

Structure and dynamics of metal sites in  
proteins: an X-ray Absorption  
Spectroscopy investigation

Lisa Giachini

A thesis submitted for the degree of Doctor of Philosophy

School of Physics

University of Bologna

March 2007

# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>X-ray Absorption of metal centres in proteins</b>	<b>5</b>
2.1	X-ray Absorption Fine Structure . . . . .	5
2.1.1	Phenomenology . . . . .	5
2.1.2	Theoretical framework . . . . .	11
2.1.3	Historical overview and some open issues . . . . .	21
2.2	BIOXAS . . . . .	22
2.2.1	BIOXAS and structural genomics . . . . .	22
2.2.2	Current status of BIOXAS . . . . .	24
2.3	Introductory remarks about metal centres in proteins . . . . .	27
<b>3</b>	<b>Experimental</b>	<b>32</b>
3.1	Instrumentation . . . . .	32
3.1.1	Introduction . . . . .	32
3.1.2	Optical apparatus . . . . .	32
3.1.3	Detection modes . . . . .	34
3.2	Sample preparation . . . . .	37
3.3	Data acquisition . . . . .	38
3.3.1	General considerations . . . . .	38

3.3.2	Radiation damage . . . . .	41
3.3.3	Estimate of metal concentration with x-ray fluorescence	43
<b>4</b>	<b>EXAFS data analysis</b>	<b>46</b>
4.1	Introduction . . . . .	46
4.2	Background subtraction . . . . .	47
4.3	Fourier filtering . . . . .	48
4.4	Simulations of theoretical signals using FEFF8.2 . . . . .	50
4.5	Multi-parameter fitting . . . . .	51
4.5.1	Introduction . . . . .	51
4.5.2	Constrained and restrained refinement . . . . .	53
4.5.3	Determinacy . . . . .	54
4.5.4	Statistical significance . . . . .	54
<b>5</b>	<b>Data analysis strategies for metalloproteins</b>	<b>58</b>
5.1	Introduction . . . . .	58
5.2	Identification of unknown metal sites in proteins using EX- AFS data . . . . .	59
5.2.1	Description of the method . . . . .	59
5.2.2	Model building . . . . .	60
5.2.3	<i>Ab-initio</i> self consistent simulations . . . . .	65
5.2.4	Fitting procedure . . . . .	67
<b>6</b>	<b>Identification of zinc sites in charge translocating membrane protein complexes</b>	<b>70</b>
6.1	Overview . . . . .	70
6.1.1	Structure and function of selected charge translocat- ing membrane protein complexes . . . . .	71
6.1.2	Zn <sup>2+</sup> inhibition . . . . .	76

6.2	The inhibitory binding site of $Zn^{2+}$ in bacterial photosynthetic reaction center . . . . .	79
6.3	The inhibitory binding site of $Zn^{2+}$ in bacterial, avian and bovine $bc_1$ . . . . .	80
6.3.1	Materials and methods . . . . .	80
6.3.2	Data collection and data analysis . . . . .	81
6.3.3	Results and discussion . . . . .	82
6.4	The inhibitory binding site of $Zn^{2+}$ in cytochrome <i>c</i> oxydase . . . . .	109
6.4.1	Materials and methods . . . . .	109
6.4.2	Data collection and data analysis . . . . .	109
6.4.3	Results . . . . .	110
6.4.4	Discussion . . . . .	113
6.5	The endogenous binding site of $Zn^{2+}$ in complex I . . . . .	117
6.5.1	Materials and methods . . . . .	117
6.5.2	Data collection and data analysis . . . . .	117
6.5.3	Results . . . . .	118
<b>7</b>	<b>Matrix effects on the structure and dynamics of heme proteins embedded in dry trehalose</b> . . . . .	<b>121</b>
7.1	Overview . . . . .	121
7.2	Cytochrome <i>c</i> in a dry trehalose matrix: structural and dynamical effects probed by x-ray absorption spectroscopy . . . . .	128
7.2.1	Materials and Methods . . . . .	128
7.2.2	Data collection . . . . .	129
7.2.3	XAFS model . . . . .	130
7.2.4	Data analysis . . . . .	133
7.2.5	Results and Discussion . . . . .	135



7.3	XAFS features of MbCO embedded in a dry trehalose matrix	151
7.3.1	Materials and methods	152
7.3.2	Data collection	152
7.3.3	Qualitative results	153
7.4	Conclusions	153
<b>8</b>	<b>Conclusions</b>	<b>157</b>
<b>A</b>	<b>Rigid body refinement for selected aminoacids</b>	<b>159</b>
A.1	Histidine	160
A.2	Aspartate	161
A.3	Glutamate	162
A.4	Asparagine	163
A.5	Glutamine	164
A.6	Lysine	165
A.7	Cysteine	166

# List of Figures

2.1	Total photon cross section in carbon, as a function of energy, showing the contributions of different processes. This figure is taken from the X-ray data booklet web site: <a href="http://xdb.lbl.gov">http://xdb.lbl.gov</a> . . . . .	6
2.2	Schematic representation of photoelectric absorption. This figure is taken from [1]. . . . .	7
2.3	Schematic view of x-ray absorption. . . . .	7
2.4	Schematic view of x-ray absorption coefficient as a function of incident photon energy. This figure is taken from [2]. . . . .	9
2.5	The relationship between the x-ray absorption edges and the corresponding excitation of core electrons. This figure is taken from [2]. . . . .	9
2.6	Pictorial representation of the quantum-interference phenomenon which causes the XAFS oscillations. This figure is taken from [1]. . . . .	11
2.7	Pictorial representation of the muffin tin potential. . . . .	14
2.8	Possible scattering paths in the (110) plane of an fcc lattice. This figure is taken from [2]. . . . .	17
2.9	Possible scattering paths in the porphyrin pyrrolic group (see also Fig. 7.8). A: Single scattering paths. B: The two most important multiple scattering paths. . . . .	18
2.10	Bonding interaction for endogenous ligands. The figure is taken from [3].	29

3.1	Transitions that give rise to the various emission lines. the X-ray data booklet web site: <a href="http://xdb.lbl.gov">http://xdb.lbl.gov</a> . . . . .	34
3.2	Fluorescence yields for K and L shells for. The plotted curve for the L shell represents an average of L1, L2, and L3 effective yields. This figure is taken from the X-ray data booklet web site: <a href="http://xdb.lbl.gov">http://xdb.lbl.gov</a> . . . . .	35
3.3	Experimental chamber with HPGe solid state detector. . . . .	36
3.4	Absorption spectrum collected for complex I. Inset: $k^3$ weighted XAFS function is shown between 14-16 $\text{\AA}^{-1}$ in order to illustrate the quality of the data. . . . .	40
3.5	Absorption spectra collected for the MbCO in solution at 300 K. The numbers indicate the sequence of the collected spectra. . . . .	44
3.6	Absorption spectra collected for the iron site of bacterial photosynthetic reaction center embedded in a dry trehalose glass (in black) and in a PVA film (in red). Measurements were performed at 300 K. . . . .	44
3.7	Fluorescence spectrum for MbCO. . . . .	45
4.1	Background subtraction for avian $bc_1$ . . . . .	49
4.2	Fourier filtering applied on avian $bc_1$ EXAFS function. Here we used two hanning windows (sills: $dk=2$ and $dr=0.2$ respectively). . . . .	57
5.1	Tetrahedral cluster used for the simulation shown in Fig. 5.1. The first shell distances are: $Zn-N = 2.00 \text{ \AA}$ ; $Zn-O = 2.06 \text{ \AA}$ . The others bond length distances and angles have been set according to [4]. . . . .	65
5.2	FEFF8.2 self consistent simulation of the cluster shown in Fig. 5.1. From the top to the bottom: total sum of paths (until $nleg=5$ ); sum of the paths belonging to carboxylates; sum of the paths belonging to histidines; sum of the paths involving different residues. . . . .	66
5.3	1-Dimensional rigid body refinement for histidine residue. . . . .	68

5.4	2-Dimensional rigid body refinement for histidine residue. . . . .	69
6.1	The photosynthetic reaction center from <i>Rhodobacter Sphaeroides</i> . . . .	71
6.2	The respiratory chain. . . . .	73
6.3	EM structures of complex I. 1. <i>N. crassa</i> . (Leonard et al., 1987) 2. <i>N. crassa</i> . (Hofhaus et al., 1991); 3. <i>N. crassa</i> (Guenebaut et al., 1997); 4. <i>E. coli</i> NDH-1 (Guenebaut et al., 1998); 5. Bovine heart (Grigorieff, 1998); 6. <i>Y. lipolytica</i> (Radermacher et al., 2006); 7 and 8. <i>E. coli</i> (Böttcher et al., 2002); 7 is inactiv form, 8 is "active" form; 9. <i>Ara-bidopsis</i> (Dudkina et al., 2005). . . . .	74
6.4	The bc <sub>1</sub> complex. . . . .	75
6.5	The cytochrome c oxidase complex. . . . .	77
6.6	XANES spectra of the bacterial, avian and bovine bc <sub>1</sub> complex (from top to bottom). Experimental spectra are shown as continuous lines; dotted lines represent simulations based on the structural parameters obtained by EXAFS analysis. . . . .	83
6.7	Experimental k <sup>3</sup> weighted EXAFS functions measured in the bacterial, avian and bovine bc <sub>1</sub> complex (continuous lines). The dashed lines represent calculated best fitting EXAFS functions corresponding to the following clusters: two His, one Lys and one Asp/Glu for the avian and the bovine complexes (the corresponding structural parameters are given in Tab. 6.2, and Tab. 6.3, model m, respectively); one His, two Asp/Glu in monodentate coordination, one Gln/Asn, two water molecules for the bacterial complex (see Tab. 6.4, model e for the corresponding structural parameters). . . . .	84
6.8	Amplitudes of the Fourier Transforms (FT) of the k <sup>3</sup> weighted EXAFS functions shown in Fig. 6.7, performed in $\Delta k = 2.5 - 14.5 \text{ \AA}^{-1}$ . . . . .	85

6.9	Sketch of the reference structural units used in multi-shell multiple scattering analysis. The values of distances and angles of amino acid residues were set according to Engh and Huber. Glutamate residue has the same structure of the aspartate provided that $C_\gamma$ and $C_\beta$ of aspartate are replaced by $C_\delta$ and $C_\gamma$ of glutamate, respectively. The structure of the glutamine residue coincides with that of the asparagine residue when $N_{\delta 2}$ , $C_\gamma$ and $C_\beta$ of asparagine are replaced by $N_{\epsilon 2}$ , $C_\delta$ and $C_\gamma$ of glutamine, respectively. . . . .	89
6.10	View of the proposed $Zn^{2+}$ binding sites in the avian (A) and bovine (B) bc1 complex. The site of Zn (represented as a grey sphere) is in the interface between cyt b (yellow) and cyt c (blue). The structure in panel A was obtained from the XRD data of the Zn-crystal of chicken bc1. Coordinates are from bczn3ref.pdb file available at <a href="http://sb20.lbl.gov/cytbc1/PDB/">http://sb20.lbl.gov/cytbc1/PDB/</a> . Stigmatellin, below the Zn site, is from the superimposed avian bc1 structure 2BCC. Coordinates of panel B are from the structure of the bovine cyt bc1 obtained in the presence of stigmatellin (1PP9). Zn ion was superimposed after alignment of this structure with that of the Zn crystal of the avian complex (see panel A).	106

- 6.11 (A)  $k^3$ -weighted  $\chi(k)$  EXAFS signals obtained from samples characterized by different Zn/COX ratios. (B) The corresponding magnitude of Fourier transform (FT). From top to bottom: trace a, sample a, characterized by 1.0 Zn/complex; trace b, sample b, 1.5 Zn/complex; trace c, sample c, 2.0 Zn/complex. The first-shell fit to the FT of sample a is shown with empty squares in Fig. 6.11B. The vertical dashed line indicates the position of the peak attributed to the binding of exogenous zinc. The amplitude of this peak increases from sample b to sample c, reflecting the progressive occupancy of a new  $Zn^{2+}$  binding site. Trace d in panel A is the EXAFS signal of the site to which exogenous  $Zn^{2+}$  binds, extracted from trace c by subtracting trace a with the appropriate weights. The fit corresponding to cluster 1 of Tab. 6.7 (i.e. three histidines and one aspartate or glutamate residue) is shown as a dashed line. See text for details. Trace d in panel B is the correspondent FT. . . . . 111
- 6.12 Space-filled view of the mass of the 13 subunits bovine cytochrome c oxidase protruding at the N side of the membrane. The picture was elaborated with the Rasmol 2.6 program using the 2OCC atomic coordinates from the PDB data bank of cytochrome c oxidase [32]. With their respective numbering the putative Zn binding residues emerging at the N surface of subunits I (red), subunit VIIc (green) and subunit III (yellow) and Asp91-I (blue) of the D channel are shown. . . . . 114
- 6.13 View parallel to the membrane of the location in subunits I (black) and VIIc (green) of the oxidase of the putative  $Zn^{2+}$  coordinating residues. The picture was elaborated as in Fig. 3A. It can be noted that the putative Zn coordinating residues, except Lys 13-I, belong to the carboxy-terminal segment of subunit-I and the NH<sub>2</sub>-terminal segment of subunit VIIC respectively. . . . . 115

6.14	From the top to the bottom: experimental (open circles) and calculated best fitting XAFS signal (superimposed continuous line) obtained using the model formed by two cysteine and two histidines; the fitting XAFS signals obtained for the models formed by 1 cysteine 3 histidines and 3 cysteine 1 histidine, respectively (continuous line). . . . .	119
7.1	a-D-glucopyranosyl-1,1-a-D-glucopyranoside. Oxygen atoms are in dark gray, carbon atoms in gray, hydrogen atoms in white. . . . .	122
7.2	Desert resurrection plant (Selaginella). Anhydrobiotic organism which can undergo anhydrobiosis. It contains large quantities (as much as 20% of the dry weight) of trehalose, . . . . .	123
7.3	Tardigrade or water bear. Like the resurrection plant it contains large quantities of trehalose and it can undergo anhydrobiosis. . . . .	124
7.4	A: Reference structural model of the Fe ligand cluster. B: Sketch of the structural units of the reference model. The values of distances and angle are the following. Heme pyrrolic unit: $\text{Fe-N}_p = 1.99 \text{ \AA}$ ; $\text{N}_p\text{-C}_{2,5} = 1.38 \text{ \AA}$ ; $\text{C}_{2,5}\text{-C}_{3,4} = 1.44 \text{ \AA}$ ; $\text{C}_{3,4} = 1.34 \text{ \AA}$ ; $\text{C}_2\text{C}_6 = 1.38 \text{ \AA}$ ; $\text{FeN}_p\text{C}_{2,5} = 109^\circ$ ; $\text{FeN}_p\text{C}_{3,4} = 163^\circ$ ; $\text{N}_p\text{C}_2\text{C}_6 = 125^\circ$ . Histidine imidazole: $\text{Fe-N}_{\epsilon 2} = 2.00 \text{ \AA}$ ; $\text{N}_{\epsilon 2}\text{-C}_{\epsilon 1} = 1.32 \text{ \AA}$ ; $\text{N}_{\epsilon 2}\text{-C}_{\delta 2} = 1.37 \text{ \AA}$ ; $\text{C}_{\epsilon 1}\text{-N}_{\delta 1} = 1.34 \text{ \AA}$ ; $\text{C}_{\delta 2}\text{-C}_\gamma = 1.35 \text{ \AA}$ ; $\text{C}_\gamma\text{-N}_{\delta 1} = 1.35 \text{ \AA}$ ; $\text{FeN}_{\epsilon 2}\text{C}_{\epsilon 1} = 128^\circ$ ; $\text{FeN}_{\epsilon 2}\text{C}_{\delta 2} = 127^\circ$ ; $\text{FeN}_{\epsilon 2}\text{N}_{\delta 1} = 163^\circ$ ; $\text{FeN}_{\epsilon 2}\text{C}_\gamma = 162^\circ$ . Methionine residue: $\text{Fe-S}_\delta = 2.29 \text{ \AA}$ ; $\text{S}_\delta\text{-C}_\gamma = 1.81 \text{ \AA}$ ; $\text{S}_\delta\text{-C}_\epsilon = 1.82 \text{ \AA}$ ; $\text{C}_\gamma - \text{C}_\beta = 1.5 \text{ \AA}$ ; $\text{FeS}_\delta\text{C}_{\epsilon,\gamma} = 109^\circ$ ; $\text{S}_\delta\text{C}_{\gamma,\beta} = 110^\circ$ . . . . .	132
7.5	Edge and enlarged preedge (inset) spectra for ferrocytochrome c in solution with glycerol. . . . .	136

7.6	k <sup>3</sup> weighted XAFS for all measured samples. sol. 1: solution without glycerol; sol. 2: solution in the presence of glycerol; PVA: protein embedded into a polyvinyl alcohol film; trehal. 1: protein incorporated into a moderately dried trehalose matrix; trehal. 2: protein embedded into an extensively dried trehalose matrix. See Materials and Methods for details concerning sample preparation. . . . .	137
7.7	Simulations based on the reference structural model using different values for the DW factors. . . . .	139
7.8	Simulations based on the reference structural model (see Fig.1). The values of the DW factors have been fixed as in model 1 (dashed line) and 2 (continuous line) (see the legend of Fig. 4). Together with the total XAFS signal (top) we show the main partial contributions. . . . .	141
7.9	The effect of different orientations of the histidine ligand on the XAFS signal. Simulations have been performed starting from the reference structural model (Fig.1) modified as follows. The imidazole ring was rotated around an axis passing through N <sub>e2</sub> and parallel to the y axis (bending movement) of 10° and 20° (bend10, bend20); the imidazole was rotated around the z axis (twisting movement) of 10° and 20° (twist10, twist20); the whole histidine residue was rotated around the y axis (in plane tilting) of 10° and 20° (tilt10, tilt20). In all cases DW factors were fixed to the values of model 2 (see legend of Fig. 7.7). . . . .	143
7.10	Experimental (continuous lines) and calculated best fitting XAFS signals (dashed lines) for the three most significant samples. From the top to the bottom: protein in solution with glycerol (sol. 2), protein embedded in the moderately dried trehalose matrix (trehal. 1) and in the extensively dried trehalose matrix (trehal. 2). See text for details. . . . .	146



7.11	Graphical representation of the porphyrin pyrrolic group fourth shell elongation in the extra dry trehalose matrix. To obtain a simple picture we maintained the planarity of the structure, keeping collinear the Fe and C <sub>3,4</sub> atoms in the elongation. With these assumptions the C <sub>2,5</sub> -C <sub>3,4</sub> and the C <sub>3</sub> -C <sub>4</sub> distances changed by 13% and 10% respectively. . . . .	151
7.12	Experimental EXAFS functions collected for MbCO. From the top to the bottom: MbCO in solution (continuous line) and embedded in trehalose (open circles) measured at 300 K, 200 K, 100 K. . . . .	154
A.1	Rigid body refinement of histidine residue. . . . .	160
A.2	Rigid body refinement of aspartate residue. . . . .	161
A.3	Rigid body refinement of glutamate residue. . . . .	162
A.4	Rigid body refinement of asparagine residue. . . . .	163
A.5	Rigid body refinement of glutamine residue. . . . .	164
A.6	Rigid body refinement of lysine residue . . . . .	165
A.7	Rigid body refinement of cysteine residue. . . . .	166

# List of Tables

2.1	Coordinated ligands in metalloproteins. The table is taken from [3]. . . . .	28
5.1	CSD parametrization atom type. This Tab. is taken from [4]. . . . .	61
5.2	Bond-length parameters derived from a statistical survey of X-ray structures of small compounds from the Cambridge Structural Database according to Engh and Huber. This Tab. is taken from [4]. . . . .	62
5.3	Angle parameters derived from a statistical survey of X-ray structures of small compounds from the Cambridge Structural Database according to Engh and Huber. This Tab. is taken from [4]. . . . .	63
6.1	Results of first shell analysis in the bovine and avian bc1 complexes ( $\Delta k = 2.5-14.6 \text{ \AA}^{-1}$ ; $\Delta R = 1-2 \text{ \AA}$ ). DW indicates the Debye Waller factor. The $1\sigma$ error on the least significant figure of the structural parameters is reported in brackets. . . . .	86
6.2	Structural results from multi-shell multiple scattering analysis in the avian bc1 complex. . . . .	92
6.3	Structural results from multi-shell multiple scattering analysis in the bovine bc1 complex. . . . .	93

6.4	Structural parameters determined by fitting the experimental data of the bacterial bc <sub>1</sub> complex to the model clusters extracted from the metallo-protein database MDB. Monodentate and bidentate binding configuration is indicated by m and b; n is the coordination number. . . . .	95
6.5	Structural parameters obtained by fitting the spectrum of the bacterial bc <sub>1</sub> complex to additional clusters containing two histidine residues. . .	97
6.6	Structural parameters derived by fitting the data obtained in the bacterial bc <sub>1</sub> complex to clusters built from the model e of Tab. 6.4 which minimizes the R factor. . . . .	98
6.7	(A) Structural results obtained for the inhibitory zinc site of cytochrome c oxidase. The 1σ error on the least significant figure is reported in brackets.	113
6.8	Structural results obtained for Complex I. The angles β and ζ' are shown in Appendix A. . . . .	118
7.1	First ligand distances for the iron site in ferrocyanochrome c obtained from the PDB. . . . .	131
7.2	Structural and dynamical parameters obtained for the five different matrices. The 1σ error on the least significant figure is reported in brackets. Δr indicates the change in distance from Fe with respect to the reference model. . . . .	145

# Acknowledgements

I would like to thank my thesis advisor Prof. Federico Boscherini for giving me the opportunity to work on such an interesting project. I am especially indebted to Francesco Francia for help, encouragement and for the extremely nice samples he managed to make! I am also very grateful to Giulia Veronesi for her agreeability and for useful discussion during the last year of my PhD. I would like to express my gratitude to Paola Serretti, Manuela Dezi and Manuela D'Alessandro for care, understanding and invaluable friendship and, of course, to my parents and to my sister for their endless support. And most of all, I would like to thank Prof. Giovanni Venturoli, for his strong guidance and inestimable assistance.

# Chapter 1

## Introduction

Metalloproteins have been receiving a great interest in the scientific community since they constitute a significant proportion of all known genomes. A particular attention is devoted to the metal sites, which are often responsible for protein function. Detailed structural data for the metal sites in a metalloprotein are essential in order to fully understand the structure-function relationships which make possible the performance of life-sustaining processes. X-ray Absorption Fine Structure (XAFS) is an ideal tool for selectively probing the local structure of metal ion in a protein. Among the advantages we recall the applicability to non crystallized samples and the accuracy of the determination of the interatomic distances (typically within a few hundredths of an Å for the first shell). Although the continuous development of advanced data simulation and analysis program has significantly improved the reliability of the technique, the use of XAFS is often still confined within an approach which combines high-resolution crystallographic information and ultra high resolution EXAFS. This kind of approach, in spite of its utility for studying subtle structural changes, makes the use of XAFS completely unhelpful in cases where crystallographic information is

not available.

In the present thesis we have developed a general analysis strategy to obtain the structure of an unknown metal site in a protein based on XAFS data. The only required information are the bond-length and angle parameters of amino acids derived from a statistical survey of x-ray structure of small compounds and some knowledge about the stereochemistry of metal sites in proteins, which can be extracted from an extensive compilation of metal coordination sites in proteins (<http://tanna.bch.ed.ac.uk>). This method has been applied for solving the structure of the zinc sites in three key enzymes of the respiratory chain (complex I, complex III and complex IV).  $\text{Zn}^{2+}$  is a strong inhibitor of the catalytic activity of these enzymes, thus the local structure determination of the metal binding sites is extremely important in order to shed light into the inhibitory mechanism. The XAFS structural determination is particularly significant since the crystallographic structure of the Zn-complexes is known only for complex III (and only at 3.8 Å resolution). We recall moreover that information on complex I structure is still extremely limited (only the hydrophilic domain has been crystallized).

Another aspect of the thesis work concerns the investigation of the matrix effects on the local structure structure and dynamics of heme proteins embedded in extra dry trehalose, a disaccharide that appears to be particularly effective in exerting a bioprotectant action. The main difficulty in such an analysis is the disentanglement between structural and dynamical aspects. We manage to obtain significant results pointing out that an extensively dehydrated trehalose matrix has strong effects on the structure and dynamics of the Fe ligand cluster.

The thesis is organized in the following Chapters:

Chapter 2 provides an overview on XAFS applied to metalloproteins. We briefly outline XAFS theory, the status of the art of BIOXAS and some aspects of the chemistry of metal sites in proteins.

Chapter 3 provides the description of the experimental techniques and sample preparation.

Chapter 4 reports the main steps of an EXAFS data analysis.

Chapter 5 provides a description of the data analysis strategies performed in this thesis, giving a particular attention to the original approach we have developed.

Chapter 6 reports the results obtained for the study of the inhibitory zinc sites in charge translocating membrane protein complexes. An introduction to the structure and function of the examined enzymes is also given.

Chapter 7 reports the results obtained for the study of heme proteins embedded in an extra dry trehalose matrix. An introduction to the impressive properties of trehalose is also given.

## Chapter 2

# X-ray Absorption of metal centres in proteins

### 2.1 X-ray Absorption Fine Structure

#### 2.1.1 Phenomenology

In the energy region of x rays (1-30 KeV) the interaction of photons with matter is governed by three distinct processes: photoelectric effect, coherent and incoherent scattering (Compton effect). The total cross section is dominated by photoelectric effect (as an example see the case of carbon illustrated in Fig. 2.1). The photoelectric effect consists in the absorption of a photon and, consequently, in the emission of an electron, usually named photoelectron, which belongs to a particular atomic core-orbital. The probability for absorption to occur is not zero when the energy of the photon is equal or bigger than the electron binding energy. A schematic representation of absorption is shown in Fig. 2.2. It is worthwhile to notice that the simultaneous conservation of energy and momentum implies that only a bound electron can absorb a photon thanks to the presence of the nucleus



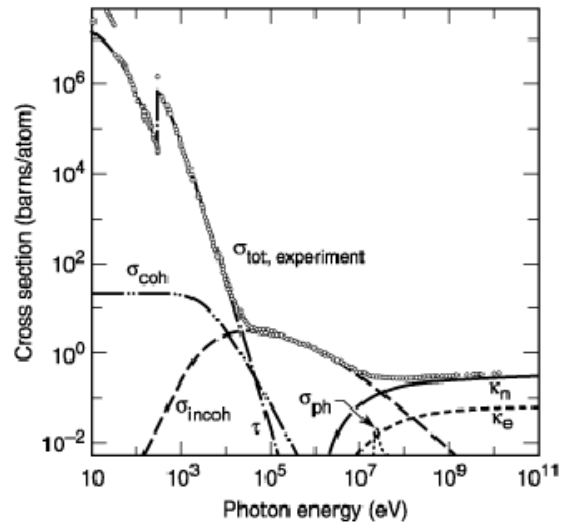


Figure 2.1: Total photon cross section in carbon, as a function of energy, showing the contributions of different processes. This figure is taken from the X-ray data booklet web site: <http://xdb.lbl.gov>.

which can pick up momentum. It is also important to underline that this process is not instantaneous having a transition rate described by the Fermi golden rule. X-ray Absorption Spectroscopy (XAS) measures the absorption

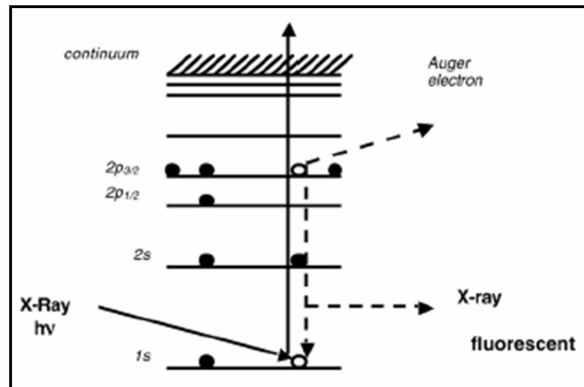


Figure 2.2: Schematic representation of photoelectric absorption. This figure is taken from [1].

of x rays as a function of x-ray energy. Experimentally it can be obtained

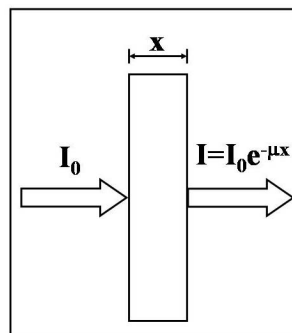


Figure 2.3: Schematic view of x-ray absorption.

from the attenuation of the beam as it traverses a sample of thickness  $x$ , which follows the relation:

$$I(E) = I_0(E)e^{-\mu(E)\cdot x} \quad (2.1)$$

where  $I_0$  and  $I$  are the intensity of the impinging and transmitting photon beam, respectively (as shown in Fig. 2.3),  $x$  is the thickness of the sample and  $\mu$  is the absorption coefficient;  $\mu$  can be related to the atomic concentration  $\varrho$  and absorption cross-section  $\sigma$  of the  $n$  constituent atoms in the following way:

$$\mu(E) = \sum_{i=1}^n \varrho_i \cdot \sigma_i(E). \quad (2.2)$$

If we measure the absorption coefficient of any kind of matter (in the solid, liquid or gaseous phase), we will observe an overall decrease in x-ray absorption with increasing energy interrupted by sharp increase at certain energies, named edges. Above the edges, a series of wiggles that modulate the absorption will appear. A schematic view of these features is shown in Fig. 2.4. The first overall trend reflects the quantum-mechanical phenomenon of x-ray absorption by atoms; the presence of an edge belongs to a given absorption atom and reflects the excitation energy of inner-shell electrons (see Fig. 2.5); the post edge wiggles are the X-ray Absorption Fine Structures and they are strictly related to the structural environment of the atom to which the edge belongs. In principle, all these features can be explained in the framework of the Fermi's Golden rule (derived within the time-dependent perturbation theory), which governs the transition rate, due to a perturbation, between initial and final many electron states:

$$W_{i \rightarrow f} = \frac{2\pi}{\hbar} |\langle \Psi_i | \mathbf{H}' | \Psi_f \rangle|^2 \delta(E_f - E_i - \hbar\omega). \quad (2.3)$$

In Eq. 2.3  $\mathbf{H}'$  is the interaction hamiltonian, due to the interaction between the atoms and the electromagnetic field and  $\hbar\omega$  is the energy of the

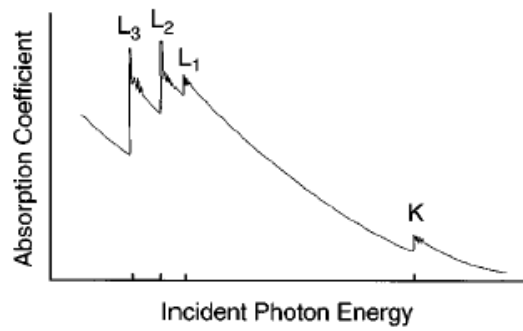


Figure 2.4: Schematic view of x-ray absorption coefficient as a function of incident photon energy. This figure is taken from [2].

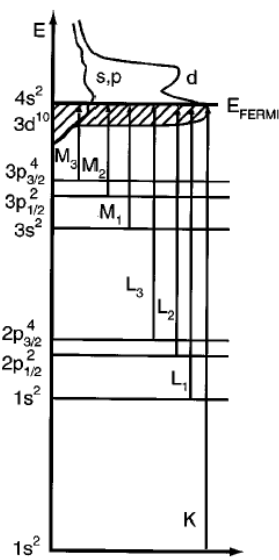


Figure 2.5: The relationship between the x-ray absorption edges and the corresponding excitation of core electrons. This figure is taken from [2].

impinging photons. The Dirac delta function, which dictates the energy conservation, implies that the initial and final states have an infinite average lifetime. If it is not the case, from the uncertainty principle follows that the correspondent energy level is not delta-function like but is better thought of as having a finite width (Lorentzian line shape). In the case of transitions to the continuum the Dirac delta function can be replaced by the density of states.

The XAFS oscillations are interpreted as a quantum-interference phenomenon. In correspondence to an absorption edge a core electron is emitted. Its energy ( $E_f$ ) is given by the difference between the photon energy ( $\hbar\omega$ ) and the core binding energy ( $E_b$ ),

$$E_f = \hbar\omega - E_b \quad (2.4)$$

Invoking the De Broglie wave-particle dualism, one can see the outgoing photoelectron as a quantum wave that propagates towards the surrounding atoms which, in turns, reflect it in the same way that any obstacles reflect water waves (see Fig. 2.6). This scattering process, physically caused by the potential generated by both nuclei and electronic charges in the system, involves simultaneously all the atoms around the absorber, as long as they are reached by the impinging photoelectron-wave. Since, as stated above, absorption and emission are not instantaneous, the backscattered waves have enough time to interact with the outgoing photoelectron wave. This interaction perturbs the final many electron states, which determines the matrix element of the transition rate, leading to an increase or a decrease of the absorption cross section depending on the relative phase of the two waves, thus depending on both photoelectron energy (or wavelength) and geometry

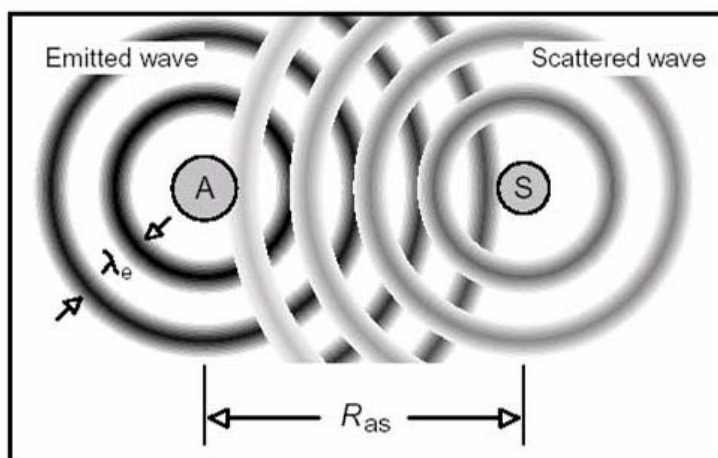


Figure 2.6: Pictorial representation of the quantum-interference phenomenon which causes the XAFS oscillations. This figure is taken from [1].

of scatterers. If we now measure the absorption coefficient as a function of photon energy, thus as a function of photoelectron energy (or wavelength), in correspondence of each energy value, the amplitude of the backscattering wave will add in a different way to the outgoing photoelectron wave, producing the observed modulations of the absorption coefficient. It is intuitive, indeed, that the XAFS oscillations must contain structural information of the absorber local environment; however the way through which this information can be extracted is quite complicated, as we will see in the next paragraph.

### 2.1.2 Theoretical framework

The purpose of this paragraph is to give some fundamentals of XAFS theory in order to better understand critical aspects of data analysis. We will confine the treatment only to general concepts. The validity and limitations of the main approximations will be discussed. For a thorough description

of theoretical aspects the reader is referred to the main review papers ([5], [6], [2]). As stated above the x-ray photoelectron interaction is treated to first order in time dependent perturbation theory and thus it is described by Eq. 2.3, where the matrix element can be written as:

$$|\langle \Psi_i | \sum_j \frac{e}{m} \cdot \mathbf{A}_j \cdot \mathbf{p}_j | \Psi_f \rangle| \quad (2.5)$$

where the index  $j$  refers to the various electrons in the system. The photon field is calculated at the position operator of each electron  $\mathbf{r}_j$ . Written in this way, the previously equation is telling us that the interaction hamiltonian acts on the many electron states, making the problem practically intractable.

A great simplification can be performed if one assumes that only one electron is involved in the process. This assumption, called one electron approximation, permits to factorize the wave functions, describing the many electron states, in the following way:

$$\Psi = \Psi_{N-1} \cdot \Psi_c, \quad (2.6)$$

where  $\psi_c$  is the wave function of the core electron and  $\psi_{N-1}$  refers to the other  $N - 1$  passive electrons. Within this approximation, the Hamiltonian will act only on the electron that makes the transition while the other  $N$  electrons will be approximated by a single configuration Slater determinant composed of one electron orbitals. For the ground state these orbitals will be obtained from a suitable self-consistent atomic calculation corresponding to the ground-state electronic configuration, while for the excited state the relaxation of the electronic charge around the core hole can be accounted for by adopting a different basis set of self-consistent orbitals, optimized for the

final state configuration. Although this approximation is extremely useful, it neglects important many-body effects and the inclusion of corrections is required to improve the agreement with experimental data, as we will see below.

The matrix element can be further simplified by assuming that the photon wavelength is bigger than the extension of the core electron wave function, thus neglecting the spatial variation of the vector potential (dipole approximation) so that  $\exp(ikr) \cong 1$ . However it should be noticed that this approximation is not a big limitation of the theory and, if necessary, multi-pole expansions for the transition operator can be inserted in the calculation. Finally we want to use an r-space representation. This can be achieved by using the commutation rule although, in this case, it is not strictly correct because we are not using exact wavefunctions (eigenstates of the hamiltonian).

