_0(E_0)} \tag{2.1}$$

where $\mu_0(E)$ is the atomic-like absorption coefficient past the edge and $\Delta \mu_0(E_0)$ is the jump in the absorption coefficient at the edge step. The more familiar function $\chi(k)$ is immediately recovered by using equation (1.4), once the edge energy E_0 has been evaluated; E_0 is empirically determined as the energy value of the first maximum of the derivative of $\mu(E)$. The normalization factor $\Delta \mu_0(E_0)$ is then derived, by taking the difference between the estrapolated pre-edge and post-edge lines at the threshold energy E_0 (see Figure2.4).

Figure 2.4: Example of background extraction on the Zn K-edge absorption coefficient of Thermolysin. The jump is evaluated from the estrapolated pre-edge (green trace) and post-edge (pur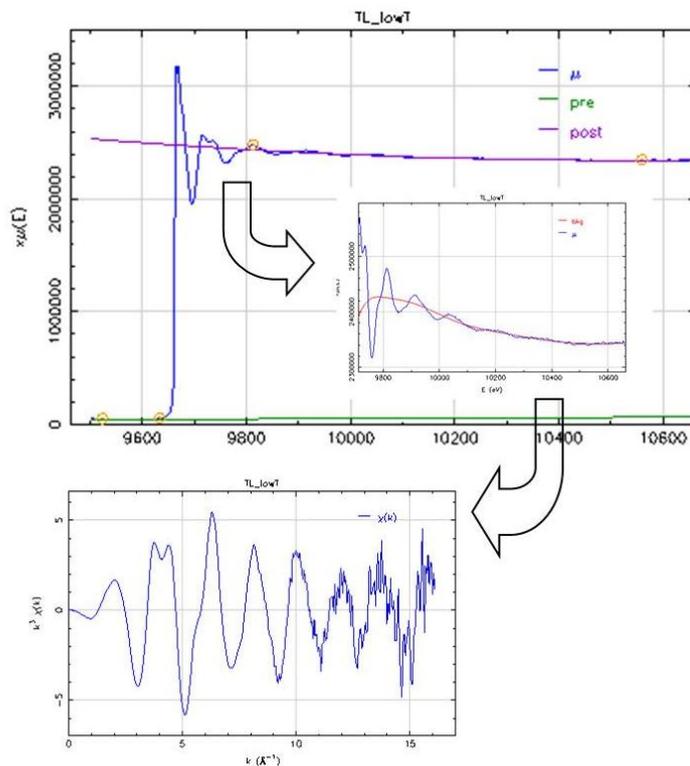ple trace) lines; a polynomial spline (red trace) mimicking the atomic background is subtracted. The k^3 weighted extracted signal is shown in the lower panel.

Finally the atomic background must be determined; this is the most trou-

blesome part of the signal extraction, since the energy dependence of $\mu_0(E)$ is unknown. AUTOBK approximates it using a cubic spline with a number of knots N_{bkg} equal to

$$N_{bkg} = 1 + \frac{2\Delta k R_{bkg}}{\pi} \qquad (2.2)$$

where Δk is the k-range of data and R_{bkg} is the lower limit of the range of significant data in the Fourier transformed signal $\tilde{\chi}(R)$. Equation (2.2) is derived from information theory [58] and it represents the number of degrees of freedom in the data below R_{bkg}. In this range, typically [0, 1] Å (see for example Figure 2.5), no contribution can be given by the first neighbors of the absorber, therefore the Fourier transformed spectrum should in principle be flat. The coefficients of the spline are then chosen in such a way to minimize the components of the signal below R_{bkg}.

First-shell analysis. The EXAFS formula (1.12) clearly shows that the EXAFS signal can be described by a sum of sine waves, each one characterized by an amplitude and a phase; as such, it can be Fourier transformed in the space of distances R in order to separate its frequency components. The resulting signal $\tilde{\chi}(R)$ presents a series of peaks that represent the contributions of each shell of atoms to the post-edge oscillations; examples are given in Figure 2.5. Nevertheless, the EXAFS Fourier Transform (FT) should not be mistaken for the Radial Distribution Function (RDF), from which it differs for several reasons: first of all, in the EXAFS FT the central R value of a peak due to a shell of atoms does not correspond to the actual distance of the atoms from the absorber. This is a consequence of the phase shift ($2\delta_c + \Phi$ in equation (1.12)) caused by the interaction between the photoelectron and the scattering atoms, and it depends on the atomic number of the atoms. Moreover, Multiple Scattering signals contribute significantly to the peaks of $\tilde{\chi}(R)$, at R positions corresponding to their effective length corrected by the phase shift, and the finite mean free path of the photoelectron causes the peaks intensity to decrease with R. A detailed description of the properties of the FT of an EXAFS signal is given in reference [59].

Concerning metal sites in proteins, the analysis of the first peak of $\tilde{\chi}(R)$ is very useful to identify the possible presence of S atoms in the metal coordination sphere. The possible donors for a metal located in a protein, when no molecules other than amino acids and water from the matrix are considered, are N, O, and S; the metal-N and metal-O equilibrium distances are comparable to each other for many metals, as confirmed by the target distances reported in Table 2.2, while they differ more significantly from the metal-S distance. In addition, the
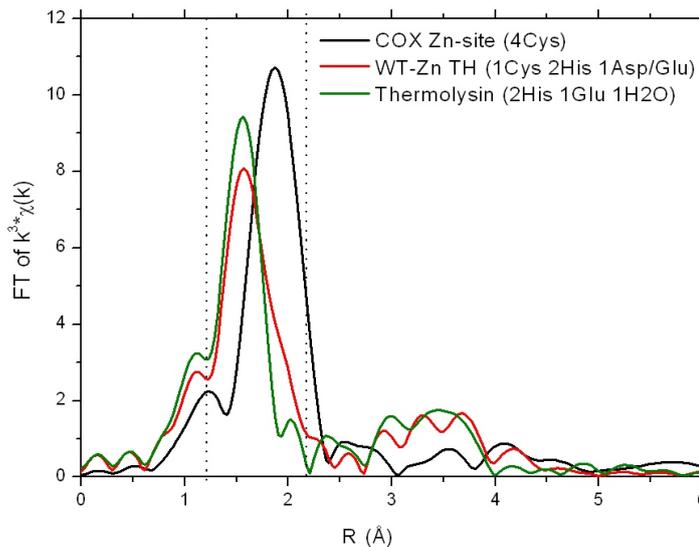
Figure 2.5: Magnitude of the Fourier transformed Zn K-edge EXAFS signal for three different protein samples: Cytochrome oxidase (COX, black trace, Zn binding motif: 4Cys), Wild Type Transhydrogenase incubated with exogenous Zn (WT-Zn TH, red trace, Zn binding: 1Cys 2His 1Asp/Glu) and Thermolysin (TLS, green trace, Zn binding: 2His 1Glu 1H_2O). Dotted vertical lines represent the first-shell fitting range relative to the WT-Zn TH spectrum.

phase shifts and backscattering amplitude produced by N/O are very similar, while they differ dramatically from the ones due to a S atom. This results in a spread in the distances for mixed first shells (i.e. containing both N/O and S atoms), causing the first peak of $\tilde{\chi}(R)$ to broaden significantly. As an example, in Figure 2.5 the FT EXAFS spectra of three tetrahedral Zn sites composed by different ligands are reported: Zn in Thermolysin is bound only to N/O and its FT spectrum shows a narrow first peak centered at ∼1.6 Å; in Cytochrome Oxidase it is bound only to S atoms and the peak is centered at ∼1.9 Å; finally, in Wilde Type Transhydrogenase it binds 3 N/O and 1 S atom, therefore the peak is centered at ∼1.6 Å, but it shows as well a tail that extends till ∼2.0 Å.

The experimental evidence illustrated above can be taken into account in a qualitative analysis based on the comparison with reference spectra, but they can also provide quantitative information by means of a first shell fit. This means that a fit of the first shell peak can be performed with any EXAFS analysis program: once the coordination number N has been estimated by XANES analysis and the R range has been properly chosen (see dotted lines in Figure 2.5), the number of S atoms n can be set as a free parameter of the fit. The number of O/N will be constrained to be N-n, but any attempt of distinguishing between the two atoms could be misleading at this stage of the analysis [46], considering their similar

scattering properties and the fact that the information content in this narrow R range is limited.

Presence of His residues. It is well known that His residue, due to its ring structure, generates multiple scattering (MS) contributions of significant amplitude that contribute at high distances (≥ 3 Å) in the FT [60, 61]; to the contrary, contributions in this region coming from other amino acids are very weak. Therefore the Fourier Transform (FT) of the experimental spectrum contains important information about the number of His residues present in the binding cluster from which the XAFS signal is generated. The amplitude of multiple scattering contributions at high R values is strongly influenced as well by two more factors: the temperature-dependent Debye Waller factors associated to the MS paths and the k-range of the data in which the FT is performed. It all suggests that the number of His residues in an unknown metal site can by determined by comparing the amplitude of high-R multiple scattering contributions in its FT spectrum with the amplitude of such contributions shown by the spectra of reference metal sites, for which the number of His is known; however it must be remembered that the comparison is allowed only when the spectra have been collected at the same temperature and Fourier transformed in the same k-range.

For mononuclear zinc proteins at room temperature, we have investigated such relationship in a systematic way by performing many theoretical simulations based on the target clusters contained in the database of Table 2.1 and by comparing these results with experimental data. The DW factors of the theoretical simulations were kept fixed at the values corresponding to T=300 K provided by DFT calculations [18]. The range used for the FT was 2-12 Å^{-1}. From these simulations and experimental data it appears that, for mononuclear zinc binding sites, the region in the FT which contains, almost exclusively, contributions coming from His residues is included between 3 and 4 (see Fig. 2.6).

Moreover we observe a systematic increase of the quantity

$$I = \int_3^4 FT[k^3\chi(k)]dk \tag{2.3}$$

with the number of His present in the cluster. The evaluation of the quantity I for the simulated EXAFS spectra of a chosen set of clusters allowed to identify a criterion that provides the estimation of the number of His bound to Zn. The relation is summarized in Table 2.3 and allows to recognize the number of His residues in an unknown Zn site by simple calculation of I, when the EXAFS
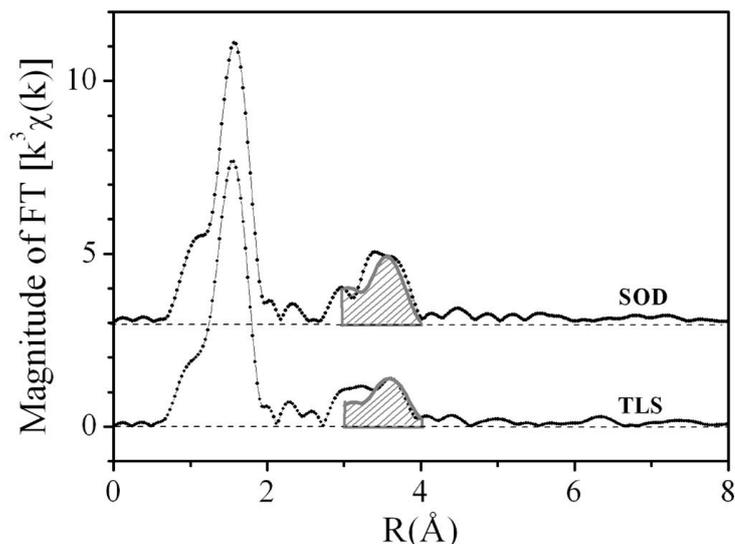
Figure 2.6: Magnitude of the Fourier Transform of experimental data collected at room temperature for Thermolysin (TLS) and Super oxide Dismutase (SOD). The shadowed areas overlapped to the spectra represent the sum of the His contributions generated in the EXAFS simulations of the binding motives *His His Glu H_2O* (TLS) and *His His His Asp* (SOD). It is noticeable that, in both cases, the contributions due to His residues cover the quasi-totality of the MS contributions in the region 3-4 Å.

spectrum is measured at room temperature and Fourier transformed in the k-range $[2, 12]$ Å^{-1}. In Figure 2.6 the FT spectra of Thermolysin (TLS) and Super oxide Dismutase (SOD), chosen as reference proteins, are reported: as shown by X-ray crystallography, and confirmed by XAFS analysis, both proteins contain a mononuclear tetrahedral Zn site, the binding motif of which is *His His Glu H_2O* in TLS [62] and *His His His Asp* in SOD [63]. The values of I relative to the experimental spectra are 1.1 for TLS and 1.5 for SOD, and they belong to the intervals predicted by simulations for a number of His equal to two and three respectively. The sum of the contributions due to His residues in theoretical calculations are overlapped to the experimental spectra in Figure 2.6 to prove that the theoretical values of I reproduce pretty well the experimental evidences; the negligible difference between the two could be due to the minor multiple scattering contributions given by non His residues.

k-space fits. After having collected information about the number and the nature of first shell ligands with the methods described above, all of the possible amino acid clusters consistent with these information can be built and fitted to the experimental signal. Depending on the requirement of each specific experi-

coordination number	number of His	interval of I values
3, 4	1	0.4 - 0.8
	2	0.8 - 1.2
	3	1.2 - 1.6
5, 6	1	0.5 - 1.0
	2	1.0 - 1.5
	3	1.5 - 2.0

Table 2.3: Correlation between the value of the integral I defined in equation (2.3) and the number of His residues present in a Zn binding motif. In the intervals reported in the third column, the lower limit is included, the upper one in excluded. The correlation was derived for T=300 K, by means of EXAFS simulations for Zn clusters characterized by a different number of His. The k-range [2, 12] Å^{-1} was used for the Fourier Transform.

ment, the clusters can be chosen between the commonest ones retrieved through database searches, or they can be built considering all permutations of the possible ligands for the metal of interest. The last option is the most demanding because a large number of putative models are taken into account, however it allows to discover eventual binding sites not encountered yet in the PDB. A very useful list of the possible amino acid ligands for each metal, derived from observations and statistical analysis in the PDB and the CDS [49, 50, 51, 52, 53], is available on the web (http://tanna.bch.ed.ac.uk/).

For the putative clusters, initial geometries can be built as described in paragraph 2.1.2 for the target clusters used for XANES simulations, and the list of spatial coordinates and atomic numbers generated can be processed through a program able to calculate *ab initio* scattering amplitudes and phases. For the studies presented in this work, we used feff8.2 [64] for *ab initio* calculations, combined with Artemis [56] to carry out the fitting procedure for each putative cluster.

Fits are performed directly in the k space, with a k weight of 3. The main problem encountered in EXAFS analysis is the presence of multiple solutions for the set of parameters to be determined, which causes the system to converge to local minima that strongly depend on the initial values; this often happens when metal-ligands distances and DW factors are set as free parameters at the same time. To attenuate this problem, we propose two possible solutions:

1. Making use of a parametrization based on DFT calculations for the Debye Waller factors as a function of the distance from the absorber.

2. Using a step-by-step procedure to fit alternatively distances and DW factors.

The first option relies on the possibility to employ the parametrization of the DWs provided by Dimakis and Bunker for amino acids bound to Zn [17, 18] and for Fe^{2+}-Porphyrin complexes [20]. According to the parametrization, Single Scattering and the most effective Multiple Scattering DW factors can be written as:

$$\sigma^2(\Delta R, T) = \sigma^2(R_0, T) + A(T)\Delta R + B(T)\Delta R^2 \qquad (2.4)$$

where R_0 is the equilibrium distance predicted by DFT calculations and provided by the authors, $\sigma^2(R_0, T)$, $A(T)$ and $B(T)$ are third order polynomials the coefficients of which are tabulated for each amino acid. Once the temperature is defined, the value in equation (2.4) is a function of the displacement from the equilibrium distance, and it can be inserted into the analysis procedure as a constrained parameter. In Chapter 5 we will also present a valuable procedure aimed to calculate DWs from the dynamical matrix obtained with DFT methods, meant to overcome the need of given values.

When *ab initio* calculated values are not available, or when the scientific issue requires to fit the DW factors, a step-by-step fitting procedure can be applied [65]; this has the effect to minimize the correlation between structural and dynamical parameters, avoiding the convergence to local minima. The steps are basically three, and they can be summarized as follows:

1. Metal-amino acid distances and bending angles are fitted in the Rigid Body Refinement scheme, starting from their target values, whereas the DWs are kept fixed at reasonable values.

2. Structural parameters are set to the best fit values from the previous step and DWs are allowed to vary.

3. DWs are set to the best fit values from the previous step and structural parameters are fitted again.

The DWs to be fixed in the first step can be derived either from the literature (when available) or from *ab initio* simulations. The fitting model can be considered appropriate only if in step 3 the system converges to the previously determined values; moreover step 3 has the function to improve the precision of structural results.

Statistical choice of the most probable cluster. The fitting criterion we used, relies on the minimization of the R factor, defined as:

$$R = \sum_{i=1,N_p} (k_i^3 \tilde{\chi}_{i_{data}} - k_i^3 \tilde{\chi}_{i_{fit}})^2 / \sum_{i=1,N_p} (k_i^3 \tilde{\chi}_{i_{data}})^2 \tag{2.5}$$

where N_p is the number of experimental points and $\tilde{\chi}$ is the EXAFS function. When different fitting models are applied to the same set of data, a confidence analysis is needed to test the statistical significance of eventual differences in the resulting goodness-of-fit. Fits are compared on the basis of the reduced chi-square, i.e. χ_ν^2, defined as [66]:

$$\chi_\nu^2 = \frac{1}{\nu} \frac{N_{ind}}{N_p} \sum_{i=1,N_p} \left(\frac{\tilde{\chi}_{i_{data}} - \tilde{\chi}_{i_{fit}}}{\sigma} \right)^2 \tag{2.6}$$

where σ is the noise in the dataset, p is the number of free parameters and $\nu = N_{ind} - p$ is the number of degrees of freedom in the fit; N_{ind} is the number of independent points defined as [66]:

$$N_{ind} = \frac{2\Delta k \Delta R}{\pi} + 2 \tag{2.7}$$

where Δk and ΔR are the intervals in the real and reciprocal space in which the fit has been performed. The noise σ can be evaluated either from Poissonian statistics, by calculating the square root of the total number of counts, or empirically as the standard deviation of the data in the high-k region, where structural oscillations are no longer detectable. Even in the presence of "good" fits, i.e. at relatively low values of the R factor, the values calculated for χ_ν^2 are usually much larger than 1. This situation is commonly encountered in XAFS analysis and attributed to small inadequacies of the model and/or to systematic experimental errors [67]. In view of this, the standard fluctuation in χ_ν^2 (which is equal to $\sqrt{2/\nu}$) can be rescaled to $\chi_\nu^2 \cdot \sqrt{2/\nu}$ [66, 67]. The comparison between two different fits of the same data set (corresponding to two different clusters, a and b) is performed according to the following criterion [67]: fit to cluster b is considered significantly better than fit to cluster a when

$$(\chi_\nu^2(a) - \chi_\nu^2(b)) \geq \sqrt{2 \left[\frac{(\chi_\nu^2(a))^2}{\nu(a)} + \frac{(\chi_\nu^2(b))^2}{\nu(b)} \right]} \tag{2.8}$$

which corresponds to the confidence level of 1σ.

2.3 Testing the method on model proteins

The XAFS analysis procedure described in this chapter consists in a collection of methods and tools originated by the efforts of several groups dedicated to XAS research; however, the way they are combined and complemented with the translation of empirical observations into quantitative criteria is original. In the last years we have applied this synergic approach, or slightly different versions of it, to investigate unknown Zn^{2+} binding sites in several charge translocating protein complexes, namely: inhibitory Zn^{2+} sites in photosynthetic Reaction Center [68], in Cytochrome bc_1 complexes [13] and in Transhydrogenase (Chapter 3 and reference [69]), and the structural Zn^{2+} site of NADH-Q oxidoreductase [70]. In the totality of these systems, our approach brought to the characterization of the metal binding motif. In order to achieve a final validation of the method, we applied it to three zinc sites whose structure was already known from protein crystallography:

1. The structural Zn site of bovine heart Cytochrome oxydase (COX)

2. The catalytic Zn site of Thermolysin (TLS)

3. The catalytic Zn site of Superoxide Dismutase (SOD)

COX is the terminal component of the respiratory chain: it catalyzes the oxidation of cyt c reduced by the cyt bc_1 complex, reducing O_2 to H_2O and pumping 4 protons across the mitochondrial membrane. There are a wide number of crystallographic structures available for bovine heart COX in different states (fully oxidized, fully reduced, azide-bound, and carbon monoxide-bound) with resolution up to 1.8 Å [71, 72, 73]. All the available crystallographic structures show the existence of an endogenous Zn^{2+} bound to subunit V_b. The local structure around the zinc ion, reported in all the crystallographic structures, is formed by 4 cysteines; this binding motif is the commonest in nature. TLS is a thermostable neutral metalloproteinase enzyme produced by the gram-positive bacterium Bacillus thermoproteolyticus. It contains a zinc site which catalyzes the hydrolysis of peptide bonds involving hydrophobic amino acids. The identification of the amino acids that bind to the catalytic zinc ion was allowed by crystallographic analyses [74] and confirmed by XAFS [75]. These studies show that the zinc cluster is formed by two His, one Glu and one water molecule (H_2O). SOD catalyzes the dismutation of superoxide into oxygen and hydrogen peroxide. Crystallographic data have shown the existence of a catalytic zinc site formed by three His and one Asp [76, 77].

2.3.1 Materials and Methods

Sample preparation. Bovine heart cytochrome c oxidase (COX), containing 10 nmoles of heme a+a$_3$ mg^{-1} protein, was purified as described in [78]. Thermolysin (TLS) and CuZn superoxide dismutase from bovine eritrocytes (SOD) were purchased from Calbiochem and Sigma-Aldrich, respectively. TLS was recrystallized as described in [79] . Measurements were performed on polyvinyl alcohol (PVA) protein films, prepared by adding 350 μL of a 10 % solution of PVA (Fluka) to 1 ml of 80 μM COX or to 1 ml of TLS suspension (15 mg/ml protein) or SOD suspension (3mg/ml protein), respectively. After mixing, the protein-PVA solutions were layered into 3×3×0.3 cm^3 teflon holders and dried under nitrogen flow until PVA films were formed.

XAFS data collection and analysis. Zn K-edge XAFS measurements were performed at the BM8 GILDA beam-line of the European Synchrotron Radiation Facility (ESRF), the specifications of which are described in Paragraph 1.3.2. The photon flux was of the order of 10^{10} ph/s and the spot size ∼1×1 mm^2. The analyzed spectrum for COX was obtained from a scan with an integration time of 15 s/point with a maximum number of counts per channel of 6 × 10^3. The analyzed spectrum for TLS was obtained from the average of two scans for a total integration time of 30 s/point. The maximum number of counts per channel was 3 × 10^4. For SOD the analyzed spectrum was obtained from the average of 3 scans for a total integration time of 45 s/point. The maximum number of counts per channel was 6 × 10^3.

The analysis was performed following the steps of the method described in this chapter, using feff8.2 for theoretical calculation of amplitudes and phases and Artemis for the fitting procedure; the putative clusters were identified amongst the ones reported in Table 2.1. The value of S$_0^2$ was calculated by feff8.2 from atomic overlap intervals, and its value was kept fixed during the fitting procedure. For each putative cluster, metal-amino acids distances and bending angles were set as free parameters of the fit, together with a shift in the energy origin ΔE_0 common to all paths; when more than one amino acid of the same nature was present in the putative cluster, they were treated in a non-degenerate mode, i.e. to each amino acid two independent structural parameters (distance from the metal and bending angle) were assigned. The Debye Waller factors were set as constrained parameters in the fit, by making use of the parametrization (2.4).

2.3.2 Results

The normalized XANES spectra for the three Zn sites are reported in Figure 2.7(a). Their WL intensities are all lower than 1.5, which indicates that the putative clusters should be searched amongst the 3- and 4-coordinated binding motives.
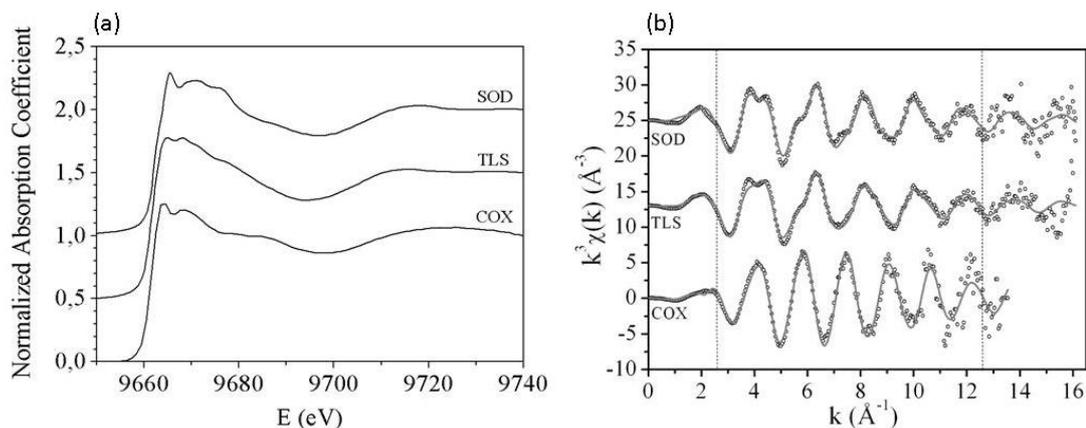
Figure 2.7: Panel (a): experimental XANES spectra for COX, SOD and TLS normalized at their value at 9740 eV. Panel (b): experimental k^3 weighted EXAFS functions (open circles) for COX, SOD and TLS. Continuous lines represent the fits obtained for the binding motives *Cys Cys Cys Cys*, *His His Glu H_2O* and *His His His Asp* for COX, TLS and SOD respectively. Vertical dashed lines indicate the fitting range 2.3-12.3 Å^{-1} used for all fits.

The analysis for COX is straightforward, since a fit of the first shell peak of its FT spectrum indicates the presence of S atoms in the Zn coordination sphere, whereas no N and/or O are detected. The only binding motif included in Table 2.1 and consistent with these information is *Cys Cys Cys Cys*; this model provides a very good k-space fit of the EXAFS function of COX, as shown in Figure 2.7(b), with an R factor of 3.7 and a reduced chi-square of 50±19. If the list of putative clusters is extended to less common or occasionally found binding motives (at least 5 PDB entries in the MDB), the cluster *Cys Cys Cys* must be taken into account as well and fitted to the data: it provides a worse fit than the previous one, with an R factor of 6.3 and a reduced chi-square of 87±31. When the significance of the difference in the goodness-of-fit is tested by means of the criterion (2.8), the actual binding motif *Cys Cys Cys Cys* is unambiguously chosen as the most probable one.