In the framework of these three approximations, we obtain the usual expression for the x-ray absorption cross section:

$$\sigma(\hbar\omega) = 4\pi^2\alpha\hbar\omega \sum_f |\langle\psi_c|\hat{\varepsilon} \cdot \mathbf{r}|\psi_f\rangle|^2 \delta(E_f - E_i - \hbar\omega) \quad (2.7)$$

where  $\hat{\varepsilon}$  is the polarization vector.

In the previous paragraph we have seen that, during x-ray absorption, the final state is modified by the waves scattered from the potential generated from neighboring atoms. It is thus clear that the determination of such potential plays a crucial role in the description of the perturbation to the final state. In spite of this an extremely simplified approximation is commonly used for the calculation of the potentials and it is sufficiently



accurate in many situations <sup>1</sup>. This consists in spherically averaging the potential around each atom and adopting a constant interstitial potential in between (muffin tin potential). For a pictorial representation of this potential see Fig. 2.7. The efficacy of such simplified model lies in the fact



Figure 2.7: Pictorial representation of the muffin tin potential.

that the photoelectron scattering depends mainly on the inner part of the atomic potential wells. However, it must be stressed that near the edge the XAFS can be, indeed, largely sensitive to the details of the potential in the interstitial region between the atoms and thus such a model can be too poor and, as a consequence, simulations can substantially differ from experimental features. Moreover this approximation completely neglects the existence of localized charge due to chemical bonds. As a result of the muffin tin approximation the atomic scattering properties are independent of the actual atomic positions. The distance of the neighbours affects only the phase factor associated with the free propagation of the photoelectron from one atomic centre to another. In this approximation, the calculation of the absorption cross-section is simplified into a multiple scattering problem of the final state wavefunction by a collection of spherically symmetric scattering

---

<sup>1</sup>However, as we will see in the next Chapters, the muffin tin approximation seems to be inadequate to mimic the XAFS features at low  $k$  values of metal sites in proteins [7], especially when the coordination geometry is tetrahedral [8]

centers.

The photoelectron scattering problem can be solved using a different formalism which appeared to be particularly convenient for XAFS, since it can naturally incorporate inelastic losses and other quasi-particle effects and avoid the necessity of explicit calculations of wave functions. This approach is based on Green function and it is considered the starting point for the multiple-scattering theory of XAFS ([9]). In position space  $G$  has the spectral representation:

$$G(\mathbf{r}, \mathbf{r}'; E) = \sum_f \frac{\psi_f(\mathbf{r}') \cdot \psi_f^*(\mathbf{r})}{E - E_f + i\Gamma} \quad (2.8)$$

where  $\psi_f$  are final states with energies  $E_f$  of the effective final state one-particle Hamiltonian, including an appropriately screened core hole described by an optical potential <sup>2</sup> and  $\Gamma$  is a net lifetime, including effects of extrinsic and intrinsic losses. Within this framework the x-ray absorption cross section is rewritten in terms of the projected photoelectron density of final states or, equivalently, the imaginary part of the one particle Green's function:

$$\sigma(\hbar\omega) = 4\pi^2 \alpha \hbar \omega \frac{1}{\pi} \text{Im} \langle \psi_c | (\hat{\varepsilon}^* \cdot \mathbf{r}') G(r, r', E) (\hat{\varepsilon} \cdot \mathbf{r}) | \psi_c \rangle \quad (2.10)$$

In this framework it has been demonstrated that the polarization averaged

---

<sup>2</sup>the optical potential  $V$  appears in the one-particle equation for the photoelectron state  $\psi$ :

$$[-\frac{1}{2}\nabla^2 + V(E)]\psi = E\psi. \quad (2.9)$$

It consists of the net Coulomb potential  $V_c$  and a self energy  $\Sigma(E)$ ,  $V(E) = V_c + \Sigma(E)$ .

cross-section for a transition to a final state of angular momentum  $l$  is:

$$\sigma(\hbar\omega) = \frac{\sigma_0}{(2l+1)\sin^2(\delta_l^0)} \text{Im} \sum_m [(\mathbf{I} + \mathbf{T}\mathbf{G})^{-1}\mathbf{T}]_{LL}^{0,0} = \sigma_0[1 + \chi(k)] \quad (2.11)$$

where  $\mathbf{I}$  is the unit matrix,  $\mathbf{T}$  and  $\mathbf{G}$  are the atomic scattering and propagator matrices, respectively.  $\mathbf{G}$  is an off-diagonal matrix and describes the free propagation among the sites,  $\mathbf{T}$  is diagonal in the site and angular momentum indices and accounts for the energy-dependent scattering properties of the atomic potential.

At sufficiently high energies above the edge (say 50 eV) the formal matrix expansion in Eq. 2.11 is convergent. This is the reason why historically the XAFS oscillations are divided into Extended X-ray Absorption Fine Structures (EXAFS) and X-ray Absorption Near Edge Fine Structures (XANES) region. The convergence of the multiple scattering series greatly simplified the calculation, so that during the last 30 years a lot of improvements have been made in the interpretation of XAFS in the EXAFS region. To the contrary, calculation of the XANES region is time consuming requiring the inversion of the matrix  $\mathbf{I} + \mathbf{T}\mathbf{G}$ . Quantitative analysis of XANES is still under development and nowadays XANES spectroscopy is normally used with a finger-printing approach, with the exception of some programs, used for molecular structures, which include a quantitative fitting procedure (for example MXAN, [10], [11], [12]). Since XANES region contains a high number of structural information, being sensitive to oxidation state and 3-D geometry, further improvement in the quantitative analysis will be particularly useful.

We now come to the better understood Extended x-ray Absorption Fine Structures. Thanks to the convergence of the multiple scattering series, the

cross section can be written as the atomic cross section multiplied by a sum of a limited number of terms ( $\chi_i$ ), each relative to a particular scattering path:

$$\sigma(\hbar\omega) = \sigma_0(\hbar\omega)\{1 + \sum \chi_2^{0i0} + \sum \chi_3^{0ij0} + \dots\} \quad (2.12)$$

Where 0 identifies the central atom, the subscript indicate the number of scattering legs and the superscripts indicate the scattering path, starting and ending on the central atom via atoms  $i$  and  $j$ . Possible scattering paths for an *fcc* lattice and a porphyrin pyrrolic group are pictorial represented in Fig. 2.8 and Fig. 2.9 respectively. Since cross section and absorption

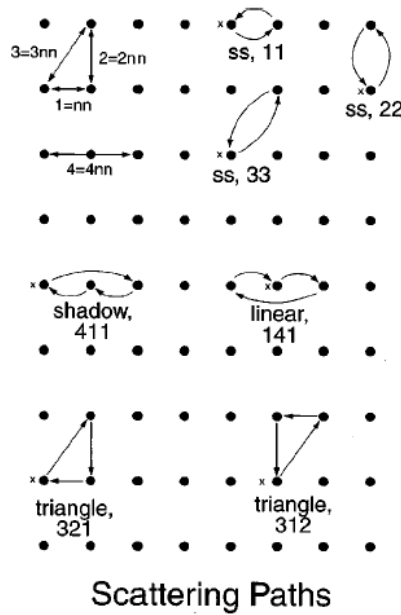


Figure 2.8: Possible scattering paths in the (110) plane of an *fcc* lattice. This figure is taken from [2].

coefficient are related by Eq. 2.2, we can also write:

$$\mu(\hbar\omega) = \mu_0(\hbar\omega)\{1 + \sum \chi_2^{0i0} + \sum \chi_3^{0ij0} + \dots\} \quad (2.13)$$

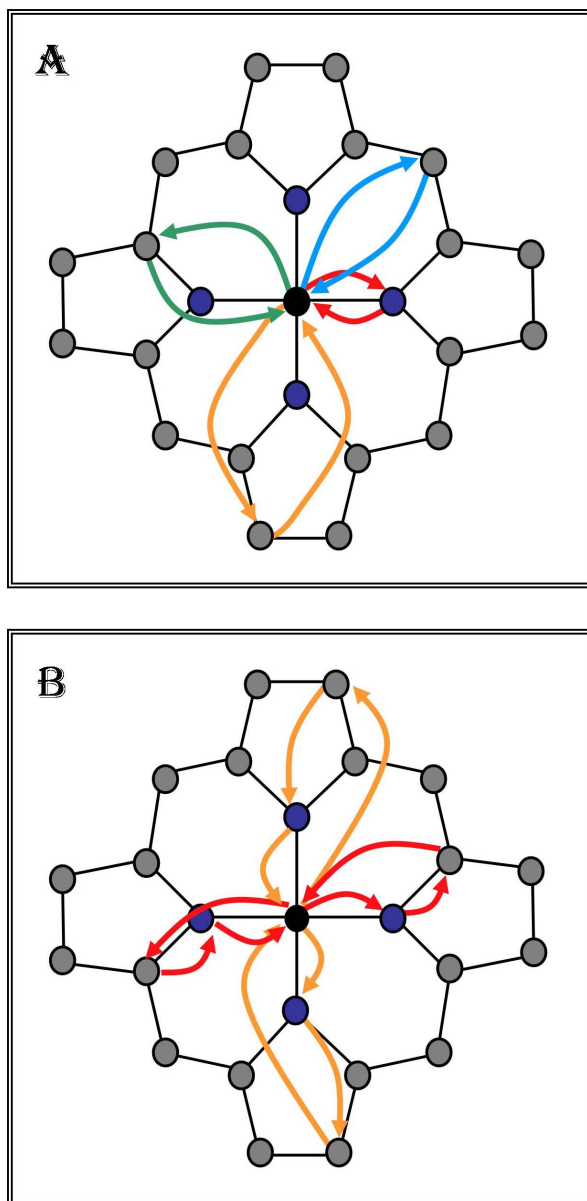


Figure 2.9: Possible scattering paths in the porphyrin pyrrolic group (see also Fig. 7.8).  
 A: Single scattering paths. B: The two most important multiple scattering paths.

We now define the EXAFS function  $\chi(k)$  as the fractional oscillation of the absorption coefficient with respect to the atomic background, as a function of the photoelectron wavevector  $k = \frac{\sqrt{2mE_f}}{\hbar}$ , calculated with respect to a guessed continuum threshold  $E$  and normalized to the atomic cross-section of the edge under consideration:

$$\chi(k) = \frac{\mu(k) - \mu_0(k)}{\Delta\mu_{k=0}} = 1 + \sum \chi_2^{0i0} + \sum \chi_3^{0ij0} + \dots \quad (2.14)$$

By solving the eq. 2.7 it can be shown that each scattering path belonging either to a SS (single scattering) or to a MS (multiple scattering) configuration contributes to the total cross section with a sinusoidal function of the form:

$$\chi_n^\Gamma(k) = \frac{f_n(k, \Gamma)}{k \cdot r_\Gamma^2} \sin[k \cdot r_\Gamma + \phi_n(k, \Gamma) + 2\delta]. \quad (2.15)$$

where  $\Gamma = \{\mathbf{r}_i\}$  is a path index which identifies the position of the set of atoms involved in the scattering process,  $f_n(k, \Gamma)$  is the backscattering function of atoms belonging to path  $\Gamma$ ,  $\Phi_n(k, \Gamma)$  is a phase term coming from the scattering of atoms belonging to path  $\Gamma$ ;  $\delta$  is the phase shift due to the absorber atom and  $r_\Gamma$  is the total scattering path length.

Since the Eq. 2.15 has been obtained in the framework of the one electron approximation, important many body effects have been neglected. In order to mimic such effects it is now necessary to insert a-posteriori some corrections. The many body effects involve indeed a lot of different processes. It is traditional to divide them in two categories: intrinsic and extrinsic processes. Intrinsic losses refer to excitations in response to creation of core hole. These losses are traditionally accounted for, phenomenologically, by a

constant many-body amplitude reduction factor ( $S_0^2$ ), the value of which is usually between 0.7-0.9. This is not strictly correct since this factor should be energy dependent. More precisely this must be described by a complex, path dependent constant given by a phasor sum. Extrinsic effects refer to losses in propagation of the photoelectron and include excitations such as plasmons, electron-hole pairs, and inelastic scattering in which the photoelectron loses energy. These effects produce a decay of the final state and they act in a time range shorter than the core-hole lifetime so that the latter process doesn't affect the spectrum. This decay can be phenomenologically described by an energy dependent mean free path ( $\lambda$ ) so that we must add in Eq. 2.15 a decay factor of the form

$$\exp\left[\frac{-r\Gamma}{\lambda}\right] \quad (2.16)$$

Finally, disorder leads to the experimentally observed damping of the EXAFS signal which cannot be explained within the model used up to now. Both thermal motion and structural disorder can be treated in the same way since the time scale of the photoabsorption process ( $10^{-15}s$ ) is faster than the atomic vibration ( $10^{-12}s$ ). The damping can be modeled by an exponential term of the form  $\exp(-k^2\sigma^2)$  (gaussian approximation), where  $\sigma^2$  is the so called Debye waller (DW) factor ([13]).

Inserting these corrections in the solution of Eq. 2.15 we obtain the following expression for the EXAFS function belonging to each possible path:

$$\chi(k)_n^\Gamma = S_0^2 \cdot \frac{f_n(k, \Gamma)}{k \cdot r_\Gamma^2} \cdot \sin[k \cdot r_\Gamma + \phi_n(k, \Gamma) + 2\delta] \cdot e^{-2k^2\sigma_\Gamma^2} \cdot e^{\frac{-r\Gamma}{\lambda}} \quad (2.17)$$

the total EXAFS function being the sum of all possible contributions, as shown in Eq. 2.14. Since the EXAFS signal is the sum of sine functions,

the argument of which is  $kr_{\Gamma}$ , a Fourier Transform (FT) of it will exhibit peaks in correspondence to the path length, and hence it can be treated as a type of radial distribution function. From the analysis of an EXAFS spectrum, which requires an adequate starting structural model, one can obtain information on local structure parameters. In particular, extended x-ray absorption fine structure is sensitive to interatomic distances, identity of neighbouring atoms and can provide a model for the radial distribution function. A detailed description of data analysis methods will be given in Chapter 4.

### 2.1.3 Historical overview and some open issues

The first observation of the XAFS phenomenon can be dated to around the year 1920 and were made at the Siegbahn laboratory in Lund (Sweden). Kronig [14] originally interpreted the XAFS oscillations in crystals as a density of states effect, due to the strong diffraction of electrons by the crystal lattice, the so called long-range order theory. Shortly thereafter, Kronig presented an alternative theory for small molecule, known as short-range order effect [15]. This theory is the basis of the theoretical framework we have explained before: the oscillatory structure is interpreted as a modulation of the matrix element in the golden rule due to the influence of neighboring atoms. It is interesting to notice that both long- and short-range -order theories can be reconciled when appropriate broadening is introduced [16]. Despite these brilliant studies and other experimental and theoretical works, a real big breakthrough in the XAFS community was possible only in the seventies thanks to the pioneering paper by Sayers et al. [5], who proposed for the first time the possibility of inversion of the EXAFS signal using a Fourier transformation algorithm, and to the development of synchrotron radiation



EXAFS facilities which greatly reduced the acquisition time and improved the quality of the spectra. A fully quantitative theory, however, was not developed until the present decade and the complete understanding of the x-ray absorption fine structures is still a challenging theoretical problem. In fact many of the developments needed to explain this phenomenon depend on the advances in our understanding of other general problems, such as ground-state electronic structure. In particular the topics in which the theory can be further enhanced are: the treatment of the many body effects, the calculation of more realistic potentials [17] and the correct modeling of DW factors. The first two aspects affect more deeply the near edge region while the last one has a macroscopic effect in the extended region of the spectrum and advances in this field would greatly help in solving structures characterized by a high broadening in atomic distances (e. g. proteins).

## **2.2 BIOXAS**

### **2.2.1 BIOXAS and structural genomics**

XAS is a powerful technique thanks to its high sensitivity to local structure. Among the advantages we remind the possibility to uniquely select the absorber atom and to perform experiments either in solution or in solid matrix. Due to these characteristics, in the past 30 years it has become a reliable tool and it has been applied in different fields (solid state physics, geology, biology, chemistry, cultural heritage materials). We now focus our attention on the so called BIOXAS, i.e. the biological x-ray absorption spectroscopy. This is, indeed, a wide field, including a large number of biological systems. Among them a particular interest has been devoted to metalloproteins, i. e. proteins that contains metal centres often responsible

of protein function. The interest towards these systems finds its motivation within the framework of the structural genomics project which consists in the determination of the three dimensional structure of all proteins of a given organism and possibly to determine protein function from their 3D structure. Protein Crystallography has been by far the protagonist among the techniques used for this aim. Its elegant simplicity, able to give a three-dimensional image of the entire macromolecular architecture, has somehow shadowed the development of XAS which, to the contrary, is based on a complicated theory and needs a well defined a-priori structural model. However it should be reminded that early XAFS experiments had a big impact on XRD anomalous scattering directly related to the X-ray absorption process. Moreover it is strongly recommended (see [18]) to perform X-ray absorption measurements together with XRD experiments in order to control the oxidation state of the sample and to monitor possible radiation damage which represent, indeed, a big problem for Synchrotron radiation experiments on biological samples (see Paragraph 3.3.2). Actually, the potentiality of XAS is not limited to a support for protein crystallography. The wide diffusion of protein crystallography has caused, as negative aspect, the misleading perception of proteins as static structures. The reality is indeed extremely different. Metalloproteins use the redox and ligand chemistry of metals to their advantage to perform varied biological functions with specificity and control. Chemical reactions mean motion, so that a protein is better thought as a dynamical object which interconvert between different conformational substates [19]. Sometimes these movements are small and confined to the prosthetic group, sometimes they can involve the entire protein, like in the case of hemoglobin. It is thus clear that to understand function it is necessary to study not only an immobilized structure but the way through which

conformational substates interconvert, i.e. the transformation of chemical energy into mechanical motion. In this respect BIOXAS is a powerful technique since it doesn't need crystalization and it can allow the determination of some structural parameters (specifically interatomic distances) with a very high precision (say about one hundredth of Å). Recently the need to study changes in structure has led to the development of time resolved techniques. In 2003, at ID09 of the European Synchrotron Radiation Facilities (ESRF), picosecond time-resolved x-ray diffraction has been performed to characterize the structure of the L29F mutant of Myoglobin as it evolves from the carboxy to the deoxy state [20]. In the same years great efforts have also been made in the field of time resolved XAS [21].

BIOXAS can be particularly advantageous when used in a synergic combination of various techniques (Protein Crystallography (XP), Small Angle X-ray Scattering (SAXS), and Neutron Scattering (NS) etc..) (see [22] and [23]). However, it can provide useful information even when other structural information are not available, as we will see in Chapter 5.

### **2.2.2 Current status of BIOXAS**

BIOXAS birth is strictly related to the development of Synchrotron Radiation sources since it needs an extremely high photon flux tunable in a continuous energy range. The feasibility of using XAFS technique for biological systems was demonstrated in the first fluorescence experiment [24]. Then the growth of the subject has been very rapid although the experimental requirements for BIOXAS are among the most stringent at the synchrotron radiation facilities (due to the high dilution of the absorbing species) and data analysis is complicated by both complexity and disorder of the structures under investigation.

Much of the early theoretical work was based on the plane-wave approximation. Subsequently much effort has been put towards the use of the curve-wave approach based on Lee and Pendry's formalism ([25], [26]). Nowadays there are several widely recognized packages available, e.g. FEFF, EXCURVE, GNXAS ([27], [28], [29]). All of them use the main approximations described in the previous paragraph: the one electron approximation, the dipole approximation, the muffin tin approximations for potentials. These *ab initio* calculations have been combined with suitable fitting procedures. The importance of multiple scattering contributions, and thus the possibility to obtain stereochemical details from Multiple Scattering (MS) analysis, was realized early on [30], [31]. Initially these contributions were considered to be important only in the case of the distal atoms of imidazole/pyrrole groups (where the photoelectron is forward scattered by the intervening N atom before being backscattered by the outer-shell atom), now days it is recognized that all atoms up to 5 Å from the absorbing center may give rise to significant MS contributions (see Chapter 5 and [1], [32]). Hence, to maintain a reasonable ratio of parameters to observations, constrained and restrained refinement has been applied using the known stereochemical information [33].

In spite of the significant advances made in recent years, XAFS data analysis has not still achieved the full potential needed for an exhaustive investigation of metalloproteins. This is essentially due to two reasons. A big limitation is due to the need of good starting models which have to be known *a priori*. This often restricts the use of XAFS to the cases in which one has complementary structural information (e.g. crystallographic structure). The second main limitation consists in the presence of possible multiple solutions that are consistent with the experimental XAFS spectra

within the uncertainty even when rigid body refinement is applied. To overcome this problem ad-hoc assumptions of the DW factors can be extremely useful. There is great interest in this issue in the scientific community, as testified by recent publications [34], [35], [36], [37], [38]. Earlier approaches rely on a local equation-of-motion method [39], [34]. This implies the knowledge of the force field constants of the system under investigation that can be found in the literature. Very recently an attempt to reduce further the number of fitting parameters has been made by calculating the phonon normal modes properties on a series of small metal aminoacid model compounds at various metal first shell scattering atom distances using Density Functional Theory (DFT) [35], [36], [37], [38]. They present an approach in which the D.W.s can be parameterized with analytical expressions of the form:

$$\sigma^2(\Delta R, T) = \sigma^2(R_0, T) + A(T)\Delta R + B(T)\Delta R^2 \quad (2.18)$$

where  $R_0$  is the equilibrium distance and  $\sigma^2(R_0, T)$ ,  $A(T)$  and  $B(T)$  are third order polynomials:

$$A(T) = A(0) + A_1 \cdot T + A_2 \cdot T^2 + A_3 \cdot T^3 \quad (2.19)$$

$$B(T) = B(0) + B_1 \cdot T + B_2 \cdot T^2 + B_3 \cdot T^3 \quad (2.20)$$

$$\sigma^2(R_0, T) = \sigma^2(R_0, 0) + C_0 \cdot T + C_1 \cdot T^2 + C_2 \cdot T^3 \quad (2.21)$$

These functions can be thus used in a fitting procedure without the need to perform complex DFT calculations. The calculated coefficients for selected amino acids coordinated to zinc can be found in [36]. At the moment a complete database of these coefficient (for all amino acids and for all metals) is not available.

## 2.3 Introductory remarks about metal centres in proteins

A large fraction (up to 50 %) of known proteins is formed by metalloproteins. This definition includes all the proteins which contain metal ions either as single atom or as part of a cluster and play a variety of life sustaining roles in the bacterial, plant and animal kingdoms. The pervasiveness of metals in biology is impressive. Approximately one-third of all proteins and enzymes require metal ions as cofactors for biological function. Among them, at least five (Fe, Na, K, Mg, Ca) appear to be crucial to every known form of life, whereas the other metals (Zn, Cr, Mn, Co, Ni, Mo, W, V, Cu, Se) are required by some organisms. Endogenous metal sites can be classified into five basic types according to their function: structural-configuration of protein tertiary and/or quaternary structure; storage-uptake, binding and release of metals in soluble form (transferrins, ferritins, metallothioneins); electron transfer-uptake, release and storage of electrons (cytochromes, iron-sulphur proteins and blue copper proteins); dioxygen binding-metal- $O_2$ . coordination and decoordination (heme proteins dioxygen carriers hemoglobin and myoglobin, hemerythrins and hemocyanins); catalytic-substrate binding, activation and turnover.

These endogenous metals and their ligands constitute prosthetic groups that usually are covalently bound to the polypeptide backbone. For a thorough description of structural and functional aspects of endogenous metal sites in biology the reader is referred to [3].

Quite often proteins bind also exogenous metal ions which can interfere in their functions, decreasing or increasing their activity (inhibitors and activators). The binding of an inhibitor, which can be either reversible or irre-

versible, can stop a substrate from entering the enzyme's active site and/or hinder the enzyme from catalysing its reaction. Irreversible inhibitors react with the enzyme modifying key amino acid residues needed for enzymatic activity, whereas reversible inhibitors bind non-covalently.

Either for endogenous or for exogenous metals, the ligands are nitrogen, oxygen and sulfur, provided by amino acids side chain. In some cases also groups which are located at the C- or N-termini of the polypeptide chain ( amido, amidato, amino carbonyl, and carboxylate) have been observed to bind endogenous metals (see table below).



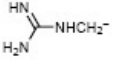
coordinating group	nomenclature <sup>a</sup> (examples)	pK <sub>a</sub> <sup>b</sup>
N-Donors		
amino: side chain	H <sub>2</sub> N <sup>+</sup> ·Lys	9–11 <sup>c</sup>
N-terminus	H <sub>2</sub> N <sup>+</sup> ·X (any residue)	
amido: backbone (–NHC(O)–)	HN <sup>+</sup> ·X (any residue)	≥ 13
side chain (–C(O)NH <sub>2</sub> )	HN <sup>+</sup> ·Asn, HN <sup>δ+</sup> ·Gln	
amidato: backbone (–N–C(O)–) <sup>–</sup>	<sup>–</sup> N·X (any residue)	
side chain (–C(O)NH) <sup>–</sup>	<sup>–</sup> N <sup>γ</sup> ·Asn, <sup>–</sup> N <sup>δ</sup> ·Gln	
imidazolyl	N·His	≥ 14 <sup>d</sup>
		
imidazolato	<sup>–</sup> N·His	
		
guanidine	H <sub>2</sub> N <sup>δ+</sup> ·Arg	> 12 <sup>e</sup>
		
O-Donors		
carbamate	O <sub>2</sub> CNH·Lys	
carboxylate: side chain	O <sub>2</sub> C <sup>γ</sup> ·Asp, O <sub>2</sub> C <sup>δ</sup> ·Glu	4–5
C-terminus	O <sub>2</sub> C·X (any residue)	
carbonyl: side chain	OC <sup>γ</sup> ·Asn, OC <sup>δ</sup> ·Gln	
backbone	OC·X (any residue)	
phenol	HO·Tyr	10
phenolate	O·Tyr	
hydroxyl	HO·X (X = Ser, Thr)	≥ 14
olate	O·X (X = Ser, Thr)	
S-Donors		
thioether	S·Met	
thiol	HS·Cys	8–9
thiolate	S·Cys	
disulfide	SS·Cys (cystine)	

Table 2.1: Coordinated ligands in metalloproteins. The table is taken from [3].

The ability of a metal ion to compete effectively with a proton in ligand binding is dictated in large measure by the strength of the metal-ligand

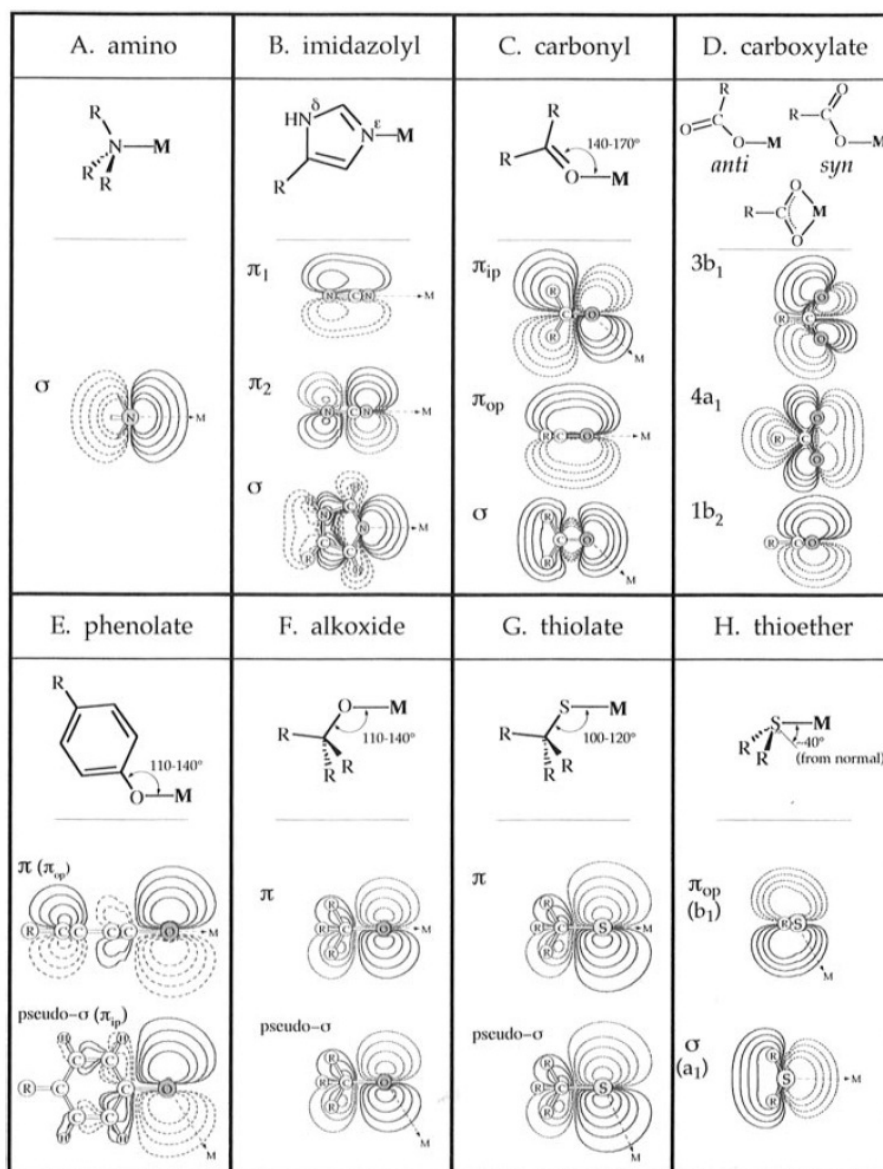


Figure 2.10: Bonding interaction for endogenous ligands. The figure is taken from [3].



bond. The metal-ligand bond is dependent on the detailed nature of the valence orbitals of the ligands as well as the effective nuclear charge and coordination number and geometry of the metal ion. Amino groups generally bind with the nitrogen and have a strong  $\sigma$ -donor interaction with the metal ion. Imidazolyl can bind through both nitrogen atoms ( $N_{\epsilon 2}$  and  $N_{\delta 1}$ ), with the metal approximately in the ligand plane and along a trigonal direction of the coordinating atom. The interaction involves mostly  $\sigma$  orbital. Amides usually coordinate metal ions through the carbonyl oxygen atom. The dominant covalent interaction involves  $\sigma$  donation. Metal ions can coordinate to carboxylate in either a bidentate or syn/anti monodentate mode. Of the monodentate possibilities, the anti arrangement is least frequently observed (see Fig...). Metal ligand bonding in metallobiomolecules is dominated by  $\sigma$ - and  $\pi$ -donor interactions with neutral ligands behaving mostly as  $\sigma$  donors. Endogenous ligands are not  $\pi$  acceptors to any significant degree.

In the present thesis we have investigated either endogenous or exogenous binding sites where the coordinated metal was iron or zinc. For some of them there was a reliable starting model (see Chapter 7) whereas for others the local structure was partly or completely unknown (see Chapter 6). In particular we have studied the following endogenous binding sites: the prosthetic heme groups of cytochrome *c* and Carboxy Myoglobin (MbCO)(see Chapter 7); the endogenous iron site of the bacterial photosynthetic reaction center (see Paragraph 3.3.2); the endogenous zinc site of cytochrome *c* oxidase (see Paragraph 6.4). We have detected the presence of an endogenous zinc binding site of complex I solving its local structure (see Paragraph 6.5). Among exogenous binding sites, we have studied: the inhibitory high affinity zinc binding site in bacterial photosynthetic reaction center pointing out

the presence of other lower affinity zinc binding sites (see Paragraph 6.2 and ref); the inhibitory zinc binding site of bacterial, avian and bovine  $bc_1$  complex (see Paragraph 6.3); the inhibitory zinc binding site of cytochrome *c* oxidase (see Paragraph 6.4). The details regarding the structure and the function of these proteins will be given in Chapter 7 and Chapter 6.

# Chapter 3

## Experimental

### 3.1 Instrumentation

#### 3.1.1 Introduction

The XAFS spectra discussed in the present work were recorded at the GILDA (General-purpose Italian Beam Line for Diffraction and Absorption) beam-line of the European Synchrotron Radiation Facility (ESRF) in Grenoble (France). A detailed description of the entire apparatus can be found in [40].

The device uses a 0.8 T Bending Magnet (BM) source operating at 6 GeV with typical currents of 200 mA. The characteristics of the beam-line are a high resolution ( $\Delta E/E \sim 10^{-4}$ ), a high flux ( $\sim 10^{11} ph/s$ ) and a few millimeter spot size (typically  $1 \times 1 mm^2$ ) on the sample. The energy range of operation is 5 – 80 keV.

#### 3.1.2 Optical apparatus

The optical apparatus is formed by a slits-filters group, a Si(311) or (111) or (511) double crystal monochromator and a pair of mirrors.

The slits-filters group performs the beam definition through a fixed slit that limits the input fan for a total horizontal acceptance of 3.6 mrad. Filters can be used in order to decrease the total power density that impinges on optical elements without adversely affecting the photon flux at the energy of interest. On GILDA the device consists in a series of 9 foils of C, Al, Cu of different thickness each optimized for several given energy intervals.

The double crystal monochromator consists in two independent Si crystals with the scattering plane perpendicular to the orbit plane of the storage ring. Both crystals rotate around two horizontal axes with mutual parallelism in order to select the proper energy, fulfilling the Bragg condition. The second crystal can move also horizontally, along the beam direction, keeping fixed the output beam height at each energy. In addition it is equipped with a sagittal bender which focuses the incoming fan at a proper distance (dynamical sagittal focusing, [41]). The rotation of the crystals is performed by a stepper motor, controlled by an optical encoder. The fine relative alignment of the two crystals is performed by a piezoelectric actuator.

Both mirrors are approximately cylindrical: the first one, placed before the monochromator, has a fixed radius of curvature and acts as a collimator; the second one, placed after the monochromator, has a variable radius of curvature and it is used to vertically focus the beam on the sample. In this way, a focused beam is available for the experiment. The mirrors also reduce the harmonic content of the beam. This is achieved using a grazing angle smaller than the critical angle for the total external reflection. In the 30-50 KeV energy range it is not possible to use the mirrors due to the extremely low incidence angles needed to obtain the total reflection<sup>1</sup>. In this energy

---

<sup>1</sup>We recall that the critical angle  $\vartheta_c$  depends on the wavelength of the impinging photons  $\lambda$ , and the sample density  $\varrho$  according to:  $\vartheta_c \propto \lambda\sqrt{\varrho}$ .

range the harmonic rejection is achieved by the classical *crystal detuning* [42], based on the fact that harmonics have a smaller Darwin width (i. e. rocking curve width) than the fundamental.

### 3.1.3 Detection modes

XAFS data acquisition on GILDA is usually performed using one of the three most common detection modes: transmission mode, electron yield and fluorescence mode. Fluorescence mode is normally employed when the sample is characterized by high dilution [24], like in our case. It is based on the isolation of the signal of the excited atom (i.e. fluorescence radiation, which originates from the relaxation of the excited state) from the total absorption of the matrix. The fluorescence photons, corresponding to a particular

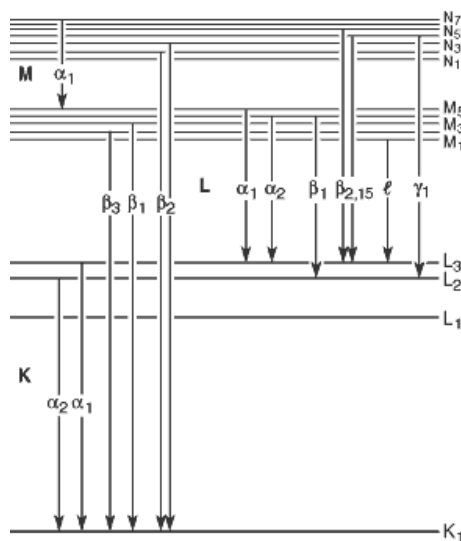


Figure 3.1: Transitions that give rise to the various emission lines. the X-ray data booklet web site: <http://xdb.lbl.gov>.

electronic transition taking place in the atoms of interest (see Fig.3.1), can be selected in energy by using an energy resolving detector. The detected fluorescence photons give a signal which is proportional to the absorption

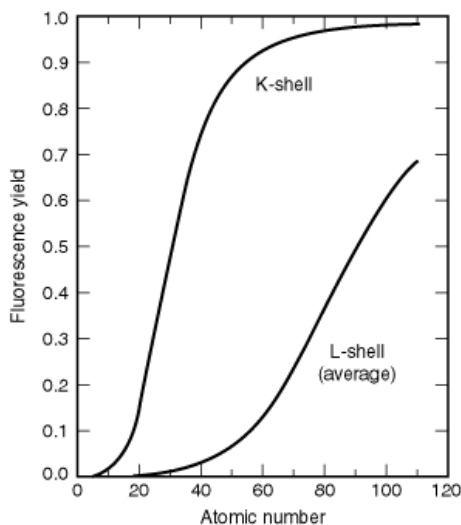


Figure 3.2: Fluorescence yields for K and L shells for. The plotted curve for the L shell represents an average of L1, L2, and L3 effective yields. This figure is taken from the X-ray data booklet web site: <http://xdb.lbl.gov>.

coefficient. In the case of interest here, an infinitely thick sample, composed of an absorbing atom  $A$  in a matrix  $M$ , and for an incidence and exit angle of  $45^\circ$ , the intensity of fluorescence radiation is:

$$I_f = \frac{I_0 \cdot \varepsilon_f(\Omega/4\pi)\mu_A(E)}{\mu_{TOT}(E) + \mu_{TOT}(E_f)} \quad (3.1)$$

Where  $E$  is the energy of the incident photon,  $E_f$  is the characteristic fluorescence energy,  $\Omega/4\pi$  is the solid angle acceptance of the detector and  $\varepsilon_f$  is the fluorescence yield, which measures the probability for emission to occur (see Fig.3.2). If the element under investigation is dilute enough, the denominator in equation does not sensibly oscillate as a function of energy and  $I_f$  is proportional to  $\mu_A$ . Fluorescence detection is popular on XAFS beamlines worldwide and has developed into a flexible experimental tool. A wide variety of x-ray detectors is available (scintillation detectors, gas ionization detectors, solid state detectors), some counting single photons, some providing only measurements of count rate or total flux, other measuring

the energy, position, and/or incidence time of each x-ray.