Concerning TLS, first shell analysis excludes the presence of S atoms. Multiple Scattering contributions in the region 3-4 Å, shown in Figure 2.6, reveal the presence of His residue; their intensity, calculated with equation (2.3), is I=1.1,

which implies that two His residues are present, by simply comparing this value with the intervals reported in Table 2.3. Matching the information, the putative Zn binding clusters singled out from the list in Table 2.1 are: *His His Asp*, *His His Glu H$_2$O* and *His His H$_2$O H$_2$O*. The R factor and reduced chi-square relative to the fit of the TLS experimental spectrum with each putative model are reported in Table 2.4: *His His Glu H$_2$O* is identified as the best fitting model, and the chi-square test confirms the significance of the difference in the goodness-of-fit.

Protein	Binding Motif	R$_{factor}$ (%)	χ^2_ν
	His His Asp	10	119 (35)
TLS	*His His H$_2$O H$_2$O*	5	69 (21)
	His His Glu H$_2$O	4	35 (11)
	His His His	10	138 (43)
SOD	*His His His H$_2$O*	7	97 (31)
	His His His Asp	5	60 (18)

Table 2.4: Values obtained for the R factor, the reduced chi-square and its standard fluctuation $\chi^2_\nu \cdot \sqrt{2/\nu}$ for the models selected for TLS and SOD.

The same analysis procedure is applied to SOD; the presence of Cys residues is excluded by first shell analysis and value of I is calculated as 1.5, which belongs to the range of values associated with the presence of 3 His residues. The putative clusters, chosen amongst the commonest patterns (more than 15 PDB entries, Table 2.1) are therefore: *His His His*, *His His His Asp*, *His His His H$_2$O*. The R factor and reduced chi-square relative to the k-space fits based on each chosen model are reported in Table 2.4; the best fitting model is again the actual one, *His His His Asp*, that is significantly better than the models *His His His* and *His His His H$_2$O* within confidence intervals of 2σ and 1σ respectively, according to the chi-square test of equation (2.8). Best fits given by the most probable model for each Zn site are shown in Figure 2.7(b).

Concerning the structural parameters resulting from these fits, comparisons with PDB structures and extension of the method in order to consider also occasionally found binding motives, we address the reader to our recent paper treating in detail the test of this analysis method [34]. With the summary presented here, we wish to demonstrate the effectiveness of the method in the characterization of unknown binding sites, in order to put the basis for the treatment of the diverse biophysical problems encountered in this and in future works. Although the efforts we put into the determination of criteria that correlate features of XAFS spectra to properties of the metal binding site were dedicated only to Zn sites, it is worthwhile to notice that the same approach could be undertaken for any

metal, leading of course to different sets of criteria.

42

Chapter 3

Inhibitory metal sites in proton-translocating proteins

Divalent metal ions, such as Zn^{2+} and Cd^{2+}, inhibit the catalytic cycle of a number of important membrane proteins, including the cytochrome bc_1 complex [80] cytochrome oxidase [81, 82], and the bacterial photosynthetic reaction center [83]. There is evidence that in each case the metal ions block proton-transfer steps. In photosynthetic reaction centers the metal-ion binding site was located at the cytoplasmic surface of the protein by x-ray crystallography. The Zn^{2+}-ligand cluster is formed by the imidazole side chains of two His residues, by the side chain of an Asp, and most likely by a water molecule [84]. The tetrahedral coordination geometry was subsequently confirmed by x-ray absorption fine structure (XAFS), which allowed a high-resolution determination of the bond lengths [68]. Cd^{2+} can also bind at this site but in an octahedral geometry by a cluster which involves the same amino acid residues and possibly four water molecules [85]. The x-ray crystallography and XAFS data provide detailed structure information on the mechanism by which Zn^{2+} and Cd^{2+} obstruct proton entry into the reaction center protein [83, 84, 85].

Biochemical studies indicate that a similar mechanism may be relevant to metal-ion inhibition of other H^+-translocating redox complexes [80, 81, 82, 86]. Structural information on the high-affinity Zn^{2+}-binding sites of cytochrome bc_1 complex [87, 13] and of cytochrome c oxidase [71, 88] are consistent with the proposed inhibition mechanism, and indicate that at least one His residue is always present in the metal binding cluster.

Low concentrations of Zn^{2+}, Cd^{2+} and other metal ions were recently shown to inhibit specific catalytic steps in transhydrogenase [89, 90]. This enzyme, found in the inner mitochondrial membrane of animal cells and in the cyto-

plasmic membrane of many bacteria, couples a redox reaction, hydride transfer between NAD(H) and NADP(H), to proton translocation through the mitochondrial membrane. Since no crystal structure exists for the transmembrane domain of the enzyme, the metal-ion binding site cannot be located by protein crystallography; however, Attenuated Total-Reflectance Fourier-Transformed Infrared (ATR-FTIR) spectroscopy studies [90] suggested that the highly conserved residue βHis91 is somehow (directly or not) involved in Zn binding. The presence of at least one His residue in the metal coordination sphere is expected, since it would confirm that the proposed mechanism of inhibition, based on competition between metal and H^+ binding to His residues, holds also for transhydrogenase.

In this chapter we present a XAFS characterization of the Zn^{2+} binding site in wild type transhydrogenase (TH) and its βHis91$\rightarrow$Lys (βH91K) mutant, aimed to bring insight into the mechanism of metal inhibition; the results were published as: Veronesi et al. *X-ray absorption studies of Zn^{2+}-binding sites in Escherichia coli transhydrogenase and its βH91K mutant*, Biochimica et Biophysica Acta (BBA) - Bioenergetics, 2010, 1797, 494 - 500 (reference [69]).

3.1 Transhydrogenase : structure and function

Transhydrogenase (TH) is an enzyme situated in animal mitochondria and bacteria which couples the redox reaction between NAD(H) and NADP(H) to proton translocation across the membrane:

$$NADH + NADP^+ + H^+_{out} \leftrightarrow NAD^+ + NADPH + H^+_{in} \qquad (3.1)$$

where H^+_{out} and H^+_{in} denote hydrogen ions outside and inside the intact coupling membrane system (for review see [91, 92]). Under physiological conditions, TH normally utilizes the electrochemical gradient generated by the respiratory electron transport chains to drive $NADP^+$ reduction. The resulting NADPH is used for biosynthesis and for reduction of glutathione, in order to limit damages caused by free radicals generated in the respiratory chain [93]. The free energies of the products of equation (3.1) are similar to those of the reactants; this allows the reaction to be driven either from left to right, using the electrochemical gradient Δp generated by the respiratory chain, or from right to left, using the redox potential difference between NADPH and NAD^+ to increase Δp. Proton translocation and redox reaction take place in a 1:1 ratio, due to the coupling of the two processes.

TH is a dimer, each of its monomers is composed by three subunits: dI and dIII bind NAD(H) and NADP(H) respectively and protrude from the membrane, while

the dII subunit spans the membrane (see Figure 3.1). The single hydride transfer
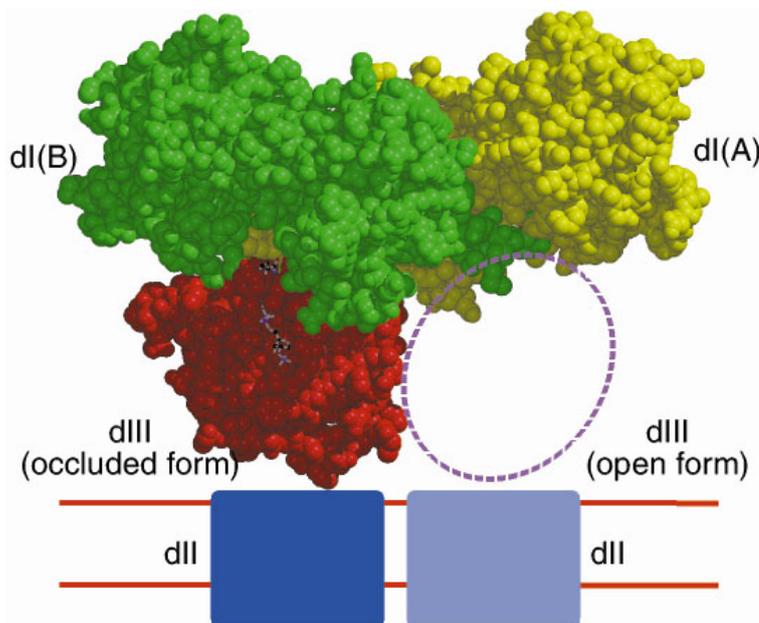
Figure 3.1: Structure of a dI_2:$dIII_1$ complex from Rhodospirillum rubrum [94].

site in the complex is at the interface between dI and dIII. Two conformational states have been identified for the enzyme:

- *open* configuration. Bound nucleotides can rapidly exchange with the ones in the solvent, but their nicotinamide rings are held apart to prevent redox reaction. When this configuration is activated, product nucleotides dissociate from the enzyme and are replaced by fresh substrates.

- *occluded* configuration. Bound nucleotides cannot exchange with the ones in the solvent and their nicotinamide rings are held together to favor hydride transfer.

The switching mechanism between the two conformational states is summarized in Figure 3.2. Proton translocation through dII activates the interconversion between the *open* and the *occluded* configuration, probably as a consequence of long range ($\geq$ 30 Å) conformational changes transmitted from the proton pathway to the redox site [91]. These conformational changes probably involve relative motions of rigid elements of dII, like transmembrane helices, but the nature of these motions is still unknown and could be clarified with the advent of a high resolution crystallographic structure of the subunit. No structure indeed is available at the moment for the whole enzyme, while several high resolution structures are available for isolated dI, dIII and for a dI_2-$dIII_1$ complex (reviewed

Figure 3.2: The binding-change mechanism of transhydrogenase: upon proton translocation, the two monomers of the enzyme switch from the open to the occluded configuration, allowing exchange of nucleotides with those in the solvent or hydride transfer between nicotinamide rings. The figure is taken from [91].
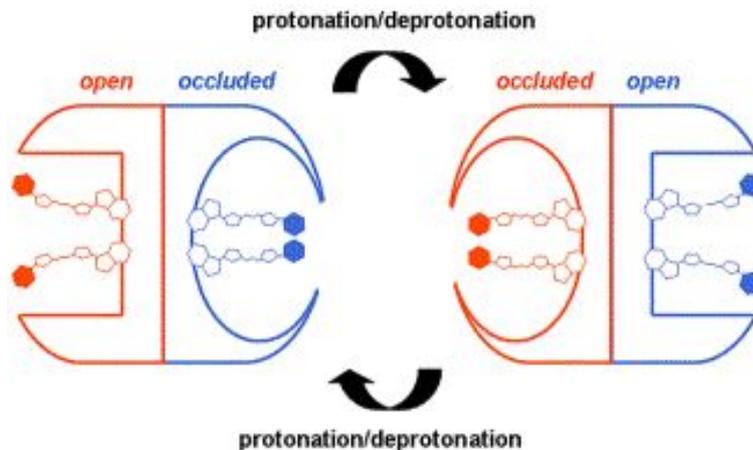
in [95]). The information available on the membrane-spanning domain dII have been provided by amino acid-sequence analysis, mutagenesys studies [96, 97], and site-specific chemical labeling [98]; they show that, depending on the species, dII has between 12 and 14 transmembrane helices per monomer. The topology of the dII component of TH is shown in Figure 3.3.

Pink-colored letters in Figure 3.3 represent the conserved residues throughout the totality of the available sequences ($\sim$ 160), while blue letters represent the highly-conserved residues. Mutagenesys studies have been carried out [99, 100, 101] in order to clarify whether these residues are involved in proton translocation: once a conserved amino acid has been mutated, the effect on the rate of reaction (3.1) is monitored. Such studies demonstrated that only a few residues are essential for the sustainment of TH activity, namely the dII residues βHis91, βAsn222, βAsp213 or those in the sequence βCys260-βSer266. This strongly argues for the involvement of these residues in the conduction of protons through dII or in the energy-transduction device itself.

3.2   Inhibition of Transhydrogenase activity

Metal ions-induced inhibition of transhydrogenase activity as a function of pH has been studied by Whitehead et al [89, 90] in membrane vesicles from both *E. Coli* and *Rhodospirillum rubrum*. Reverse transhydrogenation (left arrow in equation (3.1)) was measured as the reduction of the NAD$^+$ analogue acetylpyridine adenine dinucleotide (AcPdAD$^+$) by NADPH, through the measurement of

Figure 3.3: Topology of the transmembrane domain dII of transhydrogenase, taken from [92]. The putative helices H1-H14 are numbered on the top row; The number on the rows underneath refer to the *E.Coli* residues numbering. Grey-colored helix H5 is missing in the *E. Coli* type and in *Rhodospirillum Rubrum*, H1 only in *Rhodospirillum Rubrum*. Residues shown in pink are invariant and 100% conserved throughout the 160 available sequences of transhydrogenase, whereas those depicted in blue are highly conserved.

the absorbance at 375 nm. The rate of cyclic transhydrogenation was monitored as well: this reaction is wholly unphysiological, and consists in the combined reduction of $NADP^+$ by NADH and oxidation of NADPH by $AcPdAD^+$; the peculiarity of cyclic reaction is that it takes place without dissociation of the nucleotides from the enzyme, therefore it does not require the interconversion between conformational states.

Many metal ions have been found to inhibit reverse transhydrogenation and to stimulate the cyclic reaction at neutral and high pH, as shown in Figure 3.4 for Zn^{2+}. The order of effectiveness of the metals is:

$$Pb^{2+} > Cu^{2+} > Zn^{2+} = Cd^{2+} > Ni^{2+} > Co^{2+}.$$

As mentioned before, inhibition of TH activity takes place as well when some specific amino acids situated in the transmembrane domain are substituted. The implication of the highly conserved residue βHis91 in proton translocation has been proved by several studies [101, 102, 103], some of which showed that the substitution of βHis91 inhibits reverse and stimulates cyclic transhydrogenation [104, 105]. This scenario suggests a mechanism of inhibition which is common to metal ions binding and to βHis91 substitution: both processes could cause the enzyme to lock in the occluded state, preventing the nucleotides to exchange with the solvent, condition that hinders the physiological TH activity but favors the
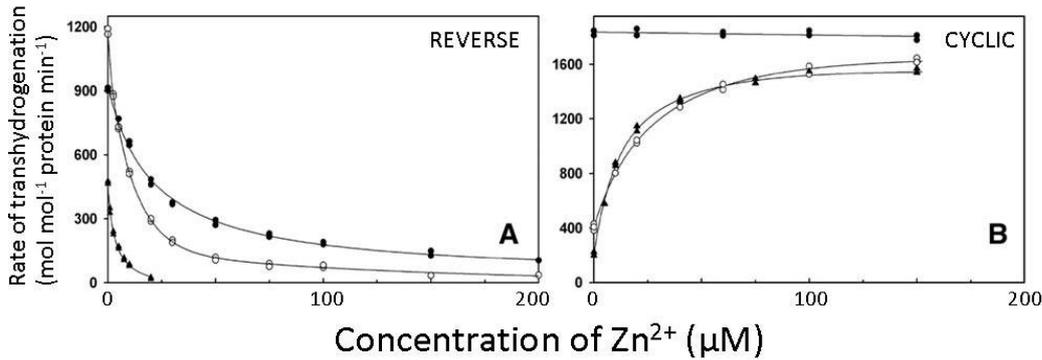
47

Figure 3.4: Effect of Zn^{2+} on (A) reverse transhydrogenase activity and (B) cyclic transhydrogenation at three different pH conditions: pH 6.2 (black circles), pH 7.2 (open circles) and pH 8.2 (black triangles). The figure is taken from [90].

cyclic reaction. The origin of the mechanism is likely to be the direct inhibition of proton translocation through the dII domain, that would prevent the conformational changes that cause interconversion between states. This hypothesis fits very well with the general mechanism of inhibition by metal ions, based on the competition between metal and proton binding.

The characterization of the Zn^{2+} binding site and its location in the enzyme would therefore help to trace the proton pathway. To address this issue, Whitehead et al [90] carried out an Attenuated Total-Reflectance Fourier-Transformed Infrared Spectroscopy (ATR-FTIR) study on wild type TH from *E. Coli* and its βH91K mutant. The study consisted in recording the differential FTIR spectra induced by Zn^{2+} perfusion for both enzymes and attribute the peaks in the spectra to perturbated vibrational modes of some specific amino acids; the amino acids detected in this way are likely to compose the Zn^{2+} binding cluster, or to be indirectly perturbated by Zn binding. In Figure 3.5 the differential FTIR spectra for wild type TH and βH91K are shown: the peak at 1602 cm^{-1} and the trough at 1556 cm^{-1}, appearing in both samples, were attributed to perturbation of the vibrational modes of carboxylic acids (Asp/Glu) upon Zn binding; moreover, the resemblance of the spectral features of TH with the ones of a poly-$_L$-His sample suggests that His residues must be involved in Zn binding.

Since the intensity of the peaks of differential FTIR spectra was found to be from 30% to 50% lower in the mutant than in the native enzyme, Whitehead et al concluded that the mutated residue is involved in Zn binding: upon βHis91→Lys substitution, indeed, spectral contributions due to amino acids perturbated by Zn binding are reduced. In this framework, our XAFS study was undertaken in order to characterize the Zn^{2+} binding site in wild type and βH91K transhydrogenase,
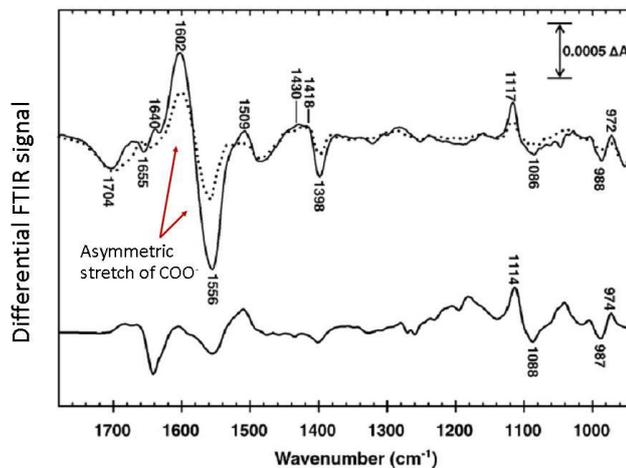
48

Figure 3.5: Zn-induced difference FTIR spectra of wild type transhydrogenase (top spectra, solid line), of its βH91K mutant (top spectra, dotted line) and of a poly-$_L$-His sample (bottom spectrum.)

and make clear whether the Zn ion influences by direct binding or not the βHis91 residue. This kind of experiment can help to trace the proton pathway through the transmembrane dII domain.

3.3 XAFS study of Zn^{2+} in native and genetically modified TH

In the following paragraphs we present an XAFS analysis of Zn^{2+}-binding sites in purified transhydrogenase complexes from *E. coli*. In order to test the possible involvement of βHis91 in coordinating the metal, we have analyzed XAFS spectra acquired with the wild-type enzyme and with the βH91K mutant. A single Zn^{2+}-binding cluster formed by one Cys, two His and one Asp/Glu residue, arranged in the same tetrahedral coordination geometry, best accounts for the XAFS spectra of both the wild-type and the mutant transhydrogenase, indicating that βHis91 does not directly participate in the binding site.

3.3.1 Materials and methods

Sample preparation. Transhydrogenase purification and preparation were carried out by the group of J.Baz Jackson, operating at the School of Biosciences in Birmingham University (UK). Wild-type E.coli transhydrogenase and its βH91K mutant, each carrying a cleavable His tag at the N-terminus of the α subunit,

were expressed from pSI4 and pSI5, respectively, and purified by Ni^{2+} chromatography, as described [90]. Precautions were taken to limit contamination of the protein by metal ions in the buffer solutions, as indicated in the earlier report. The tag was removed from both proteins with Factor Xa, and the transhydrogenases were separated from the cleavage enzyme and products by gel permeation chromatography [90]. Protein quantities were determined using the bicinchoninic acid assay [106], and are given below in mol of "dI-dII-dIII monomer". The proteins were stored in 50 mM HEPES-KOH, pH 7.5, 0.2 M NaCl, 0.05% Anapoe35, 25% w/v glycerol at -20 oC, as described [90]. Following storage the proteins (0.10 μmol wild-type protein in 20 mL of storage buffer, and 0.11 μmol βH91K in 15 mL) were dialysed against 50 mM HEPES-KOH, pH 8.2, 2 mM MgCl2 0.05% Anapoe35. The two proteins were concentrated to a volume of approx 1.5 mL using Vivascience 100k cut-off filters, and then further dialysed against similar buffer before concentrating to approx 0.6 mL. The wild-type and mutant transhydrogenases were each mixed with appropriate stock solutions to give 1.0 mL of 45 μM protein, 3.3% w/v polyvinyl alcohol (PVA, Mr $\approx$ 130,000, Fluka), 50 mM HEPES-KOH, pH 8.2, 2 mM MgCl$_2$, 0.05% Anapoe35. Separate 1.0 mL samples, identical to these but for the presence of 36 μM added ZnCl$_2$, were also prepared. The four solutions were each transferred to 3×3×0.3 cm^3 Teflon holders, and dehydrated under dry nitrogen flow, incurring a volume decrease of about 10 fold. The incorporation of membrane complexes at high concentration into PVA films yields samples that are stable and easy to handle [68, 13]. Thermolysin (TLS) and superoxide dismutase (SOD) from bovine erythrocytes were purchased from Calbiochem and Sigma-Aldrich, respectively. TLS was recrystallized as described in [79]. Films were prepared by adding 350 μl of a 10% w/v solution of PVA to 1 mL of 20 mM Ca-acetate, pH 7.5 containing 15 mg TLS and to 1 mL of 50 mM Tris-HCl, pH 7.5 containing 3 mg SOD. The solutions were dehydrated under nitrogen as described above.

Transhydrogenation assay. Rates of reverse transhydrogenation by the purified, detagged protein were measured by recording the reduction of acetyl pyridine adenine dinucleotide (AcPdAD$^+$, 200 μM) by NADPH (200 μM) in 50 mM HEPES, pH 7.2, 2 mM MgCl$_2$ (see [90]). Rates of cyclic transhydrogenation were measured from the reduction of AcPdAD$^+$ (200 μM) by NADH (200 μM) in the presence of NADP$^+$ (50 μM) in the same buffer. The reaction rates were similar to those measured under similar conditions in previous work form Whitehead et al [90].

XAFS data collection and analysis. Experimental Zn K-edge spectra of the protein films were collected at beamline GILDA-BM08 [23] of the European Synchrotron Radiation Facility, the specifications of which are described in paragraph 1.3.2. The spot size was approximately 1×1 mm^2 and the flux on the sample 3×10^{10} photons s^{-1}. Data were collected at 80 K; the total integration time was 45 s/point for the two samples with added Zn and 90 s/point for samples without added Zn. In order to check for possible modifications in the local structure of the Zn site caused by irradiation during the measurement, we compared 5 consecutive spectra which had been acquired on the same position of the samples containing the wild-type transhydrogenase or its βH91K mutant. The first XANES spectrum, which required an acquisition time of 20 minutes, was indistinguishable from the spectra acquired subsequently, leading to exclude radiation damage. No changes were also detected in the EXAFS region between consecutive spectra, even in the case of a maximum total exposure of 8 hours. EXAFS analysis was performed by means of the *IFEFFIT* package [55] and its graphic interfaces Athena and Artemis [56], following the steps described in Chapter 2. Theoretical scattering amplitudes and phase shifts were calculated by means of the ab initio code FEFF8.2 [1], including Self Consistent Field calculations for the potential in a radius of 4.8 Å around the absorber. All of the multiple scattering signals arising from up to five single scattering events relative to the same amino acid and with an effective length $\leq$5 Å were taken into account in the fitting procedure. The amplitude reduction factor S$_0^2$ was calculated from atomic overlap integrals by the program: its value was fixed to 0.95 during the analysis. Input clusters for FEFF8.2 were built using MOLDRAW [107] for visualization and manipulation, setting first neighbor distances to the target values provided by M. M. Harding (see [53] and references therein) and interatomic distances and angles relative to each amino acid to the values provided by Engh and Huber [48]. Initial Debye Waller (DW) factors values were derived from the parameterization provided by Dimakis and Bunker [18]. Fits were performed directly in k space, with a k weight of 3, in the range 2.5-12.5 Å^{-1}, minimizing the R factor. The Rigid Body Refinement scheme was applied [4].

3.3.2 Results

PVA films of intact, purified, detergent-dispersed transhydrogenase from *E.coli* were prepared in the absence and presence of ZnCl$_2$. From titrations of the inhibitory effect of Zn^{2+} on the rate of reverse transhydrogenation the Kd for the metal ion is 2.5 μM at pH 8.2 [90]. Assuming for the moment that there is only

one metal-ion binding site per protein monomer then, in the solutions used to prepare the films (45 μM protein, 36 μM $ZnCl_2$), 68% of the Zn^{2+} sites would be occupied and the concentration of free Zn^{2+} (or Zn^{2+} bound to water molecules) would be 5.4 μM. During dehydration of the film the volume reduction will have led to an increase in binding, and a corresponding decrease in the amount of free Zn^{2+}. This should ensure that the contribution of free metal ion to the XAFS spectra is negligible (and see below). XAFS spectra were acquired in films containing wild-type transhydrogenase and its βH91K mutant in the absence (samples designated WT and BH91K, respectively) and presence of added Zn ions (WT-Zn and BH91K-Zn, respectively - Figures 3.6 and 3.7).
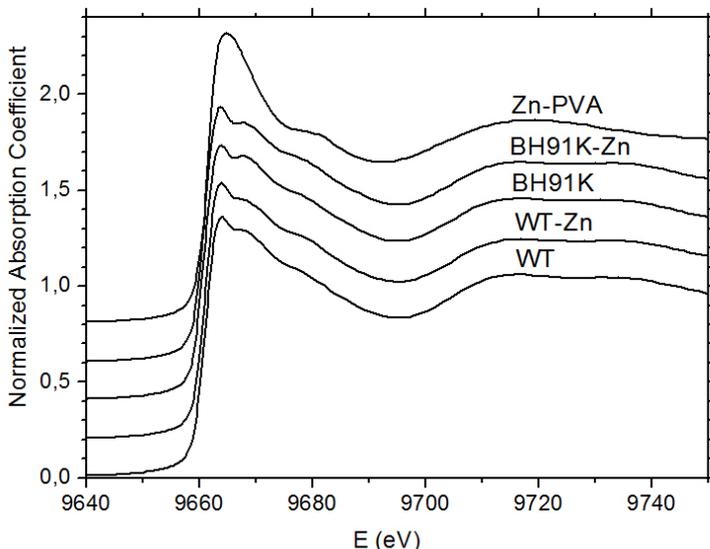

Figure 3.6: Experimental XANES spectra. K-edge Zn XANES spectra for wild-type transhydrogenase without and with added Zn^{2+} (WT and WT-Zn, respectively), for its βH91K mutant without and with added Zn^{2+} (BH91K and BH91K-Zn, respectively), and for a reference sample obtained by embedding $ZnCl_2$ in a PVA matrix (Zn-PVA). All spectra were measured at 80 K.

There was much lower but nevertheless significant Zn K_α fluorescence in the two samples prepared without added Zn^{2+} indicating a low level of contaminating metal ion in both the wild-type and βH91K proteins. Because the film thickness was not homogeneous, and because fluorescence is strongly affected by experimental geometry, these data cannot be used to evaluate accurately the amount of contaminating Zn^{2+}. We therefore used a functional assay of the purified protein in solution to show that the amount of the bound metal ion is indeed very low. Firstly, we confirmed that the inhibitory effect of added Zn^{2+} (50 μM) on

reverse transhydrogenase, and the stimulatory effect of the metal ion on cyclic transhydrogenation, were relieved (by 85%) upon subsequent addition of 200 μM EDTA [89, 90]. This shows that the EDTA effectively removes Zn^{2+} from the inhibitory site on the enzyme. Then in duplicate samples of purified transhydrogenase in the absence of added Zn^{2+}, we found that 200 μM EDTA neither stimulated the rate of reverse transhydrogenation nor inhibited the rate of the cyclic reaction. This shows that the amount of contaminating Zn^{2+} bound to the transhydrogenase under assay conditions is less than the error in the rate measurements (approx 5%, see [90]).

The XANES regions of the spectra (Figure 3.6) and the EXAFS functions (Figure 3.7) were very similar for all four samples indicating just a single class of Zn^{2+}-binding sites. Evidently the sites are similar in the wild-type protein and the
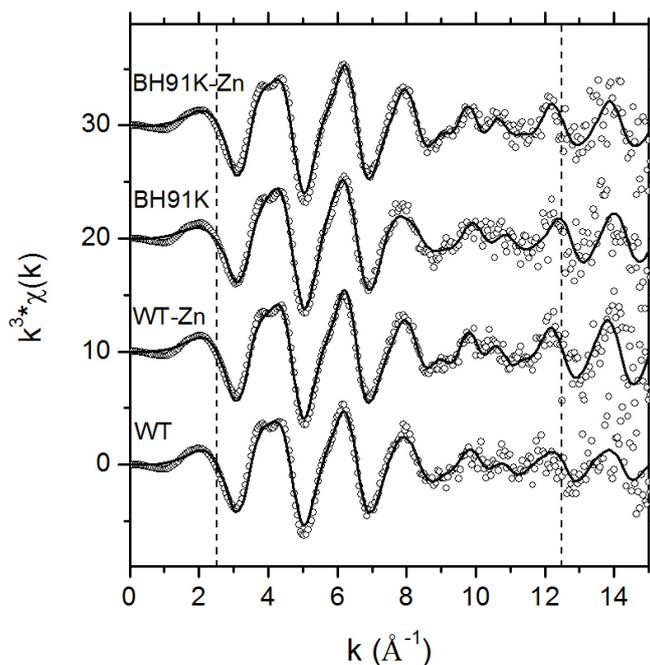

Figure 3.7: The experimental EXAFS spectra and their fits based on the 1Cys 2His 1Asp/Glu model. Zn K-edge EXAFS spectra for wild-type transhydrogenase without and with added Zn^{2+} (WT and WT-Zn, respectively), and for its βH91K mutant without and with added Zn^{2+} (BH91K and BH91K-Zn, respectively), measured at 80 K (circles). For each spectrum the continuous line shows the corresponding k-space best fit. Dashed lines indicate the fitting range 2.5-12.5 Å^{-1}

βH91K mutant and, moreover, the contaminating Zn^{2+} populates a small fraction of the sites occupied by the added metal ion. In Figure 3.6 we also show the XANES spectrum of a sample made of $ZnCl_2$ embedded in PVA in the absence of

protein (Zn-PVA). The higher white line intensity and the lack of inflection points after the white line and up to 9675 eV make this markedly different from the XANES spectra of the transhydrogenase samples. This difference indicates that no significant spectral contribution from PVA-bound Zn is present in the spectra of the transhydrogenase samples, suggesting that essentially all the metal ions detected by XAS in the transhydrogenase samples are bound to the protein and that contributions from unbound ions dispersed in the PVA matrix are negligible. Information on the coordination number of Zn^{2+} in the transhydrogenase samples can be obtained by comparing the observed XANES spectra with our XANES simulations data base (see Paragraph 2.1.2). For all TH samples, the white line intensity is 1.4, that indicates that the coordination number could be either 3 or 4 according to the criterion defined in Paragraph 2.1.3. The coordination number of 3 has been found in PDB structures, therefore it was taken into account in the systematic approach to XAFS analysis described in Chapter 2; however, such a coordination is most often ascribed to a interactions of the metal with solvent molecules that are not visible in the electron density map or to the crystallization procedure employed [108], and they are interpreted as 4-coordinated sites where one ligand is lost because of perturbations due to sample treatment. We can assume therefore that the Zn in all TH samples is four-coordinated, and neglect coordination 3 unless the analysis lacks of consistency.

The relative number of S atoms and N/O atoms in the Zn coordination sphere can be determined by an R-space first-shell fit, as described in Paragraph 2.2 In this calculations the number of S atoms, n, was a free parameter of the fit, while the total number of Zn neighbors was set to 4. The remaining (4-n) atoms could be either N or O; Zn-S and Zn-N/O distances were also free variables of the fit and their convergence to reasonable values was checked. The fourth and last free parameter in these fits was a common shift in the energy origin (ΔE_0) of the included scattering paths. DW factors were fixed to values calculated from the parameterization provided in [20]. The Fourier transformed (FT) EXAFS spectrum for WT-Zn is shown in 3.8 (empty circles), together with its first-shell fit performed in the R region 1.0-2.2 Å (continuous line). According to the fit results, indicating the presence of 0.9±0.1 S atoms, and consequently 3.1±0.1 N/O, we expect Zn^{2+} to bind 1 Cys and 3 more ligands to be identified most probably from His, carboxylic acids and water molecules [47]. We propose that the S atom can indeed be assigned to a Cys residue on the basis of a Metalloprotein Database and Browser (MDB) search: where S atoms provide a ligand for Zn^{2+}, they are always located in Cys residues, never Met.

An inspection of the 3-4 Å region of the FT spectrum can provide information on the number of His residues present in the cluster, as argued in Paragraph 2.2; the quantitative criterion proposed, however, refers only to Zn sites in proteins measured at room temperature, for which we had a wide collection of experimental observations. For TH, measured at 80 K we cannot use such a criterion, nevertheless a qualitative comparison with reference samples can serve the purpose: we compared the transhydrogenase FT spectrum with those of reference samples recorded in similar experimental conditions, and in which the number of His residues per Zn^{2+} cluster is known. The chosen reference proteins were Thermolysin (TLS) and Superoxide Dismutase (SOD). Structural information about
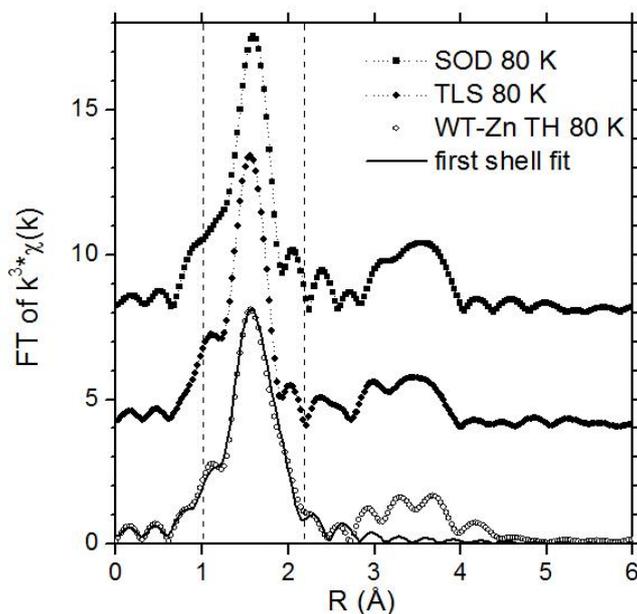

Figure 3.8: Experimental Fourier Transformed EXAFS spectra of model proteins and wild-type transhydrogenase, and its first-shell fit. The Fourier Transformed K-edge Zn spectra of wild type transhydrogenase with added Zn^{2+} (open circles), superoxide dismutase (filled squares), and of thermolysin (filled circles), embedded in PVA films, are shown. The spectra were collected at 80 K and transformed in the k range 2.5-12.5 Å^{-1}. The continuous line shows the first-shell fit for the transhydrogenase spectrum; the fitting range 1-2.2 Å is indicated by the dashed lines.

the Zn sites in the two proteins are given in Paragraph 2.3; we recall that Zn^{2+} is 4-coordinated in both cases, and the number of coordinated His residues is 2 in TLS, 3 in SOD. The FT spectra reported in Figure 3.8 clearly show that the amplitude of the MS contribution in the 3-4 Å region of the WT-Zn transhydrogenase spectrum is comparable to that of TLS. Measuring the amplitude of these

contributions by numerical integration in the region 3-4 Å we obtain a value of 1.20 for the WT-Zn spectrum, very close to the value of 1.29 calculated for TLS and markedly different from the value of 1.85 obtained for SOD. We can therefore assume that Zn^{2+} in the WT-Zn sample is bound to 2 His residues, like the Zn^{2+} of TLS. The fourth ligand could be either a carboxylic acid residue or a water molecule, but since ATR-FTIR experiments indicated a change in the vibrational modes of Asp/Glu residues upon Zn^{2+} binding [90], it is reasonable to assume the presence of an Asp/Glu in the Zn^{2+} coordination shell.

We therefore chose a 1Cys 2His 1Asp/Glu coordination shell as a starting model. In order to minimize the correlation between the fitting parameters, the fit to the EXAFS function in k space was performed in the step-by-step procedure described in Paragraph 2.2. In steps (1) and (3) the free parameters of the fit were three first-shell distances (one for each different amino acid residue), and the bending angle of Asp/Glu. In step (2) the free parameters were first-shell Single Scattering (SS) DWs and a common value, σ_{MS}^2, for all of the Multiple Scattering (MS) paths. When fitting the EXAFS spectra measured in the experimental samples the distances obtained in step (3) always converged to values that were in close agreement with those found in step (1), proving the robustness of the fitting procedure. In all fitting steps a shift in the energy origin (ΔE_0) was included as a free parameter. The fitting model was chosen and tested on the WT-Zn data, and then extended to that from the other experiments. The fits for all four samples are shown in Figure 3.7 as continuous lines. The agreement between the fitting curve and the experimental signal is very good, particularly in the WT-Zn and BH91K-Zn samples, where the signal-to-noise ratio (S/N) is higher as a consequence of the higher Zn fluorescence counts. The best-fit first-shell distances (i.e. the output of step 3 of the fitting procedure) relative to the model, 1Cys 2His 1Asp/Glu, are reported in Table 3.1.

The corresponding DW factors (i.e. the output of step 2) are given in Table 3.2. First-shell distances are the same within the error in all samples except in BH91K, which has a slightly shorter first-shell average distance. The fact that the same model provides a good fit for WT and WT-Zn, as well as for BH91K and BH91K-Zn, confirms that the contaminating Zn^{2+} in the samples with no exogenous metal ion populates a fraction of the same site resolved in the presence of added Zn^{2+}.

Since the same cluster of residues nicely fits the signals measured in samples containing the native and the mutant transhydrogenase, we are led to think that the Zn^{2+} binding site is indeed the same, i.e. that the mutated βHis91 is not

	ΔE_0 (eV)	Zn-S (Å)	Zn-$N_{\epsilon 2}$ (Å)	Zn-$O_{\delta 1/\epsilon 1}$ (Å)	α (°)	R_{factor} (%)
WT	-2.7 (1.3)	2.288 (0.013)	2.016 (0.016)	1.972 (0.025)	112 (3)	7.2
WT-Zn	-2.4 (1.0)	2.290 (0.009)	2.018 (0.014)	1.969 (0.015)	114 (2)	6.1
BH91K	-4.6 (1.5)	2.268 (0.011)	1.976 (0.013)	1.99 (0.03)	106 (1)	7.3
BH91K-Zn	-2.0 (0.8)	2.287 (0.007)	2.018 (0.010)	1.968 (0.015)	113 (2)	4.8

Table 3.1: Best-fit first-shell distances. First-shell distances were obtained by fitting the EXAFS spectra of the four samples to the 1Cys 2His 1Asp/Glu model described in the text. The parameter α is the angle Zn-$O_{\delta 1}$-C_γ or Zn-$O_{\epsilon 1}$-C_γ of the Asp/Glu residue respectively, and whose starting value is set to 105°. A common distance for Zn-$N_{\epsilon 2}$ has been assigned to the two His residues, and this therefore indicates the average of the two actual distances. The values in brackets are the 1σ errors.

	σ^2_{Zn-S} (10^{-3}Å^2)	σ^2_{Zn-N} (10^{-3}Å^2)	σ^2_{Zn-O} (10^{-3}Å^2)	σ^2_{MS} (10^{-3}Å^2)	R_{factor} (%)
WT	3.9 (1.0)	3.5 (2.5)	5 (5)	7.6 (2.2)	7.2
WT-Zn	3.1 (0.7)	4 (3)	2 (4)	4.4 (1.1)	6.2
BH91K	2.3 (0.8)	3.1 (1.4)	4.0 (2.0)	6.9 (2.5)	7.2
BH91K-Zn	2.3 (0.6)	2.9 (1.4)	3.6 (2.3)	4.5 (1.0)	4.8

Table 3.2: Best-fit Debye Waller factors. First-shell Single Scattering Debye Waller factors and a Multiple Scattering DW common to all MS paths, as determined by fitting the four EXAFS spectra with the 1Cys 2His 1Asp/Glu model. The subscripts relative to $N_{\epsilon 2}$ of His and $O_{\delta 1}$/$O_{\delta 2}$ of Asp/Glu have been omitted. The values in brackets are the 1σ errors.

involved in the binding cluster. If it were, the spectrum of the BH91K-Zn sample would differ from that of WT-Zn, and it would reveal a Lys residue instead of one of the His ligands in the coordinating cluster. In order to test quantitatively this possibility we have fitted the BH91K-Zn spectrum with the model, 1Cys 1His 1Lys 1Asp/Glu. Figure 3.9 compares the best fit to the two models. It clearly shows that the Lys-containing cluster gives a poorer fit to the spectrum. This is particularly evident when comparing the FT signals (lower panel): the 1Cys 1His 1Lys 1Asp/Glu model (indicated with LYS in the figure) gives a worse fit to the experimental data than the 1Cys 2His 1Asp/Glu model (indicated with HIS), both in the first shell and in the 3-4 Å region where the His contributions are predominant. In fact, the R factor of the fit increases from 4.8% to 7.1% when a His is replaced with a Lys. The inadequacy of the LYS model is further illustrated by the best-fit value found for the DW factor common to all MS paths, i.e. σ^2_{MS}=0.000±0.002 Å^2. The DW factors and the amplitude reduction factor S^2_0 are correlated fit parameteres; the value of S^2_0=0.95 was calculated and kept fixed during the analysis, which could affect the resulting values of the DWs. Although
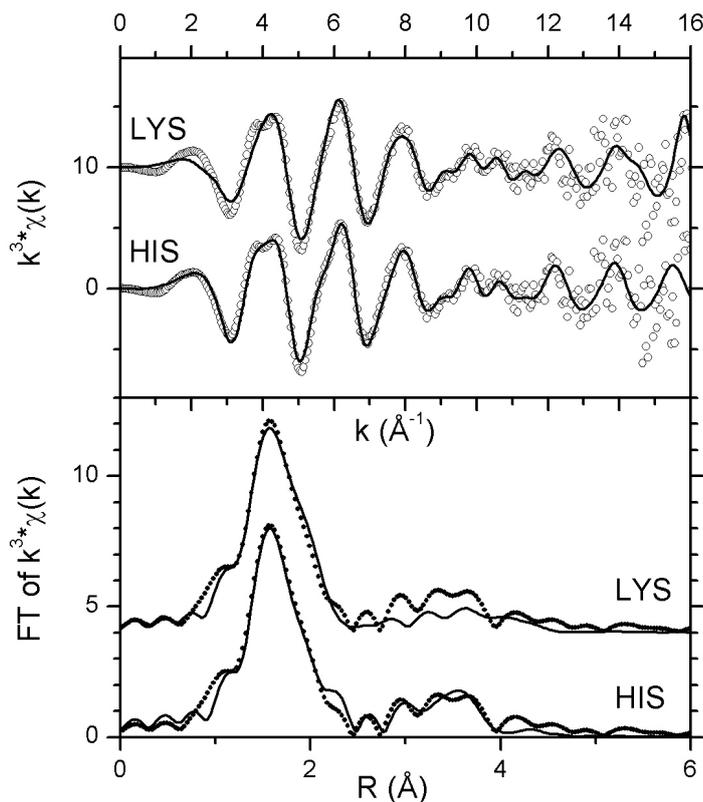
Figure 3.9: The EXAFS spectrum of Zn in the βH91K mutant of transhydrogenase fitted with two different models. Upper panel: the experimental Zn K-edge EXAFS spectrum of the BH91K-Zn sample (open circles) and its k-space fits (continuous lines) based on the 1Cys 1His 1Lys 1Asp/Glu model (indicated by the tag LYS), and the 1Cys 2His 1Asp/Glu model (indicated by HIS). Lower panel: Fourier Transforms of the experimental spectra (black dots) and of the fits (continuous lines) shown in the upper panel.

a value of 0.95 is quite reasonable for S_0^2 we have tested the effect of varying this value on the DW factors. We fitted the BH91K-Zn spectrum to the 1Cys 1His 1Lys 1Asp/Glu model setting S_0^2 to 0.8, a lower but still reasonable value. This procedure led again to the best fit value $\sigma_{MS}^2 = 0.000 \pm 0.002$ Å², and to a worse fit, as indicated by the R factor of 8.6%. The result of a null multiple scattering DW factor, independent of the choice of the S_0^2 value, would lack physical meaning, indicating that the MS contributions relative to the LYS model are too weak to reproduce the experimental spectrum, and that they could fit only if they were not damped (i.e. associated to a null Debye Waller factor). The differences in the R factors and in the reduced chi squared values (53±16 and 130±40 respectively for the HIS and LYS models) are at the limit of resolution from a statistical point of view. However, all the evidence reported above strongly supports the idea

that the Zn^{2+}-binding site does not involve a Lys residue. Moreover, if one of the two His residues present in the Zn^{2+}-binding motif of wild type transhydrogenase were replaced by another ligand in the mutant, we would encounter differences in the XANES regions of the spectra. However, these regions are nearly identical to one another (Figure 3.6). Finally, we must in principle consider also the possibility that in the βH91K mutant Zn^{2+} loses one of its ligands and results in a coordination number of 3. In fact, if βHis91 were one of the Zn^{2+} ligands in the wild type protein, its substitution with a Lys would introduce a positively charged amino acid into the coordination sphere, which would clearly have a lower propensity to bind Zn^{2+}. This possibility can however be excluded because a decrease in the coordination number would cause a significant decrease in the EXAFS oscillation amplitude and in the XANES white-line intensity which we do not observe at all in the BH91K spectra. In conclusion, XANES and EXAFS analysis identifies the same Zn^{2+}-binding cluster, formed by 1Cys, 2His, and 1Asp/Glu residue in the native transhydrogenase and in its βH91K mutant (see Figure 3.10). This strongly argues against an involvement of βHis91 in the Zn^{2+}-binding site that is detected by XAFS.

3.3.3 Discussion

The Zn K-edge XAS spectra of purified E. coli transhydrogenase described above were interpreted in terms of a single Zn^{2+}-binding site. The existence of two (or more) sites with different structures and binding affinities is unlikely: it would contradict the finding that spectra obtained with low levels of contaminating Zn^{2+} have very similar features to those obtained with much higher concentrations of the metal ion. Thus, even small differences in binding affinity would lead to significant changes in the relative site occupation, to which XAS is rather sensitive. The existence of two structurally different sites with identical Zn^{2+}-binding affinities is not ruled out by these observations but would be difficult to reconcile with the internal consistencies of the XAS analysis, notably the convergence of results from different fitting procedures.

It is likely therefore that there is indeed only a single site in transhydrogenase that binds Zn^{2+} in the concentration range of our experiments; the XAS experiments indicate that the site is a tetrahedral cluster formed by the imidazole rings of two His residues, and the side chains of a Cys and an Asp/Glu (Figure 3.10). We propose that Zn^{2+} inhibition of proton translocation by transhydrogenase [89, 90] is a consequence of the metal ion binding to this site. In the absence of metal ions, one or more of the amino acid residues in the cluster
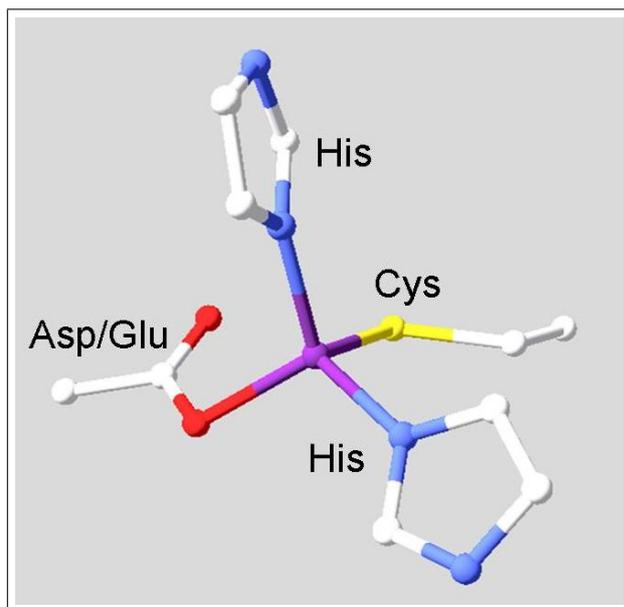
Figure 3.10: The proposed Zn^{2+}-binding cluster. Wireframe structure of the Zn^{2+}-binding cluster which gives the best fit of the experimental XAFS data obtained for both the wild-type transhydrogenase and its βH91K mutant.

may be involved in proton transfer, and Zn^{2+} would inhibit by competing with H^+ binding to the site. The inhibitory Zn^{2+}-binding sites in reaction centers of Rhodobacter sphaeroides [84, 68], and in avian and bovine cytochrome bc1 complexes [13] also have two His and one Asp/Glu residues, and these sites too are thought to be involved in proton transfer and to bind the metal ion competitively.

Mutation of βHis91 to Lys was found to have somewhat similar effects on transhydrogenase reactions to treatment of the enzyme with Zn^{2+}. An analysis of these effects led to the conclusion that both the amino acid substitution, and metal-ion binding, interfere with proton translocation. In other experiments the addition of Zn^{2+} to wild-type transhydrogenase led to changes in the ATR-FTIR difference spectrum that were attributable to effects on His and Asp/Glu residues in the protein [90]. In the βH91K mutant the amplitude of the signals due to Asp/Glu, and particularly those due to His, in the Zn^{2+}-induced difference spectrum were decreased prompting the suggestion that βHis91 in wild-type transhydrogenase might be a ligand for the Zn^{2+}. However, in the present study we found that the Zn XANES and EXAFS spectra were very similar in the wild-type and in the βH91K mutant. The same binding cluster of 1Cys, 2His and 1Asp/Glu residue provides the best fit for both sets of data. An alternative model, a 1Cys, 1His, 1Lys, and 1Asp/Glu Zn^{2+}-binding cluster, yielded a worse fit to the measured spectra of the mutant protein. The inadequacy of the mutated cluster was particularly evident in the spectral region dominated by the His contributions. It

is very unlikely that βHis91 participates directly in Zn^{2+}-binding.

It seems that, although βHis91 and the Zn^{2+}-sensitive site both function in the proton-translocation pathway of transhydrogenase, they are spatially separate. Indeed, it was noted [14] that the low residual rate of reverse transhydrogenation in βH91K was further inhibited by Zn^{2+}, and that the elevated rate of cyclic transhydrogenation in the mutant was further stimulated by metal ions. To reconcile the present XAFS results with the effects of the βHis91→Lys mutation on the Zn^{2+}-induced FTIR difference spectra we propose the following. The binding of Zn^{2+} to a site in transhydrogenase dII formed by 1Cys 2His and 1Asp/Glu, causes conformational changes that are transmitted to the somewhat more distant βHis91 residue. The FTIR difference spectra detect changes in all His residues that result from Zn^{2+} binding, including βHis91 and one or both of the His residues which form the binding site. When βHis91 is mutated into a Lys, Zn^{2+} binding still perturbs the vibrational bands of the His residue(s) that act as ligand(s), but the absence of the contribution from βHis91 causes the observed decrease in the overall His FTIR signal.

Chapter 4

The Fe^{2+} site of Photosynthetic Reaction Center

The photosynthetic Reaction Center (RC) of purple bacteria is the pigment-protein complex that initiates solar energy conversion. It was, in 1982, the first membrane protein the crystal structure of which was determined by X-ray diffraction, yielding to Hartmut Michel, Johann Deisenhofer and Robert Huber the Nobel prize in 1988. Moreover, being the smallest unit capable of performing light-driven electron transfer, RC became a model system in the study of the relation between the electron transfer process and protein conformational dynamics. The RC activity, indeed, has been demonstrated to be conformationally gated, but the cause and the nature of the conformational changes are still debated. An Fe^{2+} atom is located along the electron pathway in the photosynthetic apparatus of RC: a structure-stabilizing role has been proposed for such atom, and no evidences of the direct participation of Fe^{2+} in the redox processes have been reported so far. Nevertheless, weak evidences of transient modifications of the Fe^{2+} environment upon light absorption have been provided, and this might mean that Fe^{2+} is involved in the conformational changes that allow or inhibit electron transfer.