Our data were collected using a 13-element high-purity germanium

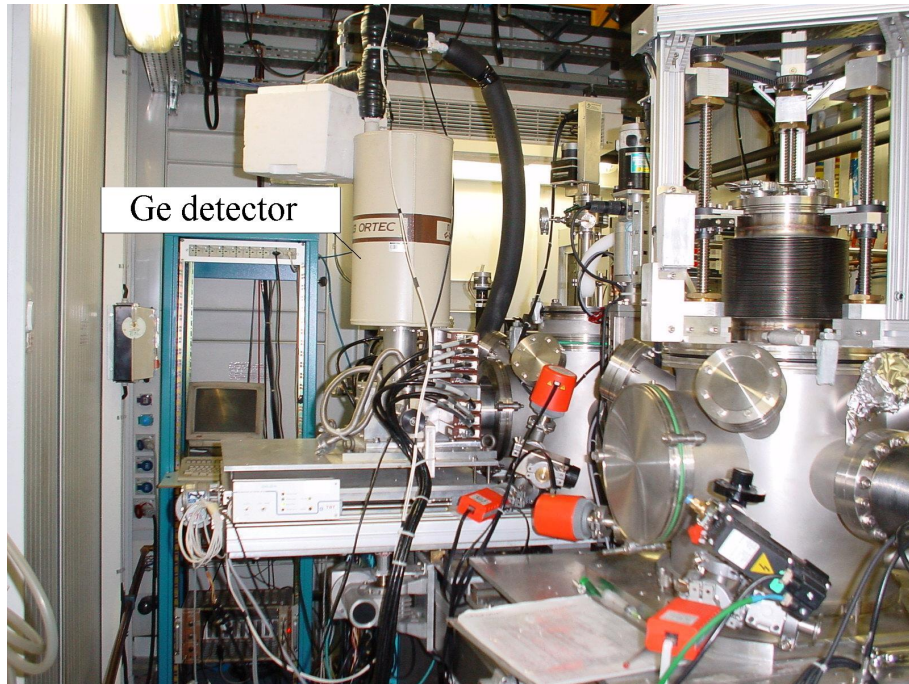


Figure 3.3: Experimental chamber with HPGe solid state detector.

model (EG&G ORTEC) equipped with fast digital electronics (see Fig. 3.3). Solid-state energy-resolving detectors can make excellent energy-resolving detectors of single photons. They are basically large, reverse-biased  $n^+ - i - p^+$  diodes. When a photon interacts in the intrinsic region, tracks of electron-hole pairs are produced. In the presence of the electric field, these pairs separate and rapidly drift to the detector contacts. The limitation of this kind of detector is the relatively high dead time of the electronics used, needing several microseconds in order to sample the voltage signal with a sufficiently low statistical noise, thus causing pulse pile-up. It has been generally recognized that, when high count rates are used, dead-time effects can

be important, the most well known effect being an artificial decrease of the amplitude of the EXAFS signal. At GILDA beamline corrections to these effects have been elaborated (see [43]). Our data have been analyzed after applying the proper corrections.

## 3.2 Sample preparation

The samples analyzed in this thesis were mostly prepared at the biochemistry laboratory, Department of Biology (University of Bologna), Italy. Bacterial bc1 was purified at the Department of Biology, University of Pennsylvania, Philadelphia, USA. Avian bc1 was purified at the Lawrence Berkeley National Laboratory, Berkeley, CA, USA. Bovine bc1 was purified at the Department of Medical Biochemistry, Biology and Physics, University of Bari, Italy. Bacterial photosynthetic reaction center was purified at the Department of Chemistry, University of Bari, Italy. The detail regarding the specific preparation of each sample can be found in the *Materials and Methods* sections of Chapters 6 and 7.

Proteins have been inserted in various environments, from solution to different solid matrices (poliviny alcohol (PVA) matrices, trehalose glasses), depending on the aim of the experiment. When the interest was the identification of the binding site (as in the case of zinc binding sites of charge translocating membrane protein complexes) we inserted the proteins into poliviny alcohol (PVA) matrices since this matrix does not interact appreciably with the protein [44] and the final films are stable and easily handled, thus particularly suitable for XAFS measurements. When we were interested specifically in the matrix effect on the local structure and dynamics of proteins, we inserted them into different environments (solution, PVA, trehalose glasses) in order to monitor the modifications induced by the sur-



rounding medium.

Since with XAFS measurements we study the average local structure and dynamics of the absorber, it is extremely important to be sure that samples are not contaminated from metals, with the same atomic number of the absorber, bound to the protein in different sites. This aspect has a great importance, especially in the case of exogenous binding sites. In fact, during the soaking processes in which metal-protein complex is formed, more than one site can, in principle, be populated. Hence it is extremely important to know the exact metal concentration. In the present work the metal concentration has been estimated using atomic absorption spectroscopy, Inductively Coupled Plasma Atomic Emission Spectroscopy (ICP-AES) and spectrophotometrical measurements. Further controls have been performed using the fluorescence counts of the 13-element high-purity Ge detector. As far as endogenous sites are concerned possible inhomogeneities in metal binding sites can arise from impurities due to the presence of other proteins containing metals. The purity of the proteins used in the present work was extensively controlled. Further details on purity, soaking process and metal concentration estimation are given in Chapters 6, 7.

### **3.3 Data acquisition**

#### **3.3.1 General considerations**

Appropriate data collection is essential in order to maximize the structure information which can be extracted from an XAFS spectrum. In particular, one of the most important requirements for accurate structural determination is to obtain XAFS data over as large a  $k$  range as possible. This im-

proves the accuracy and precision of the determination of the bond lengths, improves the resolution of shells of atoms that surround the absorbing atom at similar distance and improves the determinacy of the analysis of the three dimensional structure (see Par. 4.5). For this aim the most crucial aspects, besides the number of scans used in the average and the weighting of the count time for the data points, concerns both the samples and the experimental apparatus.

As far as experimental apparatus is concerned the intensity of the beam and the sensitivity of the detection system can largely improve the signal-to-noise ratio, thus increasing the analyzable k-range. It is important to notice that, because of the photoreduction problem (see the next paragraph), the x-ray flux should be limited to  $10^{11}$  photons/s, typical of a second-generation source or of a bending magnet beamline, like GILDA beamline. This fact makes the detector knowledge the most crucial area for the further development of the application of XAFS to dilute samples. As mentioned above we used a hyper pure Ge detector equipped with fast digital electronics able to correct for dead-time. Eq. 3.1 shows that the fluorescence counts depend on the incident photon flux  $I_0$ , on the absorption coefficient of the investigated atom  $\mu_A$ , on the probability of emission  $\varepsilon_f$  and on the geometric acceptance of the detector  $\Omega/4\pi$ . Thus considering that  $I_0 \sim 10^{11}$ ,  $\mu_A \sim 10^{-3} - 10^{-5}$ ,  $\varepsilon_f \sim 0.5$  (for Fe and Zn) and  $\Omega/4\pi \sim 0.04$ , we obtain  $I_f \sim 10^3 - 10^5$ . It is worthwhile to notice that the geometric acceptance of the detector decreases 25 times the value of the fluorescence counts. This is the reason why developments in detector area are so important.

As far as samples are concerned we recall that the concentration of the absorbing atoms  $\varrho_A$  contributes to the measured fluorescence signal through  $\mu_A = \sigma_A \varrho_A$  where  $\sigma$  is the atomic cross section.

All the aspects illustrated above are important in order to increase the signal-to-noise ratio ( $S/N$ ). Typical values of  $S/N$ , found for the spectra analyzed in this thesis, are between 0.001 – 0.003 (with total integration times between 20-100 s/point). These estimates were calculated neglecting completely possible systematic errors in the data, thus assuming that the uncertainty at point  $E_i$  is all due to random noise. The random noise is estimated by the rms value of  $\mu(E)$  over a range of high  $k$  values, where the XAFS oscillations are assumed to be indistinguishable from the random noise. This estimate was found consistent with a poissonian statistics. In Fig. 3.4 we show, as an example, the absorption spectrum collected for complex I (see Paragraph 6.5), obtained from an average of three different spectra for a total integration time of 60 s (180000 total counts). In the inset we show the  $k^3$  weighted XAFS function in the range 12-16  $\text{\AA}^{-1}$  in order to illustrate the quality of the data.

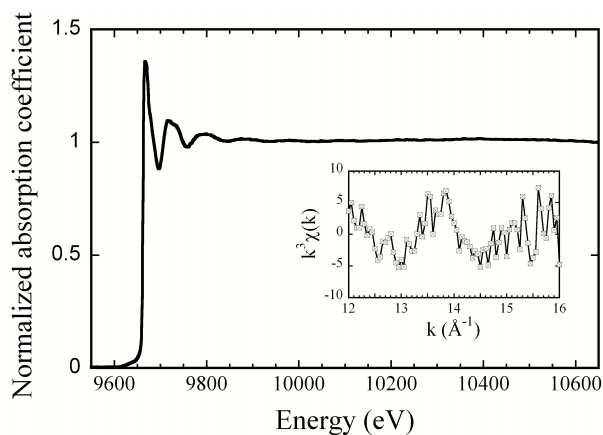


Figure 3.4: Absorption spectrum collected for complex I. Inset:  $k^3$  weighted XAFS function is shown between 14-16  $\text{\AA}^{-1}$  in order to illustrate the quality of the data.

However, in our experience, random noise does not represent the lim-

iting factor in obtaining XAFS data over a large  $k$  range. Sometimes at high  $k$  values distortions of the spectrum are present, the origin of which is not at all trivial. Some of them may be originated from inhomogeneities of the samples. Sometimes inhomogeneities of the matrices in which proteins are embedded can result in a variation of the background scattering of X-rays, thus causing abrupt discontinuities. It is thus important to align the beam in a region of the sample characterized by a high homogeneity and to control the alignment in each scan. In our experience it is useful to change frequently the investigated spot in order to avoid artificial effects due to distortions. Another thing to consider is the temperature at which measurements have to be performed. Low temperature measurements are in general desirable because of the reduction of the extent of photodamage and because of the reduction of the thermal displacements which result in a minor damping of the EXAFS signal. However, low temperature measurements in frozen solutions are extremely delicate, having two main problems: the formation of ice crystals that can result in Bragg reflections that give rise to glitches and in an increase of the background scattering of X-rays. These problems can be avoided using non-liquid sample, as for example Polyvinyl alcohol (PVA) films, which interact very weakly with the protein and in our experience have been extremely suitable for XAFS measurements.

### **3.3.2 Radiation damage**

Radiation damage is one of the most problematic area in all synchrotron based X-ray techniques applied to biological samples. It arises primarily from the production, caused by photoelectrons, of secondary lower energy electrons which can induce ionization of molecules, e.g. water. The electron that is released can reduce the absorbing atom and, hence, result in analysis

of a coordination environment that differs from that originally present. This reduction process is much more important than photooxidation, the effect of which has been found only in the case of decarboxylation of acidic residues (Glu and Asp) [45]. This is probably due to the ability of the electron to readily diffuse or tunnel through the medium to reach the absorbing atom [1]. The reduction process is particularly noticeable for metalloproteins, where most oxidized forms of metalloproteins readily show photoreduction (Cu(II), Fe(III), Mo(VI) etc.). However, most reduced forms, such as Fe(II) and Mo(IV), are extremely stable in the beam and, hence, reliable structures are more easily obtainable. Generally, photoreduction is more serious for species in solutions than in the solid state, and for metalloproteins than for coordination complexes.

The probability for the secondary effects to occur is strictly related to the x-ray dose, whereas the damage seems to be independent of the x-ray wavelength [46]. In this respect, we recall that the x-ray dose normally used in protein crystallography ( $2 \cdot 10^7 \text{grays}$ )<sup>2</sup> is much higher than that used in BIOXAS experiments. A recent study [47] has shown that the damage of the metal site occurs at a dose that is more than one order of magnitude lower than the dose that results in loss of diffractivity, which is the most common criterion used to measure x-ray damage. Thus a careful evaluation of the structural intactness of the active site(s) by spectroscopic techniques can validate structures derived from crystallography and it can be a valuable complementary method before structure-function correlations of metalloproteins can be made on the basis of high-resolution x-ray crystal structures.

In our measurements radiation damage has been extensively monitored

---

<sup>2</sup>1 gray = energy deposited per mass = 1 Joule/Kg

by performing iterated scans. As far as Zn sites are concerned no radiation damage effects were detected. As far as Fe is concerned, we performed all the measurements with the Fe in the reduced state to avoid photoreduction. However, in few cases we observed still some modifications in the spectra due to the impinging x-ray flux. Subtle modifications have been detected for the MbCO sample in solution measured at room temperature (see Fig. 3.5). Impressive changes have been detected for the iron site of the bacterial photosynthetic reaction center. In particular, the white line was greatly reduced when the protein was in solution (or in PVA film ) as compared to that observed when the protein was embedded in extra dry trehalose matrix. Preliminary analysis and XANES simulations (not shown) suggest that this decrease of the white line, together with the observed modifications of the EXAFS signal (see Fig. 3.6), was caused by the loss of at least two ligands. Interestingly this phenomenon does not occur when the protein is embedded in extra dry trehalose matrix , suggesting a particular efficacy of this sugar in preventing radiation damage. A deep investigation of the structural changes of the iron site in the two different environments is still in progress.

### **3.3.3 Estimate of metal concentration with x-ray fluorescence**

An incident beam of x-rays of energy  $E > E_0$  induces the emission of the characteristic x-rays as shown in Fig. 3.2, the emission intensity of which being dictated by Eq. 3.1. This unique signature of an element can be used to estimate the metal concentration of exogenous metals. In fact, if the protein under investigation contain an endogenous metal  $M$ , present in a fixed amount, then the ratio between the exogenous and the endogenous

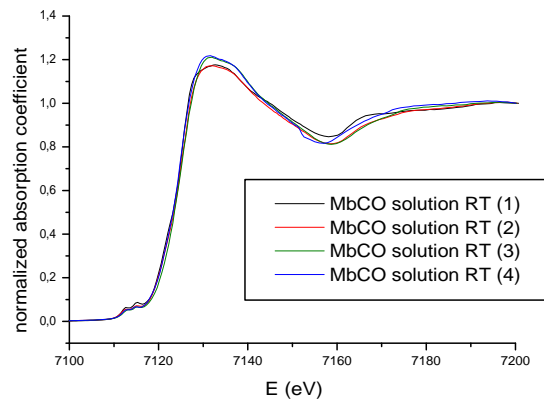


Figure 3.5: Absorption spectra collected for the MbCO in solution at 300 K. The numbers indicate the sequence of the collected spectra.

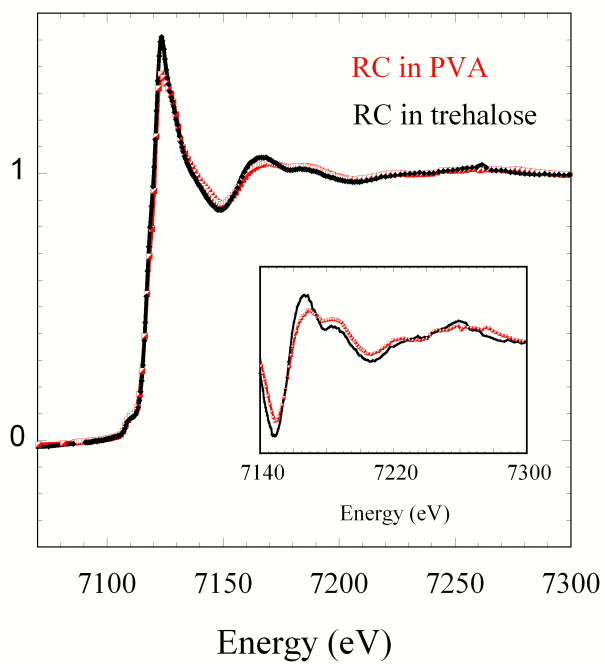


Figure 3.6: Absorption spectra collected for the iron site of bacterial photosynthetic reaction center embedded in a dry trehalose glass (in black) and in a PVA film (in red). Measurements were performed at 300 K.

metal  $m$  concentrations can be estimated through the formula:

$$[m]/[M] = \frac{A_m \sigma_M \epsilon_M}{A_M \sigma_m \epsilon_m} \quad (3.2)$$

. where the areas  $A_m$  and  $A_M$  result from the integration of the fluorescence peaks relative to the exogenous and the endogenous metal respectively. In our experiments we used x-ray fluorescence (XRF) to check the metal concentration estimated using other standard methods (ICP AES). As an example see Fig 4.1.

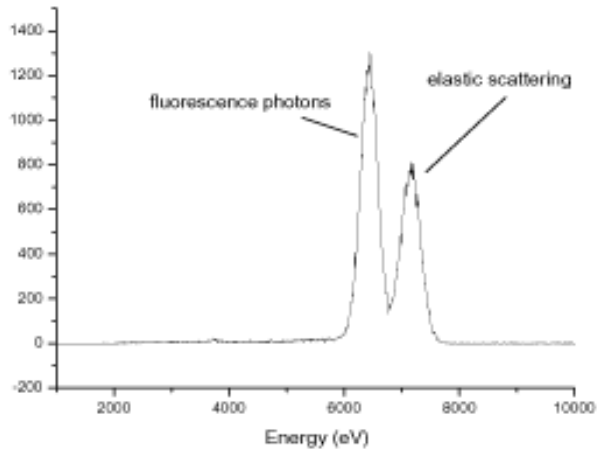


Figure 3.7: Fluorescence spectrum for MbCO.



## Chapter 4

# EXAFS data analysis

### 4.1 Introduction

The aim of an EXAFS analysis is to determine the local structural parameters by minimizing the difference between an observed EXAFS spectrum and the theoretical EXAFS calculated from a model of the absorbing site. The analysis is based on three steps: background (i. e. atomic absorption cross section) subtraction, *ab-initio* simulation of theoretical signals on the basis on a well defined starting model and multi-parameter fitting. Many Software packages are available for this analysis (UWXAFS [48], EXCURVE [28], GNXAS [29], etc..).

In our work the UWXAFS package has been used. EXAFS oscillations were extracted from the raw data using the AUTOBK program [49] as implemented in the ATHENA package [50], using a linear function for the pre-edge region and a cubic spline to mimic the atomic background. Theoretical amplitude and phase shift functions were calculated using the *ab-initio* code FEFF 8.2 [27]. The data were analyzed using the FEFFIT

program as implemented in ARTEMIS package (version 0.8.006) [50] using as minimization algorithm a modified Levenberg-Marquardt method.

Fits were performed directly in k-space to avoid truncation and loss of information due to fourier filtering.

Fourier filtering was used mainly to check the correctness of the background subtraction and to inspect the data in a qualitative way.

One of the crucial features in obtaining accurate three-dimensional structures is to minimize the number of variables with respect to the number of independent data points. This is achieved by placing constraints and restraints on parameters that are fitted in the calculations. In the present thesis constraints were often used within the rigid body refinement approach.

The analysis of EXAFS can yield the number and types of backscattering atoms, accurate absorber-backscatterer distances and disorder parameters (Debye-Waller factors). When multiple scattering effects are important, also some angular parameters can be obtained.

## 4.2 Background subtraction

Background subtraction is carried out through three steps: a pre-edge background removal using a linear function; a normalization to the edge jump; a post-edge background removal. The last procedure consists in subtracting from  $\mu(E)$  a smoothly varying background function  $\mu_0(E)$ , which approximates the absorption from the isolated embedded atom. The function used to approximate the post-edge background is a third-order polynomial spline whose knots are equally spaced in k space, the flexibility of the spline being determined by the number of knots. The spline and its first two derivatives have to be continuous at the knots and one degree of freedom is associated

to each knot; the background function is not required to go through the experimental curve at the knots. The basic idea of the program is that no signal in the Fourier transform of the EXAFS function can have a physical meaning below a quite small threshold value of  $R$ , so that the free coefficients of the spline are chosen in order to minimize the  $R$  components below a fixed  $R$  value, named  $R_{bkg}$ . This means that the method chooses the spline to best fit only the low-frequency components of  $\mu(E)$ . The value of  $R_{bkg}$  can be chosen by the user, a good choice being about half the distance of the first shell peak, and determines the maximum number of knots i.e. the stiffness of the spline. From information theory the number of independent points for a sinusoidal function is [48]:

$$N_{ind} = \frac{2\Delta R \Delta k}{\pi} + 2 \quad (4.1)$$

where  $\Delta R$  and  $\Delta k$  are the  $k$  and  $R$  ranges of the analyzed data. If the fit is performed from  $R = 0$  to  $R = R_{bkg}$ , as in the case of the background removal, the number of independent points is thus:

$$N_{ind} = \frac{2R_{bkg} \Delta k}{\pi} + 2 \quad (4.2)$$

Since each knot is associated with one degree of freedom, the maximum number of knots is the integer part of  $N_{ind}^{bkg}$  in the previous equation; the program uses a number of knots equal to  $N_{ind}^{bkg} - 1$ . The extracted EXAFS function is usually multiplied by a factor  $k^n$  to compensate for the rapid attenuation of the EXAFS with increasing  $k$ . The arbitrary exponent  $n$  is called the  $k$ -weighting.

### 4.3 Fourier filtering

The Fourier transform of a  $k$ -weighted EXAFS spectrum is a complex function of the distance thus it is related to a radial distribution function. How-

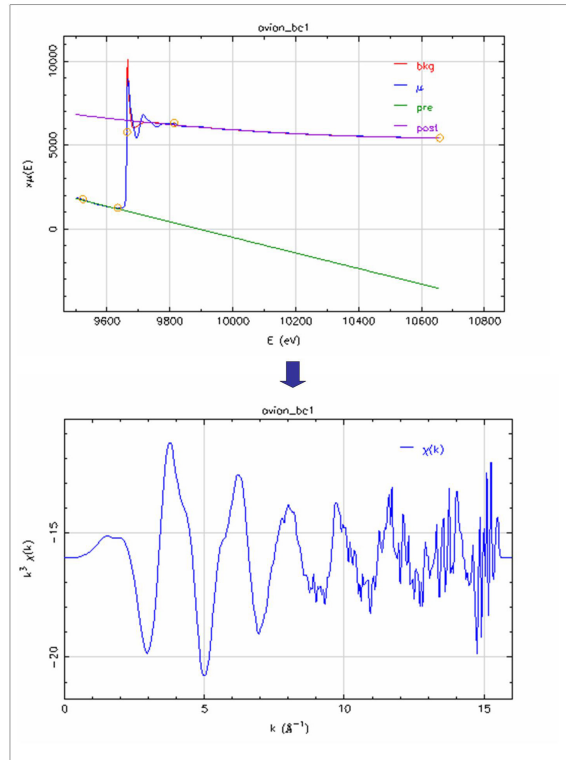


Figure 4.1: Background subtraction for avian  $bc_1$ .

ever, because of the presence of a phase factor, depending from both the central and the backscattering atoms, the Fourier transform does not peak exactly at the distances of the backscattering atoms. A 'phase' correction relating the peak positions to the true interatomic distances is required. This correction depends on the wave number. The Fourier transform acts as a frequency filtering of the original EXAFS by applying a window function prior to the forward or inverse Fourier transformation. It is thus possible to eliminate noise or unwanted contributions of the EXAFS. However, the use of filtering may distort the EXAFS contribution of one or more scatterers. For this reason in the present thesis we prefer not to use fourier filtering in quantitative analysis.

## 4.4 Simulations of theoretical signals using FEFF8.2

In order to analyze the extracted EXAFS function we need to build a starting model, on the basis on which a *ab-initio* simulation of the absorption cross section is performed. FEFF 8.2 code [27] implements a Green function approach according to Eq. 2.11. A Dirac-Fock atomic code [51] is used to obtain free electronic densities, which are successively overlapped (following the Mattheiss prescription [52] in order to obtain a total electron density and to construct Coulomb potential. Then the exchange-correlation potential is added. In FEFF8.2 there are five different options. By default the program uses a ground-state Von Barth-Hedin potential [53] and the Hedin-Lundqvist (HL) energy-dependent self-energy correction [54]. The total potential is approximated with a Muffin Tin (MT) one (see Paragraph 2.1), which is flat in the interstitial region and otherwise it is equal to the sum of overlapping, spherically symmetric potentials:

$$V_{tot}(\vec{r}) = V_{mt} + \sum_i V_i(|\vec{r} - \vec{R}_i|) \quad (4.3)$$

The MT radii, determined using the Normann prescription, are kept fixed. The Dirac equation in the spinor relativistic approximation is then solved [55] to obtain radial wave functions and partial wave shifts, that are necessary to construct the single particle Green function (see Eq. 2.8). Finally, the Green function is calculated in the multiple scattering approach: paths with negligible amplitude are rejected and amplitude and phases for all paths inside a definite cluster (the radius of which can be chosen by the user) and formed by a number of legs  $\leq n$  where  $n$  can be chosen by the user, are recorded in different files as a function of the energy. It is possible to insert self consistent corrections in the calculation. This is achieved adding the Self Consistent Field (SCF) card in the input file. If this card is inserted,

the program implements an iterative procedure of the calculations of the potentials. In the initial iteration the radii are fixed according to the Norman criterium. On the basis on these radii a charge density and the corresponding fermi energy are calculated. Then the local fermi level at each site by requiring no charge transfer is computed and a new density, including charge transfer is achieved. Self consistency is reached when the two fermi levels and the two densities are equal. For XANES simulations it is necessary to insert the full multiple scattering (FMS) card besides the SCF one. This consists in the numerical inversion of the matrix  $I + GT$  (see Eq. 2.11).

## 4.5 Multi-parameter fitting

### 4.5.1 Introduction

The aim of the fitting procedure is to optimize the match between the calculated and the observed EXAFS ( $\chi_{calc}$  and  $\chi_{obs}$  respectively) by varying the free parameters. This is done by minimizing the difference between the two EXAFS functions multiplied by the weight  $k^n$ ,

$$\sum_1^N [k^n(\chi_{obs}(k) - \chi_{calc}(k))]^2 dk. \quad (4.4)$$

Where the sum is over the experimental points  $N$ . In FEFFIT the minimization is performed using a modified Levenberg-Marquardt method. The variables commonly used are: a variation in the energy origin ( $\Delta E_0$ ), a number of distance variations to account for structural modification of the starting model, and some DW factors, properly grouped in shells, to account for the effect of thermal motion of the atoms and/or structural disorder. In some cases can be extremely convenient to fix *a priori* the DW factors to avoid local minima (see the next Chapter). When multiple scattering effects are important, also angular variables can be inserted (see the next Chapter).

The amplitude reduction factor can be either free or fixed to a reasonable value (for example, calculated from FEFF8.2 from atomic overlap integrals). In the present thesis we used the value calculated by FEFF8.2 as a set parameter. The size of the  $k$  interval in which the fit is performed is important and it is normally chosen to be as large as the data reasonably allows. The upper limit is chosen on the basis of noise (or hypothetical distortions) in the XAFS whereas the lower limit is chosen in order to eliminate the region in which the EXAFS series is not convergent. It has been demonstrated [56] that the convergence of the path summation gives the same result as the matrix-inversion method with terms higher than about fifth order only being required near a strong low-energy resonance (see also [7]). Thus, in principle, we should be able to reproduce the entire spectrum by using all the contributions until the fifth order; however, the lower-energy part of the EXAFS function for metalloproteins is difficult to be correctly reproduced due to the inadequacy of the spherically symmetric muffin-tin approximation to the potentials [7], [8].

For the same reason for metalloproteins it is convenient to choose a  $k$ -weighting equal to three. It is worthwhile to notice that in doing this we choose to optimize the match between the calculated and the observed EXAFS (see Eq. 4.4) in such a way that the discrepancies between the model and the data are considered more important for high  $k$  values. This means that we are privileging some structural parameters, since different parts of the spectrum are sensitive to different structural parameters. In particular the simulation shown in Fig. 5.2) shows that the MS paths involving different residues, thus containing information about three-dimensional structure of the site, have a contribution only in the low  $k$  region of the spectrum.

Another parameter to be taken into account is the number of legs that

are included into the fit of the data. For most systems some paths involving up to five legs make significant contributions (say 20 % of the importance of the most important scattering paths) thus the cut-off of five-leg paths is normally the compromise made in order to minimize the calculation time while still having a very good fit to the experimental data. While the inclusion of multiple scattering paths often does not affect the accuracy of the determination of the bond lengths compared to single scattering analysis, it is essential in order to provide meaningful information on the nature of the donor groups.

#### **4.5.2 Constrained and restrained refinement**

A crucial factor in the fitting procedure is to ensure that there are more independent data points than there are variables. Thus it is extremely important to minimize the number of variables. The minimization is achieved by using constraints and restraints on structural and dynamical parameters. Constraints are not allowed to vary during the fitting procedure, whereas restraints are placed on parameters such that a bond length or angle is allowed to vary only within a reasonable value that is based on structural information from other sources. In the present thesis only constrained refinement has been applied. It is obvious that this procedure is extremely successful when the system is highly symmetric and constraints are used to make equivalent or very similar bond lengths and bond angles identical (for example, the heme structure). However it is possible to use it also in an extremely asymmetric metal center in proteins by treating the aminoacid residues as rigid units (see the next chapter). In principle, also the Debye-Waller factors can be calculated and included in the fitting procedure as constraints in order to reduce the number of variables further. Actually this



is quite complicated since both static and dynamic disorder are described by the same damping factor. The static disorder is in general unknown and cannot be calculated. Dynamic disorder in molecular system can be calculated as described in section 2.2.2.

### 4.5.3 Determinacy

The determinacy of the system is defined as  $N_i/p$  ([1]), where  $p$  is the number of the fitted parameters and  $N_i$  is the estimated number of independent data points collected in the XAFS data set (see Eq. 4.1) . Determinacy in MS analysis of XAFS has a different meaning to the ratio of observations versus variables in X-ray crystallographic refinements. In fact XAFS analysis, at variance with XRD, relies on robust starting models, incorporating information from a variety of sources (for example bond lengths and angle parameters of the amino acids residues that has been obtained from XRD structures of related proteins, as we will see in the next Chapter). For this reason a high ratio of observables/variables for a MS XAFS analysis would be two (and typically much lower) [1].

In the MS XAFS analyses performed in this thesis the reliability of the fitting procedure has been assured by a high determinacy of the system ( $N_i/p$  between 2 to 9).

### 4.5.4 Statistical significance

Three numbers can be used to evaluate the goodness of a fit:  $\chi^2$ , reduced  $\chi^2$  (which we will indicate with  $\chi_r^2$ ) and the  $R$  factor.

$\chi^2$  and  $\chi_\nu^2$  are defined respectively as:

$$\chi^2 = \frac{N_{ind}}{N} \sum_{i=1, N} \left[ \frac{f_i^{exp}(k) - f_i^{theor}(k)}{k^n \cdot \sigma_i} \right]^2 \quad (4.5)$$

and

$$\chi_\nu^2 = \frac{1}{\nu} \cdot \chi^2 \quad (4.6)$$

where  $f_i^{exp}$  and  $f_i^{theor}$  are the experimental and simulated EXAFS functions multiplied by the factor  $k^n$  and  $n$  is the  $k$ -weighting (see Paragraph 4.2);  $\sigma_i$  is the uncertainty at point  $E_i$ ;  $N_{ind}$  is the number of independent points (see Eq. 4.1);  $\nu$  is the number of the degrees of freedom of the system, defined as  $N_{ind} - p$  ( $p$  being the number of the free parameters). The pre-factor  $N_{ind}/N$  in the expression of  $\chi^2$  is introduced in order to remove the arbitrariness of the calculation of the number of evaluations  $N$  in which the sum is performed.

The R factor is defined as:

$$\frac{\sum_{i=1}^N \{ [Re(f_i^{exp}(k) - f_i^{theor}(k))]^2 + [Im(f_i^{exp}(k) - f_i^{theor}(k))]^2 \}}{\sum_{i=1}^N \{ [Re(f_i^{exp})]^2 + [Im(f_i^{exp})]^2 \}} \quad (4.7)$$

This goodness of fit index is currently used (see the FEFFIT manual, [57]), since the estimation of the uncertainty  $\sigma^2$  in EXAFS data is notoriously problematic, making the use of chi-square statistic usually meaningless. In fact, by assuming that the uncertainty is all due to random noise,  $\chi_\nu^2$  values much larger than 1 are usually obtained, also in the presence of good fits (i. e. at relatively low values of the R factor). This situation is attributed to inadequacies of the model and/or to systematic experimental errors [48], [58].

However, the R factor is not a proper statistical measure, as it doesn't incorporate any measure of data quality. This makes the comparison between

two fits in terms of R factor extremely delicate, since we cannot associate a standard deviation to this goodness of fit index. Usually to compare two fits the chi-square statistics is thus used [58]. In view of the difficulties of the estimation of the uncertainty in the data, the standard fluctuation in  $\chi_\nu^2$  ( $\sqrt{\frac{2}{\nu}}$ ) is rescaled to  $\sqrt{\frac{2}{\nu}} \cdot \chi_\nu^2$  [58]. Then the comparison between two different fits of the same data set is performed by means of  $\chi_\nu^2$  according to the following criterion. Fit to cluster  $b$  is considered significantly better than fit to cluster  $a$  when:

$$(\chi_\nu^2(a) - \chi_\nu^2(b)) \geq 2\sqrt{2\left[\frac{(\chi_\nu^2(a))^2}{\nu(a)} + \frac{(\chi_\nu^2(b))^2}{\nu(b)}\right]}. \quad (4.8)$$

We have followed these criteria to compare the fits obtained for the different starting models in Chapter 6. It is worthwhile to notice, however, that when the k-weighting is bigger than 0 the R factor and the  $\chi_\nu^2$  are no longer proportional. As a consequence, the fit is obtained by minimizing the R factor (thus privileging the high k values region of the spectrum) while the comparison is done weighting in the same way the whole fitting range.

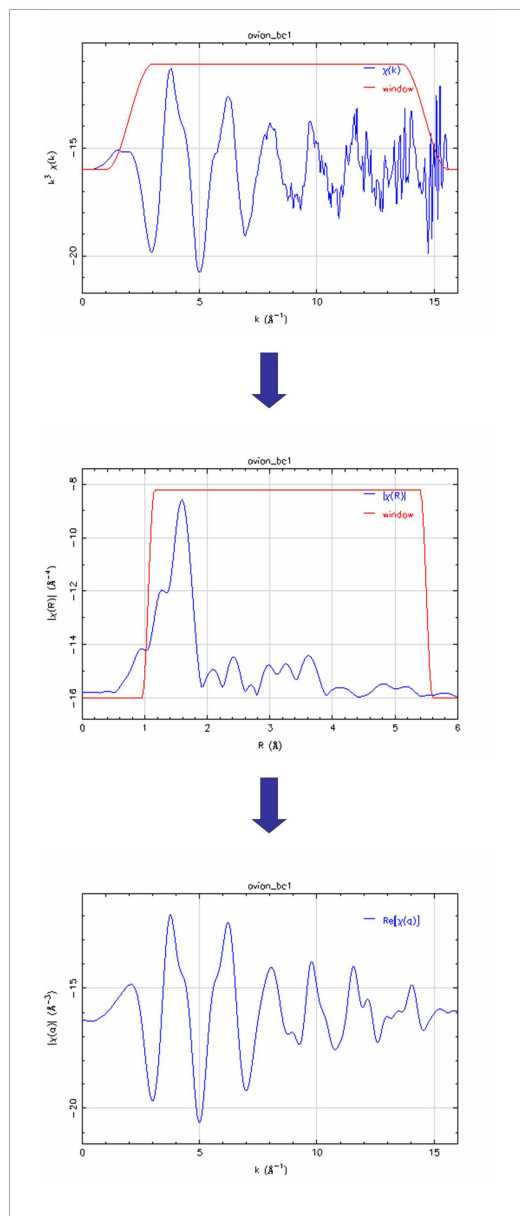


Figure 4.2: Fourier filtering applied on avian  $bc_1$  EXAFS function. Here we used two hanning windows (sills:  $dk=2$  and  $dr=0.2$  respectively).