The Fe^{2+} site of RC is clearly a very interesting probe to investigate electron transfer, therefore we present in this chapter the XAFS studies performed on such system, both in static and in pump-probe time-resolved mode. Part of the results has been published as: Veronesi et al. *The Fe^{2+} Site of Photosynthetic Reaction Centers Probed by Multiple Scattering X-ray Absorption Fine Structure Spectroscopy: Improving Structure Resolution in Dry Matrices* Biophys. J., 2008, 95, 814-822 (reference [109]).

4.1 Structure and function of RC

The bacterial photosynthetic RC is made up of three subunits, L, M and H, and of the following cofactors: four bacteriochlorophylls, two bacteriopheophytins (H_A and H_B), two quinones (Q_A and Q_B) and one non-haem iron atom (Fe^{2+}); a detailed review of the structure and function of bacterial RC is given in reference [110]. Two of the four bacteriochlorophylls form the special pair (usually labeled as P or D), that works as the primary donor by ejecting an electron upon solar light absorption. The three dimensional structures of RCs from *Rhodobacter sphaeroides* (see Figure 4.1) and *Rhodopseudomonas viridis* are similar, as well as the cofactor arrangement, except for the presence in *Rps viridis* of a four-haem cytochrome on its periplasmic site. XRD structures [111, 112, 113, 114] reveal
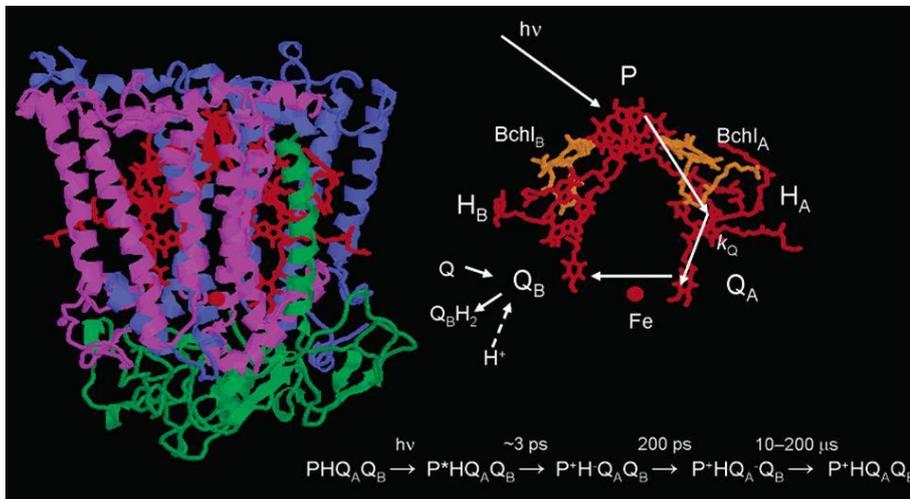

Figure 4.1: XRD structure of the RC from *Rb. sphaeroides*. The photosynthetic apparatus, situated in the deep interior of the protein, is reported on the right, together with white arrows indicating the electron pathway. In the bottom of the figure, the redox reactions performed by the protein and their characteristic time scales are indicated.

the presence of 11 transmembrane α-helices in the complex, five each for the L and M subunits and one in the H subunit; the L and M subunits compose a cylindrical core in which the photosynthetic apparatus is enclosed. The cofactors are arranged along two branches, A and B (L and M in *Rps viridis*), approximately related to each other by a two-fold symmetry axis perpendicular to the membrane plane and crossing the Fe^{2+} site; however, the redox reactions chain takes place spontaneously along the A branch, where the rate of electron transfer is at least 20 times larger than along the B branch. This is ascribed to small deviations from the 2-fold symmetry [115].

4.1.1 Photochemistry of RC

The energy conversion operated by bacterial RC starts with charge separation in the bacteriochlorophyll special pair (indicated as D in Figure 4.2); a sequence of redox reactions involving the RC cofactors delivers the electron to the primary quinone Q_A, situated $\sim$25 Å away from the special pair (step 1 in Figure 4.2). The special pair is then reduced by a cytochrome (step 2) and the charge of Q_A is passed to the secondary quinone acceptor Q_B (step 3). Absorption of a second photon activates another turnover of the RC photochemistry, leading to double reduction of Q_B; subsequently, two protons are uptaken in order to form an hydroquinone that leaves the RC (step 7) and is replaced by an exogenous neutral quinone (step 8), allowing the cycle to start again. A detailed description of the photochemical cycle can be found in reference [110].
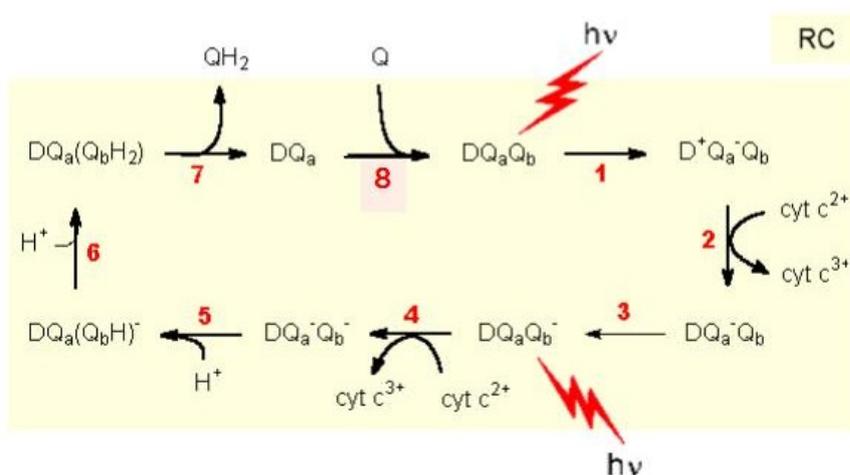

Figure 4.2: The photochemical cycle operated by photosynthetic RC from purple bacteria.

Step 1 takes place in $\sim$ 200 ps, while the intermediate acceptor bacterio-pheophitine H_A is reduced in a time interval of the order of 3 ps: the short time intervals characterizing the first steps of the cycle are crucial to avoid recombination of the electron with the bacteriochlorophyll special pair.

The solar light wavelengths able to initiate the RC turnover are determined by the electronic structure of the photosynthetic apparatus formed by the cofactors; the overall UV/Vis absorption spectrum of RC from *Rb. Sphaeroides* is shown in Figure 4.3. Three main peaks are evident in the long-wavelengths region of the spectrum, each one originated by the prevalent contribution of a different cofactor: the peak at $\sim$750 nm has been empirically attributed to the bacteriopheophytines, the one at $\sim$800 nm to monomeric bacteriochlorophylls and the peak at $\sim$860 nm
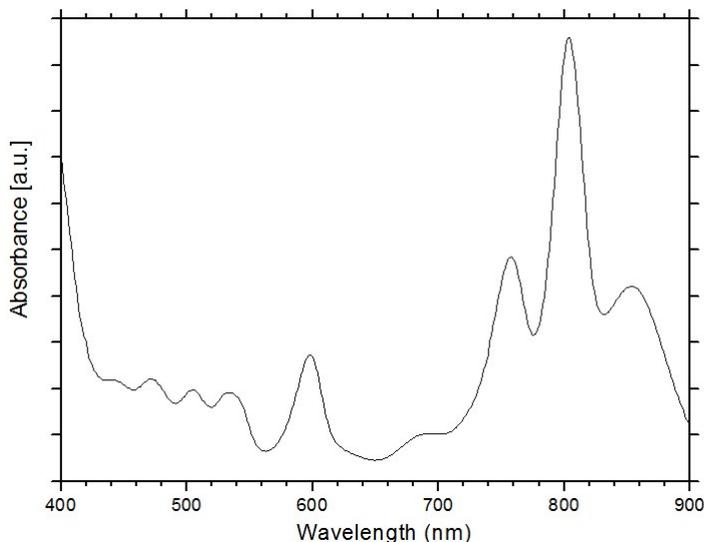
Figure 4.3: UV/Vis absorption spectrum of reaction centers isolated from *Rb. Sphaeroides.*

to the special pair. A quantitative study of the RC absorption bands can be found in reference [116].

4.1.2 Conformational gating and the role of the Fe^{2+} site

Independent experimental evidences showed that the electron transfer from Q_A to Q_B is rate limited by the dynamics of a conformational change in the protein [117, 118]; this conformational change is thought to be central for a full understanding of the general mechanism of electron-transfer coupled to proton uptake in energy transducing proteins. A possible explanation has been proposed by Stowell et al. [111] on the basis of the XRD structure of RC frozen in dark and in the light: in the light-induced charge-separated $D^+Q_AQ_B^-$ state, Q_B^- is located approximately 5 Å from the Q_B position in the charge-neutral (DQ_AQ_B) state, and has undergone a 180° twist around the isoprene chain. This was interpreted as the light-induced switch from a distal-inactive to a proximal-active position, the latter bringing Q_B close enough to Q_A to allow electron transfer (see Figure 4.1). Subsequent evidences based mainly on vibrational spectroscopy argued against this model, rather supporting the existence of a single Q_B site independently on temperature and illumination conditions [119, 120]. Alternative models have been proposed for the gating process, including protonation and/or changes in H-bonds patterns, protein relaxation, or a combination of the two processes [121, 122, 123]. It is clear that, in spite of the efforts dedicated to the

investigation of the rate-limiting step in RC's photochemistry, the interpretation of this scenario is still debated and no conclusive evidence has been reported so far.

The Fe^{2+} ion is located between the two quinones (Figure 4.1) and binds the $N_{\epsilon 2}$ atoms of 4 His residues and the 2 oxygens ($O_{\epsilon 1}$, $O_{\epsilon 2}$) of the Glu-M234 residue in a bidentate configuration (see Figure 4.5), as confirmed by XRD structures [111, 112, 113, 114]; His-M219 is H-bonded to Q_A, and His-L190 to Q_B, when the secondary quinone is in the proximal position. This arrangement gives rise to a quinone-His-Fe "bridge". Due to its position and tight coupling with the surroundings, the Fe^{2+} atom can therefore serve as a probe for the local structural changes associated with the final electron-transfer steps of the RC, hence it has been extensively studied: a pioneering work made by Debus et al. [124] showed that Fe-depleted RCs could still carry out electron transfer, but with slower rate constant and lower efficiency; however, when RCs were reconstituted with other divalent metal ions, included Zn^{2+} which cannot undergo a change in its oxidation state, they recovered essentially the same electron transfer properties of the native protein. In a more recent time, Hermes et al. performed a time-resolved XAS study of the Fe^{2+} [125], yielding no evidence of a change in the Fe oxidation state upon light absorption. It all strongly argues against the involvement of Fe^{2+} in the redox reactions chain; the atom is therefore expected to be a passive spectator of the electron transfer between the two quinones and of the eventual conformational changes connected to it.

4.2 Using XAFS to improve structural resolution in dry matrices

XAFS represents a method of choice to detect subtle modifications of the Fe^{2+} site, nevertheless relatively few XAFS studies have been performed on such site. Two pioneering works [126, 127] first revealed the presence of six atoms in the Fe^{2+} coordination shell, a result subsequently confirmed by XRD. These early analyses were based on the comparison with spectra of model Fe compounds and provided an average distance between Fe^{2+} and first-shell atoms. A more recent XAFS study was aimed to resolve temperature- and light-induced structural changes in the Fe^{2+} site [128]; this analysis, however, did not consider the Multiple Scattering (MS) contributions to the XAFS signal. Moreover, none of the three XAFS studies performed so far took explicitly into account possible heterogeneities in the first shell distances, yielding at most a common distance between Fe^{2+} and

the four His nitrogen atoms, and a single average distance for the two oxygen atoms of the Glu residue coordinating in a bidentate configuration [128].

We conducted a XAFS study on the Fe^{2+} site of RC embedded in different matrices: a non-interacting PVA film and a strongly dehydrate trehalose matrix. The study brought to the determination of high-resolution Fe-ligands bond length and underlined the effect of a strongly interacting matrix on a non-solvent-exposed portion of the protein such as the Fe^{2+} ligating cluster.

4.2.1 Materials and methods

Sample preparation. The RC was isolated and purified from *Rb. sphaeorides* R-26, a carotenoid-less spontaneous mutant, according to Gray et al. [129]. EDTA was present at 10 μM in all buffers used during the reaction center isolation. In order to remove traces of exogenous proteins and metals, the purified RC suspension was additionally flowed through a DEAE-Sephadex (Sigma) column, dialyzed for 15 hours, at 4 °C, against 10mM Tris buffer, pH 8.00, 0.025% LDAO, 10 μM EDTA, supplemented with 5 gr Chelex resin (Chelex 100, Bio-Rad) / 100mL, and concentrated to 60 μM by ultafiltration (VIVASPIN 50 kDa cut-off, VIVASCIENCE, Hannover, Germany). The magnesium/iron molar ratio in the RC final sample, determined by Inductively Coupled Plasma Atomic Emission Spectroscopy, was 4.9±0.5. A previous, similar metal analysis, performed by atomic absorption spectroscopy on a highly purified reaction center preparation, yielded a Mg/Fe ratio equal to 5.2±0.3 [130]). Considering 4 bacteriochlorophyll molecules for each reaction center, the iron content of our preparation was slightly sub-stoichiometric with respect to the RC, indicating that no exogenous iron was present in the samples used for XAFS measurements. Polyvinyl alcohol (PVA) - RC films were prepared starting from 1.4 mL of 24 μM RC, 2.5% PVA solution (PVA 130000 MW, Fluka), dried under nitrogen flow into a 3.0×3.0×0.3 cm Teflon holder. The PVA-RC film was folded several times prior to be exposed to the X-ray beam. Trehalose (>99% purity) was purchased from Sigma and used without any additional purification. RC-trehalose dried glasses were prepared as follows: 800 μL of 60 μM RCs were concentrated to 170 μL under nitorgen flow, raising the RC concentration to 282 μM. A sugar/RC molar ratio equal to 10^3 was achieved by direct addition of solid trehalose into the RC solution. The trehalose-RC solution was layered onto a sintered boron nitride support (Good-fellow Cambridge, Huntingdon, England) and extensively dried under nitrogen flow.

The content of residual water in the PVA and trehalose matrices can be evalu-

ated by near-infrared (NIR) spectroscopy from the area of the combination band of water in the 1930-1960 nm spectral region, using the RC absorption band at 802 nm as an internal standard [131]. In the case of the PVA sample, this estimate was performed directly on the sample used for XAFS measurements. For the trehalose sample, a direct determination was prevented by the boron nitride support. Parallel measurements on trehalose matrices of the same composition, layered on a transparent optical glass, showed that the extensive dehydration treatment resulted in a similar content of residual water in the PVA film and in the trehalose matrix, i.e., $\sim 4 \times 10^3$ water molecules per RC protein. Further desiccation under nitrogen flow at room temperature did not reduce the water content further, in agreement with previous determinations [132, 131, 133]. To test the structural and functional integrity of the protein after XAFS measurements, both the PVA and trehalose samples were redissolved with the proper amount of water. For both samples, visible-NIR spectra taken after irradiation were indistinguishable from that of the RC solution before preparation of the dehydrated matrices. Moreover, no alteration of primary photochemistry was detected, as evaluated from the extent of the primary donor (P^+) photo-oxidized by a laser pulse and from the unaffected lifetime of the primary $P^+Q_A^-$ charge-separated state. The experimental details of these time-resolved optical spectroscopy measurements were given elsewhere [132, 131].

XAFS data collection and analysis. Fe K-edge measurements were performed at the GILDA-BM08 beam-line [23] of the European Synchrotron Radiation Facility (ESRF). A Si(111) double crystal monochromator employing dynamical sagittal focusing was used; the photon flux was of the order of 10^{11} photons per second and the spot size was $\sim 1 \times 1$ mm^2. Data were collected in fluorescence mode. Samples were measured at room temperature in the energy range 6900-7910 eV; the final spectra are given by the average of multiple scans for a total integration time of 60 s/point for each sample. The molecular graphics program MOLDRAW [107] was used to build the structural model of the Fe^{2+} site. Signal extraction and k-space fits were performed accordingly to the procedures described in Chapter 2.2, by means of the *IFEFFIT* package [55] and its graphic interfaces Athena and Artemis [56]. Theoretical amplitudes and phase shifts were calculated using FEFF8.2 [1], in which scattering potentials are calculated by overlapping the free atom densities in the muffin-tin approximation and then adding the Hedin-Lundqvist form for the exchange potential. The amplitude reduction factor, S_0^2, was estimated as 0.94 by the program for the starting model

and was kept fixed during the fitting procedure. All of the multiple scattering signals constituted by up to five scattering paths involving atoms belonging to the same residue and with an effective length ≤ 5 Å were taken into account. The fits were performed directly in k-space, in the range 2.5-10 Å^{-1} and with a k weight of 3, following the step by step procedure described in Paragraph 2.2, even if no parametrization of the DWs is available for non-heme Fe clusters (recall that in the first step of the fitting procedure the DWs are fixed to the values provided by Dimakis and Bunker's parametrization [18, 17, 20] and only structural parameters are fitted): in order to assign the DW factors, we performed ab initio simulations grouping the atoms in three shells, depending on their distance from Fe^{2+}. The values providing the oscillation amplitude most similar to the experimental one (i.e. 0.005 Å^2 , 0.009 Å^2 and 0.007 Å^2 for the three shells respectively) were chosen.

4.2.2 Results

In the near-edge region, our spectra (see Figure 4.4) are very similar to the ones reported in previous studies which indicated six ligands in the Fe^{2+} coordination shell [126]. Considering this, and the XRD information about the Fe ligation
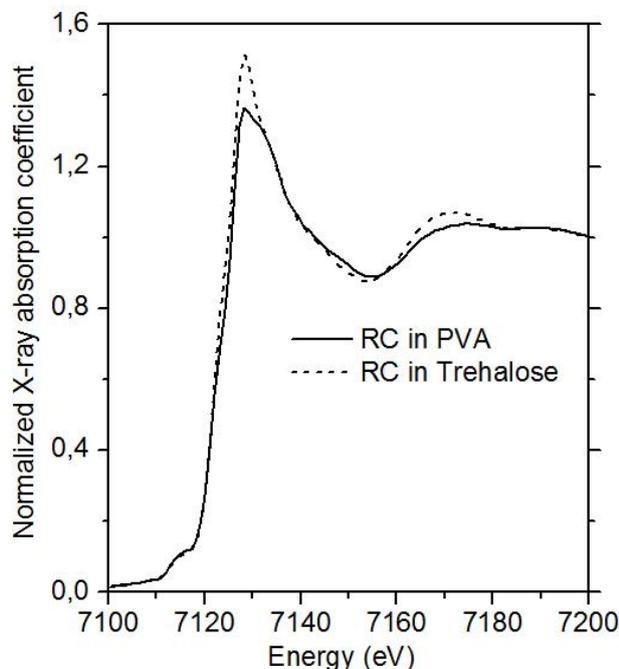

Figure 4.4: Normalized XANES region of the spectra measured for RC embedded in a PVA film (continuous line) and in a strongly dehydrate trehalose matrix (dashed line).

pattern reported in paragraph 4.1.2, the distorted octahedral cluster shown in Figure 4.5 was built and used as a starting structural model. Structures of single amino acids and first shell distances were derived from the statistics described in Paragraph 2.1.2. The vertices of the basis of the octahedron are the $N_{\epsilon 2}$ atoms of His-M219 and His-L190 and the two oxygen atoms of Glu-M234; the corresponding amino acids lie in the plane defined by this basis (see the caption of Figure 4.5 for numeric values of the structural parameters); the two remaining vertices are occupied by the $N_{\epsilon 2}$ of His-M266 and His-L230. For Glu in bidentate conformation, a distribution of Fe-O_ϵ distances is found, indicating a correlation between Fe-$O_{\epsilon 1}$ and Fe-$O_{\epsilon 2}$ bond lengths (see the database MESPEUS, MEtal Sites in Proteins at Edinburgh UniverSity [45] and reference [53]). In order to build a reliable starting model, the coordination geometry of the Glu residue was parametrized using the Fe-$O_{\epsilon 1}$ distance and the angle α (see Figure 4.5, insert); to set their starting values, the Fe site for a number of XRD structures of photosynthetic RCs with resolution higher than 2.4 Å (PDB codes: 1DXR, 1E6D , 1EYS, 6PRC, 1AIJ), found through searching the MDB, was considered and the average values of Fe-$O_{\epsilon 1}$ (2.12 Å) and α (94°) were set as the starting one. The Fe-$O_{\epsilon 2}$ distance for the Glu residue follows from the Fe-$O_{\epsilon 1}$ distance and α parameters, according to the equation:

$$(Fe-O_{\epsilon 2})^2 = (Fe-O_{\epsilon 1})^2 + (O_{\epsilon 1}-O_{\epsilon 2})^2 - 2[(Fe-O_{\epsilon 1})(O_{\epsilon 1}-O_{\epsilon 2})]cos(\alpha-\beta) \quad (4.1)$$

where β is the $O_{\epsilon 2}\hat{O}_{\epsilon 1}C_\delta$ angle (see Figure 4.5, inset). Its value (29°), as well as the distance (2.20 Å) appearing in (4.1), is derived from the Engh and Huber's survey of amino acids structures [48] and kept fixed all over the analisys. The XAFS oscillations measured in the PVA and in the trehalose matrices are shown as dotted lines in Figure 4.6 (a) and (b), respectively. The corresponding best fitting XAFS functions are represented as continuous lines. The figures also show the main contributions to the best fitting signal, coming from both single and multiple scattering (the two curves at the bottom of each panel report the multiple scattering contributions with greatest amplitudes). In both spectra, three main oscillations are evident at 4, 6 and 8 Å^{-1}, but their shape and relative amplitudes differ in the two cases, presumably as a consequence of a different arrangement of first-shell atoms in the two different matrices. The structural refinement was carried out in the Rigid Body Refinement scheme [4], initially using as fitting structural parameters: (i) a common shift in the energy origin for all paths; (ii) a distance variation for Fe-$O_{\epsilon 1}$; (iii) a Glu bending angle $\Delta\alpha$ (see Figure 4.5, insert); (iv) a Fe-$N_{\epsilon 2}$ distance variation for the four Fe-coordinating
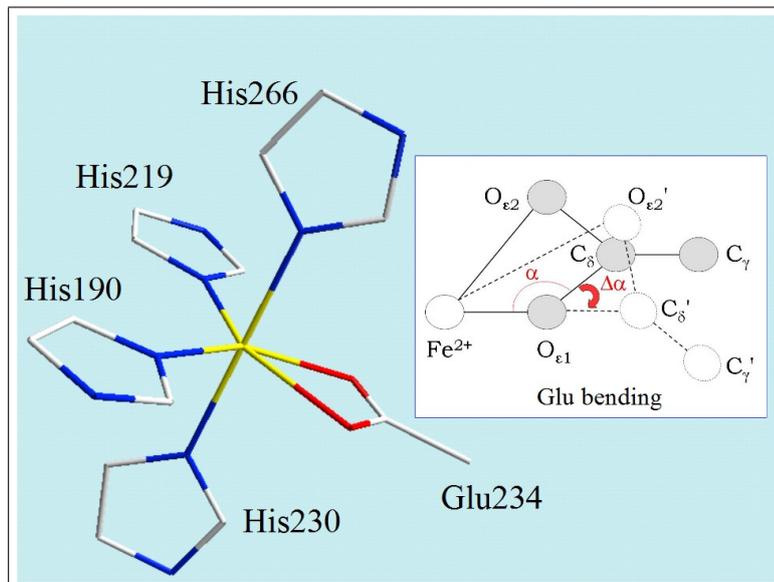
Figure 4.5: Reference structural model of the Fe^{2+} ligand cluster. The inset shows the bending of the Glu residue, i.e. a rigid rotation by an angle $\Delta\alpha$ around an axis through $O_{\epsilon 1}$. The angle α was set to 94°, and the Fe^{2+}-$O_{\epsilon 1}$ distance to 2.12 Å; the $O_{\epsilon 2}$ atom of Glu is thus placed at 2.34 Å from Fe^{2+}. The two His residues in plane with the Glu residue are placed symmetrically, so that the angle between the two Fe^{2+}-$N_{\epsilon 2}$ bonds of the two His, and the angles between each Fe^{2+}-$N_{\epsilon 2}$ bond and the direction Fe^{2+}-C_{δ} of the Glu are equal (120°). The target Fe^{2+}-$N_{\epsilon 2}$ distance for the His residues was 2.16 Å (see text for details).

histidines. The fit was initially performed with a model in which only a single, common Fe-$N_{\epsilon 2}$ average distance was considered for the four amino acids (this will be referred to as the *4 His* model).

The relatively poor results obtained with this approach (as testified by the high R factor of the fit, see Table 4.1), suggested us to introduce different Fe-$N_{\epsilon 2}$ distances for the different coordinating His residues, i.e. an heterogeneity in the His cluster. We considered in particular two fitting models in which two distinct Fe-$N_{\epsilon 2}$ distances were introduced. In the first, which we call *2+2 His* model, we have considered two groups, each composed by two His residues, and we assign to each group an independent distance from Fe^{2+} (Fe-$N_{\epsilon 2}(1)$ and Fe-$N_{\epsilon 2}(2)$). In the second, which we call the *3+1 His* model, the His cluster is divided in one group of three residues, for which we consider one common, average distance from Fe^{2+} (Fe-$N_{\epsilon 2}(1)$), and one single His residue, which is allowed to move independently (Fe-$N_{\epsilon 2}(2)$). The quantitative results obtained for each model, i.e. the first-shell distances, the bending angle $\Delta\alpha$, and the first-shell Debye-Waller factors, are reported in Table 4.1, together with the goodness-of-fit indicators, the R factor and the reduced chi-square (χ_ν^2). The fitting criterion relies in fact on
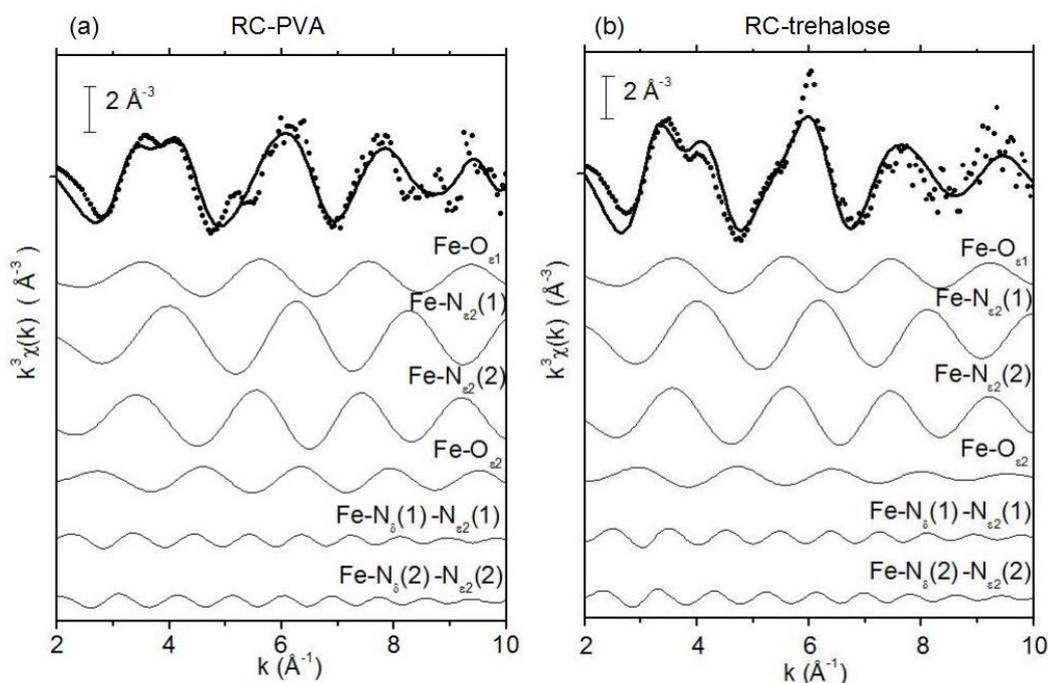
Figure 4.6: Experimental k^3 weighted XAFS functions (dots) measured in RCs embedded into (a) the PVA film and (b) in the dry trehalose matrix. For both spectra we report its best fit to the *2+2 His* model (continuous bold lines), and the corresponding dominating contributions coming from single and multiple scattering. The values of best fitting parameters are given in Table 4.1.

the minimization of the R factor, and the statistical choice of the most probable cluster on the reduced chi-square test (see Chapter 2.2).

Matrix	Model	Fe-N$_{\epsilon 2}$(1)	Fe-N$_{\epsilon 2}$(2)	Fe-O$_{\epsilon 1}$	$\Delta\alpha$	σ^2	ΔE_0	R (%)	χ^2_ν
		(Å)	(Å)	(Å)	(°)	(10^{-3} Å^2)	(eV)		
	4 His	2.14 (2)	-	1.94 (3)	6 (2)	6 (2)	7 (2)	34 (10)	46 (14)
PVA	*2+2 His*	2.18 (4)	1.99 (3)	2.09 (3)	3 (2)	4 (1)	5 (3)	18 (6)	13 (4)
	3+1 His	2.16 (3)	1.97 (6)	2.02 (3)	5 (2)	5 (2)	6 (2)	25 (8)	33 (11)
	4 His	2.16 (2)	-	1.98 (3)	5 (3)	6 (1)	7 (2)	29 (9)	22 (7)
Treha.	*2+2 His*	2.17 (5)	2.04 (2)	2.15 (7)	2 (2)	5 (1)	5 (2)	21 (6)	12 (4)
	3+1 His	2.12 (2)	1.98 (4)	2.19 (5)	1 (2)	5 (1)	6 (2)	23 (7)	21 (7)

Table 4.1: Structural and dynamical parameters determined in the PVA and in the trehalose matrix. The bending angle of the Glu residue is indicated by $\Delta\alpha$, and the first-shell Debye-Waller factor by σ^2. For each RC-matrix system the fit was performed according to three structural models, characterized by a different grouping of the four His residues (see text for details). The number reported in bracket is the 1σ error on the least significant digit calculated by the fitting program.

4.2.3 Discussion

Removing degeneracy in the His coordination. The introduction of multiple scattering contributions allowed to better resolve the structure of the histidine cluster in the Fe^{2+} coordinating shell, partially removing the degeneracy present in the previous studies. Our analysis revealed in fact that the four His residues of the cluster can be split up into two groups, each consisting of two histidines, characterized by a different average distance from Fe^{2+}: 2.18 Å and 1.99 Å respectively for RCs embedded in a weakly interacting medium (PVA). Such a variation cannot be predicted from XRD data of RC crystals, even at the maximal resolution (1.87 Å) attained at present [114]. Crystallographic structures of the RC show, however, differences and asymmetries in the environment experienced by the four His residues which coordinate the Fe^{2+} atom. In particular, His-M219 appears to be hydrogen bonded to the primary quinone acceptor Q$_A$ (bond length 2.84 Å [114]). A second histidine (His-L190) is at hydrogen bond distance (2.71 Å [114]) from the Q$_B$ molecule, when the quinone moves to the proximal configuration. These hydrogen bond interactions could be responsible for the heterogeneity resolved by us in the histidine Fe-N$_{\epsilon 2}$ distances. In this

respect, we observe that the occupancy of the proximal Q_B position appears to be 65% even in the dark XRD structure at pH 8.0 [114], so that almost two of the four His residues may be expected to be hydrogen bonded under our measuring conditions. The suggestion that the two different Fe-$N_{\epsilon 2}$ distances observed stem from the interaction of two His residues with the bound quinones is in line with the observation that only one of the distances is changed when the RC is embedded in a strongly interacting matrix (trehalose glassy matrix), rather than in the weakly interacting PVA film (see below). Our model also allowed to remove degeneracy in Fe^{2+} coordination by the two oxygen atoms [128]. According to the crystallographic model, a Glu residue (Glu-M234) is bonded to Fe^{2+} in bidentate configuration. For this residue, our analysis yields in PVA a coordination distance Fe-$O_{\epsilon 1}$ = 2.09 Å with $\Delta\alpha$=3°, which correspond to Fe-$O_{\epsilon 2}$ = 2.43 Å. The average of these two Fe-O distances (2.26 Å) differs markedly from the Fe-O distance of 1.88 Å determined in a previous XAFS study [128]. We note that this latter value is hardly compatible with distances encountered in data banks for bidentate carboxylate groups (see the database MESPEUS and ref [45]). Exploration of metal coordination by carboxylates in bidentate configuration has revealed, in the case of Co, Cu, and Zn an inverse proportionality between the O_1 and O_2 distances from the metal [53]. Figure 4.7 shows a correlation plot between the Fe-O_1 and Fe-O_2 distances, extracted from the MESPEUS Database. Only distances obtained from XRD studies at a resolution higher than 1.5 Å have been included. It appears that, also in the case of Fe, the two distances are inversely related. An average Fe-O distance of 1.88 Å [128] is clearly incompatible with the data clustering of Figure 4.7, to which, at variance, the Fe-$O_{\epsilon 1}$ and Fe-$O_{\epsilon 2}$ distances obtained by fitting our data to the *2+2 His* model, nicely conform, both in trehalose and in PVA (see closed symbols in Fig. 4.7).

Finally we note that from the first shell distances resolved in the present work (Table 4.1) an average coordination distance of 2.14±0.02 Å is obtained in the PVA film. Such a value is in excellent agreement with those determined in the pioneering studies by Eisenberger et al. [127] and Bunker et al. [126], i.e. 2.10±0.02 Å and 2.14±0.02 Å, respectively.

Matrix effect. In a previous study [65], performed on cyt c, we have shown that XAFS is a sensitive tool in revealing both structural and dynamical local effects induced on protein metal sites by the incorporation in strongly dehydrated trehalose matrices. In the case of cyt c, such effects consist in: (1) the induction of severe distortions of the metal coordinating geometry, with respect to the local
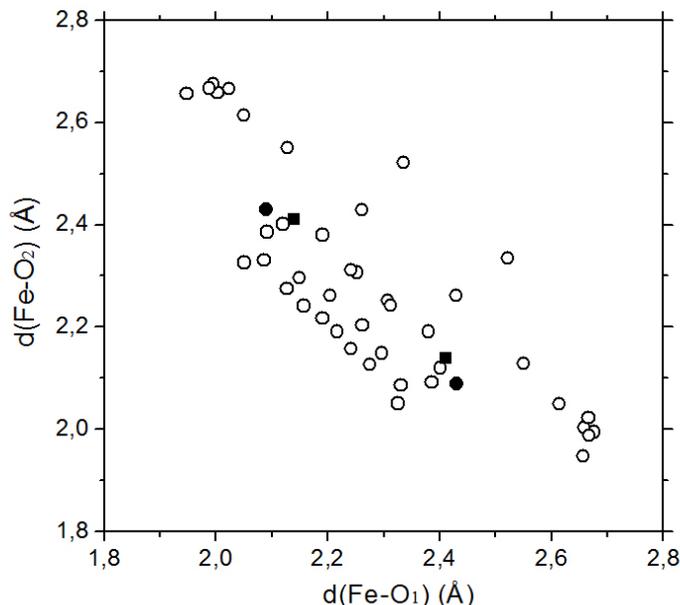
Figure 4.7: Relationship between the Fe-O_1 and Fe-O_2 distances in bidentate coordination with Asp and Glu residues. Data shown as open circles were extracted from the MESPEUS database, developed at Edinburgh University [45], selecting XRD protein structures at resolutions higher than 1.5 Å. Solid symbols correspond to distances determined in the PVA (circles) and in the trehalose matrix (squares) in this study (see Table 4.1 and text for details). The labels O_1 and O_2 can be permuted, so that each carboxylate group is shown twice in the plot, which is symmetrical about the diagonal line, d(Fe-O_1) = d(Fe-O_2).

metal site structure observed in liquid solutions and in PVA films; (2) a decrease of the DW factors, indicating a strong reduction, at the local level, in the static and/or dynamical disorder of the protein.

In view of these results, in the present work, we have performed a comparative analysis of the Fe^{2+} site in RCs embedded in PVA films and in extremely dehydrated trehalose matrices. The aim was to ascertain whether or not the trehalose effects observed on the local structure/dynamics of a small, soluble protein (like cyt c) were also detectable in the case of a large membrane protein, like the RC. By analyzing the RC Fe^{2+} site, we have found that two of the coordinating His residues are placed on average further on from the Fe^{2+} atom in the trehalose matrix, as compared to the PVA film. The Fe-$N_{\epsilon 2}$(2) distance, 1.99 Å in PVA, increases in fact to 2.04 Å in trehalose (Table 4.1). Interestingly, the other two coordinating His residues, instead, do not change their average distance Fe-$N_{\epsilon 2}$(1), strongly suggesting a tighter bond. We propose to identify these latter two His residues with His-M219 and His-L190, which could show a firmer attachment to their position, being also hydrogen bonded to the quinones Q_A and Q_B, respec-

tively. In trehalose, also the Glu residue undergoes a sizeable displacement with respect to the position in PVA, as inferred from the stretching of the Fe-$O_{\epsilon 1}$ bond (see Table 4.1). Although the Fe-$O_{\epsilon 1}$ distance increases from 2.09 Å in PVA to 2.15 Å in trehalose, the Fe-$O_{\epsilon 2}$ distance does not vary appreciably (2.41 Å and 2.43 Å in trehalose and PVA, respectively), because the value of $\Delta\alpha$ changes from 3° in PVA to 2° in the trehalose matrix. It turns out, in summary, that, as observed in the case of the heme Fe of cyt c, the incorporation into a dehydrated trehalose matrix causes the selective elongation of some of the coordinating bonds.

Although the elongations observed in the first-ligand distances are 3-4 times smaller in the RC Fe^{2+} site, as compared to the cyt c iron, it is noteworthy that a detectable, local deformation occurs at a site of the RC which is well shielded by the protein matrix from the surrounding water-trehalose matrix. This argues in favor of a strong protein-water-trehalose interaction at the RC surface, the structural effects of which propagate to the interior of the protein over a distance of at least 15 Å. This value is inferred from the minimum distance of His-M266 and His-L230 from the surface water molecules identified by XRD [114] (pdb file 2J8C). These evidences of long-range structural effects due to the strong interaction between the RC surface and the water-trehalose matrix are consistent with the large effects induced on electron transfer by incorporation of the RC into a dried trehalose matrix [132, 133] and with the much less severe effects observed in PVA films [131, 133].

The first-shell DW factors determined in the PVA film and in the dehydrated treahlose matrix are the same (within the experimental error, see Table 4.1). This is in contrast with what observed for the heme Fe of cyt c; in this case, in fact, the incorporation in trehalose leads to a dramatic decrease of the DW factors [65]. The behavior observed in the RC can be rationalized when considering that both static (structural) and dynamical (vibrational) disorder contributes to the DW factors. Since the incorporation into an extensively dehydrated trehalose matrix is expected to reduce drastically protein dynamics also in the case of the RC (see ref [133], and references therein), the lack of trehalose effects on the DW factors suggests that the values of the DW factors, both in PVA and in the trehalose matrix, are largely dominated by the static (structural), local heterogeneity of the Fe^{2+} site, being the dynamical (vibrational) contribution essentially negligible. In line with this interpretation, Eisenberger et al. [127], by examining the temperature dependence of the DW factor, concluded that the disorder probed by their XAFS analysis was predominantly static (structural) in nature. They estimated a static disorder parameter ≥ 0.01 Å^2. The DW factors we found (see Table 4.1)

are even smaller than this estimate of the static component only, indicating that the vibrational contribution has to be negligible.

This notion is further supported by a recent molecular dynamics study [134] in which a coarse-grained model of the wild type *Rb. sphaeroides* RC was used to obtain a rigidity profile of the protein, made up of force constants describing the ease of displacements of each residue with respect to the rest of the structure. The result of this local flexibility study was that regions around the Fe^{2+} center exhibit the largest force constants, and that the Fe^{2+} ligands are among those yielding the highest values. It is not surprising, therefore, that the dominating contribution to the measured DW factors comes from static disorder and that, as a consequence, a possible reduction in the dynamics, induced by incorporation in the trehalose matrix, has no effect on the measurable DW factors. At the same time, the high local rigidity of the Fe^{2+} coordinating residues, and of the surrounding protein region, can explain the sizeable structural distortions observed in the Fe^{2+} site when the RC is embedded in the trehalose matrix.

We propose that locking the RC surface to the water-trehalose matrix, and hardening the embedding structure upon decreasing the content of residual water, strains the protein surface, most likely through the formation of water-mediated hydrogen bonds between the sugar molecules and surface protein groups. Such a perturbation, due to the rigidity of the interposed protein region, is expected to propagate to the buried Fe^{2+} site, and to be strong enough to distort the geometry of the Fe^{2+} site. Our suggestion that the two His residues, which move further from the Fe^{2+} atom in trehalose, are His-M266 and His-L230 is consistent with the values of the force constants calculated by Sacquin-Mora et al [134] for these residues ($\sim$200 kcal mol^{-1}Å^{-2} and $\sim$306 kcal mol^{-1}Å^{-2}, respectively), which are systematically lower than those evaluated for the other two His residues (i.e. $\sim$330 kcal mol^{-1}Å^{-2} for His-L190 and $\sim$540 kcal mol^{-1}Å^{-2} for His-M219, which forms a strong H-bond with Q_A).

4.3 Time-resolved XAFS experiments

In spite of the amount of studies aimed to clarify the details of RC's photochemistry, the role of the Fe^{2+} site in the photosynthetic apparatus is still debated, as introduced in Paragraph 4.1.2. A time-resolved XAFS study is ideal to fucus right on this site, however its feasibility is strongly compromised by the difficulty of obtaining a high enough signal-to-noise ratio in pump-probe conditions. Only one time-resolved XAFS study has been reported so far [125] and it pro-

vided weak evidences of a conformational change in the Fe^{2+} ligands cluster in a characteristic time of the order of 500 μs in Q_B-reconstituted RCs.

We present here a time-resolved XAFS study of the Fe^{2+} site in Q_B-depleted RCs, where the electron transfer to Q_A is followed by recombination of the electron with the bacteriochlorophyll special pair.

4.3.1 Materials and methods

Sample preparation.   The RC was isolated and purified from *Rb. sphaeorides* R-26 as described in Paragraph 4.2.1.

The PVA film was prepared starting from a solution containing RC 100 μM, 0.025% LDAO, 10 mM O-phenantroline and and PVA 2.5% in 10 mM TRIS buffer, pH 8.00. 2 μl of the solution were layered on a pure SiO_2 disk, into a circular track with internal and external radii of 4.6 cm and 5.0 cm respectively (see Figure 4.8), delimited by paraffin wax. The sample was dried overnight under N flow, then the paraffin was removed.

Figure 4.8: Sample for the time resolved experiment: RCs from R-26 *Rhodobacter sphaeroides* are embedded in a PVA matrix and layered over a quartz CD.

Experimental setup.   XAS spectra were collected at the the NW14A beamline [135] of the Photon Factory Advanced Ring (PF-AR). The 6.5 GeV synchrotron PF-AR operates in single-bunch mode, delivering X-ray pulses at a frequency of 749 kHz with a pulse duration of about 100 ps. The X-ray beam was

conveyed by an undulator with period length of 36 mm (U36), able to provide an intense monochromatic X-ray source by use of a double crystal monochromator and a focusing mirror, covering the energy range 5-20 keV. Laser pulses were delivered by a Ti:sapphire laser system with 800 nm fundamental wavelength. The repetition rate was 473 Hz, and the laser power 22.5 mW; the laser spot size was ~ 1.9 mm^2, so the resulting energy density was 25 J/mm^2. The X-ray spot size was 0.9 mm^2 and it was completely included in the area irradiated by the laser.

The geometry of the experiment is shown in figure 4.9: the sample is mounted on a rotating step-motor and forms an angle of 45° with the incoming X-rays; the laser direction is displaced of 10° with respect to the X-rays. The photo-multiplier detector is placed at 90° with respect of the X-ray beam and 45° with respect to the sample plane; a 3 μm Mn high-pass filter was used to eliminate the fluorescence due to elastic scattering.
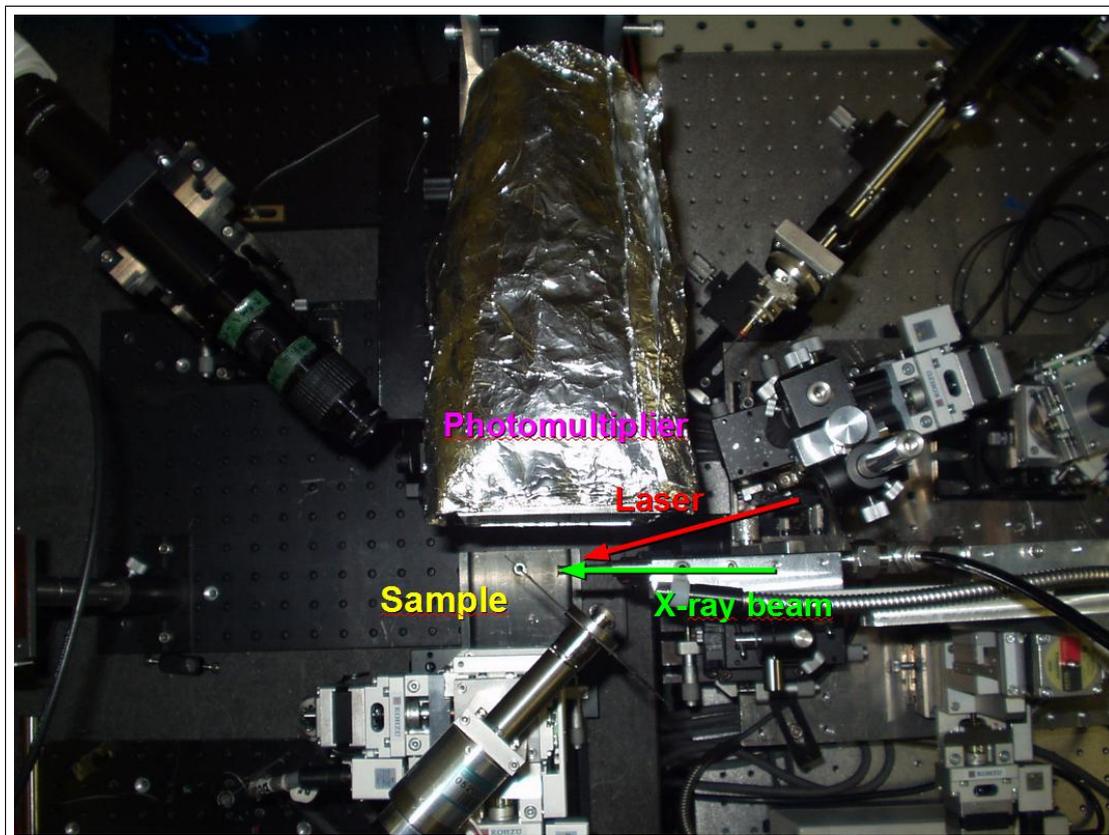

Figure 4.9: Top view of the experimental setup for the pump-probe time resolved experiment at the beamline NW14A of the PF-AR synchrotron. The sample is mounted over a step motor that allows it to rotate at the chosen frequency. The X-ray beam forms an angle of 45° with the sample, and the angle between laser and X-ray beams is 10°.

The rotation frequency of the sample disk is determined by the photochem-

istry of the protein: since the system takes at least 300 ms to relax after a pho-toexcitation, this is the minimum time that must elapse between two following excitations of the same spot (i.e. the revolution time of the disk). Moreover, the disk rotation step must be such that two following X-ray beam and laser spots do not overlap; to calculate the minimum frequency we must take into account the laser spot size in the direction of the rotation movement (the X-ray spot is smaller) and consider this as the minimum possible step. In summary, the range of allowed frequencies was determined as follows:

- **Maximum rotation frequency** determined by the RC relaxation time: 1 revolution in 300 ms $\Rightarrow$ 1/300=3.3 Hz

- **Minimum rotation frequency** determined by the laser spot size:

$$\nu_{min} = 1/(r * (289/s)) \tag{4.2}$$

where 289 mm is the length of the probed circumference, s is the laser spot size in mm and r is the repetition time in seconds.

The lower limit of the frequencies interval was calculated as 2.2 Hz, therefore we chose as the disk rotation frequency the central value of 2.7 Hz.

Synchronization between X-rays from PF-AR and the laser pulse was achieved by means of a 1 kHz chopper as described in reference [135]; the timing chart for an example of a NW14A time-resolved experiment is depicted in Figure 4.10.

The chosen delay times between laser and X-ray pulses ranged from 10 μs to 500 μs. The general settings chosen for the experiment are summarized in table 4.2.

Energy intervals	7100-7170 eV
Energy steps	1 eV
Integration time	4 s
Filter	Mn
Laser wavelength	800 nm
Laser energy density	25 J/m^2
Delay time	10-200-500 μs
Disk rotation frequency	2.7 Hz

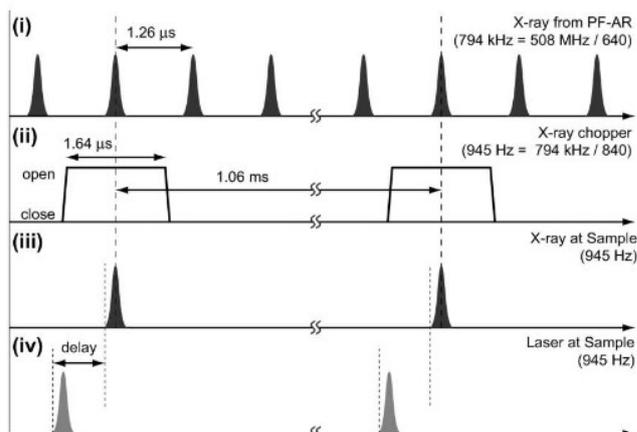Table 4.2: Chosen settings for RC pump-probe measurements

Figure 4.10: Timing chart of the synchronization system, operated by means of the X-ray chopper, between 794 kHz X-rays and 945 Hz laser pulse in 1:1 ratio. The figure is taken from reference [135].

4.3.2 Results

Three series of spectra were collected, corresponding to the delay times of 10 μs, 200 μs and 500 μs; the 10 μs time scale was never explored before for RCs, while the other two time scales have been investigated: evidences exist in literature of a voltage change across the protein happening 200 μs after the main charge separation event [136]; this voltage change could be ascribed to the formation of a transient state involving the quinone Q_A and, if so, it would have an influence on the Fe coordination shell. It is interesting to notice that the voltage change was observed also in RC samples where the quinone Q_B was removed, that is in our experimental conditions. The 500 μs time scale, instead, has been studied with XAFS [125], bringing no evidence of a change in the Fe^{2+} oxidation state.

Raw spectra were averaged and the S/N ratio was evaluated: the signal is given by the difference between the white line maximum and the pre-edge absorption intensity, while the noise is evaluated as the standard deviation of the last 30 points from a polynomial curve. In table 4.3 the calculated S/N are reported.