## Chapter 5

# Data analysis strategies for metalloproteins

### 5.1 Introduction

In the present thesis we have used different analysis strategies depending on the aim of the experiment.

When the goal was to discover subtle structural distortions and changes in the dynamical properties of a known local structure produced by different environments, we used a well defined starting model (see chapter 7). Using this model we tried either to quantify the difference in structural parameters between the reference structure and the structure under investigation or to extract dynamical information on the metal site in proteins embedded in different matrices. In doing this, we needed to release both structural and dynamical parameters in the fitting procedure. In this case, to avoid local minima, we used a step-by-step method in which structural and dynamical parameters were alternatively fixed. This procedure has been used elsewhere

(see [32] and [59]). The details regarding this method can be found in chapter 7.

A completely different approach has been used to identify unknown metal clusters, i.e. to identify the number and the nature of the metal coordinating amino-acids. In that case the main problem was how to select, within a very numerous number of possibilities, a limited number of starting models to be fitted to the experimental data. We developed a general analysis strategy to obtain the structure of an unknown metal site in a protein based on EXAFS data. The only required information are the bond-length and angle parameters of amino acids, which can be derived from Engh and Huber [4] and some knowledge about the stereochemistry of metal sites in proteins that can be extracted from an extensive compilation of metal coordination sites in proteins (<http://tanna.bch.ed.ac.uk/> and <http://metallo.scripps.edu/>). This compilation is derived from protein crystal structures in the Protein Data Bank(PDB) which have been determined by diffraction methods. Target interatomic distances are based partly on distances in the Cambridge Structural Database (CSD), and partly on those in protein structures determined at or near atomic resolution. The method is explained in detail in the next paragraph.

## **5.2 Identification of unknown metal sites in proteins using EXAFS data**

### **5.2.1 Description of the method**

The method is based on:

- 1) A preliminary first shell analysis;

2) The building of a series of clusters made of all combinations of amino-acid residues compatible with both EXAFS first shell results and coordination chemistry information;

3) *Ab-initio* simulations of the selected clusters;

4) A series of fits to the experimental data using rigid body refinement and fixing the DW according to DFT or to an efficient local equation-of-motion method (see Paragraph 2.2).

The method was tested using two zinc sites in two metalloproteins whose structure was known from XRD (the inhibitory zinc binding site of the bacterial photosynthetic reaction center and the endogenous zinc binding site of cytochrome *c* oxidase). Subsequently, it was used to identify zinc sites in charge translocating membrane protein complexes of the respiratory chain as described in the next chapter.

### 5.2.2 Model building

To build a set of possible starting model(s) for the metal cluster we use the information provided by both first shell analysis and metal coordination database. First of all we need some information about the first neighbours. First shell analysis can provide the coordination number of the ligands and their atomic number together with their distances from the absorber. To build the starting models for the first shell analysis it is sufficient to select the possible ligands, their number and their geometry by looking at the database and then combine them in  $N$  possible clusters using as starting distances those found in the database. If this leads to a very large number of possibilities one can perform preliminary fits in order to exclude some combinations (for example not distinguishing, in a first step, between  $N$

and O atoms). When one has the information about the first neighbours of the metal atom each ligand can be associated with one or more amino acid residues by examining the metal coordination sites database. Having selected the residues possibly present in the binding clusters, the coordinates of all atoms have to be set. This is done accordingly to the following criteria: as first neighbours distances one can use the values found by first shell analysis; the intra-ligand distances and angles can be fixed following the values reported by Engh and Huber [4] which are the values normally used for PX refinement (see tables below). These values have been derived from a

Table 1. CSD parametrization atom types

Atom type	Description
C	Carbonyl C atom of the peptide backbone
CSW*	Tryptophan C <sup>γ</sup>
CW*	Tryptophan C <sup>δ2</sup> , C <sup>ε1</sup>
CF*	Phenylalanine C <sup>γ</sup>
CY*	Tyrosine C <sup>γ</sup>
CY2*	Tyrosine C <sup>ε</sup>
CS*	Histidine C <sup>γ</sup>
CN*	Neutral carboxylic acid group C atom
CH1E	Tetrahedral C atom with one H atom
CH2E	Tetrahedral C atom with two H atoms (except CH2P, CH2G)
CH2P*	Proline C <sup>γ</sup> , C <sup>δ</sup>
CH2G*	Glycine C <sup>α</sup>
CH3E	Tetrahedral C atom with three H atoms
CR1E	Aromatic ring C atom with one H atom (except CR1W, CRH, CRHH, CR1H)
CR1W*	Tryptophan C <sup>ε2</sup> , C <sup>η2</sup>
CRH*	Neutral histidine C <sup>ε1</sup>
CRHH*	Charged histidine C <sup>ε1</sup>
CR1H*	Charged histidine C <sup>δ2</sup>
N	Peptide N atom of proline
NR	Unprotonated N atom in histidine
NP	Pyrrole N atom
NH1	Singly protonated N atom (His, Trp, peptide)
NH2	Doubly protonated N atom
NH3	Triply protonated N atom
NC2	Arginine N <sup>η1</sup> , N <sup>η2</sup>
O	Carbonyl O atom
OC	Carboxyl O atom
OH1	Hydroxyl O atom
S	S atom
SM*	Methionine S atom
SH1E	Singly protonated S atom

\* Atom types marked with an asterisk are new (non-XPLOR) types.

Table 5.1: CSD parametrization atom type. This Tab. is taken from [4].

statistical survey of x-ray structure of small compounds from the CSD. The 3-D geometries are set accordingly to those reported for metal binding sites determined by high resolution protein crystallography. Normally the geom-



Bond type	$\sigma$	Bond length ( $\text{\AA}$ )
C5W-CW	0.018	1.433
CW-CW	0.017	1.409
C-CH1E	0.021	1.525
C5-CH2E	0.014	1.497
C5W-CH2E	0.031	1.498
CF-CH2E	0.023	1.502
CY-CH2E	0.022	1.512
C-CH2E	0.025	1.516
CN-CH2E	0.019	1.503
C-CH2G	0.018	1.516
C5W-CR1E	0.025	1.365
CW-CR1E	0.016	1.398
CW-CR1W	0.021	1.394
CF-CR1E	0.021	1.384
CY-CR1E	0.021	1.389
CY2-CR1E	0.024	1.378
C5-CR1H	0.011	1.354
C5-CR1E	0.011	1.356
C-N	0.016	1.341
C-NC2	0.018	1.326
C5-NH1	0.011	1.378
CW-NH1	0.011	1.370
C-NH1	0.014	1.329
C-NH2	0.021	1.328
C5-NR	0.017	1.371
C-O	0.020	1.231
CN-O	0.023	1.208
C-OC	0.019	1.249
CY2-OH1	0.021	1.376
C-OH1	0.022	1.304
CH1E-CH1E	0.027	1.540
CH1E-CH2E	0.020	1.530
CH1E-CH3E	0.033	1.521
CH1E-N	0.015	1.466
CH1E-NH1	0.019	1.458
CH1E-NH3	0.021	1.491
CH1E-OH1	0.016	1.433
CH2E-CH2E	0.030	1.520
CH2P-CH2E	0.050	1.492
CH2P-CH2P	0.034	1.503
CH2E-CH3E	0.039	1.513
CH2P-N	0.014	1.473
CH2G-NH1	0.016	1.451
CH2E-NH1	0.018	1.460
CH3E-NH1	0.018	1.460
CH2E-NH3	0.030	1.489
CH2E-OH1	0.020	1.417
CH2E-S	0.020	1.822
CH2E-SM	0.034	1.803
CH2E-SH1E	0.033	1.808
CH3E-SM	0.059	1.791
CR1E-CR1E	0.030	1.382
CR1E-CR1W	0.025	1.400
CR1W-CR1W	0.019	1.368
CR1E-NH1	0.021	1.374
CRH-NH1	0.020	1.345
CRHH-NH1	0.010	1.321
CR1H-NH1	0.011	1.374
CRH-NR	0.013	1.319

Table 5.2: Bond-length parameters derived from a statistical survey of X-ray structures of small compounds from the Cambridge Structural Database according to Engh and Huber. This Tab. is taken from [4].

Angle type	$\sigma$	Angle (°)	Angle type	$\sigma$	Angle (°)
C5W-CW-CW	1.2	107.2	CH3E-CH1E-CH3E	2.2	110.8
CW-C5W-CH2E	1.4	126.8	CH3E-CH1E-NH1	1.5	110.4
C5W-CW-CR1E	1.0	133.9	CH3E-CH1E-OH1	2.0	109.3
CW-CW-CR1E	1.0	118.8	C-CH2E-CH1E	1.0	112.6
CW-CW-CR1W	1.0	122.4	C5-CH2E-CH1E	1.0	113.8
CW-C5W-CR1E	1.6	106.3	CF-CH2E-CH1E	1.0	113.8
CW-CW-NH1	1.3	107.4	C5W-CH2E-CH1E	1.9	113.6
CH1E-C-N	1.5	116.9	CY-CH2E-CH1E	1.8	113.9
CH1E-C-NH1	2.0	116.2	C-CH2E-CH2E	1.7	112.6
CH1E-C-O	1.7	120.8	C-CH2G-NH1	2.9	112.5
CH1E-C-OC	2.5	117.0	C-CH2G-NH3	2.9	112.5
CH2E-C5-CR1E	1.3	129.1	CH1E-CH2E-CH1E	3.5	116.3
CH2E-C5-CR1H	1.3	131.2	CH1E-CH2E-CH2P	1.9	104.5
CH2E-CF-CR1E	1.7	120.7	CH1E-CH2E-CH2E	2.0	114.1
CH2E-C5W-CR1E	1.5	126.9	CH1E-CH2E-CH3E	2.1	113.8
CH2E-CY-CR1E	1.5	120.8	CH1E-CH2E-OH1	2.0	111.1
CH2E-C-N	2.1	118.2	CH1E-CH2E-S	2.3	114.4
CH2G-C-N	2.1	118.2	CH1E-CH2E-SH1E	2.3	114.4
CH2E-C5-NH1	1.5	122.7	CH2E-CH2E-CH2E	2.3	111.3
CH2E-C-NH1	2.1	116.5	CH2E-CH2P-CH2P	3.2	106.1
CH2G-C-NH1	2.1	116.4	CH2P-CH2P-N	1.5	103.2
CH2E-C-NH2	1.5	116.4	CH2E-CH2E-NH1	2.2	112.0
CH2E-C5-NR	1.5	121.6	CH2E-CH2E-NH3	3.2	111.9
CH2E-C-O	2.0	120.8	CH2E-CH2E-SM	3.0	112.7
CH2G-C-O	2.1	120.8	CY2-CR1E-CR1E	1.8	119.6
CH2E-C-OC	2.3	118.4	CW-CR1E-CR1F	1.3	118.6
CH2G-C-OC	2.3	118.4	CW-CR1W-CR1W	1.3	117.5
CR1E-CY2-CR1E	2.0	120.3	CF-CR1E-CR1E	1.7	120.7
CR1E-CY-CR1E	1.5	118.1	CY-CR1E-CR1E	1.5	121.2
CR1E-CF-CR1E	1.5	118.6	C5-CR1E-NH1	1.0	106.5
CR1W-CW-NH1	1.5	130.1	C5-CR1H-NH1	1.0	107.2
CR1E-C5-NH1	1.0	105.2	C5W-CR1E-NH1	1.3	110.2
CR1H-C5-NH1	1.0	106.1	C5-CR1E-NR	2.3	109.5
CR1E-CY2-OH1	3.0	119.9	CR1E-CR1E-CR1W	1.3	121.1
N-C-O	1.4	122.0	CR1W-CR1W-CR1E	1.3	121.5
NC2-C-NC2	1.8	119.7	CR1E-CR1E-NH1	1.8	120.0
NC2-C-NH1	1.9	120.0	NH1-CRHH-NH1	1.0	108.4
NH1-C-O	1.6	123.0	NH1-CR1E-NR	1.3	111.7
NH2-C-O	1.0	122.6	C-N-CH1E	5.0	122.6
OC-C-OC	2.4	122.9	C-N-CH2P	4.1	125.0
C-CH1E-CH1E	2.2	109.1	CH1E-N-CH2P	1.4	112.0
C-CH1E-CH2E	1.9	110.1	C-NH1-CH1E	1.8	121.7
C-CH1E-CH3E	1.5	110.5	C-NH1-CH2G	1.7	120.6
C-CH1E-N	2.5	111.8	C-NH1-CH2E	1.5	124.2
C-CH1E-NH1	2.8	111.2	C-NH1-CH3E	1.7	120.6
C-CH1E-NH3	2.8	111.2	C5-NH1-CRHH	1.7	109.3
CH1E-CH1E-CH2E	1.7	110.4	C5-NH1-CRHH	1.7	109.0
CH1E-CH1E-CH3E	1.7	110.5	CW-NH1-CR1E	1.8	108.9
CH1E-CH1E-NH1	1.7	111.5	CRHH-NH1-CR1H	1.0	109.0
CH1E-CH1E-OH1	1.5	109.6	CRHH-NH1-CR1E	1.3	106.9
CH2E-CH1E-CH3E	3.0	110.7	C5-NR-CR1E	1.0	105.6
CH2E-CH1E-N	1.1	103.0	CR1E-NR-CR1E	3.0	107.0
CH2E-CH1E-NH1	1.7	110.5	CH2E-SM-CH3E	2.2	100.9
CH2E-CH1E-NH3	1.7	110.5	CH2E-S-S	1.8	103.8

Table 5.3: Angle parameters derived from a statistical survey of X-ray structures of small compounds from the Cambridge Structural Database according to Engh and Huber.

This Tab. is taken from [4].

etry is tetrahedral for a coordination number of four, trigonal bipyramidal or square based pyramidal for a coordination number of five and octahedral for a coordination number of six. Distorsions from these ideal geometries can be found, however the EXAFS signal is weakly sensitive to this parameter due to the weakness of the multiple scattering paths involving different residues. As far as carboxylate groups are concerned two different coordination modes, i.e. monodentate and bidentate binding have to be considered. Also in the case of histidine residues two configurations have to be taken into account, depending on which nitrogen atom ( $N_{\delta 1}$  or  $N_{\epsilon 2}$ ) binds the metal. In our case, however, it turned out that it can be difficult to discriminate between these two binding modes during the fitting procedure. To the contrary the change between monodentate and bidentate arrangement of carboxylate groups produce a big difference in the EXAFS signal due to the change in distance of the  $O_{\delta 2}$  atom. As far as the orientation of the aminoacid residues with respect to the metal-first ligand direction is concerned the values found in the Metal data bank (MDB) (<http://metallo.scripps.edu/>) database can be used. Since this parameter can have a great influence in the EXAFS signal we suggest to allow them to vary when a large spreading of these values is encountered in crystallographic structures (see below).

We have used MOLDRAW (P. Ugliengo, D. Viterbo and G. Borzani J. Appl. Cryst. 21, 75 (1988) "MOLDRAW: Program for the Graphical Manipulation of Molecules on Personal Computers") to build the clusters following the procedure explained above. This program can read FEFF files and it can export directly to a FEFF file the atomic coordinates.

### 5.2.3 *Ab-initio* self consistent simulations

We use FEFF8.2 (see previous Chapter) to perform *ab-initio* simulations. We perform the calculation of multiple scattering paths until five scattering legs following [1]. The amplitude reduction factor can be calculated by FEFF8.2, by atomic overlap integrals and can be kept fixed during the fitting procedure. The DW can be fixed on the basis on DFT (see Paragraph 2.2 and [36]) or calculated using the SIGEM program as implemented in FEFF8.2, based on local equation-of-motion method (see Paragraph 2.2). For the analysis of the enzymes of the respiratory chain (see Chapter 6) we have used a simplified model obtained by grouping the values found on the basis on DFT at room temperature and for an equilibrium structure in shells of similar atoms. We show, as an example, the simulation obtained from a

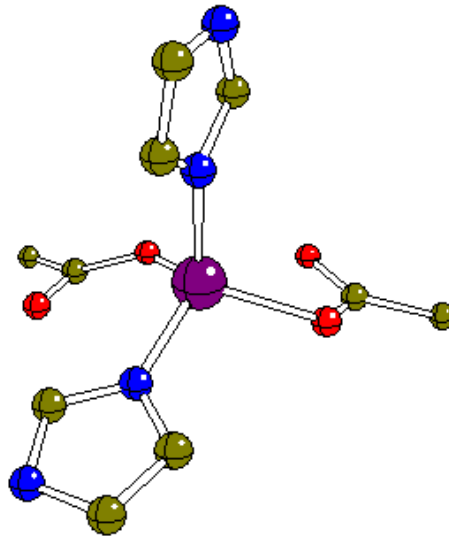


Figure 5.1: Tetrahedral cluster used for the simulation shown in Fig. 5.1. The first shell distances are: Zn-N = 2.00 Å; Zn-O = 2.06 Å. The others bond length distances and angles have been set according to [4].

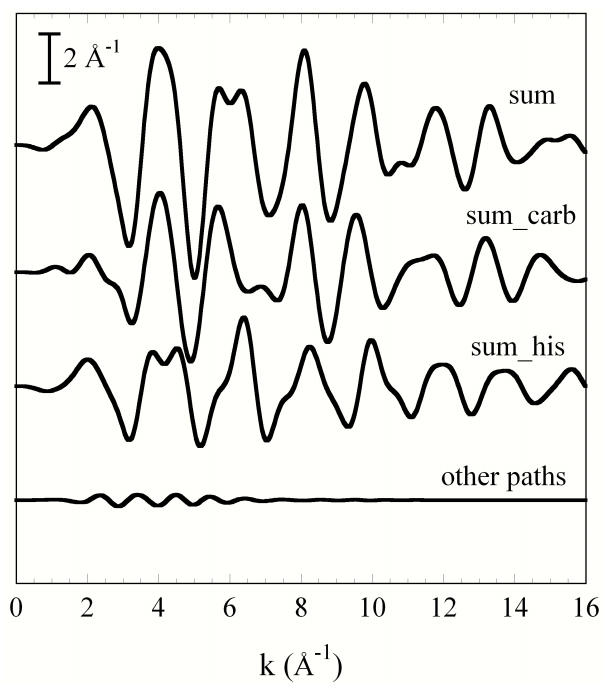


Figure 5.2: FEFF8.2 self consistent simulation of the cluster shown in Fig. 5.1. From the top to the bottom: total sum of paths (until  $nleg=5$ ); sum of the paths belonging to carboxylates; sum of the paths belonging to histidines; sum of the paths involving different residues.

tetrahedral Zn cluster composed by two histidines (Ne2 binding mode) and two carboxylic acids (monodentate binding mode) (see Fig. 5.1). This cluster is quite common among Zn metalloproteins (see, for example, PDB codes: 1DY0, 1HFE, 1QM6 etc..). In Fig. 5.2 we show the SCF simulation together with the signals obtained by summing separately the paths belonging to histidine, those belonging to carboxylic acids and those involving different amino acids. As can be shown the paths belonging to different residues give a very low contribution to the total signal.

#### 5.2.4 Fitting procedure

All FEFF8.2 scattering paths, simulated as explained above, can be imported in Artemis. In order to make the system well over-determined, rigid body refinement is applied. This can be done inserting constraints (*set* and *def* variables) in Artemis according to the following considerations.

The various aminoacids composing the entire cluster are seen as structural units with limited degrees of freedom. Closed rings, as for example histidine imidazole, are sufficiently well described by fixed structural units, thus all structural parameters, both distances and angles are set according to Engh and Huber. Carboxylic acids are better described using one degree of freedom, i.e. the  $O_{\delta_1} - \hat{C} - O_{\delta_2}$  angle, since the position of  $O_{\delta_2}$  has a great influence on the EXAFS signal.

The various structural units are free to move with respect to the metal atom. In a first approximation the movement can be forced to be only along the axis connecting the metal and the first neighbour, i. e. 1-Dimensional movement (see Fig 5.3). This is achieved by releasing the distance between

the metal  $A$  and the coordinating atom of the residue  $B$ , so that the distance of each atom  $C$  of the residue can be related to the variation of this distance simply by applying the Carnot theorem,

$$\overline{AC}^2 = \overline{AB}^2 + \overline{BC}^2 - 2 \cdot \overline{AB} \cdot \overline{BC} \cdot \cos \hat{ABC}. \quad (5.1)$$

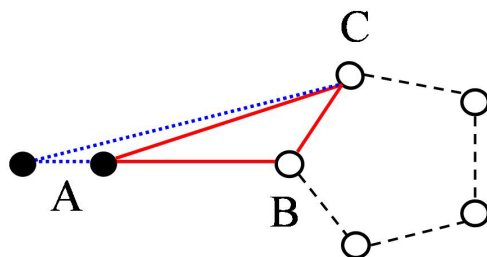


Figure 5.3: 1-Dimensional rigid body refinement for histidine residue.

The EXAFS signal can be also sensitive to the orientation of the coordinating groups with respect to the metal, since this movement changes either single scattering paths involving atoms besides the first shell or multiple scattering paths involving the atoms of the same structural units. In this case the movement is forced to be on the plane determined by the metal, the first and the second neighbours (i.e. 2-Dimensional). This is achieved inserting another degree of freedom, i. e. the angle determined by the metal, the first and the second neighbours, usually named *bending angle*. Also in this case we use the Eq. 5.1, releasing not only the distance  $\overline{AB}$  but also the angle  $\hat{ABC}$ . As an example we show the case of histidine (see Fig. 5.4). In our experience, the EXAFS signal is particularly sensitive to this parameter when ring structures are present, in which multiple scattering paths have a great influence in the entire signal and there is no other angular parameter to consider because the ring moves rigidly.

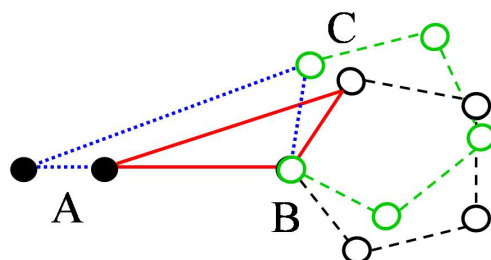


Figure 5.4: 2-Dimensional rigid body refinement for histidine residue.

In principle, the movement of the structural units with respect to the metal should be 3-dimensional, thus also rotations and tilting movements of the single residues with respect to the metal should be considered. However either rotations or tilting movements change only MS paths involving different residues. These contributions are in general very weak and it is the reason why it is so difficult to have 3-D information about the metal cluster from an EXAFS analysis. The contribution of these paths to the total signal depends on the structure of the system under investigation. For the clusters we have considered up to now they have a negligible effect on the total signal (see Fig. 5.2). However, it should be noticed that tilting movements cause changes in the geometry of first shell ligands. As a consequence they can have an effect in the *ab initio* calculations, especially at low  $k$  values.

In Appendix A we show the different degrees of freedom for selected amino acids about which one can have information from an EXAFS analysis.



## Chapter 6

# Identification of zinc sites in charge translocating membrane protein complexes

Part of the results illustrated in this Chapter have been published in

*Giachini et al, Biophysical Journal 88: 2038-2046. 2005.*

*Francia et al, FEBS Letters 581: 611-616. 2007.*

### 6.1 Overview

Zn<sup>2+</sup> metal ions bind specifically to a number of charge translocating membrane protein complexes, strongly inhibiting electron and/or proton transport. Here we report XAFS studies of the local structure of bound Zn<sup>2+</sup> in a membrane-bound pigment protein complex (the bacterial photosynthetic reaction center, RC) and three key enzymes of the respiratory chain (The NADH-quinone oxidoreductase, or complex I; the cytochrome (cyt) bc<sub>1</sub> complex, or complex III; the cytochrome c oxidase, or complex IV).

The objective is to determine at high resolution the local structure of the metal ion, and, possibly, the location of the binding sites in order to clarify the structural origin of the inhibitory effects observed.

We now describe the structure and function of the investigated proteins and some details about the mechanisms of zinc inhibition.

### 6.1.1 Structure and function of selected charge translocating membrane protein complexes

#### Reaction center

The bacterial photosynthetic reaction center (RC) is a membrane-bound pigment-protein complex that initiates the conversion of light into chemical energy (see Fig. 6.1). It catalyses a series of light-driven electron and proton

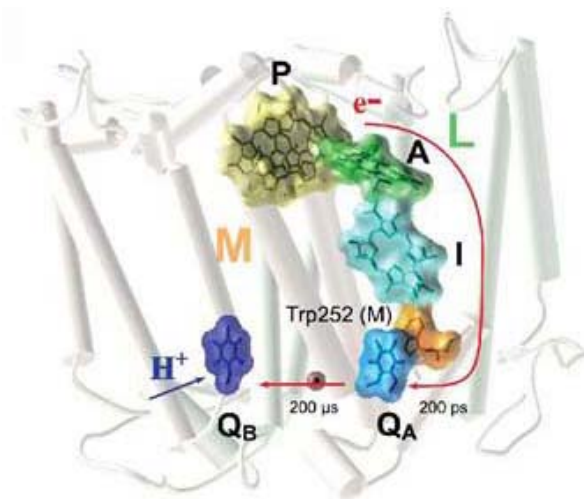


Figure 6.1: The photosynthetic reaction center from *Rhodospirillum rubrum*.

ton transfer reactions resulting in the formation across the bacterial membrane of a proton gradient utilized for ATP synthesis [60]. In the RC of

*Rhodobacter (Rb.) sphaeroides*, following absorption of a photon, electron transfer proceeds from the first excited singlet state of the primary donor P (a bacteriochlorophyll dimer) through a series of intermediate acceptors (a bacteriopheophytin and a primary quinone acceptor  $Q_A$ ) to a loosely bound secondary quinone  $Q_B$ , yielding the membrane-spanning charge separated state  $P^+Q_B^-$ . Upon re-reduction of  $P^+$  by its physiological electron donor (a soluble cytochrome  $c_2$ ), a new photoexcitation of P leads to the full reduction and protonation of  $Q_B^-$  to  $QH_2$ , which dissociates from the  $Q_B$  site and acts as a mobile electron and proton carrier from the RC to the  $bc_1$  complex [61], [62], [60]. This protein complex can be considered as a model system when examining more complex redox enzymes.

### **Key enzymes of the respiratory chain**

The respiratory chain, located in the inner mitochondrial membrane, is formed by a series of membrane-associated electron carriers which contribute to the cellular respiration catalyzing the electron transfer from the NADH to the molecular oxygen (see Fig. 6.2). This reaction results in the formation across the membrane of a proton gradient utilized for ATP synthesis. The main components of the respiratory chain are: The NADH-quinone oxidoreductase (complex I), the cytochrome  $bc_1$  complex (complex III) and the cytochrome  $c$  oxidase (complex IV).

The NADH-quinone oxidoreductase (complex I) is the point of entry for the major fraction of electrons that traverse the respiratory chain in mitochondria. Coupled to the electron transfer, protons are pumped from the

### Mitochondrial Electron Transport Chain

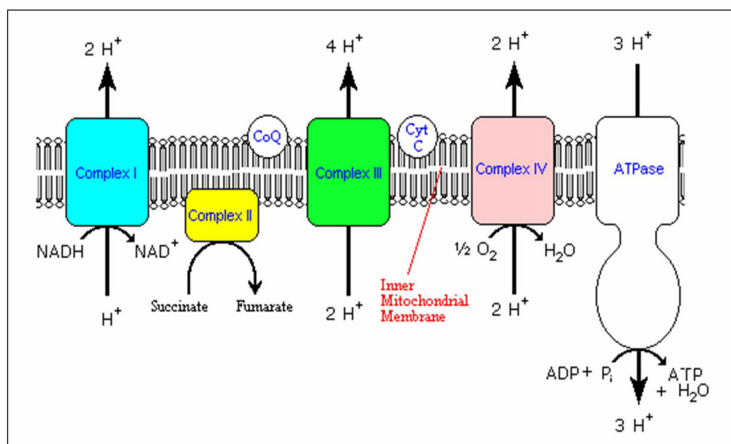
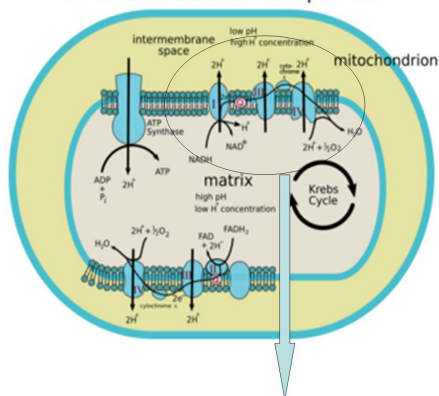


Figure 6.2: The respiratory chain.

matrix side to the inter-membrane space. Complex I is central to energy trasduction, an important source of cellular reactive oxygen species and its disfunction is implicated in neurodegenerative and muscular diseases and in aging. The information on its structure and mechanism of action is still limited, since it is the most intricate membrane-bound enzyme known to date, being composed of at least 45 different polypeptides, a non-covalently bound flavin mononucleotide (FMN) molecule, 2 binuclear and 6 tetranuclear iron-sulfur clusters. The 3D structure consists of 2 major segments:

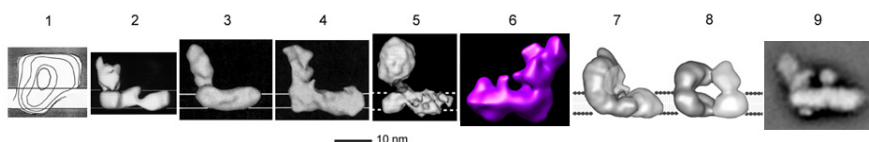


Figure 6.3: EM structures of complex I. 1. *N. crassa*. (Leonard et al., 1987) 2. *N. crassa*. (Hofhaus et al., 1991); 3. *N. crassa* (Guenebaut et al., 1997); 4. *E. coli* NDH-1 (Guenebaut et al., 1998); 5. Bovine heart (Grigorieff, 1998); 6. *Y. lipolytica* (Radermacher et al., 2006); 7 and 8. *E. coli* (Böttcher et al., 2002); 7 is inactiv form, 8 is "active" form; 9. *Arabidopsis* (Dudkina et al., 2005).

the peripheral segment, which protrudes into the mitochondrial matrix (or bacterial cytosol) and the membrane segment. The only available information on the 3D structures of complex I is low resolution Electron Microscopy analyses (see Fig. 6.3). Recently Sazanov and Hinchliffe reported the crystal structure of the hydrophilic domain of complex I from *Thermus thermophilus* at 3.3 Å resolution [63].

The cytochrome (cyt) *bc*<sub>1</sub> complexes are multi-subunit integral membrane proteins which catalyze the transfer of electrons from a hydroquinone derivative (QH<sub>2</sub>) to a soluble electron carrier (a c-type cyt). This transfer is coupled to the pumping of protons across the energy transducing membrane

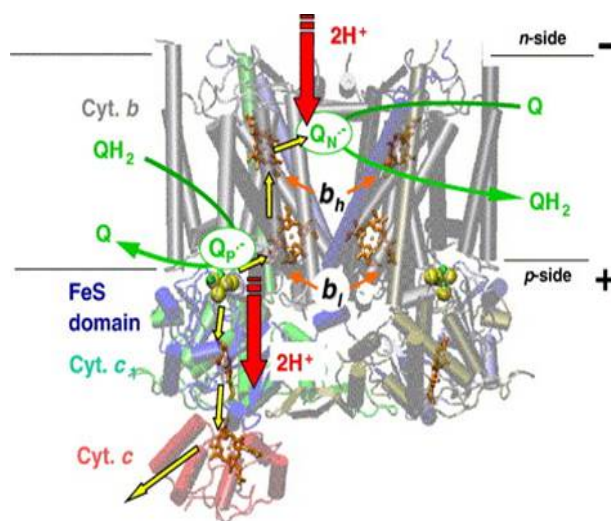


Figure 6.4: The  $bc_1$  complex.

and the resulting electrochemical potential of protons drives the synthesis of ATP via a chemiosmotic circuit. Several X-ray diffraction (XRD) structures of the mitochondrial  $bc_1$  complex have been accumulated during the last decade; they have been supplemented recently by the XRD structure of the simpler bacterial  $cyt\ bc_1$  at 3.8 Å resolution [64]. The catalytic core of all  $bc_1$  complexes comprises four redox-active metal centers: one heme  $c$ , bound to the  $cyt\ c_1$  subunit, one  $Fe_2S_2$  center in the Rieske iron-sulfur protein (ISP) and two  $b$ -type hemes ( $b_L$  and  $b_H$ ) bound to a single  $cyt\ b$  subunit. According to the modified Q-cycle mechanism [65], redox coupled  $H^+$  translocation by the  $cyt\ bc_1$  involves two catalytic sites facing the two opposite sides of the energy transducing membrane: the  $Q_o$  site, at which  $QH_2$  oxidation is coupled to proton release and the  $Q_i$  site, where quinone reduction is coupled to  $H^+$  uptake. A key feature of the Q-cycle is the bifurcation of the electron transfer chain at  $Q_o$ : upon ubiquinol oxidation one electron is delivered to the high-potential chain, reducing in sequence the iron-sulfur center and  $cyt\ c_1$ ; the second electron is transferred to quinone

or semiquinone at the  $Q_i$  site, via the two low potential hemes of cyt b. Electron transfer to cyt  $c_1$  involves a large movement of the ISP hydrophilic extension [66]. Reduction of the ISP by  $UQH_2$  occurs in a conformation docked onto cyt b, which places the  $Fe_2S_2$  cluster in contact with the  $Q_o$  site. A subsequent movement of the ISP head domain towards cyt  $c_1$  facilitates electron transfer to the latter redox partner. Although the availability of XRD structures has greatly contributed to understand and better define the catalytic mechanism, the molecular mechanisms of ubiquinol oxidation at the  $Q_o$  site and the associated proton transfer events are still unclear and intensively debated [67], [68], [69].

Cytochrome c oxidase is the terminal component of the respiratory chain: it catalyses the oxidation of cyt c reduced by the  $bc_1$  complex, reducing  $O_2$  to  $H_2O$  and pumping 4 protons across the mitochondrial membrane. Cytochrome c oxidase has four redox centers: a binuclear  $Cu_A$ , bound to subunit II, heme a, heme a3 and  $Cu_B$ , all bound to subunit I. Definition of the proton transfer pathways is an open question, crucial for understanding the complex catalytic mechanism of the enzyme. XRD structures of cyt oxidase have helped identify two putative channels for  $H^+$  uptake, referred to as the D and K channels. The roles of these channels are still debated in terms of the number and destination of the protons they conduct and how the individual proton uptake events are coupled to electron transfer [70].

### 6.1.2 $Zn^{2+}$ inhibition

In the reaction center (RC) from *Rhodobacter sphaeroides* stoichiometric binding of  $Zn^{2+}$  or  $Cd^{2+}$  inhibits proton transfer to the secondary photoreduced quinone acceptor,  $Q_B$  [71]. A high-affinity binding site for  $Zn^{2+}$  and

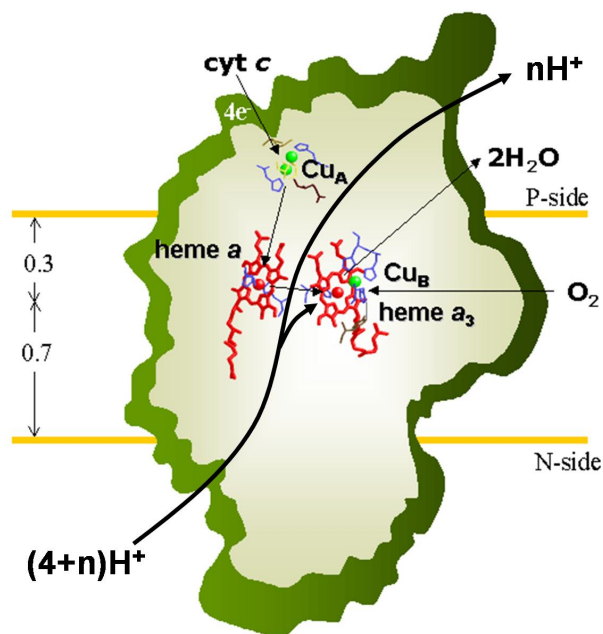


Figure 6.5: The cytochrome c oxidase complex.

$\text{Cd}^{2+}$  has been located by XRD at the cytoplasmic surface of the reaction center [60]. The  $\text{Zn}^{2+}$  ligand cluster is formed by the imidazole side chains of two histidines (HisH126, HisH128) and by the side chain of an aspartic acid (AspH124). A water molecule was also proposed to interact with  $\text{Zn}^{2+}$ . We have confirmed this tetrahedral coordination geometry with an EXAFS study [72].  $\text{Cd}^{2+}$  involves the same cluster of residues and possibly three water molecules in a octahedral geometry [60]. By examining the pH dependence of  $\text{Cd}^{2+}$  binding in native and mutated RCs, it was clearly shown that inhibition of proton transfer by the metal ion is due to competition with protons for binding to HisH126 and HisH128, thus hampering the function of these residues as proton donor/acceptors along the proton pathway to the  $\text{Q}_B$  site [73]. As a consequence, the structural definition of Zn binding site allowed determination of the entry point of  $\text{H}^+$  and contributed to the full definition of the proton pathway from the aqueous phase to the acceptor



$Q_B$  molecule [74]. This mechanism of Zn inhibition, i.e. competitive block of an entry or exit proton pathway, has been suggested to be responsible of Zn inhibitory effects observed in other redox enzymes, including the redox enzymes of the respiratory chain [73]. If this is correct, location of Zn binding sites and resolution of their local structure becomes a powerful tool in tracing proton pathways in an entire class of membrane proteins.