The S/N ratio is proportional to the square root of the number of spectra, in agreement with the Poissonian statistics. In order to compare the spectra collected before and after the laser pulse for each delay time, we chose a common normalization procedure aimed to rescale the data with the smallest possible manipulation: the average of the first 10 points was subtracted as a constant to

Delay (μs)	N. of averaged spectra	S/N
10	89	48
200	139	56
500	175	62

Table 4.3: Signal to noise ratios for the 3 series of spectra characterized by a different delay between laser and X-ray pulses.

the whole spectrum, then the average of the last 5 points was set to 1. With this normalization, we do not consider that the background due to scattering from the disk increases with energy (we subtract a constant value instead of a line with non-zero slope); nevertheless, the energy dependence of the background is reasonably the same before and after the laser and this normalization allows a more reliable comparison than after an arbitrary background subtraction. When the delay time is 10 μs or 200 μs, spectra before and after the laser pulse do not seem to show differences within the noise, as shown in figure 4.11.
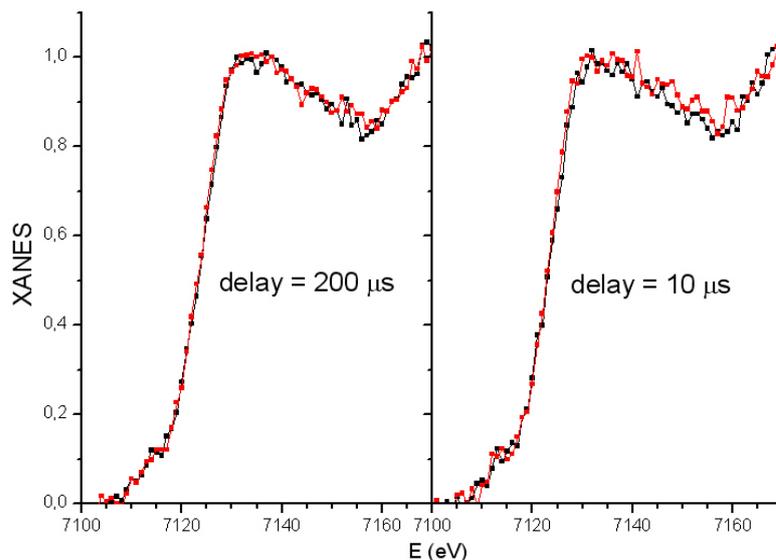

Figure 4.11: Final spectra for the delay times of 200 and 10 μs resulting from the average of 139 and 89 spectra respectively. The normalization procedure is described in the text. The red spectrum is collected before the laser pulse, while the black spectrum after it.

The spectral features that are expected to change as a consequence of conformational changes in the Fe cluster or Fe oxidation are:

- The pre-peak centered at about 7114 eV. The pre-peak is known to be deter-

mined by $1s \rightarrow 3d$ transitions [39], forbidden in the dipole approximation, therefore its intensity increases when the Fe shell gets further from spherical geometry and the dipole approximation fails. This feature is visible also in the XANES spectrum shown by Hermes et al. [125] and in the static spectra that we previously measured at ESRF, shown in figure 4.4 (continuous line). Some small changes are detected in Fe pre-peak of metalloproteins upon Fe oxidation, for an example see Figure 2 of the paper from Rich et al. [37] where Fe-XANES of reduced and oxidized forms of Myoglobin and Hemoglobin are compared. We notice that in these proteins Fe oxidation causes a small decrease in the pre-peak intensity and only in the case of Hemoglobin a displacement to higher energies (of about 2 eV).

- The edge position at 7124 eV (chosen as the central value of the main peak of the derivative). The edge is expected to shift of 2-3 eV upon Fe oxidation, while a smaller shift could be ascribed to conformational changes [37].

- Post edge features. Conformational changes like ligands displacement or rotation cause a change in the relative intensity of post-edge features, as can be demonstrated for example by means of XANES simulations.

For the 500 μs delay spectra shown in figure 4.12, a difference between dark and photoexcited structures is just visible in the pre-peak region. In particular it is noticeable that the spectrum measured 500 μs after the laser pulse (black curve in figure 4.12) has a less evident pre-peak feature if compared with the dark structures and with photoexcited structures measured at a different delay (red curves in figure 4.11). Nevertheless, a more detailed data analysis is needed to make clear whether it is a spectral change or an artifact.

Data filtering. In order to detect whether the spectral change at 500 μs delay is real, we eliminated the high frequency noise components from the spectrum. Filtering was performed by a Fourier low-pass filter, choosing as a cutoff frequency 0.2 eV^{-1}. It means that we consider as noise, and therefore neglect, oscillations whose period is less than 5 eV. Since filters often introduce distortions in the data, we must first of all control if the trend of experimental XANES spectra are conserved after filtration. In figure 4.13 the experimental XANES are reported, together with the corresponding filtered signals. We see that the filtered spectra overlap very well to the experimental ones, especially in the region 7110-7130 eV where eventual changes due to photoexcitation should be present. We can therefore focus on the comparison between filtered data.
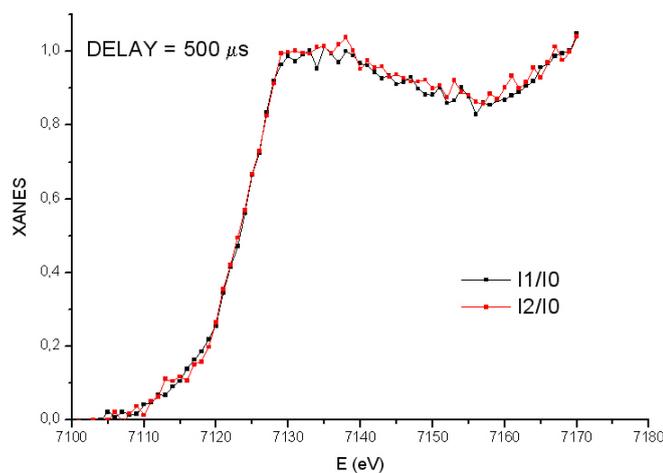
Figure 4.12: Fe-XANES spectra of RC collected before (red line) and 500 μs after (black line) illumination with laser pulse.

Filtered spectra exhibit exactly the same trend on the whole energy range, except for the region 7110-7120 eV, where the pre-peak is present (figure 4.14). A comparison between the derivatives of the two signals makes clear that only in the 7110-7120 eV region the XANES spectra are different both in the features position and in their relative amplitudes (see inset of figure 4.14), while for the rest of the energy range the derivatives nearly overlap. The features appearing in the post edge region are indeed centered at the same energy values and they only slightly differ in their relative amplitude.

If the same filtering operation is performed on the 10 μs and 200 μs data set, the spectra collected before and after photoexcitation do not show any difference in the 7110-7130 eV energy region. In this case the filter brings to the uniformity of the spectra, confirming that if any difference was present it was due to the high frequency noise, to the contrary of what we saw in the 500 μs data set where the differences were evidenced after filtration.

Simulations. In order to understand what conformational changes could be associated to changes in the pre-peak region, this region was simulated making use of the finite difference method provided by FDMNES [5]. The elements of the transition matrix are calculated by the program in the dipole approximation. As an example, two structures were taken into account, differing only in the position of the Glu-M234 residue with respect to Fe. The starting model for the Fe cluster with Glu in bidentate arrangement (see figure 4.5) is the one determined in our previous work [109] described in Paragraph 4.2, where the Fe-O first shell
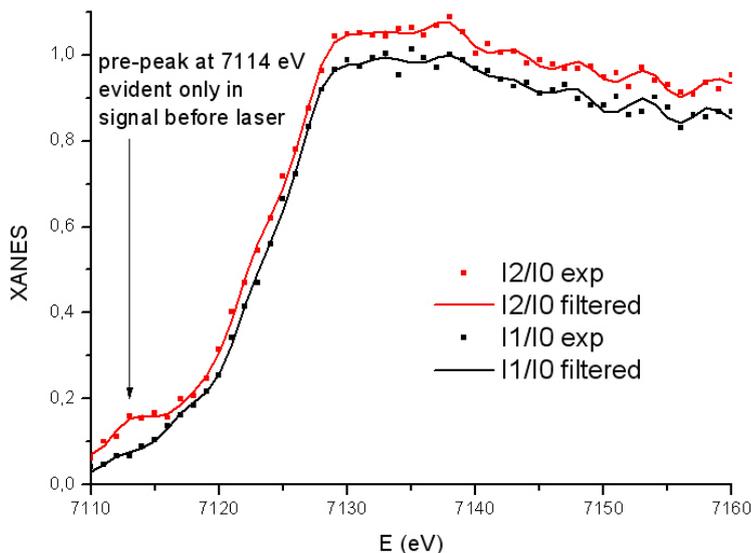
Figure 4.13: Normalized experimental signals collected before (red dots) and 500 μs after (black dots) the laser pulse. On each experimental signal the corresponding filtered spectrum is reported (red line before laser, black line after laser).

distances are $Fe - O_{\epsilon 1} = 2.09$ Å and $Fe - O_{\epsilon 2} = 2.41$ Å; the deformation taken into account is a rotation of the Glu residue in such a way that $Fe - O_{\epsilon 1}$ does not change and $Fe - O_{\epsilon 2} = 3$ Å, according to monodentate configurations reported by Harding [53].

Since in monodentate configuration one of the first shell ligands is put further from Fe than in bidentate, this conformational change corresponds to a deviation from spherical symmetry. The choice of this conformational change is also due to the findings of Hermes et al. [125], that hypothesize a change in Glu configuration from bi- to mono-dentate upon laser excitation in Q_B reconstituted Reaction Centers. Simulations are shown in figure 4.15. The simulations show that a conformational change bringing the system to a less symmetric configuration (Glu passing from bi- to mono-dentate, respectively red and black curves in figure 4.15) causes an increase in the prepeak amplitude. A subtle shift of the edge to lower energies is also present, as we can appreciate considering the position of the peaks in the derivative spectra (see inset of figure 4.15). This shift is estimated by the simulations to be about 0.3 eV, therefore undetectable with our energy resolution (1 eV). These considerations on simulated spectra are in line with what was experimentally found by Hermes et al [125]. It is important to notice that a change in the Fe oxidation state would cause a shift of the edge to higher energies
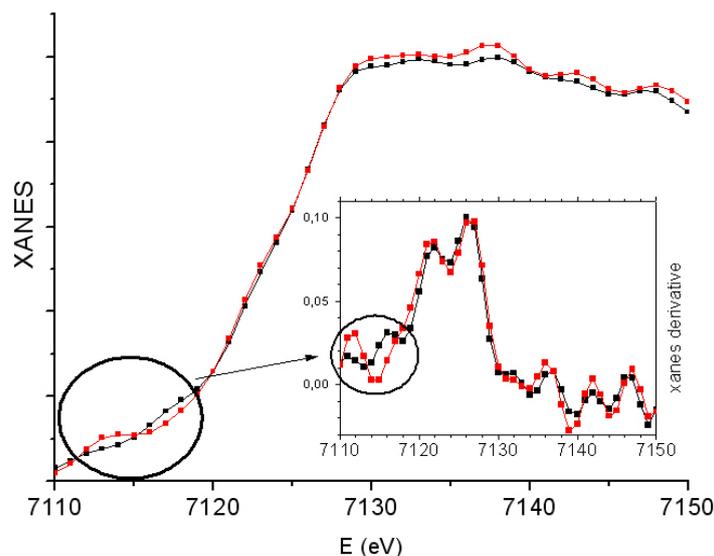
Figure 4.14: Comparison between RC Fe-XANES collected before (red curve) and 500 μs after (black curve) excitation with laser pulse, and their derivatives (inset, same color code). Signals were filtered using a low-pass Fourier filter to eliminate high frequency (≥ 0.2 eV^{-1}) noise contributions.

of about 2-3 eV and would be clearly detectable in this experiment.

The change of XANES spectra encountered in this experiment at the time scale of 500 μs after photoexcitation consists in the lowering of the pre-edge feature and is not accompanied by a shift in the edge position (that means that if a shift is present, it must be smaller than 1 eV, therefore we can take into account conformational changes and exclude Fe oxidation). This trend in the spectra can be related to conformational changes following from the presence of the electron at the Q$_A$ site that lead to a transitory more symmetric Fe cluster configuration.

4.3.3 Conclusions and future developments

In the 500 μs data set the pre-peak feature centered at 7114 eV encountered in the dark structure becomes indistinguishable from the noise after photoexcitation. To check if this difference in spectra before and after illumination is real and to put it in evidence with respect to the noise background, a Fourier low-pass filter was applied to the data, in such a way to eliminate high-frequency oscillations due only to noise. The same filter was applied to all of the collected data sets, revealing that the pre-peak feature is evident in all of them except for the spectrum collected 500 μs after laser excitation. Moreover, considering dark and
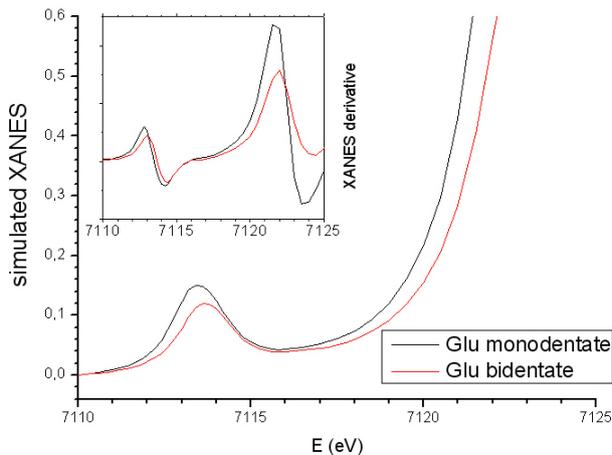
Figure 4.15: Pre-peak region of simulated Fe cluster of RC where the Glu residue is in bidentate (red line) or monodentate (black line) configuration.

illuminated structures at 500 μs delay, we notice that the derivatives of filtered spectra are nearly identical, except for the region 7110-7120 eV, where minima and maxima appear at different positions, indicating a different distribution of pre-edge features. We are therefore allowed to assert that a difference in the Fe coordination shell exists between dark and photoexcited structure 500 μs after light absorption. Since no energy shift is visible we can assume that the Fe^{2+} does not change its oxidation state. As widely discussed in literature and confirmed in this specific case by FDM simulations, the closer the coordination shell of the irradiated metal ion is to spherical symmetry the lower is the amplitude of its XANES pre-peak. Therefore we can infer that 500 μs upon light absorption, the presence of the ejected electron in the surroundings of the Fe atom causes a transitory change in the local symmetry in the direction of a more spherical structure. These conformational changes could bring for example to a temporary higher uniformity of first shell distances. It is noticeable that Hermes et al. [125] revealed a conformational change in the Fe cluster at the same time scale, but this change could be explained assuming a distortion in the direction of a less spherical structure, to the contrary of what we found. The two findings are not contradictory, since the RCs they studied were reconstituted in Q_B, while our sample does not contain the second quinone. It is therefore likely to find different conformations of the Fe coordination shell, depending on the different reactions taking place around the Fe site ($Q_A^- \rightarrow Q_B$ electron transfer in the reconstituted sample, recombination of the photoexcited electron to the bacteriochlorophyll dimer in our sample).

To ascertain these results, however, it is necessary to record spectra with a S/N ratio high enough to appreciate changes in the spectral features without application of a filter. We propose therefore to improve the experimental setup by introducing a 2-dimensional Soller slit between the sample and the photomultiplier, in order to reduce the background due to scattering from the quartz disk. Moreover the X-ray flux can be enhanced by removing the 1 kHz chopper and employing only a heat-load chopper: in this configuration, the sample would be irradiated with pulse trains of 10 μs duration, made of 10 single X-ray pulses each and delivered every 1 ms. The combination of these improvements is expected to increase the S/N of a factor $\sim$20 and to allow the spectral changes to emerge by reducing the magnitude of the noise.

Chapter 5

Matching DFT calculations and XAFS experiments: the application to MbCO

The analysis procedure introduced in Chapter 2 and applied throughout this work, relies on the existence of a parameterization for the DWs that helps disentangling the structural parameters from the dynamical ones, in the case where the metal binding site is completely unknown and the simultaneous determination of the two sets of parameters would lead to multiple solutions. However, the DWs parametrization is provided by Dimakis and Bunker only for Zn sites [17, 18] and, in a simplified version, for the Fe-porphyrine complex [20], while it is desirable to get calculated DWs for all metal-amino acid bonds, in order to extend our analysis method. When the parametrization is not available, indeed, it is a common practice to assign reasonable starting values to the DWs and fix such values to allow the determination of structural parameters alone. This could bring a bias in the fitting procedure, as a consequence of the correlation between parameters.

The problem of static an dynamical disorder is an open issue in EXAFS analysis, since the approaches implemented in data analysis codes are often unsuitable to account for Multiple Scattering (MS) contributions and non-thermal components; for example FEFF8 employs the Debye Model, which provides exact values only for Single Scattering DWs in homogeneous systems. Therefore, the need of matching computational techniques with MS XAFS analysis for the determination of disorder parameters arose, and was handled either with Density Functional Theory (DFT) methods or with Molecular Dynamics (MD). The former approach was practiced by Dimakis and Bunker and provided the parameterizations used in this work; the latter avoids explicit determination of DWs by generating different

sets of atomic configurations with MD and calculating then the averaged EXAFS spectrum over the different configurations [137, 138, 139]. However, the quantitative determination of the DWs is highly desirable, since it carries important information about the dynamical properties of the system. The existing studies that match *ab initio* calculations with XAFS on biological systems are revised in [28]: they treat the determination of geometric end electronic properties of the investigated system, but they do not address to the protein dynamics.

The problem of multiple solutions in EXAFS analysis is particularly evident in a system like the Fe site of MbCO, where the MS contributions due to the heme plane overwhelm the weaker contributions due to axial ligands; moreover, the strong correlation between structural and dynamical parameters requires the a priori knowledge of one set in order to determine the other. It all makes the heme an ideal test for the realization of techniques aimed to improve the capabilities of XAFS analysis. For such reasons we tested an original approach for the calculation of DW factors in protein metal sites, in collaboration with the research group led by J.J. Rehr, which provided us with a program able to extract the DWs from the dynamical matrix for a given structure [19]. The DFT optimized structure of MbCO and the corresponding dynamical matrix were therefore calculated by means of Quantum Espresso (QE), the DWs were extracted, and the EXAFS spectrum generated by setting the calculated parameters was compared with the experimental one. This study allowed to extend the validity of the approach proposed by Vila et al [19] also to metalloproteins and to put emphasis on the role of Multiple Scattering contributions.

5.1 Structure and function of Mb

Myoglobin is a single-chain globular protein, expressed in cardiac myocites and oxidative skeletal muscle fibers, that reversibly binds O_2 or other small ligands by its heme residue. Its main function is to store oxygen in muscles, nevertheless it also serves as buffer for the intracellular concentration of O_2 and facilitates O_2 diffusion; for a complete review of Mb functions see reference [140]. Mb is made up of 153 amino acids arranged in 8 α-helices, that fold around a central core where the heme prosthetic group is embedded, as shown in Figure 5.1; two His residues are in contact with the heme site: His93, called the proximal His, that is a direct donor for the Fe atom, and His64, called the distal His, the position of which is responsible for the binding/release mechanism of small molecules at the Fe site [141]. The small molecule bound to the heme Fe^{2+} can be O_2, when
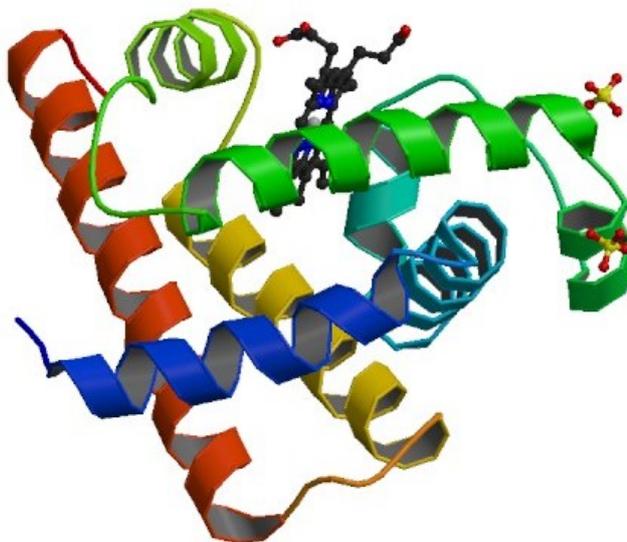
Figure 5.1: Backbone and prosthetic group of Mb. The backbone consists of 8 α-helices that fold around a central pocket containing the heme group.

the protein performs its standard physiological role, but also CO (Carboxymyo-globin), NO, water or hydroxyde. The choice of the molecule that binds to the Fe site is driven both by metal-ligand affinity and by the conformation of the ligand binding pocket delimited by His64.

Since its first crystal structure was solved, which was achieved by John Kendrew in 1958 as the first structure of a macromolecule, several XRD structures have become available, with increasing resolution up to 1.1 Å (see as examples the PDB entries 1MBO [142], 1BZR [143] and 1A6G [144]).

5.2 Heme as a model system

Heme is the functional site of Mb, since the binding and release of molecules takes place at its Fe^{2+} site. The iron atom binds permanently 4 N from the pyrrolic rings and a fifth N belonging to His93, while the sixth bond is available for external molecules (see figure 5.2). When the Fe coordination is complete, the atom lies on the heme plane, while if the sixth ligand is missing (deoxy-Mb) it undergoes a displacement from the plane of ~0.3-0.4 Å [144]. The ligand-binding mechanism of Mb has been studied in detail, in particular by following the photolysis of MbCO, the most stable and most easily photolyzed adduct [145]; a 150 ps time-resolved XRD study brought insight into photodissociation, by correlating the

binding/release mechanism to a set of molecular motions that take place within the protein [146]. Although the structure-to-function relationship is in this case
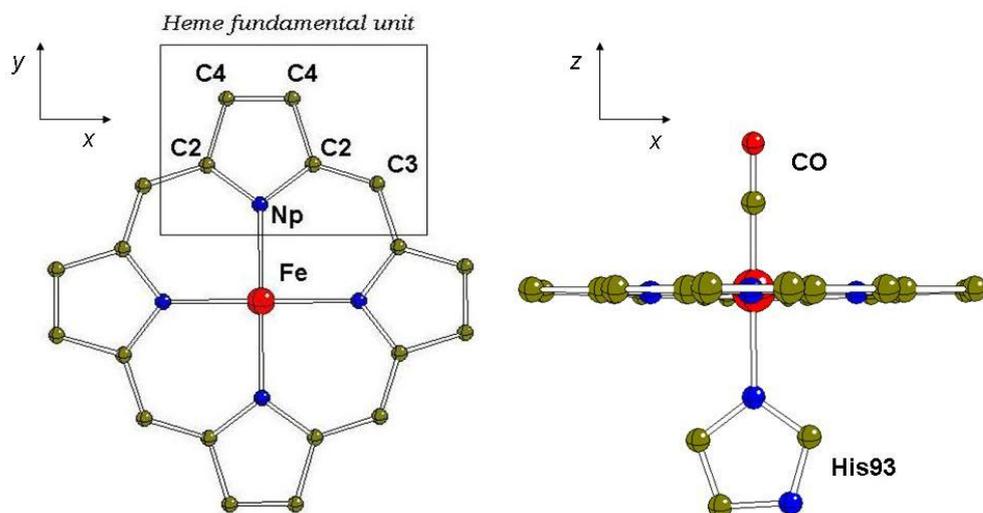

Figure 5.2: Graphical representation of the heme plane. For the atoms of the heme fundamental unit, the chosen nomenclature is reported on the picture. Np indicates the pyrrolic nitrogens (first shell), while to each carbon atom a number is associated to indicate the shell it belongs to (2-3-4).

much clearer than for the great majority of proteins, Mb and its Fe^{2+} site are still intensively studied with a wide variety of techniques, either to clarify the debated points (related for example to Mb dynamics [147, 148, 149]) or as a model system, for example to investigate protein-matrix interactions (reviewed in [133]) or to test original experimental setups or data analysis softwares [6] on macromolecules.

The series of XAFS studies on the Fe^{2+} site of Mb started in the early 80s with the pioneering works by Chance, Powers et al. [150, 151, 152], centered on the investigation of MbCO, its photoproduct Mb*CO and deoxy-Mb, and provided high resolution structures for the Fe coordination shell geometry in the three systems. Afterwards, many studies confirmed these early findings, except for the CO geometry, that was found to be closer to perpendicularity to the heme plane than suggested by Chance et al., both by XAFS [153] and by diffraction studies at atomic resolution [144].

XAFS was also applied to the refinement [154] of the heme structure in met-Mb (where the iron atom is oxidized to Fe^{3+}) and deoxy-Mb using XRD structures as starting point; based on constrained and restrained multiple scattering analysis, this study brought to the determination of the heme internal bond lengths

with an error of 0.02-0.03 Å, and served as a case study for the applicability of XAFS to the refinement of protein metal sites. Concerning MbCO, such a refinement of the heme internal bond length was never reported, however some XAFS studies dealt with the determination of Fe- first ligands bond lengths [155, 153]. One noteworthy result is that a limitation of EXAFS analysis was found, consisting in the impossibility to uniquely position the CO molecule on the basis of EXAFS data alone, because of the presence of a consistent number of local minima [155].

In the next sections, we will test the applicability of *ab initio* Density Functional Theory simulations to the determination of geometry and dynamical parameters in the Fe-heme of MbCO; the atoms of the heme plane will be referred to by using the labels depicted in Figure 5.2, where the numbers associated to C atoms refer to the shell they occupy with respect to the central Fe atom. Atoms belonging to the axial ligand, instead, will be indicated with the subscripts His and CO respectively.

5.3 Comparison between DFT calculations and experimental EXAFS spectra

As reported in previous paragraphs, the Fe-heme of MbCO appears as a widely studied but nevertheless challenging system; we decided to refer to it in order to test the efficacy of DFT calculations of statical and dynamical parameters relative to protein metal sites.

5.3.1 Materials and methods

Experimental. Horse-heart Mb was purchased by Sigma (Sigma, St.Louis, MO) and used without further purification. It has been dissolved ($\sim$10 mM) in a solution containing trehalose 200 mM (Hayashibara Shoij, Okayama, Japan) and a phosphate buffer 20 mM (pH 7 in H_2O). The solution has been equilibrated with CO and reduced by anaerobic addition of sodium-dithionite.

Fe K-edge XAFS spectra of MbCO have been collected on the beamline GILDA-BM08 of ESRF, employing a Si(111) monochromator; the temperature was set to 80 K and maintained by N flow. The flux on the sample was of the order of 10^{11} ph/s. Five spectra were collected in fluorescence mode, each one with an integration time of 10 s/point, changing the irradiatiated spot after two spectra. All of the spectra showed the same XANES, independently on the point

of the sample and of the order in which they were recorded; it suggests that no radiation damage occurred.

***Ab initio* Simulations.**   The structural relaxation towards the equilibrium structure was performed through a DFT approach (for a complete review see [156]) in the Born-Oppenheimer approximation, by means of the QuantumESPRESSO (QE) code. The code works within the plane wave approach, therefore the system had to be included in a super-cell of an hypothetic super-lattice, in order to simulate periodicity. At the same time, interaction between neighboring copies of the system must be avoided, therefore we chose a super-cell of $20 \times 20 \times 15$ Å^3. In the whole we took into account 48 atoms, consisting in the 25 atoms of the heme plane, the 7 atoms belonging to the axial ligands (depicted in Figure 5.2), and 16 H atoms necessary to saturate the bonds of the heme outer C shell and of the His residue. Ultrasoft pseudopotential with a Perdew-Burke-Ernzerhof exchange-correlation term were used.

Once the equilibrium structure was reached, the dynamical matrix matrix was calculated within the perturbation theory. The elements $D_{\alpha i}^{\beta j}$ of the dynamical matrix represent the coupling force between the displacement of atom α in direction i and atom β in direction j. For our purposes the elements $D_{\alpha i}^{\beta j}$ were calculated at the Gamma point, i.e. where the wavevector $\vec{q}=0$. The consistent part of the job described so far has been developed by Cristian degli Esposti Boschi; calculations were performed on the computing resources supplied by the Consortium of Universities CINECA.

Once the eigenvalues and eigenfunctions of the normal modes had been extracted, DW factors for the paths of interest were calculated by making use of a program [19] developed by H.J.Krappe (Hahn-Meitner-Institut, Berlin) and the research group led by J.J. Rehr (University of Washington, Seattle).

EXAFS analysis.   Theoretical amplitudes and phase shifts for XAFS analysis were calculated with feff8.2, the self-consistent mode was chosen to compute the potential and all paths with a maximum of 4 legs were taken into account. More than 200 scattering paths were generated, therefore a step-by-step procedure was necessary to identify the contributions to the final spectrum due to positions and DWs of the different shells of atoms; the overall analysis strategy, carried out with Artemis, follows these steps:

1. From a first shell fit in the R region 1-2 Å the amplitude reduction factor S_0^2 is extracted. The best fit value is then kept fixed during the following

steps.

2. The overall comparison in k space is performed, including all shells, initially setting only the Single Scattering (SS) DWs to the calculated values and fitting a common value σ^2_{MS} for all of the Multiple Scattering (MS) paths.

3. The MS paths with greatest amplitude are identified, their DW factors are calculated and fixed in the k space comparison; a common DW factor for the remaining MS paths is fitted.

In the steps above, a shift in the energy origin is always allowed (ΔE_0), while neither the interatomic distances nor the calculated DWs are allowed to vary.

5.3.2 Results

The analysis procedure outlined in the previous section is aimed to check the applicability of our DFT-based technique to the determination of both structural and dynamical parameters in MbCO; this step-by-step approach was adopted in order to disentangle the contributions to the final spectrum due to structural parameters, single scattering DWs and multiple scattering DWs. In step 1, the first shell fit, first shell distances and DWs were set to the calculated values reported in Table 5.1. In the table we report also the DWs extracted from Dimakis and Bunker's parametrization for Fe^{2+}-porphyrine complexes at the same temperature (80 K); we notice that the values calculated by us are systematically lower of about 10%. The resulting fit is reported in Figure 5.3 (red trace), the corresponding best fit values for the two free parameters are S_0^2=0.75±0.16 and ΔE_0=-6.8±2.4 eV; the other fit in the figure is obtained by allowing the Fe-C distance to vary, starting from a different position than the QE-calculated one (the reason for this test will be explained below).

The best fit value for S_0^2 was then fixed and the fit described in step 2 was performed, yielding the value of σ^2_{MS}=0.0034±0.0007 Å^2 for the DW factor common to all MS paths, and ΔE_0=-8.9±0.9. The main oscillation frequency of the spectrum is very well reproduced by the fit, which suggests that the calculated structural parameters are consistent with the data. On the other hand, the amplitudes of higher frequency oscillations (around 7 Å^{-1} for example) are not well reproduced: one might ascribe this discrepancy to the strong approximation assumed in step 2, that is, assigning a common DW factor to all of the MS paths. In such case the discrepancy might be reduced when step 3 is performed, since the calculated DWs are expected to influence in a different way the different MS

	Fe-Np	Fe-N$_{His}$	Fe-C$_{CO}$	Fe-C2	Fe-C3	Fe-C4
Distance (Å)	2.012 (1)	2.1011	1.7345	3.049 (1)	3.428 (1)	4.218 (1)
σ^2 (10^{-3}Å^2)	2.409 (4)	2.9464	1.9769	2.842 (8)	2.967 (6)	2.88 (1)
σ^2_{DB}(10^{-3}Å^2)	2.684	-	-	3.150	3.206	3.211

Table 5.1: Fe-atoms distances in MbCO resulting from structural optimization of the Fe-heme cluster, and DW factors (σ^2) at the temperature of 80 K extracted from the dynamical matrix. Calculations were performed with Quantum Espresso (QE). For the atoms of the heme plane, the average distances of all atoms belonging to the same shell are reported (atoms labels are depicted in Figure 5.2); values in brackets are the standard deviation relative to the last significant digit. The DWs are compared with the values resulting from Dimakis and Bunker's parametrization [20] for an Fe^{2+}-Porphyrin complex at 80 K (σ^2_{DB}).

contributions. However, once the greatest amplitude MS paths are identified and their DW factors are fixed to the values extracted from the dynamical matrix (see Table 5.2), the fit does not improve. The fit resulting from step 3 of the analysis

n	Path	Amplitude	Degeneracy	σ^2 (10^{-3}Å^2)
1	Fe-C$_{CO}$-O$_{CO}$-C$_{CO}$	170	1	1.9654
2	Fe-C$_{CO}$-O$_{CO}$	145	2	1.9654
3	Fe-Np-C2	324	16	2.6848
4	Fe-C4-Np	236	16	2.8277
5	Fe-C2-C4-Np	232	16	2.8244
6	Fe-Np-C2-Np	180	8	3.0290
7	Fe-C4-C2-Np	176	16	3.1854
8	Fe-C4-C2	176	16	2.9614
9	Fe-Np-C4-Np	166	8	2.8701
10	Fe-C2-C3	90	16	2.7663

Table 5.2: Most important MS paths generated with feff8.2. Path 1 and 2 refer to the CO molecule while the following ones to the heme plane. The DWs at the temperature of 80 K are extracted from the dynamical matrices calculated with QE and their values are reported in the last column of the table.

procedure, almost indistinguishable from the one resulting from step 2, is shown in Figure 5.4 (red trace); the corresponding best fit values are ΔE_0=-8.9±0.9 and σ^2_{MS}=0.006±0.004 Å^2.

We attribute the origin of the discrepancy at 7 Å^{-1} to the contribution of the CO molecule to the overall spectrum: as shown in Figure 5.4, when the CO paths are generated starting from the QE-calculated structure and DWs, their sum (indicated with CO-QE) presents high amplitude oscillations that are exactly in phase with the overall simulated EXAFS in the region 6.5-8 Å^{-1}, and are therefore likely to be responsible for the too high oscillation amplitude shown by
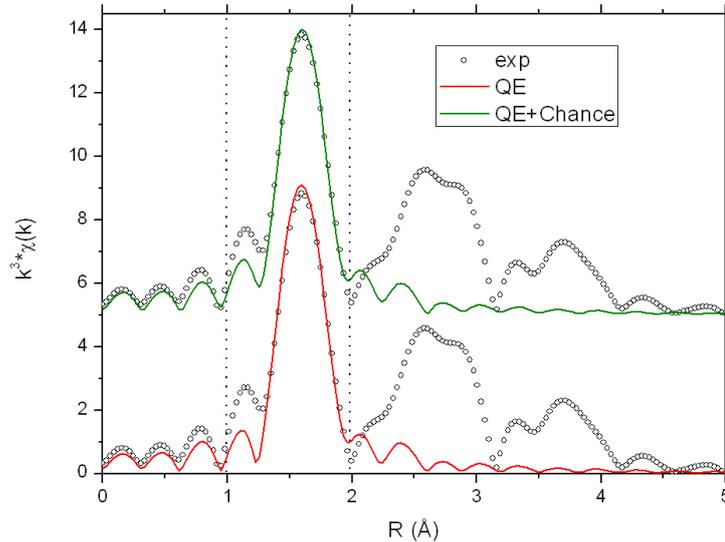
Figure 5.3: First shell fits of the XAFS spectra of MbCO measured at 80 K. One fit (red trace) is based on the first shell distances and DWs calculated with QE, the only free parameters are S_0^2 and ΔE_0. The other one (green trace) differs only in the Fe-C distance, the starting value of which is taken from two early XAFS works by Chance et al. [150, 151] and which is then allowed to vary during the fit.

the simulated spectrum with respect to the experimental one. However, besides this deviation, the overall agreement between fit and experiment is very good, in particular if we recall that no structural or dynamical degrees of freedom have been fitted so far, but the only free parameters were a shift in the empirically determined energy origin and the amplitude reduction factor. We chose therefore to allow also the CO parameters (see Figure 5.5) to vary, starting from different initial configurations.

Two hybrid structures were built, where the structural and dynamical parameters for the heme plane and for the proximal His (His93) were set to the QE results, while two different CO structures were adopted: one was taken from the PDB structure 1A6G [144], the other from two early XAFS work performed by Chance et al. [150, 151]. The structural parameters for the CO molecule in the two models are reported in Table 5.3; initial values for the DWs relative to C and O were set to the reasonable value of 0.003 Å^2, while the ones relative to the MS paths were set to 0.005 Å^2.

Figure 5.4 shows that the sum of the CO contributions changes dramatically depending on the model adopted; therefore we applied a slightly different version
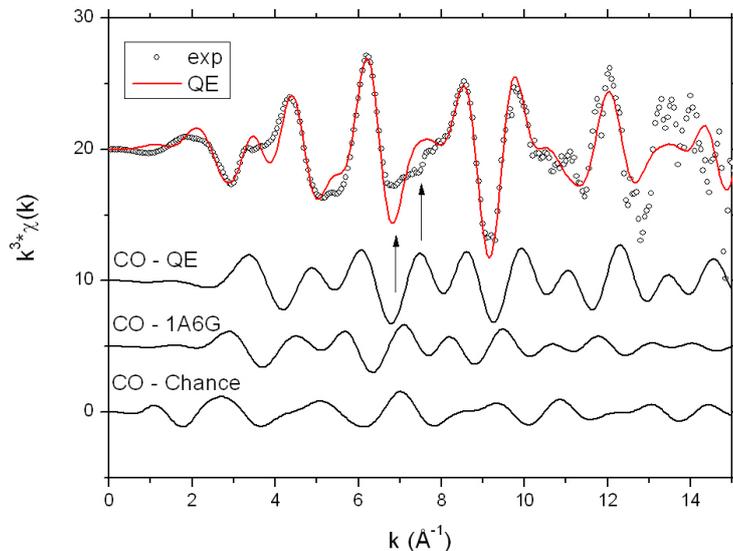
Figure 5.4: Comparison between the spectrum originated by the structure and DWs calculated with QE (red trace) and eperimental data (open circles). The three black curves below represent the sum of the paths originated by the CO atoms, from top to bottom: 1) when CO structure and DWs are calculated with QE; 2) when CO structure is taken from an XRD structure (PDB code 1A6G [144]); 3) when CO structure is taken from early XAFS works [150, 151].

of the analysis procedure, starting with the two hybrid models in turn. The only differences with respect to the procedure described in Paragraph 5.3.1 are that some free parameters are added in the three steps: a shift in the C atom position in step 1, variations in both Fe-C and Fe-O distances in step 2 and the single scattering DWs relative to C and O (σ_C^2 and σ_O^2) in step 3. Once the Fe-C and Fe-O bond length are determined, the angle α can be calculated by means of the Carnot theorem:

$$cos\alpha = \frac{(Fe - C)^2 + (C - O)^2 - (Fe - O)^2}{2(Fe - C)(Fe - O)} \tag{5.1}$$

where the C-O bond length is fixed to 1.167 Å, as found by our DFT calculation in agreement with data reported in literature [157].

The aim of this data treatment is to check that the discrepancy encountered when the QE model is employed is due only to a wrong estimation of CO position and DWs, and whether a good agreement can be achieved by allowing the CO parameters to vary keeping all the rest fixed to the calculated values. The QE+1A6G model led to the same discrepancy encountered for the pure QE
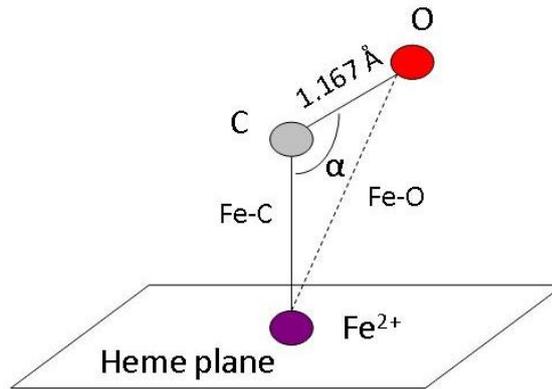
Figure 5.5: CO molecule geometry and degrees of freedom. The CO molecule, in the Rigid Body Refinement scheme is described by two degrees of freedom. We chose to refine independently the Fe-C and Fe-O bond lengths; the C-O bond length is fixed to the value defined by QE, so the angle α is determined from the two free variables by means of the Carnot theorem.

	Fe-C (Å)	α (°)
1A6G	1.82 (2)	171 (3)
Chance et al.	1.93 (2)	127 (4)
QE	1.734	179.9
This work	1.78 (2)	135.5 (1)

Table 5.3: Values of the structural parameters relative to the CO molecule in MbCO determined by X-ray crystallography (1A6G [144]), early XAFS works [150, 151], our DFT calculation performed with QE and the fit of the EXAFS spectrum based on the QE+Chance model (see text). α is the $Fe - \hat{C} - O$ angle; the values in bracket are the errors on the last significant digit.

model, while the QE+Chance model nicely fits the experimental data (Figure 5.6, blue trace). Since the CO position determined by Chance et al. differs from the other models mainly because of its marked bending, it all suggests that the CO molecule is actually bent, such that it is impossible to properly fit the spectrum without taking it into account in the starting model.

The first-shell fit based on the QE+Chance starting model is reported in Figure 5.3; the Fe-C distance was allowed to vary in this step, starting from the value of 1.93 Å indicated by Chance et al. [150, 151] and converged to the value of 1.78 Å, much closer to the one calculated with QE (see Table 5.3). The other best fit parameters were $S_0^2=0.82\pm0.08$ and $\Delta E_0=-7\pm1$. The value of S_0^2 is now higher than the one determined before, when the fitting model was based only on QE calculations: this could be due to a correlation between this parameter and the DW factor relative to the C atom. The latter was in this case fixed to the

reasonable value of 0.003 $Å^2$, since we are not allowed to use the value extracted with QE, considering that we introduced a different CO structure.

The resulting CO geometry, after the whole analysis procedure based on the QE+Chance model, is reported in table 5.3, and the corresponding fit in Figure 5.6 (blue trace): it is noteworthy that the Fe-C distance that best fits the data is much lower that the starting value of 1.93 Å, while the bending angle α is close to the one determined by Chance et al. [150, 151] than to the QE-calculated one. The reason is ascribed to the fact that in QE simulations the heme was treated as an isolated system, while its actual environment is the protein backbone; this approximation seems reasonable for the whole heme structure and for the proximal His, while it does not hold for the CO molecule, which is the most affected by the surroundings, as a consequence of its interaction with the distal His (His64) [148]. This interpretation clarifies also the fact that the DWs relative to the CO seemed underestimated by DFT calculations, if we take into account that the Fe-C bond is actually much weaker than what one would expect for an isolated system. The calculated CO position, moreover, gives rise to collinear MS paths, that are the most effective paths since their backscattering amplitude, present in the EXAFS formula (1.12), is maximized ; this could explain the considerable influence of the CO paths on the overall spectrum that produced the discrepancy at 7 $Å^{-1}$, and the improvement achieved when the QE+Chance model, with a marked CO bending, is employed.

The third step of the analysis provides best fit values for σ_C^2 and σ_O^2 that are within the interval of reasonable values, but the common σ_{MS}^2 that minimizes the R factor is the unreasonable value of 0.3 $Å^2$; this is clearly due to the low influence of this parameter on the fit, such that the analysis program tends to cancel its contribution. Therefore, a fit where only the 8 MS paths relative to the heme atoms reported in Table 5.2 were included; for the included paths, the DWs were fixed to the values extracted from the dynamical matrix, except for the SS paths σ_C^2 and σ_O^2 that were allowed to vary, together with ΔE_0. Since the remaining MS paths were ignored, no additional degrees of freedom are needed. The fit results are reported in Table 5.4, and the graphical output in Figure 5.6 (red trace).

Surprisingly, the resulting fits are almost identical independently on the number of MS paths taken into account. It is remarkable that the MS paths relative to CO, that caused the mismatch at 7 $Å^{-1}$ in the first phase of the analysis, turn out to be unnecessary to fit the de data; it suggests that their contributions are negligible with respect to the heme ones, eight of which are enough to describe

σ_C^2	σ_O^2	ΔE_0
(10^{-3}Å^2)	(10^{-3}Å^2)	(eV)
12 (7)	0 (2)	-9.3 (0.6)

Table 5.4: Best fit values obtained by fitting the MbCO EXAFS spectrum with the model QE+Chance and including only the 8 MS paths involving the heme atoms and with highest amplitude.

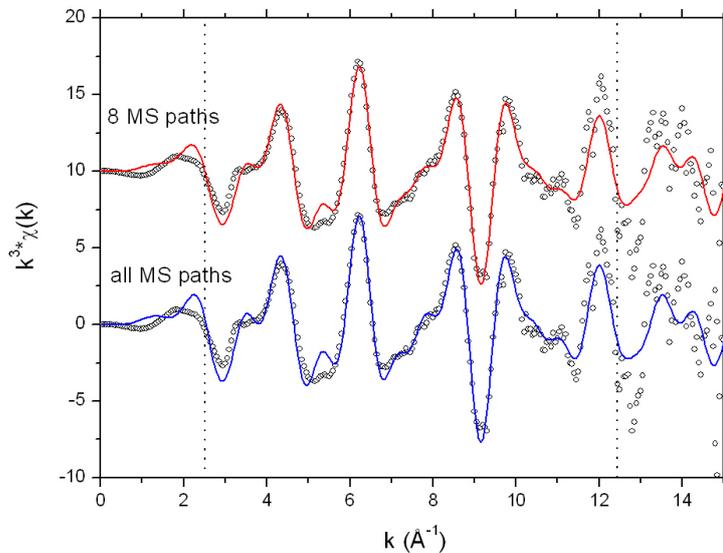the X-ray absorption spectrum of the system with good approximation.

Figure 5.6: Fit to the experimental MbCO spectrum measured at 80 K (empty circles) with models based on QE calculations for His93 and for the atoms of the heme plane, and on the works by Chance at al [150, 151] for the CO molecule. The fit can be performed by taking into account all of the MS paths (bottom, blue trace), or only the most representative 8 MS paths (top, red trace).

5.3.3 Discussion

The structural and dynamical parameters calculated with QE for the Fe cluster in MbCO generate an EXAFS spectrum in very good agreement with the data, except for the CO molecule, whose calculated geometry and DW factors cause the spectrum to differ from the experimental one. The agreement, that holds for the great majority of the atoms involved in the cluster, is remarkable, since our DFT simulation did not take into account the heme environment, i.e. the protein backbone and the solvent. This means that the heme structural and dynamical properties are completely determined by its internal bonds, and that

interactions with the surroundings do not play any significant role. Moreover, the good agreement is achieved also when only almost collinear MS paths relative to the heme plane are taken into account (we included 8 of them in the fit), because the contributions due to axial ligands become negligible if compared to them; this stresses the importance of providing a reliable high-resolution starting model for the heme plane to be employed in XAFS studies on the Fe-cluster of MbCO, otherwise the determination of CO and His93 structure and DWs is prevented by the predominance of the heme MS contributions in the spectrum.

For each atom of the cluster, the determination of DWs is as crucial as the determination of the equilibrium structure, because of the strong correlation between the two variables in EXAFS analysis. This study revealed that DFT calculations are a valuable tool to provide equilibrium structure and DW factors, even in a system the spectrum of which is dominated by Multiple Scattering, as is the heme. The method used to calculate the DWs, based on the DFT calculation of the dynamical matrix from which the projected phonon spectra is extracted, has never been tested before on a protein system and its effectiveness was fully demonstrated only for Single Scattering paths [19]; here we have successfully tested the method for a paradigmatic protein system with major influence of MS.

The only discrepancy between calculations and experiment resides in the determination of structure and DWs relative to the CO molecule. Such discrepancy might be reasonable, since the CO molecule is the one with the highest mobility into the protein, and its binding is not so tight as for the heme atoms; moreover, its interaction with the distal histidine His64, not included in the DFT-simulated cluster, is ascertained. Considering this, we reckon that the approximation made when considering the heme as an isolated system is too strong for the CO molecule, to the contrary of what we found for the rest of the cluster. Nevertheless, the problem can be unraveled by fitting the CO parameters once the rest of the structure is fixed.

This combined approach between EXAFS fitting and DFT calculations brings advantages with respect to the independent use of the two techniques: XAFS analysis is facilitated because the problems of multiple solutions and correlation between parameters are partially solved, while DFT calculations do not need to include a too high number of atoms, because corrections due to interactions with the surroundings can be introduced by means of the EXAFS fit.

Conclusions

The work we reported was meant to improve the capabilities of BioXAS analysis and to employ it as a reliable tool to investigate key topics in structural biology. Therefore we presented first of all a systematic method of analysis obtained by gathering tools issued by the expertise of several research groups and by matching them in the light of systematic experimental observations and *ab initio* simulations; it all contributed to give the guidelines of an aware XAFS analysis procedure, that might be useful for non-expert users in alternative to "black box" approaches. The procedure was fully developed and tested for mononuclear Zn sites; in this case, the dynamical parameters relative to Zn-amino acid bonds are available in literature, while for other metals they are not. Therefore, a possible step towards the generalization of the analysis procedure consists in the *ab initio* calculation of such parameters. In this regard, we adopted a way to extract them from the dynamical matrix calculated with Density Functional Theory and successfully tested it on Fe-MbCO.

Thanks to this data analysis procedure, or to adaptations of it to the specific subject under study, we faced diverse topics of biophysical relevance. One of those is the competition between metal and proton binding which was proposed as the basic mechanism for the inhibition of proton-translocation activity by metal ions; the study reported here, regarding Transhydrogenase, belongs to a wider series of studies on energy-transducing proteins, that reveal the constant presence of His residues in the coordination sphere of the inhibitory metal, supporting the proposed mechanism of inhibition.

A different but as well current topic in biophysics is the effect exerted on proteins by strongly dehydrate sugar matrices. We reported here of a XAFS study on the Fe^{2+} site of photosynthetic Reaction Center embedded in different matrices, which, besides allowing to improve structural resolution of the Fe-ligands distances, demonstrated how XAFS can be employed to study the short-scale repercussions into the deep interior of the protein of an interaction taking place at the protein surface. A further step on the study of the Fe^{2+} site of RC was

also attempted, consisting in a time-resolved pump-probe experiment; the results reported in this work argue for the presence of a conformational change in the Fe^{2+} coordination sphere in the 500 μs time scale.

List of Publications

Below we list the publications yielded from this work:

- Veronesi, G.; Giachini, L.; Francia, F.; Mallardi, A.; Palazzo, G.; Boscherini, F.; Venturoli, G. *The Fe^{2+} Site of Photosynthetic Reaction Centers Probed by Multiple Scattering X-ray Absorption Fine Structure Spectroscopy: Improving Structure Resolution in Dry Matrices* Biophys. J. (2008), 95, 814-822.

- Giachini, L.; Veronesi, G.; Francia, F.; Venturoli, G.; Boscherini, F. *Synergic approach to XAFS analysis for the identification of most probable binding motifs for mononuclear zinc sites in metalloproteins* J. Synchrotron Rad. (2010), 17, 41-52.

- Veronesi, G.; Whitehead, S. J.; Francia, F.; Giachini, L.; Boscherini, F.; Venturoli, G.; Cotton, N. P.; Jackson, J. B. *X-ray absorption studies of Zn^{2+}-binding sites in Escherichia coli transhydrogenase and its $\beta H91K$ mutant*, Biochimica et Biophysica Acta (BBA) - Bioenergetics (2010), 1797, 494 - 500.

Acknowledgments

I would like to thank Prof. Federico Boscherini and Prof. Giovanni Venturoli, who made this work possible. Special thanks must be addressed to Lisa Giachini, who patiently taught me how to move the first steps into this scientific field, and to Francesco Francia for his excellent work and wise precepts.

For all the rest I would like to express my gratefulness to my family, to Mauro, and to my friends.