$Zn^{2+}$  is a well established inhibitor of the bovine mitochondrial cytochrome  $bc_1$  [75], in which it competes with  $H^+$  ions. In bacterial  $bc_1$  complexes  $Zn^{2+}$  has been recently shown to decelerate specific electron transfer steps and the generation of transmembrane voltage [76]. These observations have been tentatively explained by proposing that  $Zn^{2+}$  binds close to the  $Q_0$  site, blocking the proton release channel(s). In agreement with this suggestion two potential inhibitory  $Zn^{2+}$  binding sites have been revealed in the XRD structure of the avian cytochrome  $bc_1$ : one of them, located in a hydrophilic area between cytochrome  $b$  and  $c_1$ , might interfere with the egress of  $H^+$  to the aqueous medium. Although a number of putative Zn ligands (including histidines) were identified, the coordination geometry could not be resolved [77].

$Zn^{2+}$  inhibits bovine heart cytochrome  $c$  oxidase [78]. Investigations on cytochrome  $c$  oxidase reconstituted in liposomes (COV) in the presence of  $ZnCl_2$ , revealed an uncoupling effect exerted by the metal on the proton pumping activity of the oxidase [79]. Uncoupling was associated with inhibition of intermediate steps of the catalytic cycle of oxygen reduction to water [79], [80]. A slow [81] inhibitory effect exerted by externally added zinc to preformed COV on the oxidase activity, dependent on the presence of a membrane potential [82], has also been observed [81], [82], [83].

Recently [84] it has been shown that  $\text{Zn}^{2+}$  inhibits also complex I by inhibiting either quinone reduction (perhaps by inhibiting proton transfer to the quinone site) or proton translocation. This opens the way for a thorough inspection of proton channels or other key structural catalytic features in this enzyme.

## 6.2 The inhibitory binding site of $\text{Zn}^{2+}$ in bacterial photosynthetic reaction center

As mentioned above, XRD data at 2.5 Å resolution have identified a  $\text{Zn}^{2+}$  binding site, indicating that the metal ion is coordinated to two imidazole nitrogens from HisH126 and HisH128 and one oxygen from AspH124. A water molecule has been proposed to interact with  $\text{Zn}^{2+}$ , determining a tetrahedral coordination. We have performed XAFS measurements on samples of Zn-RC complexes characterized by a 0.9 and 1.9 atom per complex respectively. The details regarding sample preparation, data collection and data analysis can be found [72]. The coordination geometry obtained in the non-crystalline sample characterized by a Zn/RC ratio close to one reproduces well the one proposed on the basis of XRD data in RC crystals soaked in the presence of the metal although the first shell distances resulting from our study are considerably contracted in comparison with those determined by XRD.

We found that incubation of RCs in the presence of a moderate stoichiometric excess of Zn over the protein, followed by gel filtration chromatography, results systematically in Zn-RC complexes characterized by Zn/RC ratios close to two (see [72] for details). This strongly suggests that the high

affinity site located by XRD is not the only  $\text{Zn}^{2+}$  binding site of the RC. Both RC-bound Zn ions can be removed by equilibrium dialysis against a metal chelating resin. The XAFS spectrum of one of such samples ( $\text{Zn}/\text{RC} = 1.9 \pm 0.2$ ) is consistent with the existence of a new, lower affinity site which involves one O and three N, not belonging to a histidine residue. The possible contribution of the lower affinity binding site to inhibition of electron and/or proton transfer should be considered in functional studies performed in the presence of excess metal ion.

### **6.3 The inhibitory binding site of $\text{Zn}^{2+}$ in bacterial, avian and bovine $\text{bc}_1$**

#### **6.3.1 Materials and methods**

The cytochrome  $\text{bc}_1$  complex was purified from chicken essentially as described in [66]. Purification of the  $\text{bc}_1$  complex from *Rhodobacter capsulatus* is described in detail in [64]. The  $\text{bc}_1$  complex from bovine-heart mitochondria was prepared as in [75]. The concentration of the avian  $\text{bc}_1$  complex was estimated spectrophotometrically using an extinction coefficient of  $60 \text{ mM}^{-1}\text{cm}^{-1}$  for the dithionite-reduced form at 558 nm versus 600 nm; analogous procedures were used for the bovine and bacterial complexes using extinction coefficients of  $70 \text{ mM}^{-1}\text{cm}^{-1}$  (at 562 nm versus 600 nm) and  $\text{mM}^{-1}\text{cm}^{-1}$  (at 560 nm versus 600 nm), respectively. To obtain Zn-cyt  $\text{bc}_1$  complexes the avian and bovine proteins were incubated at a concentration of  $80 \mu\text{M}$  in 20 mM glycylglycine, pH 7.5, 0.01 dodecylmalto-side supplemented with  $\text{ZnSO}_4$  at a molar ratio of 0.9 Zn/ $\text{bc}_1$ . Incubation was performed in a final volume of 1.2 ml for 15 hours in ice. The same procedure was used for the bacterial complex except that the final incuba-

tion volume was 3.4 ml and the protein concentration 28  $\mu\text{M}$ . A slightly sub-stoichiometric Zn/bc<sub>1</sub> ratio was chosen to maximize the occupancy of high affinity binding site(s), while minimizing the possibility that additional, lower affinity sites are populated. Following incubation, samples were applied into a Sephadex G-25 column (PD10 Pharmacia) and eluted with the same buffer without zinc. Zn stoichiometries were re-determined on the basis of the Zn and Fe content of the samples, measured by Inductively Coupled Plasma-Atomic Emission Spectroscopy (ICP-AES). The concentration of cyt bc<sub>1</sub> complexes was estimated from the Fe content assuming 5 Fe atoms per complex (2 cyt b hemes, 1 cyt c<sub>1</sub> heme and the Fe<sub>2</sub>S<sub>2</sub> center). ICP-AES analysis yielded the following Zn/bc<sub>1</sub> stoichiometries:  $0.79 \pm 0.07$ ,  $0.80 \pm 0.03$  and  $0.94 \pm 0.03$  in the bacterial, avian and bovine complex respectively. Stigmatellin was added to all samples from a 10 mM stock solution (ethanolic) at a ratio of two molecules per cyt bc<sub>1</sub> complex. For XAFS measurements the Zn-cyt bc<sub>1</sub> suspensions described above were supplemented with 2.5 w/v polyvinyl alcohol (PVA) (Fluka, Buchs, Switzerland, Mw  $\approx$  130000) and dehydrated under N<sub>2</sub> flow. Following this procedure Zn-bc<sub>1</sub> complexes are embedded at high concentration in dry PVA films. Following XAFS measurements the portion of each PVA-Zn-bc<sub>1</sub> film exposed to x-rays was redissolved and diluted for spectrophotometric measurements. Essentially the same spectra were observed in the Zn-bc<sub>1</sub> suspensions before incorporation into PVA matrices and after XAFS measurements.

### 6.3.2 Data collection and data analysis

Zn K-edge measurements were performed at the BM 8 GILDA beam-line of the European Synchrotron Radiation Facility (ESRF) in Grenoble, France. Measurements are described in Paragraph 3.1; data analysis was performed

using ARTEMIS package according to Chapter 4 and 5. Samples were measured at room temperature in the region 9500-10660 eV. For each sample we collected a minimum of 3 spectra in order to monitor possible modifications caused by the exposure to X-ray flux. Such modifications were not detected. The final spectra were obtained from the average of multiple scans for a total integration time of 80 s/point for the avian bc1, of 60 s/point for the bovine bc1 and 40 s/point for the bacterial bc1.

### 6.3.3 Results and discussion

#### Comparing Zn XAFS spectra in the different cyt bc<sub>1</sub> complexes

Fig. 6.6 (continuous lines) shows the near edge region of the spectra recorded in the samples containing the bacterial, the avian and the bovine cyt bc<sub>1</sub> complexes treated with Zn<sup>2+</sup>, as described under Materials and Methods. While avian and bovine bc1 exhibit almost the same spectral features, this is not the case for the bacterial complex, for which we observe an enhancement of the white line and some modifications in the region around 9670 eV and around 9685 eV. These differences in the XANES indicate that the local structure around zinc must be significantly different in the bacterial bc1 compared to avian and bovine complexes. The experimental EXAFS functions and their Fourier Transforms (FT) are shown in Fig. 6.7 and 6.8, respectively. Again a strong similarity appears between the bovine and avian complexes while significant differences are present in the bacterial bc<sub>1</sub>. In the latter, the first peak of the FT, contributed by first ligands, is shifted to longer distances and its amplitude is enhanced (see Fig. 6.8). When considering more distant ligands, stronger contributions are observed over the 2–3 Å range in the FT amplitude of the bacterial bc1 as compared to the avian and bovine case, while it is conspicuous that all three samples

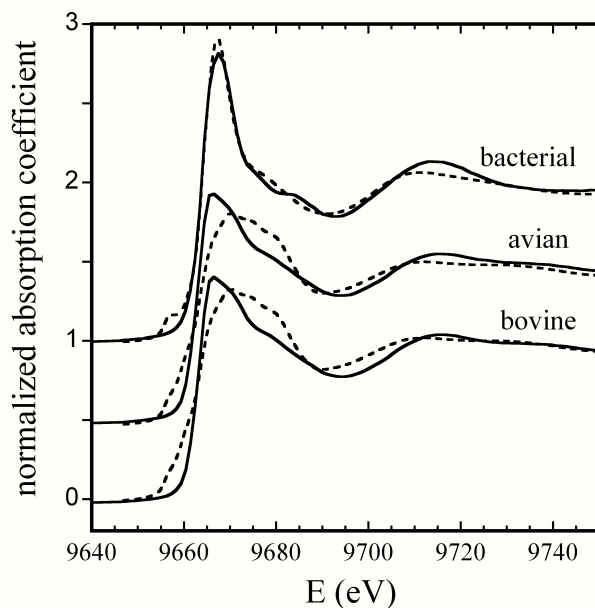


Figure 6.6: XANES spectra of the bacterial, avian and bovine bc1 complex (from top to bottom). Experimental spectra are shown as continuous lines; dotted lines represent simulations based on the structural parameters obtained by EXAFS analysis.

show the triple peak feature in the range 2.9 - 4 Å characteristic of histidine residues [85].

#### First shell analysis: the avian and bovine bc<sub>1</sub>

It is well known that, in proteins, zinc can bind 4, 5 or 6 ligands and that these ligands can be oxygen, nitrogen or sulphur [86]. A number of preliminary first shell fits, performed in R space indicated that zinc binds 4 N or O atoms. Any attempt to insert in the model sulphur atoms failed to reproduce the experimental data. As far as the coordination number is concerned, we found that, when this parameter was changed to 3 or 5, the R-factor increased appreciably. Moreover when we allowed the coordination number to vary, using Zn-N or Zn-O scattering paths, it converged to a value of 4. To properly distinguish between nitrogen and oxygen atoms

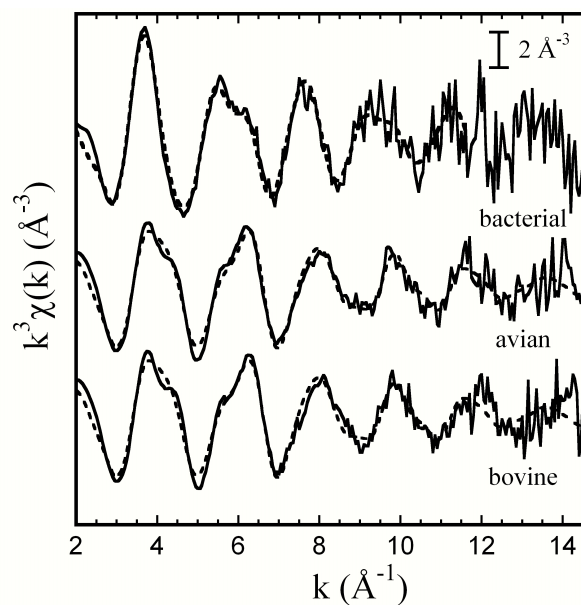


Figure 6.7: Experimental  $k^3$  weighted EXAFS functions measured in the bacterial, avian and bovine  $bc_1$  complex (continuous lines). The dashed lines represent calculated best fitting EXAFS functions corresponding to the following clusters: two His, one Lys and one Asp/Glu for the avian and the bovine complexes (the corresponding structural parameters are given in Tab. 6.2, and Tab. 6.3, model m, respectively); one His, two Asp/Glu in monodentate coordination, one Gln/Asn, two water molecules for the bacterial complex (see Tab. 6.4, model e for the corresponding structural parameters).

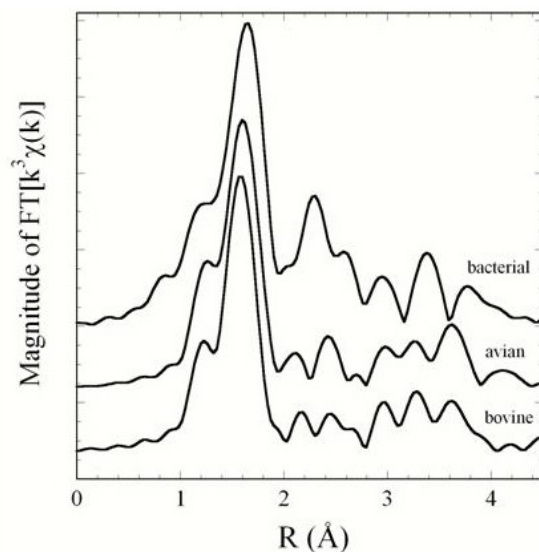


Figure 6.8: Amplitudes of the Fourier Transforms (FT) of the  $k^3$  weighted EXAFS functions shown in Fig. 6.7, performed in  $\Delta k = 2.5 - 14.5 \text{ \AA}^{-1}$ .

we have considered all structural models obtainable from different combinations of N and O atoms, keeping the coordination number fixed to a value of 4 (see Tab. 6.1). This implies 5 different cluster types. All of them were built in a tetrahedral geometry, which is the geometry commonly found in tetra-coordinated zinc clusters in proteins [86], using a starting distance of  $2 \text{ \AA}$ . We have considered as free parameters the Zn-N and Zn-O distances, the energy shift and a single DW factor. Since with the fitting range used the number of independent points in the data was 9, the model was sufficiently over-determined, having a ratio of  $\sim 2$ . As shown in Tab. 6.1, both in the bovine and in the avian complex, the best fit (minimum R-factor) is obtained for 3N and 1O atom as ligands. The cluster formed by 1N and 3O atoms also yielded a reasonable (although larger) R-factor. This combination, however, can be excluded because the distance found for the nitrogen is far bigger ( $\sim 10\%$ ) than the value expected ( $2.05 \text{ \AA}$ ) for Zn-N bond length



Avian bc <sub>1</sub> complex				
model	Zn-N (Å)	Zn-O (Å)	DW (N/O) (10 <sup>-3</sup> Å <sup>2</sup> )	R factor (%)
4 N	2.03(1)		7(1)	1.7
3N 1O	2.02(1)	2.14(3)	3.0(8)	0.8
2N 2O	1.99(87)	2.04(76)	5(70)	1.9
1N 3O	2.22(4)	2.01(1)	4.7(7)	1.1
4 O		2.00(1)	7(1)	2.6

Bovine bc <sub>1</sub> complex				
model	Zn-N (Å)	Zn-O (Å)	DW (N/O) (10 <sup>-3</sup> Å <sup>2</sup> )	R factor (%)
4 N	2.01(2)		6(2)	3.7
3N 1O	2.01(1)	2.13(4)	2.8(7)	1.2
2N 2O	1.97(6)	2.03(8)	5(88)	3.7
1N 3O	2.21(5)	2.00(1)	4.6(9)	1.9
4 O		1.99(2)		5.1

Table 6.1: Results of first shell analysis in the bovine and avian bc<sub>1</sub> complexes ( $\Delta k = 2.5\text{-}14.6 \text{ \AA}^{-1}$ ;  $\Delta R = 1\text{-}2 \text{ \AA}$ ). DW indicates the Debye Waller factor. The  $1\sigma$  error on the least significant figure of the structural parameters is reported in brackets.

in a tetra-coordinated Zn cluster (<http://tanna.bch.ed.ac.uk/>). Conversely, the distances obtained for the cluster formed by 3N and 1O atom perfectly match the ones evaluated from databases of protein crystal structures [86].

### **First shell analysis: the bacterial bc<sub>1</sub>**

The zinc binding site in the bacterial bc<sub>1</sub> is characterized by a high broadening in the first neighbour distances (see Fig. 6.8). In such cases, the first shell analysis can provide only limited information with high precision (i.e. distance) while a considerable uncertainty is associated to other structural parameters, such as coordination number and atomic number (e.g. oxygen and nitrogen atoms can hardly be distinguished). For the Zn binding site of the bacterial complex, a number of first shell fits, performed in R space, indicated that zinc binds 5-6 N or O atoms. In fact any attempt to insert in the model sulphur atoms and to decrease the coordination number led to unacceptable fit.

### **Discriminating between putative ligand clusters by multi-shell, multiple-scattering analysis**

In an attempt to define the clusters of amino acid residues forming the Zn<sup>2+</sup> binding sites, multi-shell multiple scattering analysis was subsequently performed. Within the restrictions of first shell analysis results, we built a series of possible clusters proceeding as described in Chapter 5:

- 1) We associated the first ligands with one or more amino acid residues by examining a database of Zn binding sites in proteins of known crystallographic structure (<http://tanna.bch.ed.ac.uk/cngroups.html>). The N<sub>δ1</sub> and N<sub>ε2</sub> of the histidine imidazole ring, the O<sub>δ1</sub> and O<sub>δ2</sub> of the aspartate (Asp) and glutamate (Glu) are by far the most common nitrogen/oxygen

ligands of zinc in proteins [86]. In some cases the  $O_{\delta 1}$  of asparagine (Asn) and glutamine (Gln) can bind this metal. In few cases also the  $N_z$  of lysine (Lys) has been observed to coordinate zinc. This coordination is rather uncommon requiring the deprotonation of the charged amino group of Lys characterized by typically high pKa values ( $\sim 10$ ). However the vicinity of carboxylic groups (which are often present in zinc binding sites) can facilitate this process. Starting from these considerations we can reasonably infer that if the first ligand is a nitrogen atom it belongs to a His or to a Lys, if it is an oxygen atom it belongs to Asp or Glu or to Asn or Gln (since EXAFS is sensitive only to the atoms inside a sphere of about 5 Å radius around the absorber, we cannot actually distinguish between Asp and Glu or between Asn and Gln). When considering the involvement of lysine in coordination, we required the concomitant presence of at least a carboxylic group. The participation of a water molecule in zinc coordination has been observed in several cases ( see <http://tanna.bch.ed.ac.uk/>). This increased the number of possible oxygen donors which had to be taken into account.

2) Having selected the residues possibly present in the binding clusters, the coordinates of all atoms were set according to the following criteria. As first neighbours distances we used those found by first shell analysis and as intra-ligand distances and angles those reported by Engh and Huber [4]. The 3-D geometries considered were those normally found in  $Zn^{2+}$  binding sites determined by high resolution protein crystallography [86], i.e. tetrahedral for a coordination number of four, trigonal bipyramidal for a coordination number of five and octahedral for a coordination number of six. As far as carboxylate groups are concerned we have considered two different coordination modes, i.e. monodentate and bidentate binding. Also in the case of histidine residues two configurations have been taken into account, depend-

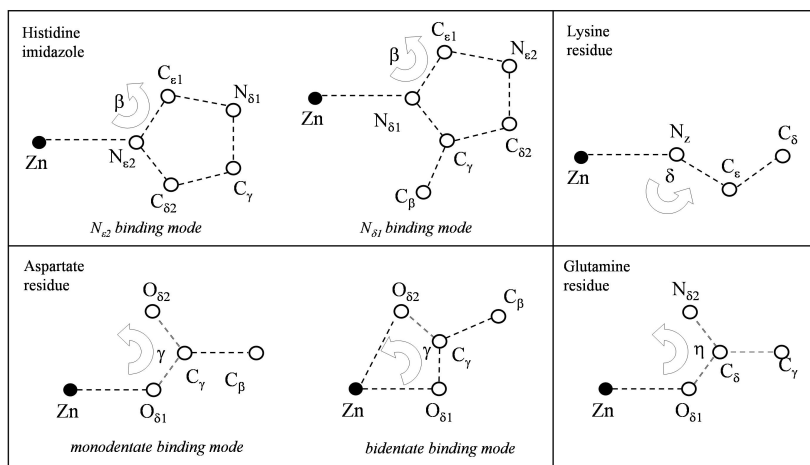


Figure 6.9: Sketch of the reference structural units used in multi-shell multiple scattering analysis. The values of distances and angles of amino acid residues were set according to Engh and Huber. Glutamate residue has the same structure of the aspartate provided that  $C_{\gamma}$  and  $C_{\beta}$  of aspartate are replaced by  $C_{\delta}$  and  $C_{\gamma}$  of glutamate, respectively. The structure of the glutamine residue coincides with that of the asparagine residue when  $N_{\delta 2}$ ,  $C_{\gamma}$  and  $C_{\beta}$  of asparagine are replaced by  $N_{\epsilon 2}$ ,  $C_{\delta}$  and  $C_{\gamma}$  of glutamine, respectively.

ing on which nitrogen atom ( $N_{\delta 1}$  or  $N_{\epsilon 2}$ ) binds the metal. The orientation of the aminoacid residues with respect to the Zn-first ligand direction was set according to the following angular values (with reference to Fig. 6.9):  $128^\circ$  and  $127^\circ$  for the Zn- $N_{\epsilon 2}$ - $C_{\epsilon 1}$  and Zn- $N_{\delta 1}$ - $C_{\epsilon 1}$  angles ( $\beta$ ) of the histidine residue;  $110^\circ$  for the Zn- $N_z$ - $C_\epsilon$  angle ( $\delta$ ) of lysine residue;  $120^\circ$  and  $90^\circ$  for the Zn- $O_{\delta 1}$ - $C_\gamma$  angles of carboxylic groups, in monodentate and bidentate binding modes respectively;  $120^\circ$  for the Zn- $O_{\delta 1}$ - $C_\gamma$  and Zn- $O_{\delta 1}$ - $C_\delta$  angle of asparagine and glutamine. Some of these angles (i.e. Zn- $O_{\delta 1}$ - $C_\gamma$  of carboxylic groups and Zn- $O_{\delta 1}$ - $C_\gamma$  angle of asparagine or Zn- $O_{\delta 1}$ - $C_\delta$  angle of glutamine) were kept fixed during the fitting procedure, whereas  $\delta$  of lysine and the  $\beta$  angles of the histidine residues were allowed to vary since a large spreading of these values has been encountered when examining crystallographic structures.

3) Built models were fitted to the experimental data and compared in terms of goodness of fit and consistency of the structural parameters obtained as described in Chapter 4 and 5. The fitting range was  $2.5$ - $14.6 \text{ \AA}^{-1}$  for avian and bovine bc<sub>1</sub> samples and  $2.5$ - $11.6 \text{ \AA}^{-1}$  for bacterial bc<sub>1</sub> sample. The free parameters were a common shift in the energy origin and, for each amino acid bound to the metal, the variation of the first ligand distance as well as an angular parameter. The fitting range used and this number of free parameters led to a  $N_{ind}/p$  ratio between 3 and 9, i.e. the problem was considerably over-determined. As angular parameters we chose: a bending angle for histidine residues (in Fig. 6.9, accounting for rotation of the imidazole ring around an axis passing through the nitrogen bound to the metal and perpendicular to the imidazole plane), the  $O_{\delta 1}$ - $C_\gamma$ - $O_{\delta 2}$  ( $\gamma$ ) angle in carboxylic groups, the  $O_{\delta 1}$ - $C_\delta$ - $N_{\epsilon 2}$  or  $O_{\delta 1}$ - $C_\gamma$ - $N_{\delta 2}$  angle in Gln and Asn residues, the Zn- $N_z$ -C angle in lysine residues ( $\delta$  in Fig. 6.9). The Debye-

Waller factor of the first-neighbours was fixed to the value obtained from the first shell analysis ( $3 \times 10^{-3} \text{ \AA}^2$ ). The Debye-Waller factors of the others shells were grouped and fixed to 5 values on the basis on DFT according to [36].

#### *Avian and bovine bc1 complexes*

We recall that for both the avian and the bovine complex first shell analysis indicated clearly  $\text{Zn}^{2+}$  coordination by 3 N and 1 O atoms. In attempting to associate each first ligand to an amino acid within the framework described above, we have now to consider three possible amino-acids combinations: 3His, 1Asp/Glu; 3His, 1H<sub>2</sub>O; 2 His, 1Lys, 1 Asp/Glu, 3 His, 1Gln/Asn. The aspartate/glutamate residues were built in a monodentate configuration, since the bidentate configuration has never been observed in the case of a total coordination number of 4 (52). However we have to consider the two binding modes ( $\text{N}_{\delta 1}$  or  $\text{N}_{\epsilon 2}$ ) for histidine residues although we recall that  $\text{N}_{\epsilon 2}$  binding mode is more frequent than  $\text{NN}_{\delta 1}$  (70% and 30% respectively) (52). This increased the number of models to 15. For all clusters the results of the fitting procedure are reported in Tab. 6.2 and Tab. 6.3 for the avian and the bovine complex, respectively. In both complexes the best fits were obtained for clusters (model i and k) formed by two histidines, one lysine and one aspartic acid. The  $\text{N}_{\epsilon 2}$  binding mode seems to be the most favoured one in general, but essentially the same R factor is obtained for the cluster in which both histidines bind  $\text{Zn}^{2+}$  by  $\text{N}_{\epsilon 2}$  (model i) and the cluster in which one histidine coordinates in the  $\text{N}_{\epsilon 2}$  and the other in the N configuration (model k). The small difference between the R factors of the

model	ligand cluster	Zn-N (Å)	Zn-O (Å)	$\beta$ (°) His	$\gamma$ (°) Asp/Glu	$\delta$ (°) Lys	R (%)	$\chi_w^2$
<i>a</i>	3 His (3N <sub>e2</sub> ) 1 Asp/Glu	2.014 (9)	2.12 (2)	130 (2)	119 (1)		13.0	31.0 (7.6)
<i>b</i>	3 His (3N <sub>s1</sub> ) 1 Asp/Glu	2.012 (9)	2.12 (2)	130 (2)	119 (1)		14.9	107.4 (26.2)
<i>c</i>	3 His (2N <sub>e2</sub> 1N <sub>s1</sub> ) 1 Asp/Glu	2.012 (9)	2.12 (2)	130 (2)	119 (1)		13.4	61.8 (15.1)
<i>d</i>	3 His (1N <sub>e2</sub> 2N <sub>s1</sub> ) 1 Asp/Glu	2.012 (9)	2.11 (2)	130 (2)	119 (1)		14.0	83.9 (20.5)
<i>e</i>	3 His (3N <sub>e2</sub> ) 1 Gln/Asn	2.014 (9)	2.12 (2)	130 (2)	125 (2)		12.7	44.4 (10.8)
<i>f</i>	3 His (3N <sub>s1</sub> ) 1 Gln/Asn	2.012 (9)	2.12 (2)	130 (2)	125 (2)		14.6	114.5 (27.9)
<i>g</i>	3 His (2N <sub>e2</sub> 1N <sub>s1</sub> ) 1 Gln/Asn	2.013 (9)	2.12 (2)	130 (2)	125 (2)		13.7	49.6 (12.1)
<i>h</i>	3 His (1N <sub>e2</sub> 2N <sub>s1</sub> ) 1 Gln/Asn	2.014 (9)	2.12 (2)	130 (2)	125 (2)		13.2	87.7 (21.4)
<i>i</i>	3 His (3N <sub>e2</sub> ) 1 H <sub>2</sub> O	2.011 (9)	2.11 (2)	133 (2)			13.4	81.2 (19.5)
<i>j</i>	3 His (3N <sub>s1</sub> ) 1 H <sub>2</sub> O	2.010 (9)	2.11 (2)	133 (2)			15.0	90.8 (21.8)
<i>k</i>	3 His (2N <sub>e2</sub> 1N <sub>s1</sub> ) 1 H <sub>2</sub> O	2.012 (9)	2.12 (2)	133 (2)			13.7	71.1 (17.1)
<i>l</i>	3 His (1N <sub>e2</sub> 2N <sub>s1</sub> ) 1 H <sub>2</sub> O	2.011 (9)	2.11 (2)	133 (2)			14.3	80.2 (19.3)
<i>m</i>	2 His (2N <sub>e2</sub> ) 1 Asp/Glu 1 Lys	2.03 (1) 1.98 (2)	2.11 (2)	130 (5)	121 (2)	127 (12)	12.2	37.8 (9.4)
<i>n</i>	2 His (2N <sub>s1</sub> ) 1 Asp/Glu 1 Lys	2.03 (1) 1.97 (3)	2.10 (2)	128 (2)	121 (1)	130 (16)	13.1	99.0 (24.5)
<i>o</i>	2 His (1N <sub>e2</sub> 1N <sub>s1</sub> ) 1 Asp/Glu 1 Lys	2.03 (1) 1.97 (2)	2.10 (2)	129 (4)	120 (1)	128 (14)	12.6	61.6 (15.3)

Table 6.2: Structural results from multi-shell multiple scattering analysis in the avian bc1 complex.

model	ligand cluster	Zn-N (Å)	Zn-O (Å)	$\beta$ (°) His	$\gamma$ (°) Asp/Glu	$\delta$ (°) Lys	R (%)	$\chi_w^2$
<i>a</i>	3 His (3N <sub>ε2</sub> ) 1 Asp/Glu	2.00 (1)	2.11 (2)	130 (2)	117 (1)		15.1	43.6 (10.6)
<i>b</i>	3 His (3N <sub>δ1</sub> ) 1 Asp/Glu	1.99 (1)	2.10 (2)	132 (1)	113 (1)		16.2	108.7 (26.5)
<i>c</i>	3 His (2N <sub>ε2</sub> 1N <sub>δ1</sub> ) 1 Asp/Glu	2.00 (1)	2.11 (2)	130 (2)	117 (1)		15.3	71.9 (17.5)
<i>d</i>	3 His (1N <sub>ε2</sub> 2N <sub>δ1</sub> ) 1 Asp/Glu	1.99 (1)	2.10 (2)	131 (2)	114 (1)		15.8	91.7 (22.3)
<i>e</i>	3 His (3N <sub>ε2</sub> ) 1 Gln/Asn	2.00 (1)	2.11 (2)	131 (1)	120 (1)		14.6	52.4 (12.8)
<i>f</i>	3 His (3N <sub>δ1</sub> ) 1 Gln/Asn	2.00 (1)	2.11 (2)	131 (2)	120 (2)		16.0	102.8 (25.1)
<i>g</i>	3 His (2N <sub>ε2</sub> 1N <sub>δ1</sub> ) 1 Gln/Asn	2.00 (1)	2.11 (2)	131 (2)	120 (2)		15.2	69.6 (17.0)
<i>h</i>	3 His (1N <sub>ε2</sub> 2N <sub>δ1</sub> ) 1 Gln/Asn	2.00 (1)	2.11 (2)	131 (2)	120 (2)		15.5	91.3 (22.3)
<i>i</i>	3 His (3N <sub>ε2</sub> ) 1 H <sub>2</sub> O	2.00 (1)	2.11 (2)	132 (2)			15.0	81.2 (19.5)
<i>j</i>	3 His (3N <sub>δ1</sub> ) 1 H <sub>2</sub> O	2.00 (1)	2.11 (2)	132 (2)			15.7	139.7 (33.6)
<i>k</i>	3 His (2N <sub>ε2</sub> 1N <sub>δ1</sub> ) 1 H <sub>2</sub> O	2.00 (1)	2.11 (2)	132 (2)			15.1	115.8 (27.8)
<i>l</i>	3 His (1N <sub>ε2</sub> 2N <sub>δ1</sub> ) 1 H <sub>2</sub> O	2.00 (1)	2.11 (2)	132 (2)			15.3	128.1 (30.4)
<i>m</i>	2 His (2N <sub>ε2</sub> ) 1 Asp/Glu 1 Lys	2.01 (1) 1.98 (3)	2.11 (2)	134 (5)	125 (2)	119 (11)	13.2	44.4 (11.0)
<i>n</i>	2 His (2N <sub>δ1</sub> ) 1 Asp/Glu 1 Lys	2.01 (1) 1.98 (2)	2.11 (3)	132 (2)			15.2	67.0 (16.6)
<i>o</i>	2 His (1N <sub>ε2</sub> 1N <sub>δ1</sub> ) 1 Asp/Glu 1 Lys	2.01 (1) 1.98 (3)	2.11 (2)	133 (3)	125 (2)	118 (9)	13.2	56.7 (14.0)

Table 6.3: Structural results from multi-shell multiple scattering analysis in the bovine bc1 complex.



different models tested is reasonable in view of the high structural similarity between the clusters. However, when the reduced chi-square statistics is considered and the crystallographic information available is taken into account (see Discussion) the binding clusters of the avian and bovine complexes can be identified with a high level of confidence.

#### *Bacterial bc<sub>1</sub> complex*

In the case of the bacterial bc<sub>1</sub> the most probable configuration indicated by first-shell analysis included a group of 5 or 6 N or O atoms. This first shell information was clearly not exhaustive leaving in principle a very large, practically unmanageable, number of clusters to scrutinize. However, a large number of *ab initio* simulations (not shown) strongly suggested that the presence of a prominent contribution in the FT amplitude over the 2-3 Å range (see fig. 6.8) can only be reproduced when residues such as carboxylic acids or Gln/Asn are included in the binding cluster. Moreover, as already noticed, the triple peak observable in the FT amplitude between 2.9 and 4 Å is diagnostic of the presence of one or more histidine residues [85]. Based on this evidence, we looked for all the clusters deposited in the metalloprotein database MDB which contain 1 or 2 his, no cysteine or methionine, and with a total coordination number of 5 or 6. Within this framework we found the following possibilities:

(a) 2 His, 2 carboxylic acids in a monodentate binding mode and 1 H<sub>2</sub>O ( PDB codes: 1AH7, 1FOJ, 1QMD, 1QH5) (N=5);

(b) 1 His, 2 carboxylic acids in a monodentate binding mode, 1 Gln/1Asn and 1 H<sub>2</sub>O (PDB codes: 1BH5, 1FRO, 2USH) (N=5);

(c) 1 His, 1 carboxylic acid in a bidentate binding mode, 1 Gln/ 1 Asn and 1 H<sub>2</sub>O (1F83) (N=5);

(d) 1 His, 3 carboxylic acids in a monodentate binding mode, 1 Gln/ 1 Asn (1USH) (N=5);

(e) 1 His, 2 carboxylic acids, 1 Gln/ 1 Asn and 2 H<sub>2</sub>O (N=6). (1BH5, 1QIP). Fitting to these clusters yields the results summarized in Tab. 6.4.

We observe that, among the clusters described above, only one contains

model	ligand cluster	Zn-N (Å)	Zn-O (Å)	$\beta$ (°) His	$\gamma$ (°) Asp/Glu	$\eta$ (°) Gln/Asn	R (%)	$\chi^2$
<i>a</i>	2 His (2 N <sub>ε2</sub> ) 2 Asp/Glu m 1 H <sub>2</sub> O n = 5	2.05 (2)	2.14 (3) 2.09 (5)	120 (5)	123 (13)		11.0	8.05 (2.3)
<i>b</i>	1 His (1 N <sub>ε2</sub> ) 2 Asp/Glu m 1 Gln/Asn 1 H <sub>2</sub> O n = 5	2.01 (4)	2.08 (6) 2.18 (9)	120 (11)	124 (3)	115 (6)	9.3	5.4 (1.6)
<i>c</i>	1 His (1 N <sub>ε2</sub> ) 1 Asp/Glu b 1 Gln/Asn 1 H <sub>2</sub> O n = 5	2.01 (7)	2.11 (7) 2.10 (8) 2.10 (17)	128 (40)	122 (2)	115 (7)	14.5	24.3 (7.4)
<i>d</i>	1 His (1 N <sub>ε2</sub> ) 3 Asp/Glu m 1 Gln/Asn n = 5	2.14 (4)	2.06 (2) 2.22 (5)	110 (14)	124 (1)	115 (4)	10.3	5.7 (1.7)
<i>e</i>	1 His (1 N <sub>ε2</sub> ) 2 Asp/Glu m 1 Gln/Asn 2 H <sub>2</sub> O n = 6	1.99 (3)	2.10 (7) 2.28 (4) 2.10 (7)	128 (17)	123 (3)	111 (6)	8.6	4.1 (1.2)

Table 6.4: Structural parameters determined by fitting the experimental data of the bacterial bc<sub>1</sub> complex to the model clusters extracted from the metalloprotein database MDB. Monodentate and bidentate binding configuration is indicated by m and b; n is the coordination number.

two histidines (model a in Tab. 6.4). A couple of coordinating histidine residues has been found in the high-affinity Zn<sup>2+</sup> inhibitory binding sites of

the photosynthetic reaction center from *Rhodobacter sphaeroides* [60] and of the bovine cytochrome c oxidase (see the next paragraph). Two histidines seem also to be present in one of the two  $\text{Zn}^{2+}$  sites found by XRD in the avian  $\text{bc}_1$  complex [87]. In view of this feature of  $\text{Zn}^{2+}$  inhibitory binding sites characterized so far in redox membrane proteins, we tested additional clusters, compatible with first shell results, containing two histidine residues. These structures were built starting from cluster a in Tab. 6.4 by considering the different coordination modes of carboxylates and by exchanging water molecules with Gln/Asn residues. Water molecules were added or eliminated in such a way to maintain a coordination number of 5 or 6. The fitting results obtained are given in Tab. 6.5. An analogous procedure was used to generate other putative clusters starting from the one which minimizes the R factor (model e in Tab. 6.4). Testing these new possibilities yielded the results shown in Tab. 6.6. For all the clusters reported in Tab. 6.4- 6.6 only the results obtained for the  $\text{N}_{\epsilon_2}$  binding mode of histidine are shown. When the corresponding clusters including  $\text{N}_{\delta_1}$  as histidine ligand were tested the respective R factors did not change appreciably. None of the additional tested clusters (Tab. 6.5 and 6.6) improves the R factor obtained for model e in Tab. 6.4.

#### *Simulations of the XANES spectra newline*

To test the reliability of EXAFS results we performed simulations of the XANES spectra based on the atomic structure of the clusters which best fit the extended spectra, following the procedures outlined above. The results obtained are reported in Fig. 6.6 as the dashed lines. For the bacterial  $\text{bc}_1$  complex, the simulation based on the most probable six-fold coordinated

model	ligand cluster n	Zn-N (Å)	Zn-O (Å)	$\beta$ (°) His	$\gamma$ (°) Asp/Glu	$\eta$ (°) Gln/Asn	R factor (%)	$\chi_w^2$
<i>a</i>	2 His (2 N <sub>e2</sub> ) 2 Asp/Glu m 1 Gln/Asn n = 5	2.05 (3)	2.11 (2) 2.18 (5)	115 (5)	124 (2)	116 (4)	10.6	7.4 (2.2)
<i>b</i>	2 His (2 N <sub>e2</sub> ) 1 Asp/Glu m 1 Asp/Glu b n = 5	2.08 (3)	2.18 (8) 2.11 (7)	137 (7)	122 (4) 121 (4)		15.4	18.4 (5.5)
<i>c</i>	2 His (2 N <sub>e2</sub> ) 2 Asp/Glu m 2 H <sub>2</sub> O n = 6	2.02 (3)	2.15 (3) 2.09 (1)	118 (4)	123 (1)		11.5	13.1 (3.8)
<i>d</i>	2 His (2 N <sub>e2</sub> ) 1 Asp/Glu m 1 Asp/Glu b 1 H <sub>2</sub> O n = 6	2.05 (5)	2.19 (6) 2.16 (9) 1.93 (7)	134 (7)	123 (4) 126 (2)		11.8	12.8 (3.9)
<i>e</i>	2 His (2 N <sub>e2</sub> ) 1 Asp/Glu m 1 Asp/Glu b 1 Gln/Asn n = 6	2.02 (3)	2.13 (10) 2.13 (9) 2.13 (10)	135 (6)	123 (11) 124 (2)		10.2	10.6 (3.3)
<i>f</i>	2 His (2 N <sub>e2</sub> ) 2 Asp/Glu b n = 6	2.07(3)	2.14 (3)	135 (5)	122 (2)		23.6	37.9 (10.8)
<i>g</i>	2 His (2 N <sub>e2</sub> ) 1 Asp/Glu b 2 H <sub>2</sub> O n = 6	2.09 (4)	2.25 (7) 2.11 (3)	135 (5)	124 (3)		17.8	34.2 (10.2)
<i>h</i>	2 His (2 N <sub>e2</sub> ) 1 Asp/Glu m 3 H <sub>2</sub> O n = 6	2.06 (2)	2.28 (5)	137 (5)	118 (3)		11.7	13.2 (3.8)

Table 6.5: Structural parameters obtained by fitting the spectrum of the bacterial bc<sub>1</sub> complex to additional clusters containing two histidine residues.

model	ligand cluster	Zn-N (Å)	Zn-O (Å)	$\beta$ (°) His	$\gamma$ (°) Asp/Glu	$\eta$ (°) Gln/Asn	R (%)	$\chi^2$
<i>a</i>	1 His (1 N <sub>e2</sub> ) 1 Asp/Glu m 1 Asp/Glu b 1 Gln/Asn 1H <sub>2</sub> O n = 6	2.01(8)	2.12(10) 2.09(10) 2.13(10) 2.09(10)	130 (32)	123(13) 125(3) 115(15)		9.6	13.4 (4.3)
<i>b</i>	1 His (1 N <sub>e2</sub> ) 1 Asp/Glu m 1 Asp/Glu b 1 Gln/Asn n = 5	2.00 (4)	2.12 (10) 2.08 (8) 2.12 (12)	132 (20)	123 (14) 122 (4)	115 (20)	11.1	12.2 (3.8)
<i>c</i>	1 His (1 N <sub>e2</sub> ) 2 Asp/Glu b 1 Gln/Asn n = 6	2.01 (7)	2.11 (7) 2.10 (8)	128 (4)	122 (2)	115 (8)	16.8	26.9 (8.0)

Table 6.6: Structural parameters derived by fitting the data obtained in the bacterial bc1 complex to clusters built from the model e of Tab. 6.4 which minimizes the R factor.

cluster with first-shell ligands in the octahedral geometry is in very good agreement with the experimental data, both in the white line intensity and in the overall spectral features. We performed a number of tests with alternative structures, all of which gave worse agreement. In particular, it is important to note that any simulation performed removing one of the six ligands, i.e. changing the coordination number from six to five, led to a decrease in the intensity of the white line. A similar observation of the strong correlation between coordination number and white line intensity has been observed for other metalloproteins, for example for the 2-His-1-carboxylate motif of tyrosine hydroxylase [8]; this provides a further proof of the reliability of the EXAFS results for the bacterial complex. The decrease of the white line intensity in the XANES spectra of the bacterial and avian bc1 complexes apparent in the experimental spectra is qualitatively compatible with the tetrahedral coordination determined from EXAFS, following the

pattern outlined above. However, in these cases the simulation of XANES spectra - while it does reproduce a reduced white line intensity - does not provide as good a quantitative agreement as for the bacterial complex. We believe this is due to a limitation of the simulation procedure rather than to an erroneous structural determination, since it is known that in a quasi-tetrahedral coordination the approximation of the scattering potential in the muffin-tin scheme is not sufficiently accurate [8].

## Discussion

### *The high-affinity Zn binding site of the avian and bovine bc1 complex*

First-shell analysis of the EXAFS data of the avian bc1 complex identifies unequivocally three N and one O atom as  $\text{Zn}^{2+}$  ligands. By subsequent multi-shell multiple scattering analysis we found that the cluster which minimizes the R factor, among the many that were tested, is formed by two histidines, one lysine and one carboxylic acid (Asp or Glu) (model i in Tab. 6.2). When comparing fits of models i,j and k on the basis of reduced-chi square statistics (see Paragraph 4.5, it appears that the model with both histidines in the  $N_{e2}$  configuration (model i) is significantly better than the one with both in  $N_{\delta1}$  (model j), while model k (one His in the N 2 and the other in the  $N_{e2}$  configuration) is not significantly worse. Cluster i is characterized by  $\chi_{\nu}^2=37.8$ , a value lower than all the other models considered, except for cluster a, formed by three histidines and one carboxylic acid, for which  $\chi_{\nu}^2=31.0$  is found (see Tab. 6.2). However, this difference in the reduced chi-square values is not statistically significant, since the condition of Eq. 4.8 is not satisfied. When comparing the fit to model a with the other fits summarized in Tab. 6.2 by means of Eq. 4.8, it

appears that the fits which are significantly worse correspond to models: b, d, e, f, g, h, j, m, o. In summary, we can discard the cluster formed by 3 His and one water molecules (for all the configurations of histidine coordination). Of the remaining three clusters, on a purely statistical basis, we can exclude only those configurations which include more than one histidine bound at  $N_{\delta 1}$ . Two  $Zn^{2+}$  binding sites have been located by XRD close to the stigmatellin binding site [87]. This study, performed in Zn-treated crystals of the chicken bc<sub>1</sub> complex in the absence of stigmatellin, located a first site (Zn01) at which  $Zn^{2+}$  appears coordinated by His121 of cyt c<sub>1</sub> and Asp253 of cyt b. His268 and Glu255, although not refining into a coordinating position, were tentatively proposed as possible additional ligands. The second site (Zn02) was found in the hydrophobic channel between the Qo site and the bulk lipid phase. Here only one potential ligand (Met125 of cyt b) could be identified. Among the three putative  $Zn^{2+}$  binding clusters defined by our EXAFS analysis (models i or k, models a or c, models l or n in Table 6.2) the former (2 His, 1 Lys and 1 carboxylic acid) fits well the crystallographic Zn01 site. In the crystallographic structure only two histidine residues are found within a radius of 10 Å from the  $Zn^{2+}$  ion, so that the other two clusters, which both include three histidines can be excluded. By combining our EXAFS analysis with XRD data [87] we identify therefore a ligand cluster characterized by a tetrahedral geometry and formed by: His121 of cyt c<sub>1</sub>, His268, Lys270 and Asp 253 of cyt b. A view of the binding site is presented in Fig. 6.10A, based on the Zn-crystal structure. Stigmatellin from the superimposed 2BCC structure is also shown below the  $Zn^{2+}$  binding site. EXAFS analysis provides high resolution structural information on this site. We found a considerable contraction of the Zn-N distance obtained for histidines (2.03 Å) as compared to the XRD Zn-N

bond length for His121 (2.9 Å). The carboxylic acid predicted by the EXAFS analysis has been identified with Asp253, rather than Glu255, since this residue appears closer to the Zn ion in the crystallographic structure ( $\text{Zn-O}_{\delta 2}(\text{Asp253})=2.7 \text{ \AA}$  as compared to  $\text{Zn-O}_{\epsilon 1}(\text{Glu255})=3.7 \text{ \AA}$ ). However, also in this case the Zn-O distance resulting from EXAFS analysis (2.10 Å) is considerably shorter than the corresponding XRD length. We can confirm the involvement of His268 in  $\text{Zn}^{2+}$  coordination, which was uncertain on a crystallographic basis, due to the large XRD Zn-N $_{\delta 1}$  distance (4.2 Å). As the fourth ligand, EXAFS analysis indicate clearly a nitrogen atom which, to a high confidence, belongs to a lysine. Such a residue (Lys270) is present is in the vicinity of Zn01. It was not considered as a possible ligand residue in the XRD study, in view of the large Zn-N $_z$  distance (6.5 Å). However, by exploring the allowed conformational motions of Lys270 and His268, we found that their N atoms can be brought at coordination distances of  $\sim 2 \text{ \AA}$ . The large Zn-N distances obtained from XRD data for both Lys270 and His268 can be due to a number of factors, including: (i) the relatively low resolution of the crystallographic structure (3.85 Å); (ii) the large B-factors obtained for the potential ligands in the crystallographic structure, when refined assuming full occupancy of the binding site, indicating a high degree of disorder; (iii) possible disruption of coordination bonds caused by x-ray irradiation, which usually involves exposure to a higher dose in XRD as compared to XAFS measurements; (iv) the amorphous state of the sample in XAFS measurements, which were performed by incorporating the protein in a PVA film.

At variance with the crystallographic study of Zn binding sites, XAFS data reported in the present paper have been obtained in the presence of stigmatellin. We choose to acquire XAFS data in the presence of this in-



hibitor to fix the ISP domain in a well defined (proximal) position. It has been reported, in fact, that, in the absence of stigmatellin, the conformation of the ISP domain depends on the Fe<sub>2</sub>S<sub>2</sub> redox state, switching from proximal to distal when the reduced ISP goes oxidized, while in the presence of the inhibitor the ISP domain is in the proximal state, independently of the redox state. However, it is unlikely that the presence of the inhibitor in the EXAFS samples and its absence in the Zn-crystal is responsible for differences between the EXAFS and XRD models on the local scale of the Zn<sup>2+</sup> binding cluster. A comparison between XRD structures of the avian bc1 complex in the absence (1BCC) and in the presence of stigmatellin (2BCC) or stigmatellin plus antimycin (3BCC) shows only limited distortions in the cluster which forms the Zn binding site. Interestingly, XAFS measurements performed on the bovine bc1 complex in the presence and in the absence of stigmatellin yielded quite similar spectra, which could be described by essentially the same structural parameters. The only difference was a slightly more pronounced damping of oscillations in the absence of the inhibitor, which suggests a larger dynamical and/or static disorder in the absence of stigmatellin. XAS measurements have been performed on samples incubated at substoichiometric Zn/bc1 ratios. The consistency of the binding cluster determined by EXAFS analysis with the crystallographic Zn01 site indicates that Zn01 is a high-affinity site. The occupancy of the lower affinity Zn02 site is presumably the result of the prolonged (one week) incubation of the crystal in ZnCl<sub>2</sub>. XAFS spectra of the avian and of the bovine complex exhibit a striking similarity both in the XANES and in the extended region. This is confirmed by multi-shell multiple scattering analysis which indicates that, also in the case of the bovine bc1 complex, the clusters which minimizes the R factor are formed by two histidines, one lysine and one carboxylic acid

(Asp or Glu) (model i and k in Table 3). Confidence analysis based on reduced chi-square statistics leads to conclusions similar to those drawn in the case of the avian complex. In particular only the cluster which is formed by three histidines and one water molecule can be considered significantly unfavoured. Inspection of the crystallographic structure of the bovine complex (1PP9) and alignment of this structure with that of the Zn-crystal of the avian bc1 complex, show in the bovine bc1 a site, strongly homologous to Zn01. Since also in this case only two histidine residues are found within a radius of 10 Å from the Zn<sup>2+</sup> position resulting from alignment of the bovine structure with that of the Zn-crystal of the avian complex, we exclude binding clusters which include three histidines. We propose a Zn<sup>2+</sup> binding cluster formed by His121, His267, Lys269 and Asp254. The location of these residues is shown in, obtained from the crystallographic coordinates of the bovine bc1 complex in the presence of stigmatellin (1PP9). Although, as expected, the resulting distances between Zn<sup>2+</sup> and ligand atoms exceeds largely the ligand bond-length determined by EXAFS analysis, also in this case the amino acid chains of the proposed cluster are allowed to accommodate rather freely to form a reasonable tetrahedral binding site. In particular His267 and Lys 269 can be adjusted so that the Zn-N distance becomes in both case lower than 2 Å. The coincidence of the bovine Zn<sup>2+</sup> binding site with Zn01 of the avian complex has a remarkable functional implication. It strongly suggests that these structurally homologous sites can be identified with the high-affinity site characterized by Link and von Jagow [88] as the inhibitory site which interferes with a proton extrusion pathway. Noteworthy this site shares with the Zn inhibitory binding site of the photosynthetic reaction center from Rb. sphaeroides the presence of two histidine residues and one carboxylic acid. This structural feature suggests that, also in the

case of the  $bc_1$  complex, Zn inhibits proton transfer because it binds to histidines involved in proton transfer steps, thereby impairing their function as proton donors/acceptors along a proton transfer pathway. This role could in principle be played also by the lysine residue (Lys270 and Lys269 in the avian and bovine enzyme, respectively) the pKa of which is expected to be lowered by the carboxylic acids present in the vicinity.

*The  $Zn^{2+}$  binding site of *Rb. capsulatus* cyt  $bc_1$  complex*

The existence of a high-affinity  $Zn^{2+}$  binding site in the cyt  $bc_1$  complex of the photosynthetic bacterium *Rb. capsulatus* is definitely proved by the observation of a well structured XAFS signal in samples characterized by a Zn/ $bc_1$  stoichiometry close to one. The bacterial complex exhibits quite distinct spectral features as compared to the avian and bovine complexes, revealing a different local structure of the binding site. First-shell analysis indicates in fact a coordination number of 5 or 6, but provides limited information on the number of N and O atoms involved in binding. As a consequence a large number of putative binding clusters has been tested when performing multi-shell multiple scattering fitting of the EXAFS signal. Among the considered possibilities, the cluster corresponding to the minimum R factor (R=8.6) is characterized by a octahedral geometry and is formed by 1 His, 2 carboxylic acids in monodentate binding mode, 1 Gln or Asn residue and two water molecules (model e in Tab. 6.4). This cluster yields also the minimum reduced chi-square among all tested models. A reasonably good, although higher, R factor (R=9.6) is obtained for a similar cluster, also characterized by a coordination number of six and formed by the same residues, in which one of the carboxylic acids binds in the biden-

tate mode and a single water molecule is involved (model a of Tab. 6.6). When the reduced chi-square values corresponding to these two clusters are compared it appears that the former (model e of Tab. 6.4) is significantly better. A systematic comparison of model e with all the other tested models indicates that the difference in reduced chi-square values is significant in all cases except for: model d in Tab. 6.5, characterized by a coordination number of six, and four clusters (models a, b, d in Tab. 6.4 and model c in Tab. 6.5), all with a coordination number of five. The hexacoordinated cluster can be ruled out on the basis of XRD structural information (see below). All the clusters with a coordination number of five, besides any statistical evaluation of confidence levels, are highly unfavoured by the results of XANES simulations. Any attempt to simulate the near edge spectral region with clusters characterized by a coordination number of five results in fact in a white line amplitude appreciably lower than the experimental one; by contrast the simulation based on the six-fold coordinated octahedral cluster which minimizes the R factor (model e in table 4) yields the experimentally observed amplitude of the white line and reproduces with a remarkable accuracy the other XANES features (see the dotted line in the upper trace of Fig. 6.6). By aligning the crystallographic structure of the bc<sub>1</sub> complex of *Rb. capsulatus* (1ZRT) with that of Zn-crystals of the avian complex, in the region homologous to that of the Zn01 site, a group of residues is found in the bacterial complex, which fits the Zn binding cluster (model e) indicated by EXAFS analysis as the most probable one. The cluster includes: His276, Asp278, Glu295 and Asn279 of the cyt b subunit (see Fig. 6.10). By exploring the allowed motion of these residues it appears that the cluster can rearrange to accommodate a coordinated Zn ion in a pseudo octahedral geometry, which includes additionally two water molecule, characterized by

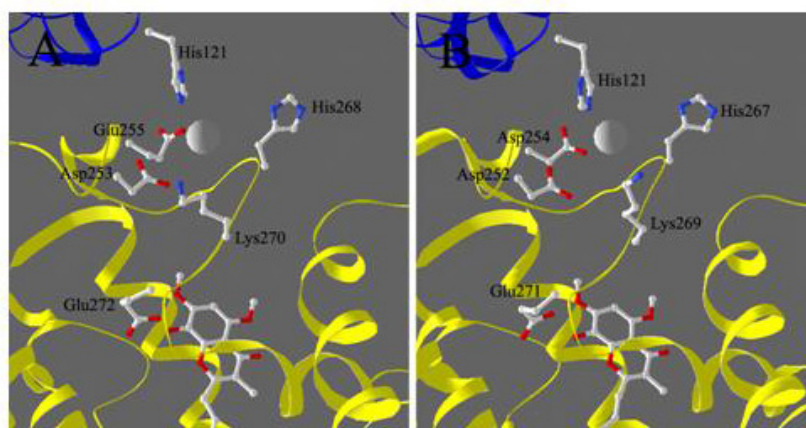


Figure 6.10: View of the proposed  $\text{Zn}^{2+}$  binding sites in the avian (A) and bovine (B) bc1 complex. The site of Zn (represented as a grey sphere) is in the interface between cyt b (yellow) and cyt c (blue). The structure in panel A was obtained from the XRD data of the Zn-crystal of chicken bc1. Coordinates are from bczn3ref.pdb file available at <http://sb20.lbl.gov/cytbc1/PDB/>. Stigmatellin, below the Zn site, is from the superimposed avian bc1 structure 2BCC. Coordinates of panel B are from the structure of the bovine cyt bc1 obtained in the presence of stigmatellin (1PP9). Zn ion was superimposed after alignment of this structure with that of the Zn crystal of the avian complex (see panel A).

the first-ligand distances reported in Tab. 6.4 (model e). We propose this cluster as the inhibitory  $\text{Zn}^{2+}$  binding site responsible of the effects observed in chromatophores from *Rb. capsulatus*. As shown in Fig. 6.10, alignment of the crystallographic structure of the  $\text{bc}_1$  complex from *Rb. capsulatus* with that of Zn-crystals of the avian complex locates the proposed binding cluster somewhat displaced with respect to the position of the  $\text{Zn}^{2+}$  bound to the chicken complex. Structural alignment could indeed suggest a contiguous, different  $\text{Zn}^{2+}$  binding site, involving His276, Asp278 (i.e. two residues belonging to the cluster we propose) and additionally His291 (see Fig. 6.10). Since the coordination number of six emerges clearly from the XANES analysis, this cluster could be completed by two or three water molecules, depending on the bidentate or monodentate binding mode of the carboxylic acid. Although there is no example of such Zn binding sites in crystallographic data bases, a cluster formed by 2 His, 1 Asp and three water molecules has been found by XRD analysis to bind  $\text{Cd}^{2+}$  in the bacterial photosynthetic reaction center of *Rb. sphaeroides*. Moreover the location of this  $\text{Cd}^{2+}$  binding site coincides with that of the high affinity  $\text{Zn}^{2+}$  binding site. In the case of  $\text{Zn}^{2+}$  coordination with two water molecules is lost and tetrahedral geometry is obtained. These features make the cluster formed by His291, His276, Asp278 and three water molecules a possible, attractive alternative to model e of Tab. 6.4. As shown in Tab. 6.5, however, such a cluster (model h) provides a worse fit to the EXAFS spectrum measured in the bacterial complex. Most importantly the difference in goodness of fit is significant. The same conclusions hold for a putative binding cluster formed by His291, His276, Asp278 in bidentate configuration and two water molecules (model g in Tab. 6.5). As mentioned above, also the hexacoordinated cluster formed by 2 His, 2 Asp/Glu (one in monodentate and the other

in bidentate configuration) and 1 Gln or Asn (model d in Tab. 6.5) deserves attention because, although it yields a worse fit than model e of Tab. 6.4, the difference in the reduced chi-squares is not highly significant. Inspection of the crystallographic structure of *Rb. capsulatus* shows, however, that no cluster formed by these residues can be located in the region identified by alignment of this structure with that of the Zn-crystal avian complex. In fact, when considering His291, His276, Asp278 as participating residues, a second carboxylic residue and a Gln (or Asn) are not found at coordinating distances, even considering large structural rearrangements. When moving to a possible contiguous cluster, involving His276, Asp278, Glu295 and Asn279, a second histidine at coordinating distances is lacking. The integration of crystallographic information available for the avian and the bacterial complex with EXAFS and XANES analyses indicates therefore as the most probable binding site a cluster formed by His276, Asp278, Glu295 and Asn279 and two water molecules. It has been reported that in *Rb. capsulatus* chromatophores the inhibitory action of  $\text{Zn}^{2+}$  depends on pH with an apparent pK 7.0 in the acidic range. This agrees with a pK 7.2 found for  $\text{Zn}^{2+}$  inhibition in the mitochondrial bovine cyt bc1 complex. Studies on  $\text{Zn}^{2+}$  inhibition in *Rb. sphaeroides* reaction centers showed a similar pH dependence, with an apparent pK 6.8, which has been attributed to the histidine residues involved in divalent metal ion binding and proton uptake. The presence of histidine residues in the  $\text{Zn}^{2+}$  binding sites identified by our analysis (one histidine in the cluster of the *Rb. capsulatus* complex and a couple in the site of the avian and bovine bc1) points to a common mechanism of zinc inhibition: zinc would compete with protons for binding to the histidines, thus hampering the function of these residues as proton donor/acceptors along the proton pathway.

## 6.4 The inhibitory binding site of $\text{Zn}^{2+}$ in cytochrome *c* oxydase

### 6.4.1 Materials and methods

Purified bovine COX contained 10 nmoles of heme a + a<sub>3</sub>/mg protein, SDSPAGE analysis revealed the complete set of 13 subunits [89]. Respiratory activity was measured polarographically in 40 mM KCl, 10 mM Hepes, 0.1 mM EDTA, pH 7.4, 40 nM aa<sub>3</sub>, 50  $\mu\text{M}$  cytochrome *c*, 25 mM ascorbate at 25 °. Oxidase vesicles (COV), were prepared by cholate dialysis as in [89]. Where indicated,  $\text{ZnCl}_2$  (200  $\mu\text{M}$ ) was added to the sonication medium and to the first two dialysis solutions, but omitted from the third. Respiratory rates, orientation of the oxidase in COV, passive proton conduction, residual soluble COX, COX absorbance changes and pH changes were measured as in [89].

### 6.4.2 Data collection and data analysis

Zn K-edge-extended X-ray absorption spectroscopy (EXAFS) measurements were performed on samples of bovine COX characterized by 1, 1.5 and 2 zinc atoms per complex respectively. The latter two samples were obtained by incubating the protein with sub-stoichiometric amounts of  $\text{ZnSO}_4$  (20 h on ice) in order to maximize the occupancy of high affinity binding site(s). Metal content, measured by inductively coupled plasma-atomic emission spectroscopy (ICP-AES), showed that the preparation was exempt from contaminating Zn. Fe content revealed by ICP-AES was consistent with the heme content of the complex evaluated spectrophotometrically. For EXAFS measurements Zn-COX suspensions were embedded in dry PVA films. EXAFS measurements were performed at the BM8 GILDA beam-line of the



European Synchrotron Radiation Facility (ESRF) in Grenoble. Measurements are described in Paragraph 3.1; data analysis was performed using ARTEMIS package according to Chapter 4 and 5. The total integration time was 60 s/point for the sample characterized by 1 zinc atom per complex and 40 s/point for the other two samples. Final fitting was performed directly in  $k$  space (with a  $k$  weight of 3), in the range  $2.5 - 10 \text{ \AA}^{-1}$ . As free parameters we used a single, common shift in the energy origin, the first ligand distances, the bending angle between the  $\text{ZnN}_\epsilon$  bond and the imidazole ring of histidines (keeping Zn coplanar with the imidazole ring). DebyeWaller factors were grouped and fixed to 5 values accordingly to [36].

### 6.4.3 Results

X-ray crystallographic analysis has shown the existence in bovine heart cytochrome *c* oxidase of an endogenous  $\text{Zn}^{2+}$  bound to subunit Vb. The structure of this site, to which  $\text{Zn}^{2+}$  is coordinated by four cysteine residues, has been studied by EXAFS [90]. We examined three samples: sample a, exempt of exogenous zinc, characterized by a stoichiometry of 1.0  $\text{Zn}^{2+}$  per complex; samples b and c, characterized by a stoichiometry of 1.5 and 2.0 Zn ions per complex, respectively. The EXAFS functions extracted from the data and the corresponding Fourier transforms are shown in Fig. 6.11. The results of first shell analysis of sample a (see fit in Fig. 2B) are consistent with a  $\text{Zn}^{2+}$  coordination by four sulphur atoms with a bond length of  $2.33 \pm 0.01 \text{ \AA}$  in agreement with previous EXAFS and crystallographic analysis.

Fig. 6.11B shows that, going from sample a to samples b and c, a second peak at shorter distances appears in the radial distribution function (its position is indicated by a vertical dashed line in Fig. 6.11B). The amplitude of this second peak increases at increasing Zn/complex ratios (from sample b

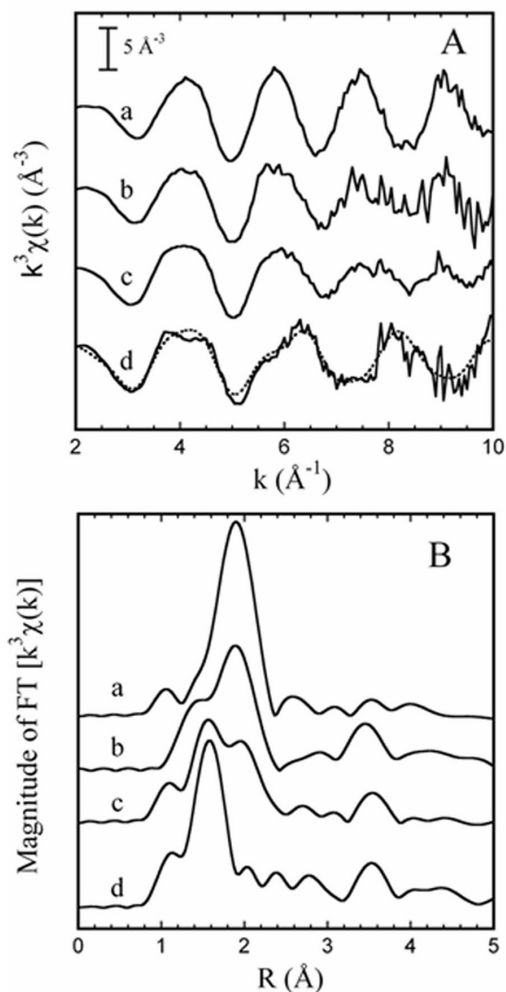


Figure 6.11: (A)  $k^3$ -weighted  $\chi(k)$  EXAFS signals obtained from samples characterized by different Zn/COX ratios. (B) The corresponding magnitude of Fourier transform (FT). From top to bottom: trace a, sample a, characterized by 1.0 Zn/complex; trace b, sample b, 1.5 Zn/complex; trace c, sample c, 2.0 Zn/complex. The first-shell fit to the FT of sample a is shown with empty squares in Fig. 6.11B. The vertical dashed line indicates the position of the peak attributed to the binding of exogenous zinc. The amplitude of this peak increases from sample b to sample c, reflecting the progressive occupancy of a new  $\text{Zn}^{2+}$  binding site. Trace d in panel A is the EXAFS signal of the site to which exogenous  $\text{Zn}^{2+}$  binds, extracted from trace c by subtracting trace a with the appropriate weights. The fit corresponding to cluster 1 of Tab. 6.7 (i.e. three histidines and one aspartate or glutamate residue) is shown as a dashed line. See text for details. Trace d in panel B is the correspondent FT.

to c), being comparable to that of the peak of the endogenous Zn binding site. This strongly suggests that a second, high affinity binding site becomes progressively populated. The EXAFS signal for this site ( $\chi_{site2}$ ) can be extracted by subtracting the signal of sample a ( $\chi_a$ ) from those, appropriately weighted, of sample b ( $\chi_b$ ), or sample c ( $\chi_c$ ) according to:

$$\chi_{site2} = (1.5\chi_b - \chi_a)/0.5 \quad (6.1)$$

$$\chi_{site2} = 2\chi_c - \chi_a \quad (6.2)$$

The EXAFS signal obtained from the sample characterized by 2 Zn/complex using Eq. 6.2, is shown in Fig. 6.11A (trace d). The similarity of the EXAFS signal extracted from sample b or from sample c using Eq. 6.1 or Eq. 6.2 (not shown) clearly indicates that a unique, additional high affinity site is occupied in our samples. First-shell analysis indicates a coordination number of four and the best fit was obtained for a cluster type formed by 3N at  $1.98 \pm 0.01$  Å and 1O at  $2.03 \pm 0.05$  Å. Among nitrogen/oxygen ligands of zinc in proteins, the  $N_{\delta 1}/N_{\epsilon 2}$  of the histidine imidazole ring and the  $O_{\delta 1}/O_{\delta 2}$  of the aspartate and glutamate residues are the most common. Coordination of  $Zn^{2+}$  by O from water has also been observed, particularly for non-structural zinc [25]. The mean Zn – N distance resulting from first shell analysis is however slightly shorter than the typical bond length reported for histidine, i.e. 2.05 Å for Zn –  $N_{\epsilon 2}$  and 2.14 Å for Zn –  $N_{\delta 1}$  bonds [86]. This suggests the presence of at least one unusually shorter Zn –  $N_{\epsilon 2}$  bond or that one of the coordinating N atoms belongs to a lysine, whose Zn – N bond length is typically 1.94 Å [86] and that the coordination with the histidine residues involves  $N_{\epsilon 2}$ . The EXAFS signal was therefore fitted to a series of putative binding clusters formed by the  $CCOO^-$  of an aspartate or glutamate residue or the O atom of a water molecule and three imidazoles

or two imidazoles plus a lysine.

Cluster	Residues	Zn–N (his) (Å)	Zn–N (lys) (Å)	Zn–O (Asp/Glu) (Å)	Bending angle (His) (°)	R factor
1	3his(3N <sub>ε2</sub> )1glu/asp	1.97(1)		2.03(3)	3(4)	14
2	3his(1N <sub>δ1</sub> , 2N <sub>ε2</sub> )1glu/asp	1.97(1)		2.03(3)	4(3)	14
3	3his(2N <sub>δ1</sub> , 1N <sub>ε2</sub> )1glu/asp	1.97(1)		2.03(3)	4(3)	15
4	3his(3N <sub>δ1</sub> )1glu/asp	1.97(2)		2.01(3)	4(3)	17
5	2his(2N <sub>ε2</sub> )1lys1glu/asp	2.00(3)	1.93(5)	2.02(5)	3(7)	15
6	2his(1N <sub>δ1</sub> , 1N <sub>ε2</sub> )1lys1glu/asp	2.00(3)	1.93(5)	2.02(5)	4(5)	15
7	2his(2N <sub>δ1</sub> )1lys1glu/asp	2.00(3)	1.93(5)	2.02(5)	5(4)	16

Table 6.7: (A) Structural results obtained for the inhibitory zinc site of cytochrome c oxidase. The  $1\sigma$  error on the least significant figure is reported in brackets.

As shown in Tab. 6.7, which summarizes the binding clusters considered, except those including a water molecule, an excellent fit is obtained in all cases. The corresponding clusters in which the aspartate or glutamate residue was replaced by water yielded equivalent fits, but the best fitting Zn – O distances were systematically shorter (1.98 – 2.02 Å) than the typical coordination distance derived from a large protein data set of tetrahedral Zn<sup>2+</sup> binding sites (2.12 Å for catalytic and 2.15 Å for structural sites [86], see also <http://tanna.bch.ed.ac.uk/>). The presence of a water molecule seems therefore unlikely. Using as a goodness of fit index the R factor, it appears that the clusters including more than one imidazole N<sub>δ1</sub> as ligand are unfavoured.

#### 6.4.4 Discussion

By EXAFS analysis we have extracted a signal attributable to a site which binds exogenous Zn<sup>2+</sup> with a tetrahedral geometry, formed by a carboxylate residue (glutamate or aspartate) and three histidines (clusters 1–4) or two histidines plus a lysine (clusters 5–7). We cannot discriminate between

these two possibilities, since the R factor favours cluster 1, while including a lysine led to a Zn-NLys and Zn-NHis distance that agrees better with the expected values [86](see Tab. 6.7). The fit corresponding to cluster 1 is shown as a dashed line in Fig. 6.11A.