Bibliography

[1] Rehr, J. J. and Albers, R. C. *Rev. Mod. Phys.* **72**(3), 621–654 Jul (2000). 4, 7, 11, 51, 69

[2] Ankudinov, A. L., Ravel, B., Stern, J. J., and Conradson, S. D. *Phys. Rev. B* **58**, 7565–7576 (1998). feff8.2. 6, 8, 9, 10

[3] Filipponi, A. and Di Cicco, A. *Phys. Rev. B* **52**(21), 15135–15149 Dec (1995). 6

[4] Binsted, N., Strange, R. W., and Hasnain, S. S. *Biochemistry* **31(48)**, 12117–25 Dec (1992). 6, 28, 51, 71

[5] Joly, Y. *Phys. Rev. B* **63**(12), 125120 Mar (2001). 6, 10, 11, 22, 25, 85

[6] Benfatto, M., Della Longa, S., and Natoli, C. R. *J. Synchrotron Rad.* **10**(1), 51–57 Jan (2003). 6, 94

[7] Smolentsev, G. and Soldatov, A. *Computational Materials Science* **39**(3), 569 – 574 (2007). 6, 22

[8] Durham, P. J. *Principles, Applications and Techniques of EXAFS, SEXAFS and XANES*. Wiley, New York, (1989). 7

[9] Filipponi, A., Di Cicco, A., and Natoli, C. R. *Phys. Rev. B* **52**(21), 15122–15134 Dec (1995). 7, 8

[10] Lee, P. A. and Pendry, J. B. *Phys. Rev. B* **11**(8), 2795–2811 Apr (1975). 8, 14, 15

[11] Benfatto, M., Natoli, C. R., Bianconi, A., Garcia, J., Marcelli, A., Fanfoni, M., and Davoli, I. *Phys. Rev. B* **34**(8), 5774–5781 Oct (1986). 8

[12] Mijovilovich, A. and Meyer-Klaucke, W. *J. Synchrotron Rad.* **10**(1), 64–68 Jan (2003). 10, 22, 24

[13] Giachini, L., Francia, F., Veronesi, G., Lee, D.-W., Daldal, F., Huang, L.-S., Berry, E. A., Cocco, T., Papa, S., Boscherini, F., and Venturoli, G. *Biophys. J.* **93**(8), 2934–2951 October (2007). 10, 24, 37, 43, 50, 60

[14] Joly, Y. *J. Synchrotron Rad.* **10**(1), 58–63 Jan (2003). 10

[15] Rehr, J. J. and Albers, R. C. *Phys. Rev. B* **41**(12), 8139–8149 Apr (1990). 12

[16] Poiarkova, A. V. and Rehr, J. J. *Phys. Rev. B* **59**(2), 948–957 Jan (1999). 13

[17] Dimakis, N. and Bunker, G. *Phys. Rev. B* **65**(20), 201103 May (2002). 13, 35, 70, 91

[18] Dimakis, N. and Bunker, G. *Phys. Rev. B* **70**(19), 195114 Nov (2004). 13, 32, 35, 51, 70, 91

[19] Vila, F. D., Rehr, J. J., Rossner, H. H., and Krappe, H. J. *Phys. Rev. B* **76**(1), 014301 (2007). 13, 92, 96, 104

[20] Dimakis, N. and Bunker, G. *AIP Conference Proceedings* **882**(1), 334–336 (2007). 13, 35, 54, 70, 91, 98

[21] Yano, J., Kern, J., Irrgang, K. D., Latimer, M. J., Bergmann, U., Glatzel, P., Pushkar, Y., Biesiadka, J., Loll, B., Sauer, K., Messinger, J., Zouni, A., and Yachandra, V. K. *Proc Natl Acad Sci U S A.* **102(34)**, 12047–52 Aug (2005). 14

[22] Matsushita, T. and Hashizume, H. *Handbook of Synchrotron Radiation*, volume 1, chapter 4: X-ray monochromators, 261. North-Holland (1983). 17

[23] D'Acapito, F., Colonna, S., Pascarelli, S., Antonioli, G., Balerna, A., Bazzini, A., Boscherini, F., Campolungo, F., Chini, G., Dalba, G., Davoli, G. I., Fornasini, P., Graziola, R., Licheri, G andMeneghini, C., Rocca, F., Sangiorgio, L., Sciarra, V., Tullio, V., and Mobilio, S. *ESRF Newsletters* **30**, 42 Apr (1998). 18, 51, 69

[24] Pascarelli, S., Boscherini, F., DAcapito, F., Hardy, J., Meneghini, C., and Mobilio, S. *J. Synchrotron Rad.* **3**, 147–155 (1996). 18

[25] Ciatto, G., D'Acapito, F., Boscherini, F., and Mobilio, S. *J. Synchrotron Rad.* **11**, 278–283 (2004). Ge detector. 18

[26] Arcovito, A., Benfatto, M., Cianci, M., Hasnain, S. S., Nienhaus, K., Nienhaus, G. U., Savino, C., Strange, R. W., Vallone, B., and Della Longa, S. *Proceedings of the National Academy of Sciences* **104**(15), 6211–6216 (2007). 19

[27] Holton, J. M. *Journal of Synchrotron Radiation* **14**(1), 51–72 Jan (2007). 19

[28] Strange, R. W. and Feiters, M. C. *Curr. Opin. Struct. Biol.* **18**(5), 609–616 October (2008). 19, 92

[29] Penner-Hahn, J. E. *Nat Struct Mol Biol* **10**(2), 75–77 February (2003). 20

[30] Haumann, M., Pospíšil, P., Grabolle, M., Müller, C., Liebisch, P., Solé, V. A., Neisius, T., Dittmer, J., Iuzzolino, L., and Dau, H. *Journal of Synchrotron Radiation* **9**(5), 304–308 Sep (2002). 20

[31] Dau, H. and Haumann, M. *Photosynthesis Research* **92**(3), 327–343 June (2007). 20

[32] Kleifeld, O., Frenkel, A., Martin, J. M., and Sagi, I. *Nat Struct Mol Biol* **10**(2), 98–103 February (2003). 20

[33] Wellenreuther, G., Parthasarathy, V., and Meyer-Klaucke, W. *J. Synchrotron Rad.* **17**(1) Jan (2010). 21

[34] Giachini, L., Veronesi, G., Francia, F., Venturoli, G., and Boscherini, F. *J. Synchrotron Rad.* **17**(1), 41–52 Jan (2010). 22, 40

[35] Benfatto, M. and Della Longa, S. *J. Synchrotron Rad.* **8**(4), 1087–1094 Jul (2001). 22

[36] Kau, L. S., Spira-Solomon, D. J., Penner-Hahn, J. E., Hodgson, K. O., and Solomon, E. I. *J. Am. Chem. Soc.* **109**(21), 6433–6442 (1987). 22

[37] Rich, A. M., Ellis, P. J., Tennant, L., Wright, P. E., Armstrong, R. S., and Lay, P. A. *Biochemistry* **38**(50), 16491–16499 (1999). 22, 84

[38] Shearer, J., Scarrow, R. C., and Kovacs, J. A. *J. Am. Chem. Soc.* **124**(39), 11709–11717 October (2002). 22

[39] Dräger, G., Frahm, R., Materlik, G., and Brmmer, O. *physica status solidi (b)* **146**(1), 287–294 (1988). 22, 84

[40] Feiters, M. C., Eijkelenboom, A. P. A. M., Nolting, H.-F., Krebs, B., van den Ent, F. M. I., Plasterk, R. H. A., Kaptein, R., and Boelens, R. *J. Synchrotron Rad.* **10**(1), 86–95 Jan (2003). 22, 26

[41] Peariso, K., Huffman, D. L., Penner-Hahn, J. E., and O'Halloran, T. V. *J. Am. Chem. Soc.* **125**(2), 342–343 (2003). 22

[42] Dau, H., Liebisch, P., and Haumann, M. *Phys. Scr.* **2005**(T115), 844–846 (2005). 22

[43] Banci, L., Bertini, I., and Mangani, S. *J. Synchrotron Rad.* **12**(1), 94–97 Jan (2005). 22

[44] Castagnetto, J. M., Hennessy, S. W., Roberts, V. A., Getzoff, E. D., Tainer, J. A., and Pique, M. E. *Nucl. Acids Res.* **30**(1), 379–382 (2002). 23

[45] Hsin, K., Sheng, Y., Harding, M. M., Taylor, P., and Walkinshaw, M. D. *J. Appl. Crystallogr.* **41**(5), 963–968 Oct (2008). 23, 71, 75, 76

[46] Clark-Baldwin, K., Tierney, D. L., Govindaswamy, N., Gruff, E. S., Kim, C., Berg, J., Koch, S. A., and Penner-Hahn, J. E. *J. Am. Chem. Soc.* **120**(33), 8401–8409 August (1998). 24, 31

[47] Alberts, I. L., Nadassy, K., and Wodak, S. J. *Protein Sci.* **7**(8), 1700–1716 (1998). 25, 54

[48] Engh, R. and Huber, R. *Acta Cryst. A* **47**, 392–400 (1991). 25, 51, 71

[49] Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**(8), 1432–1443 Aug (1999). 25, 34

[50] Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**(7), 857–867 Jul (2000). 25, 34

[51] Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **57**, 401–411 (2001). 25, 34

[52] Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**(5), 872–874 May (2002). 25, 34

[53] Harding, M. M. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **62(Pt 6)**, 678–82 Jun (2006). 25, 34, 51, 71, 75, 86

[54] Binsted, N. Technical report, CCLRC Daresbury laboratory, (1998). 28

[55] Newville, M. *J. Synchrotron Rad.* **8**(2), 322–324 Mar (2001). 28, 51, 69

[56] Ravel, B. and Newville, M. *J Synchrotron Radiat.* **12(Pt 4)**, 537–41 Jul (2005). 28, 34, 51, 69

[57] Newville, M., Livins, P., Yacoby, Y., Rehr, J. J., and Stern, E. A. *Phys Rev B Condens Matter.* **47(21)**, 14126–14131 Jun (1993). 29

[58] Stern, E. A. *Phys. Rev. B* **48**(13), 9825–9827 Oct (1993). 30

[59] Kelly, S. D., Hesterberg, D., and Ravel, B. *Methods of Soil Analysis. Part 5: Mineralogical Methods*, chapter 14: Analysis of soils and minerals using X-ray Absospion Spectroscopy, 387–464. Soil Science Society of America, Inc., Madison, Wisconsin, USA (2008). 30

[60] Co, M. S., Scott, R. A., and Hodgson, K. O. *J. Am. Chem. Soc.* **103**(4), 986–988 February (1981). 32

[61] Strange, R. W., Blackburn, N. J., Knowles, P. F., and Hasnain, S. S. *J. Am. Chem. Soc.* **109**(23), 7157–7162 November (1987). 32

[62] Holmes, M. A. and Matthews, B. W. *J. Mol. Biol.* **160**(4), 623 – 639 (1982). 33

[63] Murphy, L. M., Strange, R. W., and Hasnain, S. *Structure* **5**(3), 371–379 March (1997). 33

[64] Ankudinov, A. L., Bouldin, C. E., Rehr, J. J., Sims, J., and Hung, H. *Phys. Rev. B* **65**(10), 104107 Feb (2002). 34

[65] Giachini, L., Francia, F., Cordone, L., Boscherini, F., and Venturoli, G. *Biophys J.* **92(4)**, 1350–60 Feb (2007). 35, 75, 77

[66] Stern, E. A., Newville, M., Ravel, B., Yacoby, Y., and Haskel, D. *Physica B* **208 & 209**, 117–120 (1995). 36

[67] Kelly, S. D., Kemner, K. M., Fryxell, G. E., Liu, J., Mattigod, S. V., and Ferris, K. F. *J. Phys. Chem. B.* **105(27)**, 6337–6346 (2001). 36

[68] Giachini, L., Francia, F., Mallardi, A., Palazzo, G., Carpen, E., Boscherini, F., and Venturoli, G. *Biophys.J.* **88**(3), 2038–2046 March (2005). 37, 43, 50, 60

[69] Veronesi, G., Whitehead, S. J., Francia, F., Giachini, L., Boscherini, F., Venturoli, G., Cotton, N. P., and Jackson, J. B. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1797**(4), 494 – 500 (2010). 37, 44

[70] Giachini, L., Francia, F., Boscherini, F., Pacelli, C., Cocco, T., Papa, S., and Venturoli, G. *FEBS Lett.* **581**(29), 5645 – 5648 (2007). 37

[71] Muramoto, K., Hirata, K., Shinzawa-Itoh, K., Yoko-o, S., Yamashita, E., Aoyama, H., Tsukihara, T., and Yoshikawa, S. *Proc. Natl. Acad. Sci. U S A.* **104**(19), 7881–7886 (2007). 37, 43

[72] Shinzawa-Itoh, K., Aoyama, H., Muramoto, K., Terada, H., Kurauchi, T., Tadehara, Y., Yamasaki, A., Sugimura, T., Kurono, S., Tsujimoto, K., Mizushima, T., Yamashita, E., Tsukihara, T., and Yoshikawa, S. *EMBO J.* **26**(6), 1713–1725 March (2007). 37

[73] Aoyama, H., Muramoto, K., Shinzawa-Itoh, K., Hirata, K., Yamashita, E., Tsukihara, T., Ogura, T., and Yoshikawa, S. *Proceedings of the National Academy of Sciences* **106**(7), 2165–2169 (2009). 37

[74] Matthews, B. W., Colman, P. M., Jansonius, J. N., Titani, K., Walsh, K. A., and Neurath, H. *Nature New Biol* **238**(80), 37–41 (1972). 37

[75] Meneghini, C. and Morante, S. *Biophys. J.* **75**(4), 1953–1963 October (1998). 37

[76] Tainer, J. A., Getzoff, E. D., Richardson, J. S., and Richardson, D. C. *Nature* **306**(5940), 284–287 November (1983). 37

[77] Hough, M. A. and Hasnain, S. *J. Mol. Biol.* **287**(3), 579 – 592 (1999). 37

[78] Errede, B., Kamen, M. D., and Hatefi, Y. In *Part D: Biomembranes*, Fleischer, S. and Packer, L., editors, volume Volume 53, 40–47. Academic Press (1978). 38

[79] Matsubara, H. In *Proteolytic Enzymes,* Perlmann, G. E. and Lorand, L., editors, volume 19 of *Methods in Enzymology*, 642 – 651. Academic Press (1970). 38, 50

[80] Link, T. A. and von Jagow, G. *J. Biol. Chem.* **270**(42), 25001–25006 (1995). 43

[81] Faxn, K., Salomonsson, L., delroth, P., and Brzezinski, P. *Biochim. Biophys. Acta - Bioenergetics* **1757**(5-6), 388 – 394 (2006). 14th European Bioenergetics Conference. 43

[82] Mills, D. A., Schmidt, B., Hiser, C., Westley, E., and Ferguson-Miller, S. *Journal of Biological Chemistry* **277**(17), 14894–14901 (2002). 43

[83] Paddock, M. L., Graige, M. S., Feher, G., and Okamura, M. Y. *Proc. Natl. Acad. Sci. U S A.* **96**(11), 6183–6188 (1999). 43

[84] Axelrod, H. L., Abresch, E. C., Paddock, M. L., Okamura, M. Y., and Feher, G. *Proc. Natl. Acad. Sci. U S A.* **97**(4), 1542–1547 (2000). 43, 60

[85] Paddock, M. L., Sagle, L., Tehrani, A., Beatty, J. T., Feher, G., and Okamura, M. Y. *Biochemistry* **42**(32), 9626–9632 August (2003). 43

[86] Klishin, S. S., Junge, W., and Mulkidjanian, A. Y. *Biochim. Biophys. Acta - Bioenergetics* **1553**(3), 177 – 182 (2002). 43

[87] Berry, E. A., Zhang, Z., Bellamy, H. D., and Huang, L. *Biochim. Biophys. Acta - Bioenergetics* **1459**(2-3), 440 – 448 (2000). 43

[88] Francia, F., Giachini, L., Boscherini, F., Venturoli, G., Capitanio, G., Martino, P. L., and Papa, S. *FEBS Lett.* **581**(4), 611 – 616 (2007). 43

[89] Whitehead, S. J., Rossington, K. E., Hafiz, A., Cotton, N. P., and Jackson, J. B. *FEBS Lett.* **579**(13), 2863 – 2867 (2005). 43, 46, 53, 59

[90] Whitehead, S. J., Iwaki, M., Cotton, N. P., Rich, P. R., and Jackson, J. B. *Biochim. Biophys. Acta - Bioenergetics* **1787**(10), 1276 – 1288 (2009). 43, 44, 46, 48, 50, 51, 53, 56, 59, 60

[91] Jackson, J. B. *FEBS Lett.* **545**(1), 18 – 24 (2003). 44, 45, 46

[92] Jackson, J. B., White, S. A., and Brondijk, T. H. C. In *Biophysical and Structural Aspects of Bioenergetics,* Wikstrom, M., editor, 376–393. Royal Society of Chemistry, Cambridge (2005). 44, 47

[93] Hoek, J. B. and Rydstrom, J. *Biochem. J.* **254**(1), 1–10 (1988). 44

117

[94] Cotton, N. P., White, S. A., Peake, S. J., McSweeney, S., and Jackson, J. *Structure* **9**(2), 165–176 February (2001). 45

[95] Mather, O. C., Singh, A., van Boxel, G. I., White, S. A., and Jackson, J. B. *Biochemistry* **43**(34), 10952–10964 (2004). 46

[96] Bizouarn, T., Fjellström, O., Meuller, J., Axelsson, M., Bergkvist, A., Johansson, C., Karlsson, B. G., and Rydström, J. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1457**(3), 211 – 228 (2000). 46

[97] Bizouarn, T., Meuller, J., Axelsson, M., and Rydström, J. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1459**(2-3), 284 – 290 (2000). 46

[98] Meuller, J. and Rydstrom, J. *J. Biol. Chem.* **274**(27), 19072–19080 (1999). 46

[99] Althage, M., Bizouarn, T., and Rydstrom, J. *Biochemistry* **40**(33), 9968–9976 (2001). 46

[100] Bragg, P. D. and Hou, C. *Arch. Biochem. Biophys.* **363**(1), 182 – 190 (1999). 46

[101] Holmberg, E., Olausson, T., Hultman, T., Rydstroem, J., Ahmad, S., Glavas, N. A., and Bragg, P. D. *Biochemistry* **33**(24), 7691–7700 (1994). 46, 47

[102] Glavas, N. A., Hou, C., and Bragg, P. D. *Biochemistry* **34**(23), 7694–7702 June (1995). 47

[103] Bragg, P. D. and Hou, C. *European Journal of Biochemistry* **241**(2), 611–618 (1996). 47

[104] Bragg, P. D. and Hou, C. *Archives of Biochemistry and Biophysics* **388**(2), 299 – 307 (2001). 47

[105] Yamaguchi, M., Stout, C. D., and Hatefi, Y. *Journal of Biological Chemistry* **277**(37), 33670–33675 (2002). 47

[106] Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M. D., Fujimoto, E. K., Goeke, N. M., Olson, B. J., and Klenk, D. C. *Analytical Biochemistry* **150**, 76–85 (1985). 50

[107] Ugliengo, P., Viterbo, D., and Chiari, G. *Z. Kristallogr.* **207**, 9 (1993). 51, 69

[108] Dokmanić, I., Šikić, M., and Tomić, S. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **64**(3), 257–263 Mar (2008). 54

[109] Veronesi, G., Giachini, L., Francia, F., Mallardi, A., Palazzo, G., Boscherini, F., and Venturoli, G. *Biophys. J.* **95**(2), 814–822 (2008). 63, 85

[110] Feher, G., Allen, J. P., Okamura, M. Y., and Rees, D. C. *Nature* **339**, 111–116 (1989). 64, 65

[111] Stowell, M. H. B., McPhillips, T. M., Rees, D. C., Soltis, S. M., Abresch, E., and Feher, G. *Science* **276**, 812–815 (1997). 64, 66, 67

[112] Allen, J. P., Feher, G., Yeates, T. O., Komiya, H., and Rees, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **85**(22), 8487–8491 (1988). 64, 67

[113] Ermler, U., Fritzsch, G., Buchanan, S. K., and Michel, H. *Structure* **2**(10), 925–936 October (1994). 64, 67

[114] Koepke, J., Krammer, E. M., Klingen, A. R., Sebban, P., Ullmann, G. M., and Fritzsch, G. *J Mol Biol.* **371(2)**, 396–409 Aug (2007). 64, 67, 74, 75, 77

[115] Michel-Beyerle, M. E., Plato, M., Deisenhofer, J., Michel, H., Bixon, M., and Jortner, J. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **932**, 52–70 (1988). 64

[116] Hughes, A. V., Rees, P., Heathcote, P., and Jones, M. R. *Biophys. J.* **90**(11), 4155 – 4166 (2006). 66

[117] Kleinfeld, D., Okamura, M. Y., and Feher, G. *Biochemistry* **23**(24), 5780–5786 November (1984). 66

[118] Graige, M. S., Feher, G., and Okamura, M. Y. *Proc Natl Acad Sci U S A.* **95(20)**, 11679–84 (1998). 66

[119] Breton, J., Boullais, C., Mioskowski, C., Sebban, P., Baciou, L., and Nabedryk, E. *Biochemistry* **41**(43), 12921–12927 October (2002). 66

[120] Breton, J. *Biochemistry* **43**(12), 3318–3326 March (2004). 66

[121] Li, J., Takahashi, E., and Gunner, M. R. *Biochemistry* **39**(25), 7445–7454 June (2000). 66

[122] Remy, A. and Gerwert, K. *Nature Structural Biology* **10**(8), 637 – 644 (2003). 66

[123] Paddock, M. L., Flores, M., Isaacson, R., Chang, C., Abresch, E. C., and Okamura, M. Y. *Biochemistry* **46**(28), 8234–8243 July (2007). 66

[124] Debus, R. J., Feher, G., and Okamura, M. Y. *Biochemistry* **25**(8), 2276–2287 April (1986). 67

[125] Hermes, S., Bremm, O., Garczarek, F., Derrien, V., Liebisch, P., Loja, P., Sebban, P., Gerwert, K., and Haumann, M. *Biochemistry* **45**(2), 353–359 (2006). 67, 78, 82, 84, 86, 88

[126] Bunker, G., Stern, E. A., Blankenship, R. E., and Parson, W. W. *Biophys. J.* **37(2)**, 539–51 Feb (1982). primo lavoro di XAFS su RC. 67, 70, 75

[127] Eisenberger, R., Okamura, M. Y., and Feher, G. *Biophys. J* **37(2)**, 523–38 (1982). 67, 75, 77

[128] Chen, L. X., Utschig, L. M., Schlesselman, S. L., and Tiede, D. M. *The Journal of Physical Chemistry B* **108**(12), 3912–3924 March (2004). 67, 68, 75

[129] Gray, K. A., Farchaus, J. W., Wachtveitl, J., Breton, J., and Oesterhelt, D. *EMBO J.* **9**(7), 20612070 July (1990). 68

[130] Feher, G. *Photochem. Photobiol.* **14**(3), 373–387 (1971). 68

[131] Francia, F., Giachini, L., Palazzo, G., Mallardi, A., Boscherini, F., and Venturoli, G. *Bioelectrochemistry* **63**(1-2), 73–77 (2004). 69, 77

[132] Francia, F., Palazzo, G., Mallardi, A., Cordone, L., and Venturoli, G. *Biophys J.* **85**(4), 2760–2775 October (2003). 69, 77

[133] Cordone, L., Cottone, G., Giuffrida, S., Palazzo, G., Venturoli, G., and Viappiani, C. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **1749(2)**, 252–81 Apr (2005). 69, 77, 94

[134] Sacquin-Mora, S., Sebban, P., Derrien, V., Frick, B., Lavery, R., and Alba-Simionesco, C. *Biochemistry* **46**(51), 14960–14968 (2007). 78

[135] Nozawa, S., Adachi, S.-i., Takahashi, J.-i., Tazaki, R., Guérin, L., Daimon, M., Tomita, A., Sato, T., Chollet, M., Collet, E., Cailleau, H., Yamamoto,

S., Tsuchiya, K., Shioya, T., Sasaki, H., Mori, T., Ichiyanagi, K., Sawa, H., Kawata, H., and Koshihara, S.-y. *Journal of Synchrotron Radiation* **14**(4), 313–319 Jul (2007). 79, 81, 82

[136] Brzezinski, P., Okamura, M. Y., and Feher, G. In *The Photosynthetic Bacterial Reaction Center II. Structure, Spectroscopy, and Dynamics.*, Breton, J. and Vermeglio, A., editors, 321–330. Plenum Press, New York (1992). 82

[137] Okamoto, Y. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* **526**(3), 572 – 583 (2004). 92

[138] Ferlat, G., Soetens, J.-C., Miguel, A. S., and Bopp, P. A. *Journal of Physics: Condensed Matter* **17**(5), S145–S157 (2005). 92

[139] Kuzmin, A. and Evarestov, R. A. *J. Phys.: Condens. Matter* **21**(5), 055401 (2009). 92

[140] Ordway, G. A. and Garry, D. J. *J Exp Biol* **207**(20), 3441–3446 (2004). 92

[141] Frauenfelder, H., McMahon, B. H., Austin, R. H., Chu, K., and Groves, J. T. *Proceedings of the National Academy of Sciences of the United States of America* **98**(5), 2370–2374 (2001). 92

[142] Phillips, S. E. V. *Journal of Molecular Biology* **142**(4), 531–554 October (1980). 93

[143] Kachalova, G. S., Popov, A. N., and Bartunik, H. D. *Science* **284**(5413), 473–476 (1999). 93

[144] Vojtechovský, J., Chu, K., Berendzen, J., Sweet, R. M., and Schlichting, I. *Biophys J.* **77**(4), 2153–2174 October (1999). 93, 94, 99, 100, 101

[145] Ye, X., Demidov, A., and Champion, P. M. *J. Am. Chem. Soc.* **124**(20), 5914–5924 May (2002). 93

[146] Schotte, F., Lim, M., Jackson, T. A., Smirnov, A. V., Soman, J., Olson, J. S., Phillips, George N., J., Wulff, M., and Anfinrud, P. A. *Science* **300**(5627), 1944–1947 (2003). 94

[147] Scherk, C. G., Ostermann, A., Achterhold, K., Iakovleva, O., Nazikkol, C., Krebs, B., Knapp, E. W., Meyer-Klaucke, W., and Parak, F. G. *Eur Byophys J* **30**, 393–403 (2001). 94

[148] Rovira, C., Schulze, B., Eichinger, M., Evanseck, J. D., and Parrinello, M. *Biophys J.* **81**(1), 435–445 July (2001). 94, 102

[149] Melchers, B., Knapp, E., Parak, F., Cordone, L., Cupane, A., and Leone, M. *Biophys J.* **70**(5), 2092–2099 May (1996). 94

[150] Chance, B., Fischetti, R., and Powers, L. *Biochemistry* **22**(16), 3820–3829 August (1983). 94, 99, 100, 101, 102, 103

[151] Powers, L., Sessler, J. L., Woolery, G. L., and Chance, B. *Biochemistry* **23**(23), 5519–5523 November (1984). 94, 99, 100, 101, 102, 103

[152] Powers, L., Chance, B., Chance, M., Campbell, B., Friedman, J., Khalid, S., Kumar, C., Naqui, A., Reddy, K. S., and Zhou, Y. *Biochemistry* **26**(15), 4785–4796 July (1987). 94

[153] Della Longa, S., Arcovito, A., Girasole, M., Hazemann, J. L., and Benfatto, M. *Phys. Rev. Lett.* **87**(15), 155501 Sep (2001). 94, 95

[154] Rich, A. M., Armstrong, R. S., Ellis, P. J., Freeman, H. C., and Lay, P. A. *Inorg. Chem.* **37**(22), 5743–5753 November (1998). 94

[155] Chance, M. R., Miller, L. M., Fischetti, R. F., Scheuring, E., Huang, W.-X., Sclavi, B., Hai, Y., and Sullivan, M. *Biochemistry* **35**(28), 9014–9023 January (1996). 95

[156] Baroni, S., de Gironcoli, S., Dal Corso, A., and Giannozzi, P. *Rev. Mod. Phys.* **73**(2), 515–562 Jul (2001). 96

[157] Sigfridsson, E. and Ryde, U. *J. Inorg. Biochem.* **91**(1), 101–115 July (2002). 100