On the basis on EXAFS results we tried to identify the inhibitory zinc site using the indications supported by the inhibition mechanisms (for details see *Francia et al, FEBS Letters 581: 611-616. 2007*). An internal binding site can be conceived to be located at the entry mouth of the D channel. Guided by the EXAFS characterization of the possible Zn-coordinating residues, we have inspected, by RasMol analysis of the X-ray crystallographic structure of bovine cytochrome c oxidase, the position of candidate residues exposed at the N surface at the entrance site of the D pathway. As shown in Fig. 6.12, there are His 503 and Glu 506 of sub-

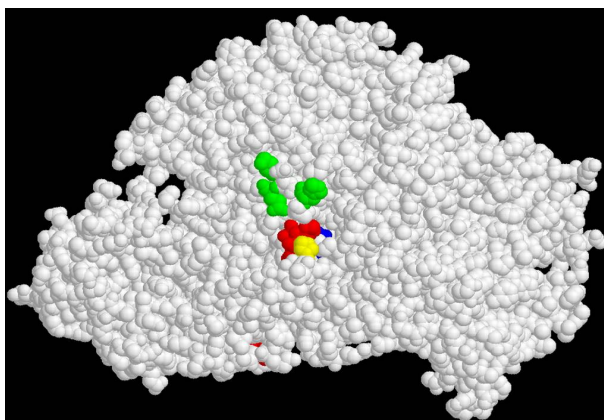


Figure 6.12: Space-filled view of the mass of the 13 subunits bovine cytochrome c oxidase protruding at the N side of the membrane. The picture was elaborated with the RasMol 2.6 program using the 2OCC atomic coordinates from the PDB data bank of cytochrome c oxidase [32]. With their respective numbering the putative Zn binding residues emerging at the N surface of subunits I (red), subunit VIIc (green) and subunit III (yellow) and Asp91-I (blue) of the D channel are shown.

unit I, His 2, Glu 5 and Lys 9 of subunit VIIc and His 3 of subunit III,



Figure 6.13: View parallel to the membrane of the location in subunits I (black) and VIIc (green) of the oxidase of the putative Zn<sup>2+</sup> coordinating residues. The picture was elaborated as in Fig. 3A. It can be noted that the putative Zn coordinating residues, except Lys 13-I, belong to the carboxy-terminal segment of subunit-I and the NH<sub>2</sub>-terminal segment of subunit VIIc respectively.

which could be candidates, in various combinations, for tetrahedral coordination of  $\text{Zn}^{2+}$ . Lys 13 of subunit I is close to the D-pathway entrance and could, even if not exposed to the N-surface in the crystal structure, participate in  $\text{Zn}^{2+}$  coordination (Fig. 6.13). Since all the residues indicated in Fig. 6.12 are located in non-structured segments of the protein backbone, conformational perturbation induced by the binding of  $\text{Zn}^{2+}$  itself are possible. Our ZnCOX complexes are embedded in a weakly interacting matrix as PVA; this situation differs from the crystallographic one, in which conformational movements can be partially hindered. A  $\text{Zn}^{2+}$  coordination involving two His N-imidazols, one Glut O-carboxylate, and one N-histidine (or one N-lysine) seems to be favoured by the EXAFS data. Recent X-ray crystallographic analysis shows  $\text{Zn}^{2+}$  coordination to His 503-I, in different combinations with residues of subunit VIIc and III shown in Fig. 6.12 and Fig. 6.13 (Shinya Yoshikawa, personal communication). Zn tetrahedral coordination with these residues at/around the entrance of the D pathway can explain the reported inhibition of the  $\text{P} \rightarrow \text{F}$  and  $\text{F} \rightarrow \text{O}$  transitions of the catalytic cycle (see *Francia et al, FEBS Letters 581: 611-616. 2007* and references therein). As shown for *P. Denitrificans* cytochrome c oxidase and for the bovine oxidase (see *Francia et al, FEBS Letters 581: 611-616. 2007*), internal  $\text{Zn}^{2+}$  binding, possibly at the site described here, results also in inhibition of proton pumping. The site is thus involved in the coupling of proton pumping with the reduction of oxygen to water.

## 6.5 The endogenous binding site of $\text{Zn}^{2+}$ in complex I

### 6.5.1 Materials and methods

NADH-quinone oxidoreductase was purified from bovine heart mitochondria according to [91]. The concentration of complex I was estimated on the basis on FMN content, determined spectrophotometrically after extraction, using an oxidized-reduced extinction coefficient of  $9.8 \text{ mM}^{-1}\text{cm}^{-1}$  at 450 nm. Evaluation of Fe content yielded a ratio  $\text{Fe}/\text{complex} = 19 \pm 2$ . This value is compatible with previous estimates ranging from 15 to 19 iron atoms per FMN group, i. e. per complex [91]. Recent XRD data of the hydrophilic arm of the complex, which contains all the redox cofactors, support the presence of 2 binuclear  $\text{Fe}_2\text{S}_2$  centers and 6 tetranuclear  $\text{Fe}_4\text{S}_4$  clusters for a total of 28 Fe atoms per complex. The lower stoichiometry found by analytical methods in the isolated preparations is likely to reflect the loss of one or more subunits during the purification [92]. ICP-AES evaluation of Zn in the untreated preparation gives a ratio  $\text{Zn}/\text{complex} = 0.96 \pm 0.10$ , suggesting that complex I contains a stoichiometric, endogenous zinc atom.

### 6.5.2 Data collection and data analysis

EXAFS measurements were performed at the BM8 GILDA beam-line of the European Synchrotron Radiation Facility (ESRF) in Grenoble. Measurements are described in Paragraph 3.1 The total integration time was 60 s/point. Data analysis was performed using ARTEMIS package according to Chapter 4 and 5. Final fitting was performed directly in k space (with a k weight of 3), in the range  $2.4 - 15.8 \text{ \AA}^{-1}$ . As free parameters we used a single, common shift in the energy origin, the first ligand distances, the



bending angle between the  $\text{ZnN}_{e2e}$  bond and the imidazole ring of histidines (keeping Zn coplanar with the imidazole ring) and the  $\zeta'$  angle of cysteine (see Appendix A). DebyeWaller factors were fixed according to [36].

### 6.5.3 Results

To analyse these data we used the same approach described in Chapter 5 and applied for the  $\text{bc}_1$  and COX complexes (see above). First shell analysis indicated the presence of N and S atoms as first neighbours. These atoms are well distinguishable in R space since their distances from the absorber differ of about 0.3 Å. In fact the mean Zn – S and Zn – N distances indicated in the metal data bank are 2.31 Å and 2.03 Å, respectively. By looking at the metal database one can thus deduce that the only possible ligands are cysteine and histidine residues. However, since the Zn – N and Zn – S EXAFS oscillations are nearly out of phase it is difficult to determine the relative number of sulfur and nitrogen scatterers from a first shell analysis [93].

Hence we performed a complete multi-shell, multiple scattering analysis for the three clusters: 1 cys 3 his, 2 cys 2 his, 3 cys 1 his. The best fit is obtained for the cluster containing 2 histidine and 2 cysteine residues.

ligand cluster	Zn-N (Å)	Zn-S (Å)	$\beta$ (°) His	$\zeta'$ (°) Cys	R factor (%)
2 His (2 N <sub>e2e</sub> ) 2 cys	2.00 (1)	2.31 (1)	121 (5)	87 (4)	33

Table 6.8: Structural results obtained for Complex I. The angles  $\beta$  and  $\zeta'$  are shown in Appendix A.

The difference in the R factor and reduced chi square values of the different

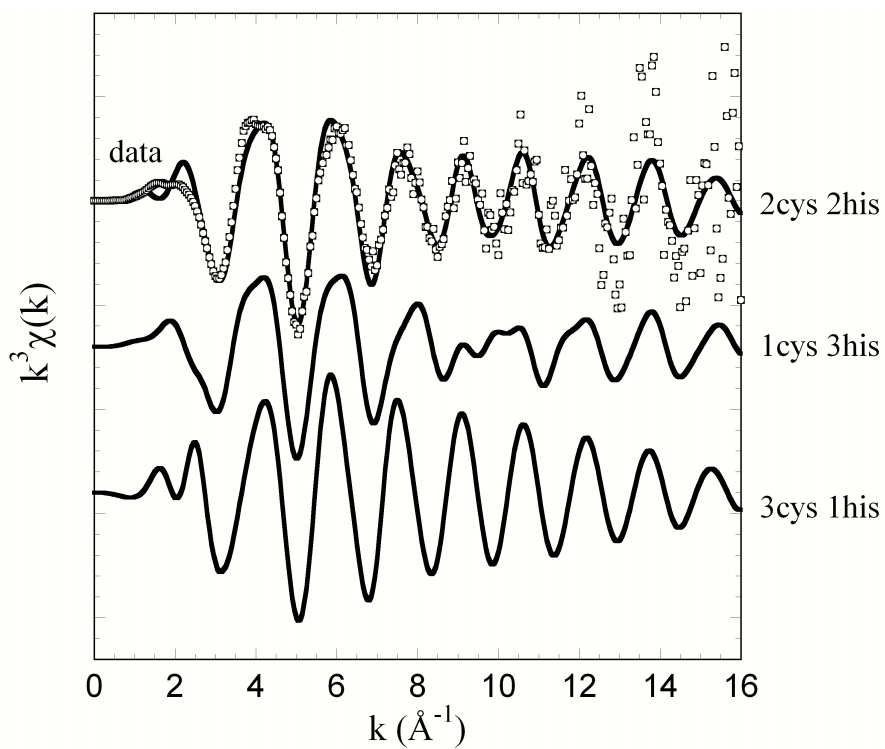


Figure 6.14: From the top to the bottom: experimental (open circles) and calculated best fitting XAFS signal (superimposed continuous line) obtained using the model formed by two cysteine and two histidines; the fitting XAFS signals obtained for the models formed by 1 cysteine 3 histidines and 3 cysteine 1 histidine, respectively (continuous line).

clusters (not shown) clearly indicated that the 2his 2 cys cluster is by far the most favorite one, as can be appreciated by looking at Fig. 6.14, where the experimental data together with the three fits are shown. The results obtained for the cluster which minimize the R-factor are reported in Tab. 6.8.

## Chapter 7

# Matrix effects on the structure and dynamics of heme proteins embedded in dry trehalose

Part of the results illustrated in this Chapter have been published in

*Giachini et al, Biophysical Journal 92: 1350-1360. 2007.*

### 7.1 Overview

Trehalose ( $\alpha$ -D-glucopyranosyl-1,1- $\alpha$ -D-glucopyranoside) is a non-reducing disaccharide consisting of two glucose molecules linked by an  $\alpha,\alpha$ -1,1 glycosidic bond (see Fig. 7.1). Trehalose was initially thought to be an uncommon sugar. It is now known that trehalose occurs in a large variety of

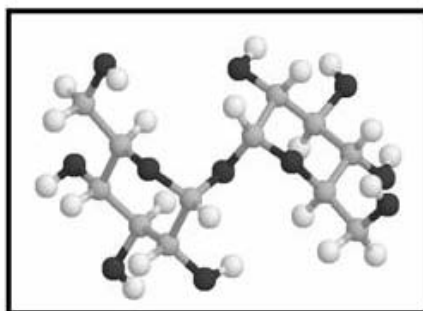


Figure 7.1:  $\alpha$ -D-glucopyranosyl-1,1- $\alpha$ -D-glucopyranoside. Oxygen atoms are in dark gray, carbon atoms in gray, hydrogen atoms in white.

organisms, such as fungi, bacteria, yeasts, invertebrates, algae and vascular plants. Trehalose is similar to sucrose, another common non-reducing disaccharide consisting of one molecule of fructose and one molecule of glucose. Compared to sucrose, trehalose has more interesting chemical-physical properties, being, for example, only mild sweet and not hydrolyzed to reducing sugars at elevated temperatures. Trehalose is an extremely stable sugar, non toxic and chemically inert, therefore it is an ideal preservative for several biomolecules as proteins, enzymes, flavours and fragrances [94], [95], [96]. From 2001 trehalose has been introduced as new human food by the European Union. However application of trehalose in food preservation is still limited due to its high commercial value. As a consequence, the use of this disaccharides is limited to high value products, such as pharmaceuticals and cosmetics.

Among the many interesting properties of this sugar, the most surprising is that it is synthesized by many organisms (see Fig. 7.2 and Fig. 7.3) which can undergo anhydrobiosis, i.e. a condition enabling them to preserve their structural and functional integrity over several years of dehydration, surviving high temperatures ( $60^{\circ}$ ) and resuming their metabolic activity

when rehydrated [97], [98]. When in anhydrobiosis such organisms contain a large (up to 20% w/w) trehalose concentration. In spite of the large at-

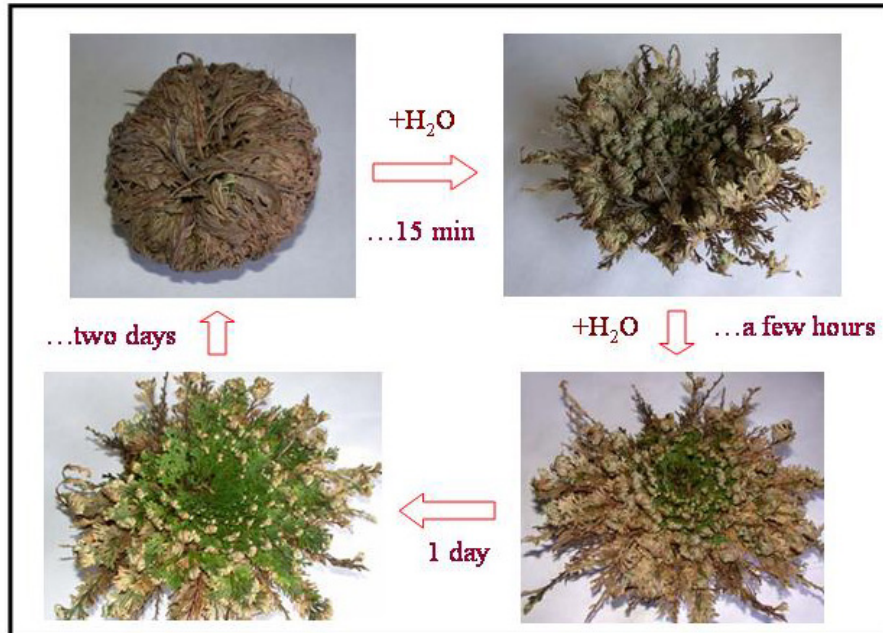


Figure 7.2: Desert resurrection plant (*Selaginella*). Anhydrobiotic organism which can undergo anhydrobiosis. It contains large quantities (as much as 20% of the dry weight) of trehalose,

tention devoted, the molecular mechanisms by which carbohydrate matrices impart dehydration and temperature stability are still matter of intensive studies; particularly elusive is the molecular basis of the peculiar efficacy of trehalose in bioprotection. According to the water replacement hypothesis, stabilization occurs via the direct hydrogen bonding between the sugar and the bio-structure [99]. In a different model (water-entrapment hypothesis) residual water is supposed to concentrate close to the surface of the embedded macromolecule in the dried trehalose matrix [100]. Viscosity effects, inhibiting dynamical processes which lead to denaturation, have been also invoked [101]. According to this view the peculiarity of trehalose has

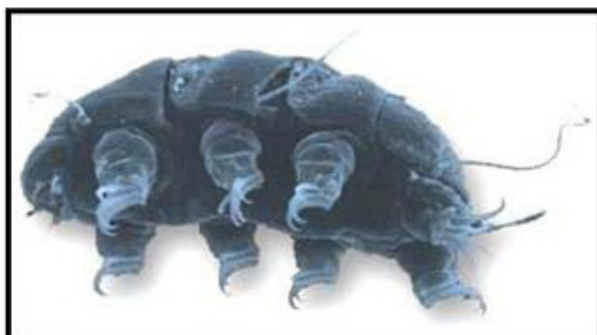


Figure 7.3: Tardigrade or water bear. Like the resurrection plant it contains large quantities of trehalose and it can undergo anhydrobiosis.

been put in relation to its rather high glass transition temperature with respect to other glass forming sugars [102]. In the recent past, a large number of experimental and Molecular Dynamics (MD) studies has focused on protein-trehalose-water matrices. These studies, mainly performed on heme proteins, definitely established that in trehalose systems of low water content internal protein dynamics are modulated by the content of residual water, being dramatically hindered in extremely dried matrices. Several spectroscopic studies, based on complementary techniques sensitive to atomic motion, support this view. Mössbauer spectroscopy measurements on carboxy myoglobin (MbCO) [103] showed that in dry trehalose matrices the large-scale quasi-diffusive motions of the Fe nuclei, which in the hydrated system appear at about 180 K [104], are largely hindered even at room temperature. An analogous effect has been detected when measuring the protein hydrogen mean-square displacements in an extremely dry trehalose MbCO sample by elastic neutron scattering [105]: again the large scale motions that in the hydrated protein appear at about 180 K [106] resulted hindered even at room temperature. A similar behaviour has been recently observed in a neutron scattering study performed on C-phycoyanin incorporated into

a trehalose dry matrix [107]. These observations have been rationalized by MD simulations performed on a MbCO-trehalose-water system which showed that: (a) the amplitude of anharmonic motions stemming from the interconversion among the protein's conformational substates is greatly reduced in a trehalose matrix of low water content [108]; (b) in such systems structures are formed in which the protein is confined within a network of hydrogen bonds connecting groups of the protein surface, water and trehalose molecules. It has been proposed that these structures play a role in coupling the structure-dynamics of the protein to that of the matrix [109]. The occurrence of a tight dynamical and structural coupling within trehalose-water-MbCO structures has been further supported by the strict correlation observed between the thermal evolution (300-20K) of the CO stretching and of the water association band in samples of MbCO embedded in trehalose matrices of different hydration [110]. In agreement with these observations, a recent study of spectral diffusion dynamics, performed on trehalose-coated cytochrome (cyt) *c* suggests that also in this heme protein diffusive-like motions are strongly reduced in trehalose matrices as compared to glycerol-water glasses and that the solvent changes some structural features of the protein [111]. These results have indicated that embedding a protein in a trehalose-water system can be a useful tool to investigate function-dynamics relationships at room temperature as a function of the rigidity of the embedding matrix. This approach has been applied to both soluble and integral membrane proteins, allowing to disentangle rigidity from temperature effects (for a review see [112]). The results of these functional studies are consistent with the proposal that in trehalose protein dynamics are essentially conditioned by a hydrogen bond network which anchors the protein surface to the surrounding matrix. In particular, studies on the kinetics of electron



transfer in bacterial photosynthetic reaction centres embedded in different amorphous matrices evidenced that the presence of trehalose in the matrix plays an essential role in hindering internal protein dynamics; in fact, it was shown that the reaction centre dynamics coupled to electron transfer are only marginally affected in reaction centre films dried in the absence of the sugar [113] or in polyvinyl alcohol (PVA) matrices [44], even at extremely low content of residual water. Studies performed in bacterial reaction centres have clearly shown that electron transfer processes, involving different cofactors and occurring in different protein regions, are differently inhibited by dehydration of the trehalose matrix [114], suggesting that specific, local protein dynamics are selectively modulated in the trehalose-water-protein system. That incorporation of the protein into a trehalose matrix can cause subtle structural and dynamical changes localized to specific protein regions is also suggested by MD studies performed in parallel on an H<sub>2</sub>O-solvated and on a trehalose-coated MbCO [108]. By examining structure and dynamics effects specifically on the heme group, it was shown that while mean thermal fluctuations of the heme heavy atoms were decreased in the amorphous system at 300 K as compared to solution, the opposite occurred to the mean square fluctuations of the heme H atoms, which exhibited a larger amplitude in trehalose. This observation, suggesting a different configuration of the heme in H<sub>2</sub>O-solvated and trehalose-coated MbCO, has been put in relation with the finding that, in MD simulations of a trehalose-MbCO system, one trehalose molecule has been found in strong interaction with the heme propionate oxygen atoms [109]. As a whole, the results summarized above point out the necessity for a deeper knowledge of matrix-induced structural and dynamical effects at the local atomic scale. X-ray absorption fine structure (XAFS) spectroscopy appears to be particularly suited

to provide such information, in view of its applicability to non-crystalline samples and of its high sensitivity in probing the local structure of a metal ion in a protein in the first few coordination shells [115], [59]. Since the XAFS function contains Debye-Waller factors, it can also yield information on thermal fluctuations and on the static disorder which contribute to the relative mean square displacements of the atoms [13]. It is worthwhile to recall that XAFS measures only relative mean square displacements between the absorbing and backscattering atoms, thus providing information distinct and complementary to that contained in the Debye-Waller factors determined by Mössbauer and by X-ray diffraction [32].

In this part of the thesis work we have compared the local structure and dynamics around heme-Fe in two extensively studied proteins (Carboxymyoglobin (MbCO) <sup>1</sup> and cytochrome *c* (cyt *c*) <sup>2</sup>) in solution, when incorporated into a weakly interacting medium (polyvinyl alcohol, PVA) and when embedded in a strongly interacting matrix such as extremely dehydrated trehalose glass. In the case of cyt *c* a more extensive study at room

---

<sup>1</sup>Myoglobin (Mb) is a small protein (17.6 kDa) consisting of a single polypeptide chain of 153 amino acid residues and a protoporphyrin IX (heme) prosthetic group. It functions in cardiac muscle to store oxygen ( $O_2$ ) and to facilitate its diffusion from blood capillaries to the mitochondria. The heme is linked to the protein by a covalent bond from the iron atom to an imidazole nitrogen ( $N$ ) of the proximal histidine. The sixth position can be occupied by  $O_2$  (Oxy-Mb),  $H_2O$  (met-Mb) or  $CO$  (MbCO).

<sup>2</sup>Cytochrome *c* (cyt *c*) is a small heme protein (MW 12 384) that functions as a biological electron-transfer agent. It consists of a single polypeptide chain and a prosthetic heme group and provides a pathway for the transfer of electrons from cyt *c* reductase to cyt *c* oxidase in the mitochondrial respiratory chain (oxidative phosphorylation). In the native protein, histidine 18 and methionine 80 of the protein chain are bound as axial ligands to produce a sixcoordinate Fe center (see Fig. 7.4).

temperature has been performed (see the next paragraph). In the case of MbCO we have performed additional measurements at three temperatures in order to compare the dynamical effects produced by the matrix with the thermal ones. The main results of these studies are summarized in the next paragraphs.

## **7.2 Cytochrome *c* in a dry trehalose matrix: structural and dynamical effects probed by x-ray absorption spectroscopy**

In this paragraph we report the results of a comparative XAFS Fe K-edge study of cyt *c* in solution, in a weakly interacting matrix (PVA film) and in two trehalose matrices characterized by a different content of residual water. We have found that embedding the protein into the trehalose matrix leads to a significant structural distortion of the porphyrin group, more pronounced in the more dehydrated system. In this strongly interacting matrix, we observed an impressive reduction of the mean square relative displacements of the atomic correlations involving the central Fe and the first shell ligands.

### **7.2.1 Materials and Methods**

Horse heart cytochrome *c* (Sigma-Aldrich, St. Louis, Missouri, USA) was dissolved in 10mM Tris buffer, pH 7.9. A molar excess of sodium ascorbate was added to fully reduce the hemic iron; the solution was eluted through a Sephadex G25 column (Amersham-Pharmacia PD10, Uppsala, Sweden) and the ascorbate free, ferrocytochrome *c* solution used to prepare all the samples. The concentration of ferrocyt *c* (3.1 mM) was evaluated spectrophotometrically on the basis of the reduced-oxidized absorption peak at

551 nm, using a differential extinction coefficient of  $19.5 \text{ mM}^{-1} \text{ cm}^{-1}$ .

Starting from a 3.1 mM ferrocyt *c* mother solution we prepared five samples characterized by different matrix environments for the protein: two liquid samples, a polyvinyl alcohol film and two amorphous trehalose matrices differing in the extent of dehydration. The first liquid sample (sol. 1) was the mother solution itself; the other liquid sample (sol. 2) was obtained by addition of glycerol to produce a 40% v/v glycerol/water mixture. Solutions were inserted into a 800  $\mu\text{l}$  plastic cell with kapton windows. The ferrocyt *c* containing PVA film (named PVA) was prepared by supplementing the mother solution with 2.5% w/v PVA ( $M_w \approx 130,000$ , Fluka, Buchs, Switzerland) and drying 1.4 ml of suspension (placed into a  $3 \times 3 \times 0.3 \text{ cm}^3$  Teflon holder) under  $\text{N}_2$  flow. Ferrocyt trehalose samples were obtained by adding 147 mg of trehalose to 250  $\mu\text{l}$  of stock solution to achieve a molar sugar/protein ratio of 500. A first sample (trehal. 1) was formed by layering the solution on a sintered boron nitride support and by dehydrating it under dry  $\text{N}_2$  flow for 3 hours at room temperature. To obtain a more extensively dehydrated sample (named trehal. 2) a sample prepared as described above was placed under high vacuum for 2 hours and further dried under nitrogen flow for 8 hours prior to x-ray measurements.

### 7.2.2 Data collection

Fe K-edge measurements were performed at the BM 8 GILDA beam-line of the European Synchrotron Radiation Facility (ESRF) with the apparatus described in Paragraph 3.1. Samples were measured at room temperature in the region 6900–8200 eV. The edge region of each sample was recorded using an energy step of 0.1 eV to verify that in all samples the prepeak occurred at the same energy and had the same amplitude characteristic of

the reduced form of cyt *c* (see Results and Discussion). Spectra are the average of multiple scans (from three to six), run consecutively on each sample.

### 7.2.3 XAFS model

It is well known that the local cluster around the iron center in cyt *c* (in both oxidized and reduced forms) is made up of a heme group, a histidine residue, and a methionine residue (see Tab. 7.1). In Table 1 we report the interatomic distances between iron and the first ligands for ferrocyt *c* from various species according to the Protein Data Bank (PDB). The significant variability in the interatomic distances reported might reflect real differences in the structure of cyt *c* from different organisms. However, it should be recalled that macromolecular crystallography and nuclear magnetic resonance (NMR) are unable to solve local structure details on the scale of 0.01- 0.1 Å. Because we are interested in subtle structural changes induced by different matrices, an accurate starting model is highly desirable; we used the iron cluster atomic coordinates of horse heart ferrocyt *c* described by Cheng et al. [116] using multiple scattering analysis (MS) of XAFS data (obtained in a water-glycerol mixture at a 10 K temperature). The geometry of the cluster is shown in Fig. 7.4. The S<sub>δ</sub> and N<sub>ε2</sub> atoms of the methionine and histidine residues are on the axis (*z*) normal to the heme plane, and the histidine imidazole lies on the *x-z* plane. The values of distances and angles are reported in the figure caption. We note that in obtaining these values, constraints were applied to keep both the porphyrin and the imidazole group planar and to maintain a high symmetry in the heme structure [116], [117].

Table 7.1: First ligand distances for the iron site in ferrocytochrome c obtained from the PDB.

PDB ID	Crystal/solution	Organism	Reference	Experimental method	Fe-N <sub>p</sub> (Å)	Fe-N <sub>ε2</sub> (Å)	Fe-S <sub>δ</sub> (Å)
1KX2	solution	Shevanella	Bartalesi et al., 2002	NMR	2.07	1.96	2.37
1GIW	solution	Horse	Banci et al., 1999	NMR	2.08	1.94	2.36
2FRC	solution	Horse	Qi et al., 1996	NMR	1.95	1.93	2.23
5CYT	crystal	Tuna	Takano et al., 1984	XRD	2.04	1.98	2.31
1CYC	crystal	Bonito	Tanaka et al., 1975	XRD	1.81	2.60	2.49

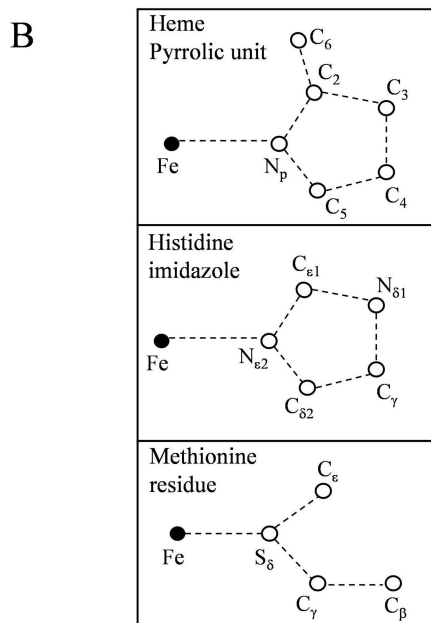
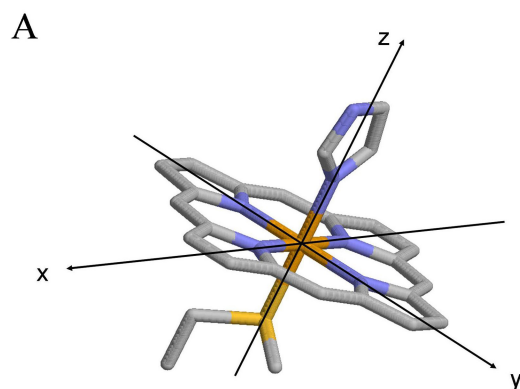


Figure 7.4: A: Reference structural model of the Fe ligand cluster. B: Sketch of the structural units of the reference model. The values of distances and angle are the following. Heme pyrrolic unit:  $\text{Fe-N}_p = 1.99 \text{ \AA}$ ;  $\text{N}_p\text{-C}_{2,5} = 1.38 \text{ \AA}$ ;  $\text{C}_{2,5}\text{-C}_{3,4} = 1.44 \text{ \AA}$ ;  $\text{C}_{3,4} = 1.34 \text{ \AA}$ ;  $\text{C}_2\text{C}_6 = 1.38 \text{ \AA}$ ;  $\text{FeN}_p\text{C}_{2,5} = 109^\circ$ ;  $\text{FeN}_p\text{C}_{3,4} = 163^\circ$ ;  $\text{N}_p\text{C}_2\text{C}_6 = 125^\circ$ . Histidine imidazole:  $\text{Fe-N}_{\epsilon 2} = 2.00 \text{ \AA}$ ;  $\text{N}_{\epsilon 2}\text{-C}_{\epsilon 1} = 1.32 \text{ \AA}$ ;  $\text{N}_{\epsilon 2}\text{-C}_{\delta 2} = 1.37 \text{ \AA}$ ;  $\text{C}_{\epsilon 1}\text{-N}_{\delta 1} = 1.34 \text{ \AA}$ ;  $\text{C}_{\delta 2}\text{-C}_{\gamma} = 1.35 \text{ \AA}$ ;  $\text{C}_{\gamma}\text{-N}_{\delta 1} = 1.35 \text{ \AA}$ ;  $\text{FeN}_{\epsilon 2}\text{C}_{\epsilon 1} = 128^\circ$ ;  $\text{FeN}_{\epsilon 2}\text{C}_{\delta 2} = 127^\circ$ ;  $\text{FeN}_{\epsilon 2}\text{N}_{\delta 1} = 163^\circ$ ;  $\text{FeN}_{\epsilon 2}\text{C}_{\gamma} = 162^\circ$ . Methionine residue:  $\text{Fe-S}_{\delta} = 2.29 \text{ \AA}$ ;  $\text{S}_{\delta}\text{-C}_{\gamma} = 1.81 \text{ \AA}$ ;  $\text{S}_{\delta}\text{-C}_{\epsilon} = 1.82 \text{ \AA}$ ;  $\text{C}_{\gamma} - \text{C}_{\beta} = 1.5 \text{ \AA}$ ;  $\text{FeS}_{\delta}\text{C}_{\epsilon,\gamma} = 109^\circ$ ;  $\text{S}_{\delta}\text{C}_{\gamma,\beta} = 110^\circ$ .

#### 7.2.4 Data analysis

The XAFS oscillations were extracted and analyzed as described in Chapter 4. The amplitude reduction factor ( $S_0^2$ ) was calculated by FEFF 8.2 from atomic overlap integrals. This gave a value of 0.94, which has been kept fixed during the analysis. In the reference structural model reported in Fig. 7.4, the most distant atom ( $C_\beta$  of methionine) is at 4.71 Å from the iron. It has been shown that in heme proteins atoms up to 5 Å from the iron center may give rise to significant MS contributions [1]. Thus, all paths involving up to five scattering processes with an effective path length of 5 Å were included for the fitting procedure [117], [1]. The fits were performed directly in  $k$  space (with a  $k$  weight of 3) to avoid truncation and loss of information through Fourier filtering [32]. The fitting range was 2.5–13 Å<sup>-1</sup> for all samples. As fitting parameters we used 1) a common shift in the energy origin,  $\Delta E$ ; 2) a number of distance variations to account for structural distortions and 3) some Debye Waller (DW) factors, properly grouped in shells to account for the effect of thermal motion of the atoms [13] and/or structural disorder. As a goodness-of-fit index, we used the R-factor (see Chapter 4). To avoid overinterpretation of the data and to obtain a reliable estimate of structural and dynamical parameters, their number was reduced as described in the following: 1), we initially used rigid body refinement [33], [59], introducing additional degrees of freedom only in a second step and only when necessary (see Results and Discussion); 2), in a first step rigid body refinement was performed with fixed values of DW factors; and 3), the DW factors of atoms at similar geometric positions were combined and successively varied only together [59], [32]. Once the ligand distances obtained with fixed DW factors provided a satisfactory fit, the coordinates were fixed, and only the DW factors were varied together with a common



shift in the energy origin [32]. DW factors were grouped in the following manner. For the first neighbors, we introduced three DW parameters (one for the four pyrrolic nitrogens ( $N_p$ ), one for the histidine nitrogen ( $N_{\epsilon 2}$ ), and the other one for the methionine sulphur ( $S_\delta$ )). The appropriateness of considering two distinct Fe-N distances with individual DW factors (one for the four pyrrolic  $N_p$  and one for the histidine  $N_{\epsilon 2}$ ) was tested by fitting the EXAFS signal to simplified models characterized by 1), a single Fe-N bond length (and correspondent DW) common to the histidine and to the four pyrrolic nitrogens, and 2), two bond lengths (Fe- $N_p$  and Fe- $N_{\epsilon 2}$ ) with a jointly refined DW factor (see Results and Discussion). For the more distant atoms, only the DW factors of the porphyrin group were allowed to vary in the fit, combined in three parameters (DW (C2) DW (C5); DW (C6); DW (C4) DW (C3)). All other DW factors were kept fixed at a value of  $0.01 \text{ \AA}^2$  (45). In rigid body or constrained refinement, the number of parameters is reduced by treating a selected set of scattering atoms as a rigid unit (see Chapter 5). Following this approach, we considered as rigid structural units the imidazole ring, the methionine residue, and the heme pyrrolic unit (Fig. 7.4). As far as the porphyrin structure is concerned, a refinement of the first ligand distances together with the rigidity of the planar pyrrolic unit and the maintenance of the system symmetry imply that we are considering as the only structural degree of freedom the Fe out-of-plane axial displacement. In the case of histidine and methionine residues, along with the first ligand distances, angular parameters defining the bending on the x-z plane and twisting of the units around the  $N_{\epsilon 2}$  or  $S_\delta$  atom (see Fig. 7.4) should also, in principle, be considered. However, for clusters containing a heme group, the XAFS signal is only weakly sensitive to the scattering paths that include atoms not belonging to the porphyrin ring. This makes

the determination of the parameters of the axial ligands rather difficult (see Results and Discussion and [1]). In a first analysis, we refined, as structural parameters, only the first ligand distance of the imidazole ring and of the methionine residue, along with that of the four pyrrolic rings. Subsequently, the bending angles of the imidazole ring and of the methionine residue were also added to the free parameters. Including these additional parameters did not lead to a decrease of the R-factor (see Results and Discussion). The structural degrees of freedom described above were sufficient to correctly reproduce the XAFS signals of the iron cluster in all samples except for the dried trehalose matrices. To account for the spectra measured in these samples, the constraint of rigid pyrrolic rings had to be released, and a minimum of three additional degrees of freedom in the heme structure were introduced (see Results and Discussion).

### 7.2.5 Results and Discussion

In Fig. 7.5 we show the near edge region of the spectrum (XANES) of the sample in solution with glycerol (sol. 2); the spectra measured in the other samples are similar in this region. XANES features have been shown to give information on the oxidation state of cyt *c*. In fact, a shift to lower energy ( $\sim 1$  eV) of the edge transition has been observed in going from ferricyt to ferrocyt *c* form [116]. Moreover, the pre-edge region exhibits a pre-peak (see inset of Fig. 7.5), due to the dipole-forbidden 1s - 3d transition, which occurs at slightly lower energy (7112 eV) in ferrous form. Also, the amplitude of the pre-peak is somewhat higher in the reduced state.

We conclude that the spectra of all our samples show the typical XANES features reported for ferrocyt *c*. The XAFS oscillations for all samples are shown in Fig. 7.6. Only small differences (mainly around  $8 \text{ \AA}^{-1}$ ) are de-

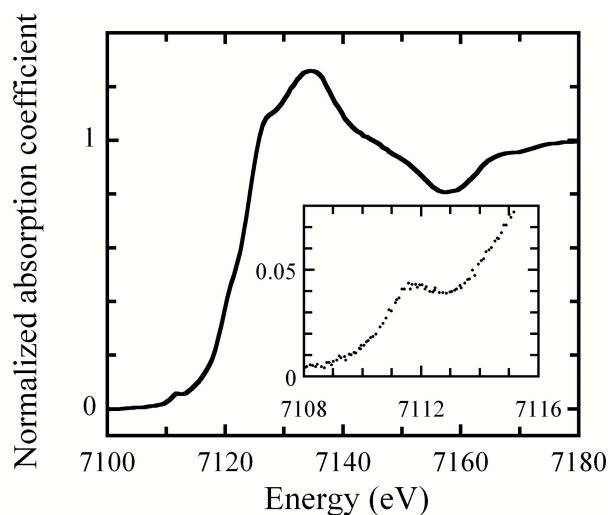


Figure 7.5: Edge and enlarged preedge (inset) spectra for ferrocyanochrome *c* in solution with glycerol.

tectable in the spectra of the solutions in the absence (sol.1) and in the presence (sol.2) of glycerol and the PVA sample. The most evident change is observed in the extra dry trehalose matrix (trehal.2). This sample exhibits a large increase of the oscillation at  $11.5 \text{ \AA}^{-1}$ . Changes are also evident around  $8 \text{ \AA}^{-1}$ , as well as at  $5 \text{ \AA}^{-1}$  where higher frequencies appear. Moreover, damping of oscillations at increasing  $k$  values is strongly reduced in the extensively dried trehalose sample. Far less evident changes are detectable in the moderately dried trehalose matrix. Some features of this sample seem to indicate an intermediate situation between the extensively dried trehalose matrix and solution or PVA matrix: the two oscillations at  $8 \text{ \AA}^{-1}$ , clearly detectable in trehal.2, start to separate in trehal.1 and the signal is more structured at high  $k$  values as compared to the solution or PVA samples.

These qualitative observations suggest that the local structure and dynamics of the iron site in ferrocyanochrome *c* is only marginally affected by the pres-

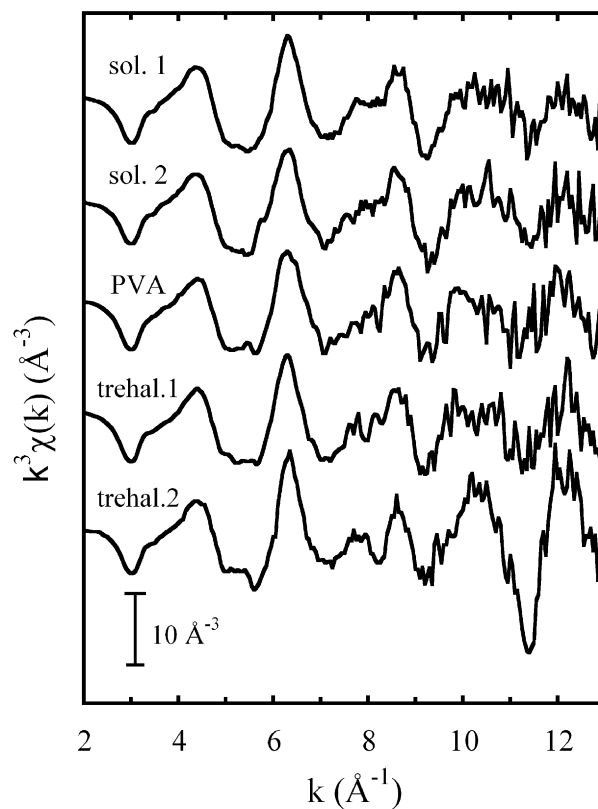


Figure 7.6:  $k^3$  weighted XAFS for all measured samples. sol. 1: solution without glycerol; sol. 2: solution in the presence of glycerol; PVA: protein embedded into a polyvinyl alcohol film; trehal. 1: protein incorporated into a moderately dried trehalose matrix; trehal. 2: protein embedded into an extensively dried trehalose matrix. See Materials and Methods for details concerning sample preparation.

ence of glycerol in solution or by incorporation of the protein into a solid PVA matrix. Therefore, we expect the XAFS signals of the solution and PVA samples not to differ substantially from a signal generated *ab initio* from the reference model structure described under Materials and Methods, obtained from XAFS data of ferrocycytochrome in a water/glycerol mixture [116]. Instead, significant differences in the structure and dynamics of the ligand cluster can be reasonably expected for the trehalose matrices. The strongly reduced damping of oscillations suggests in fact that incorporation into an extensively dried trehalose matrix affects substantially the DW factors. At the same time, some of the spectral changes detected in the trehalose matrices (see above) may result from significant structural modifications of the ligand cluster, induced by a strong protein-matrix interaction.

When analyzing these matrix effects it is important to discriminate between changes in the XAFS signal determined by variations in the DW factors or in the interatomic distances. We found *ab initio* XAFS simulations to be particularly useful to disentangle dynamical from structural contributions and proceeded as follows. Simulations were performed by fixing the atomic coordinates according to the reference structural model (see Materials and Methods). As far as values of the DW factors are concerned, a limited number of models was considered. In the first one, all the DWs were set equal to zero, thus without any treatment of disorder; in the second one, DW values were selected on the basis of literature results [32], [116], [117] in order to mimic the dynamical behaviour expected for a protein embedded in a non-interacting medium at room temperature:  $DW(N_p) = 2.5 \times 10^{-3} \text{ \AA}^2$ ;  $DW(N_{e2}) = 3.5 \times 10^{-3} \text{ \AA}^2$ ;  $DW(S_\delta) = 5.0 \times 10^{-3} \text{ \AA}^2$ ;  $DW(C_{2,3,4,5}) = 3.0 \times 10^{-3} \text{ \AA}^2$ ;  $DW(C_6) = DW(C_{\epsilon1}, C_{\delta2}, C_\gamma, N_{\delta1}) = 4.0 \times 10^{-3} \text{ \AA}^2$ . The value of the other single and multiple scattering DWs was  $0.01 \text{ \AA}^2$ .

Besides these two models we built other three sets of DW factors as follows: first, starting from model 2, we fixed to zero the DW factors of the first ligands only, leaving those of the other atoms unchanged (model 3); then we introduced an additional partial damping in the relative mean square displacements of more distant shells, decreasing the DW factors of  $C_{2,3,4,5}$ , of  $C_6$  and of  $C_{\epsilon 1}, C_{\delta 2}, C_{\gamma}, N_{\delta 1}$ ) by  $2.0 \times 10^{-3} \text{ \AA}^2$  in two steps of  $1.0 \times 10^{-3} \text{ \AA}^2$  and the remaining DW factors by  $4.0 \times 10^{-3} \text{ \AA}^2$  in two steps of  $2.0 \times 10^{-3} \text{ \AA}^2$  (model 4 and 5, respectively). The results of these five simulations are shown in Fig. 7.7.

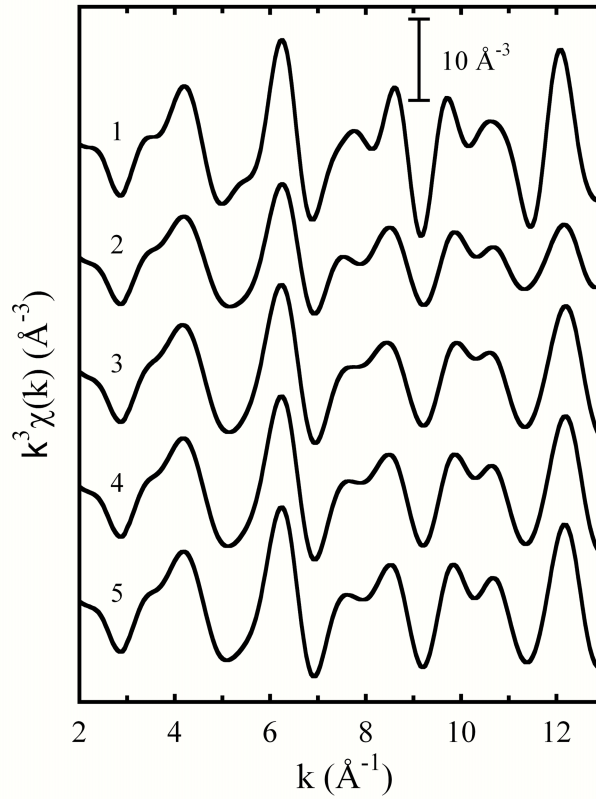


Figure 7.7: Simulations based on the reference structural model using different values for the DW factors.

Changing the DW factors as described, besides affecting the damping of

the main oscillation, causes limited modifications mainly in the region where higher oscillations appear (8 and 10  $\text{\AA}^2$ ). A comparison with experimental signals (cf. Fig 7.6.) shows that model 2 reproduces more reasonably the main features of the spectra measured in solution (sol.1, sol.2) and in the PVA matrix (PVA). None of the simulated signals, however, exhibits the clear separation of the two peaks at 8  $\text{\AA}^2$  which is observed experimentally in the extensively dried trehalose matrix (trehal.2). Moreover, in all the simulations the amplitude of the negative peak at 9  $\text{\AA}^2$  is comparable or larger than that of the peak at 11  $\text{\AA}^2$ . Just the opposite occurs in the spectrum measured in the extra-dry trehalose sample, which exhibits a peak at 11  $\text{\AA}^2$  largely exceeding in amplitude that at 9  $\text{\AA}^2$ . It appears therefore that a decrease of the DW factors, which of course can account for the decreased damping of oscillation observed in the trehalose matrices, is unable to reproduce some prominent features of the spectrum measured in the extensively dried sugar sample. These observations suggest that in trehalose the protein experiences, along with an enhanced local order, also significant structural distortions. Analysis of the simulated spectra in terms of the partial contributions arising from the different scattering paths can provide useful hints in identifying putative structural changes responsible for the spectral modifications observed in the trehalose samples. Fig. 7.9 shows the simulation performed on the basis of the reference structure using the DW factors of model 1 (dashed line) and model 2 (continuous line) described above, together with the eight main partial contributions. They all come from the heme structure. A quantitative inspection of the relative amplitude of all the contributions confirms that the spectrum is dominated by paths involving atoms of the heme group. This appears to be a general property of heme proteins [1]. It suggests that the XAFS signal is only

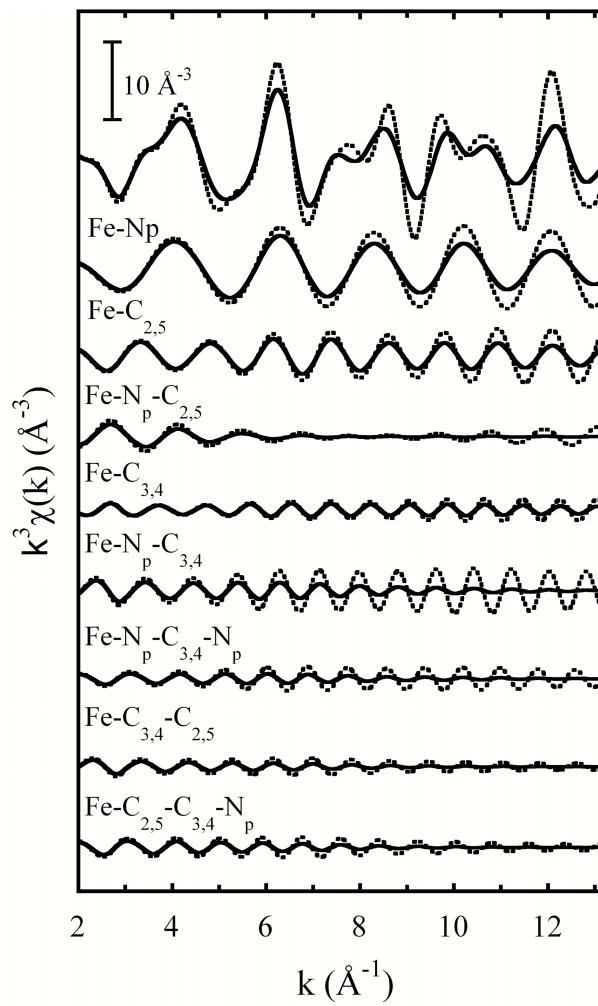


Figure 7.8: Simulations based on the reference structural model (see Fig.1). The values of the DW factors have been fixed as in model 1 (dashed line) and 2 (continuous line) (see the legend of Fig. 4). Together with the total XAFS signal (top) we show the main partial contributions.



weakly sensitive to structural modifications involving atoms not belonging to heme structure. As to the axial ligands histidine and methionine, we noticed that the first ligand contributions involving these residues have an amplitude which is comparable to that of some MS heme paths, while the single scattering and MS contributions involving the most distant atoms of the histidine and methionine are all below 15% relative to the contribution of greatest amplitude. Effects due to changes in the orientation of the histidine residue relative to the heme group have been explored by performing a series of simulations in which bending of the imidazole group, in-plane tilting of the histidine and twisting around the heme axis by angles of  $10^\circ$  and  $20^\circ$  were considered independently. These simulations are shown in Fig. 7.8. In the case of twisting movements moderate alterations are induced at low  $k$  values. Interestingly, bending and twisting movements induce small spectral changes again in the region around  $8 \text{ \AA}^{-1}$ . However, we found that all the orientation changes tested had only moderate effects on the spectrum and in no case were they able to reproduce the characteristic features of trehalose matrices. Also when these simulations were performed trying different sets of DW factors (not shown) the same conclusion was reached. Similar tests were performed for the methionine residue and it was found that the effect of the angular distortions were even weaker (not shown). When examining the relative contribution of the different scattering paths for the reference model structure (see Fig 7.9.) it is clear that the main oscillation of the spectrum is due to the four pyrrolic nitrogens of the porphyrin group. Since this oscillation does not show significant phase changes in the different samples (cf. Fig. 7.6), we infer that the heme first ligand distances should not change markedly also when the protein is embedded in the trehalose matrices. To explain the spectral alterations observable in the sugar samples

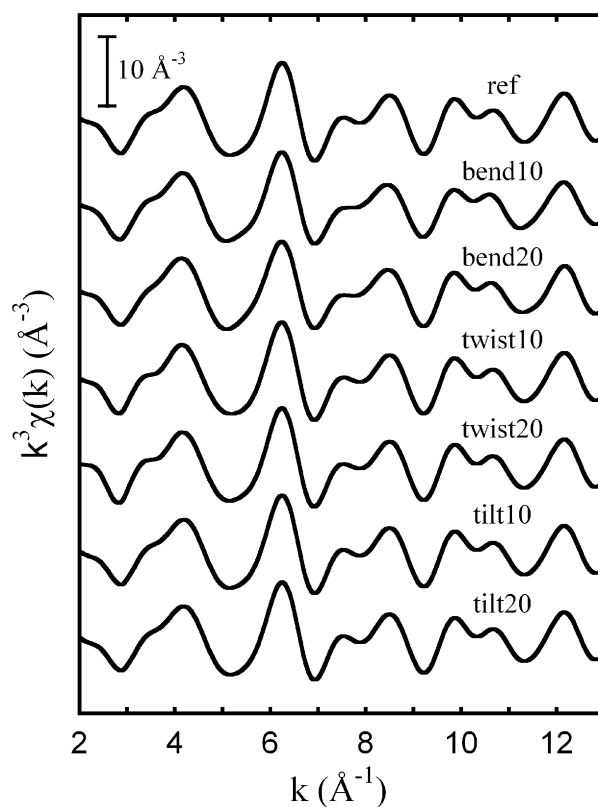


Figure 7.9: The effect of different orientations of the histidine ligand on the XAFS signal. Simulations have been performed starting from the reference structural model (Fig.1) modified as follows. The imidazole ring was rotated around an axis passing through  $N_{e2}$  and parallel to the y axis (bending movement) of  $10^\circ$  and  $20^\circ$  (bend10, bend20); the imidazole was rotated around the z axis (twisting movement) of  $10^\circ$  and  $20^\circ$  (twist10, twist20); the whole histidine residue was rotated around the y axis (in plane tilting) of  $10^\circ$  and  $20^\circ$  (tilt10, tilt20). In all cases DW factors were fixed to the values of model 2 (see legend of Fig. 7.7).

we are left with the possibilities that they are caused by changes of histidine or methionine first ligand distances and/or by structural distortions of the higher shells of the heme group. Since the spectrum is essentially determined by the heme structure, the possibility that a variation of the first ligand distance of the axial residues can account by itself for the strong spectral alterations which characterize the extensively dried trehalose sample is quite unlikely. This was confirmed by the finding that any attempt to fit the signal measured in the extensively dried trehalose matrix keeping the rigid body approach resulted in unacceptably high R-values. Summarizing, the results of these *ab initio* simulations, yield the following guidelines in selecting appropriate fitting procedures: a) a rigid body refinement analysis, based on the reference model described under Materials and Methods, is expected to account adequately for the spectra in solution and in PVA; b) neither changes in disorder parameters nor the considered rotations of histidine and methionine seem sufficient to mimic the prominent features observed in the extensively dried trehalose matrix. Besides possible variations in the first ligand distance of histidine and/or methionine, stereochemical distortions of the heme group have to be considered, thus releasing the constraint of rigid body refinement. In view of this last indication, when analyzing spectra obtained in the trehalose matrices (trehal.1 and trehal.2), we inserted three additional degrees of freedom grouping together the porphyrin atoms ( $C_{2,5}$ ,  $C_6$ ,  $C_{3,4}$ ) in three shells along with the first one and letting the correspondent distances from the iron vary independently. The detailed fitting procedures have been described under Materials and Methods.

In Tab. 7.2 we report the structural parameters and the corresponding R-factors we have determined for the five samples following the procedures described above. Good fits were obtained for all samples. In Fig. 7.10 the

Table 7.2: Structural and dynamical parameters obtained for the five different matrices. The  $1\sigma$  error on the least significant figure is reported in brackets.  $\Delta r$  indicates the change in distance from Fe with respect to the reference model.

	Sol. 1	Sol. 2	PVA	Trehal. 1	Trehal. 2
Structural parameters - interatomic distances ( $\text{\AA}$ )					
Fe-N <sub>p</sub>	1.99 (2)	1.99 (2)	1.99 (2)	1.99 (1)	1.991 (9)
Fe-N <sub>e2</sub>	2.10 (5)	2.10 (5)	2.10 (6)	2.20 (5)	2.29 (5)
Fe-S <sub>8</sub>	2.29 (4)	2.29 (4)	2.29 (5)	2.24 (5)	2.32 (2)
$\Delta r$ C <sub>2,5</sub>	-	-	-	0.01 (1)	0.02 (2)
$\Delta r$ C <sub>6</sub>	-	-	-	0.00 (3)	0.01 (6)
$\Delta r$ C <sub>3,4</sub>	-	-	-	0.04 (7)	0.16 (4)
Disorder parameters - Debye-Waller factors ( $10^{-3} \text{\AA}^2$ )					
DW (N <sub>p</sub> )	3.6(9)	3.2(9)	3.1(9)	2.3(8)	0.0(8)
DW (N <sub>e2</sub> )	10*	20(8)	30(40)	2(4)	0(5)
DW (S <sub>8</sub> )	20(20)	20(10)	30(30)	10(5)	1(2)
DW (C <sub>2,5</sub> )	4(2)	5(2)	5(2)	4(2)	4(2)
DW (C <sub>6</sub> )	8(8)	10(20)	8(9)	2(3)	4(6)
DW (C <sub>3,4</sub> )	10(10)	7(7)	6(6)	7(8)	2(4)
Goodness of fit					
R factor (%)	15	18	14	17	21

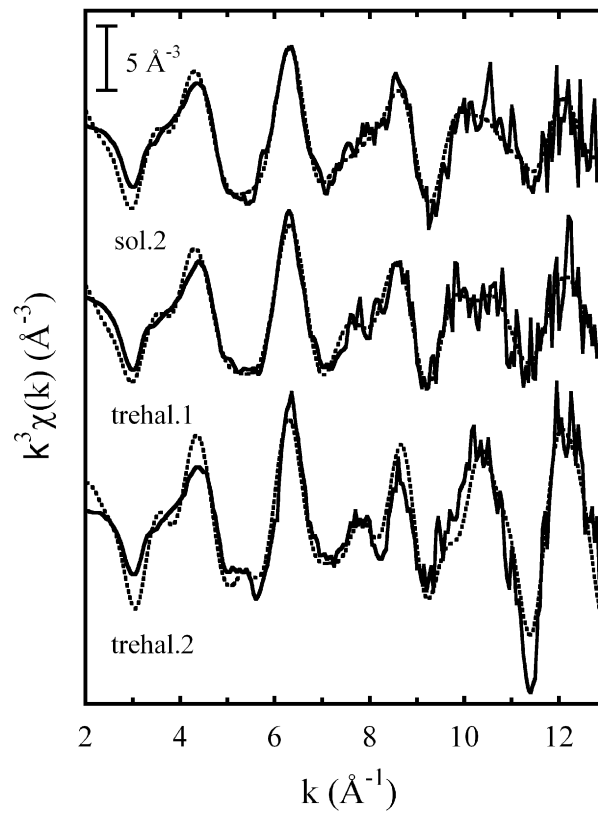


Figure 7.10: Experimental (continuous lines) and calculated best fitting XAFS signals (dashed lines) for the three most significant samples. From the top to the bottom: protein in solution with glycerol (sol. 2), protein embedded in the moderately dried trehalose matrix (trehal. 1) and in the extensively dried trehalose matrix (trehal. 2). See text for details.

three most significant ones are shown. For the samples not embedded in trehalose matrices the best fits were obtained using rigid body refinement keeping the values of the bending angles  $\text{FeN}_{\epsilon 2}\text{C}_{\delta 2}$  and  $\text{FeS}_{\delta}\text{C}_{\epsilon,\gamma}$  for the histidine and methionine residues, respectively (see Fig. 7.4), fixed to the ones of the reference model. Allowing the histidine and methionine bending angles to vary as free parameters did not improve the fit, confirming the validity of the reference model. Inspection of the parameters listed in Tab. 7.2 shows that there are no significant structural differences between samples in solution with and without glycerol and in the PVA matrix. When the rigid body refinement constraint was released, allowing the three distances of Fe from  $\text{C}_{2,5}$ ,  $\text{C}_6$  and  $\text{C}_{3,4}$  to vary, fitting yielded essentially the same values for the first shell distances and no significant variation was obtained in the three distances introduced as additional free parameters. The subtle differences observable between the solution and PVA samples in the region around  $8 \text{ \AA}^{-1}$ , are possibly due to small differences in the dynamical properties of the proteins (as the simulations in Fig. 7.7 suggest) or to small structural distortions, both involving degrees of freedom not considered in the described fitting procedure. Since these differences are close to the noise level, no further attempt to reproduce them was made. At variance with what obtained in the solution and PVA samples, the parameters listed in Tab. 7.2 reveal considerable structural changes in the two trehalose matrices examined. While the  $\text{Fe-N}_p$  distance does not change (as expected), a progressive elongation of the distance between Fe and the  $\text{N}_{\epsilon 2}$  atom belonging to the imidazole ring is clearly detectable going from solution to the moderately dried ( $\Delta r\text{-N}_{\epsilon 2} = 0.1 \text{ \AA}$ ) and to the extensively dried trehalose matrix ( $\Delta r\text{-N}_{\epsilon 2} = 0.2 \text{ \AA}$ ). Moreover, in trehalose matrices we detected an elongation of the fourth shell ligands of the porphyrin group which reaches

0.16 Å in the extra dry sample. Since the second ( $C_{2,5}$ ) and the third shell ( $C_6$ ) distances remain essentially unchanged (see  $r$  values in Table 2) the fourth shell elongation is the result of an increase of heme intra-ligand distances ( $C_2-C_3$ ,  $C_4-C_5$ ,  $C_3-C_4$ ) of about 10-13% (see Fig. 7.11). This is a considerably high value since the maximum variation of distances for XAFS restrained refinement is usually taken to be about 10% [33], [4]. The considerable structural distortions of the Fe ligand cluster revealed by the XAFS analysis in the extensively dried trehalose matrix are in line with a number of experimental observations and with MD simulations, indicating that the trehalose matrix, upon decreasing its water content, interacts strongly with the embedded protein, significantly perturbing the energy landscape and function of the protein [109], [110], [114], [118]. On the basis of these studies, it has been proposed that the matrix-protein coupling is essentially promoted by a network of hydrogen bonds connecting groups at the protein surface, residual water molecules and trehalose molecules [109]. It is therefore conceivable that through these interactions the matrix induces conformational changes in the protein which in turn result in structural distortions of the heme group at the local atomic scale. Alternatively, or most likely concomitantly, the heme could be involved in a direct interaction with the matrix. Interestingly in MD simulations performed in a non-liquid trehalose-MbCO-water system [109] one trehalose molecule has been found in persistent interaction with the heme propionate oxygen atoms. As stated in this work, this strong interaction may well induce a different configuration of the heme in the trehalose matrix with respect to the water-protein system. This interaction could be largely responsible for the elongation of the heme fourth shell detected by our analysis. In any case, our results clearly show that the protein-matrix structural interaction extends from the

protein surface to a significant length scale towards the protein interior. The effect of trehalose on the system disorder is even more dramatic (see Tab. 7.2). While we obtained comparable values for the DW factors of first ligands for the two samples in solution and in the PVA film, the values of these parameters decrease progressively in the moderately and extensively dried trehalose matrices. In the latter, extremely low values were obtained for the pyrrolic nitrogens and for the  $N_{e2}$  of histidine. Also the DW value of the methionine sulphur undergoes a strong reduction in trehalose with respect to the value exhibited in solution and in PVA. It is worthwhile to notice that in solution and PVA the DW values of the porphyrin  $N_p$  are much smaller than those of axial ligands  $N_{e2}$  and  $S_\delta$ , indicating that relative motions normal to the heme plane are greater than inside the heme plane. A similar difference, although less pronounced, has been reported in a previous XAFS study of ferrocyanide *c* performed at 10 K (38). An analogous relation between the XAFS DW factors of the histidine  $N_{e2}$  and of the heme pyrrolic N atoms has been obtained for met-myoglobin [32]. Since the DW factor measures the relative mean square displacement between the absorbing and backscattering atoms only, a substantial contribution to the dramatic reduction of the Debye-Waller factors could arise from a higher correlation in the motions of the cluster atoms when the protein is embedded in dry trehalose matrices. This increased motional coherence agrees with the results of MD simulations performed on MbCO in 89% (w/w) trehalose-water [108]. In fact, it was found that the deviations of the mean-square fluctuations from the linear temperature dependence above  $\sim 270$  K have a similar amplitude for all the heme heavy atoms, thus suggesting that already in plasticized trehalose systems the heme group, including the iron, performs quasi-diffusive motion as a whole. XAFS studies of DW factors as a function of



temperature performed on met-myoglobin [32] show that the decrease of the DW factors going from 260 K to 40 K is far less dramatic than that we observed upon incorporation of the protein into a dry trehalose matrix at room temperature. Even at 40 K pyrrolic nitrogens have a DW factor of  $1.9 \times 10^{-3} \text{ \AA}^2$  [32]. The authors suggest that static disorder, resulting from a conformational structural heterogeneity of the protein, strongly contributes to the values of the DW factors at low temperature. This hypothesis is corroborated by a comparison with relative mean square displacements calculated from normal modes which turned out to be much smaller than the experimental ones. On this basis, we should infer that the static contribution to the disorder is also significantly decreased by the trehalose matrix. This indicates that the trehalose matrix drastically alters the protein energy landscape, hindering the protein internal dynamics and also promoting only some conformational substates at the level of local structure. We discussed here only the DW values obtained for the first ligands. The values found for the more distant shells, although quite reasonable, are less reliable due to the approximations of the model. We recall, in fact, that for higher order shells, only the DW factor of the porphyrin group were allowed to vary (see Materials and Methods). Therefore caution is needed, in particular in the case of samples inserted in trehalose. The values of the fixed DW factors ( $0.01 \text{ \AA}^{-1}$ , see Materials and Methods), might no longer be reliable in the trehalose matrices. If this is the case, since multiple scattering contributions are very important at distances greater than  $3 \text{ \AA}$ , differences in their DW factors can lead to a poor estimate of the free DW values. Hence, it is rather risky to make conclusions based on these values. However, it should be noticed that in the extra dry trehalose matrix also the relative mean square displacements of the porphyrinic outer shells are lower when

compared to those found in solution. Interestingly, in this matrix the fourth shell value ( $DW(C_{3,4})$ ) seems to be lower than that related to the second shell ( $DW(C_{2,5})$ ), at variance with all the other matrices, where the opposite is true. The larger decrease in  $DW(C_{3,4})$  might reflect the strong interaction with the matrix which is responsible for the structural change localized at the same  $C_{3,4}$  atoms.

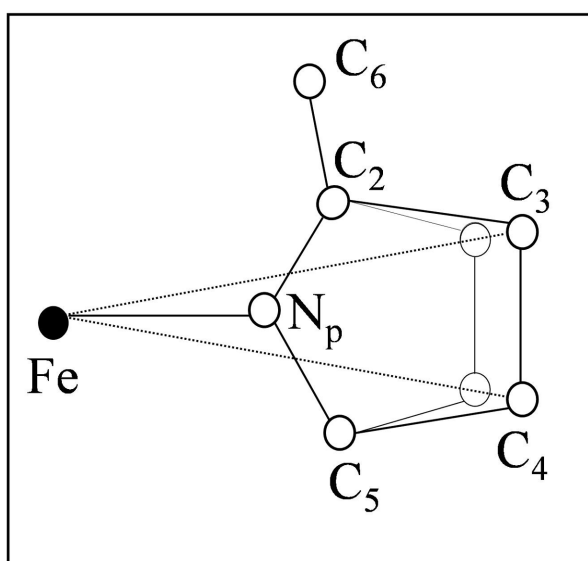


Figure 7.11: Graphical representation of the porphyrin pyrrolic group fourth shell elongation in the extra dry trehalose matrix. To obtain a simple picture we maintained the planarity of the structure, keeping collinear the Fe and  $C_{3,4}$  atoms in the elongation. With these assumptions the  $C_{2,5}$ - $C_{3,4}$  and the  $C_3$ - $C_4$  distances changed by 13% and 10% respectively.

### 7.3 XAFS features of MbCO embedded in a dry trehalose matrix

In the present experiment we have performed Fe K-edge X-ray absorption spectroscopy measurements in MbCO samples at three temperature (100,

200, 300 K) in order to compare structural and dynamical effects induced by trehalose with those produced by temperature lowering.

### 7.3.1 Materials and methods

Lyophilized ferric horse myoglobin was purchased from Sigma (Sigma, St. Louis, MO) and used without further purification. Trehalose from Hayashibara Shoj (Hayashibara Shoj Inc., Okayama, Japan) was used after recrystallization from aqueous solutions. For sample preparation, myoglobin was dissolved ( $5 \cdot 10^{-3} M$ ) in a solution containing  $2 \cdot 10^{-1} M$  trehalose and  $2 \cdot 10^{-2} M$  phosphate buffer (pH 7 in water). The solution was equilibrated with CO and reduced by anaerobic addition of sodium dithionite ( $10^{-1} M$ ). MbCO-trehalose samples were obtained by layering the solution (2 aliquots of 0.8 mL) on a capton support and by dehydrating it under a CO atmosphere in a silica gel desiccator. In order to reach a high dehydration further drying proceeded at  $60^\circ$  for about 12 hours under vacuum (100 mbar).

### 7.3.2 Data collection

Fe K-edge measurements were performed at the BM 8 GILDA beam-line of the European Synchrotron Radiation Facility (ESRF) with the apparatus described in Paragraph 3.1. Samples were measured at three temperatures (100, 200, 300 K) in the region 6900–8200 eV. Spectra are the average of multiple scans (from three to six), run consecutively on each sample. Subtle modifications of XAFS features in consecutive scans were detected for the measurements performed in solution at 300 K (see Paragraph 3.3.2).

### 7.3.3 Qualitative results

In Fig. 7.12 we show the experimental EXAFS functions collected for MbCO. We observed significant differences between samples in trehalose (Fig. 7.12, open circles) and in solution (Fig. 7.12 continuous lines). These differences are much greater at 300 K, indicating that the incorporation in a dried trehalose matrix produces, at room temperature, effects similar (but not equal) to those caused by lowering the temperature in solution. In particular the damping of the EXAFS function obtained for the samples inserted in trehalose matrices is comparable at all temperatures and it is less pronounced than that observed at 100 K in solution. This supports the notion that incorporation into a trehalose glass, affecting the protein energy landscape, strongly hinders the dynamics. Moreover, alterations in the EXAFS functions, probably due to subtle structural modifications, are observable in the trehalose matrix around  $8 \text{ \AA}^{-1}$ . A detailed analysis is in progress.

## 7.4 Conclusions

The reported XAFS analysis of cytochrome c in different solvent/matrix systems indicates that incorporation of the protein into an extensively dehydrated trehalose matrix has strong effects on the structure and dynamics of the Fe ligand cluster. While essentially the same structural and disorder parameters account for the XAFS spectra measured in aqueous solution, in a glycerol-water system and when cytochrome c is embedded into a dried PVA film, incorporation into the sugar matrix results from the structural point of view in: (i) an increase of the distance between Fe and the Ne2 atom of the histidine imidazole ring; (ii) a distortion of the heme pyrrolic units revealed by a significant elongation of the fourth-shell distance (Fe-C<sub>3,4</sub>).

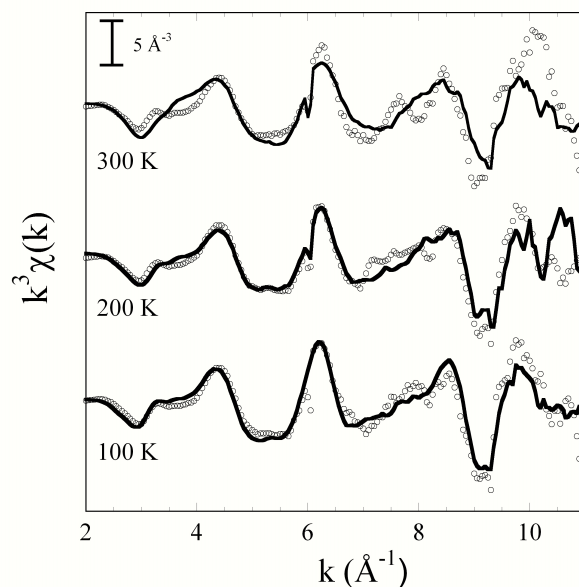


Figure 7.12: Experimental EXAFS functions collected for MbCO. From the top to the bottom: MbCO in solution (continuous line) and embedded in trehalose (open circles) measured at 300 K, 200 K, 100 K.

Comparison between a moderately and an extremely dehydrated trehalose matrix shows that these structural distortions increase strongly when the amount of residual water in the matrix is reduced. In fact the elongation of the (Fe-N<sub>ε2</sub>) and of the (Fe-C<sub>3,4</sub>) distances increase from 0.1 and 0.04 Å to 0.2 and 0.16 Å respectively when going from the less dried to the extremely dried trehalose matrix. This result is in line with a number of experimental observations, performed on heme proteins and on photosynthetic reaction centers, which clearly indicate a critical dependence of structural/dynamical and functional effects upon the hydration state of the embedding trehalose matrix [112]. It is noteworthy that, at variance with the large effects observed in the sugar matrices, incorporation of the protein in a dehydrated PVA film does not lead to any relevant change in the structural and disorder parameters of the Fe ligand cluster as compared to solution. This finding,

indicating a very weak structure/dynamics coupling between the protein and the PVA matrix, is consistent with the observation that incorporation of bacterial photosynthetic reaction centers into fully dehydrated PVA films very weakly affects the kinetics of electron transfer processes as compared to trehalose matrices [44]. The occurrence of structural distortions of the heme group when a heme protein interacts with a trehalose matrix has been suggested by spectral diffusion studies in cytochrome *c* [111] and inferred on the basis of MD simulations of a water-trehalose-MbCO amorphous system [109], [108]. In the case of *cyt c* our analysis leads to identify and resolve these proposed structural effects at the atomic scale. The obtained elongation of the distance between Fe and the coordinating histidine N atom, as well as the strong distortion of the heme pyrrolic geometry, testify of a very tight structure/dynamics coupling between the matrix and the protein in trehalose systems. Indeed, disorder parameters obtained from XAFS analysis evidence a parallel severe decrease in the relative mean-square displacements of the Fe first ligand atoms, which becomes dramatic in the extensively dried trehalose matrix. This finding is in full agreement with the results of experimental studies and MD simulations performed on trehalose-coated MbCO, all indicating a drastic hindering of thermal fluctuations. Our data highlight this effect on the local Fe structure. The values of the DW factor obtained by us in solution or in the PVA compare favorably with those reported in a previous XAFS study performed on small crystals of met-myoglobin [32]. However, the decrease of the DW observed in met-myoglobin upon lowering the temperature to 40 K is relatively limited [32] and smaller than that we observe upon incorporation of the protein in the dried trehalose matrix at room temperature. This is in line with what we see in MbCO, where the damping of the EXAFS function for the sample inserted in dehydrated tre-

halose matrix is less pronounced than that obtained at 100 K in solution. Following Sherk et al. [32], we propose that, also in the case of cytochrome *c*, static structural heterogeneities, resulting from a large number of coexisting protein conformations, give a substantial contribution to the values of the DW factors obtained in solution or in the PVA film. Although we do not have still quantitative results for MbCO we believe that analogous considerations can be made also for this heme protein, since the qualitative features of XAFS spectra are in full agreement with the results obtained for cyt *c*. The dramatic reduction of the relative mean-square displacement found in the extensively dried trehalose would therefore reflect not only a reduction of dynamics, possibly resulting from an increased coherence in the motions of the cluster atoms, but, to a considerable extent, also a decreased static disorder. If this is the case, it appears that incorporation into an extremely dried trehalose matrix, by drastically affecting the energy landscape of cytochrome *c*, selects a restricted number of conformational substates at the level of local structure.

## Chapter 8

# Conclusions

In this thesis we have investigated with XAFS two biophysical issues: in the first one the aim was to identify zinc sites in three key enzymes of the respiratory chain in order to shed light into the inhibitory mechanism exhibited by this metal; in the second one the goal was to detect possible structural/dynamical modifications produced by an extra dry trehalose matrix on the iron site of two heme proteins. Such modifications can be important in order to fully understand the molecular mechanism which makes trehalose so special in exerting a bioprotectant action.

The approaches used for studying these two aspects were different: in the first case we wanted some information about the zinc local structure which was partly or completely unknown, whereas in the second case the analysis relied on a well defined starting model and we were more interested in dynamics.

We obtained significant results in both the cases.

As far as matrix effects are concerned we found that a strongly dehydrated trehalose matrix hinders the protein local dynamics and induces



changes in the atomic environment around the metal atom.

As far as zinc binding sites are concerned, we have detected the presence of an endogenous zinc binding site in bovine complex I solving its local structure. We have solved the local structure of an exogenous, inhibitory zinc binding site of bovine cytochrome c oxidase; on the basis on the XAFS local structure, using the indications supported by the inhibition mechanism we attempted to locate the cluster. We have solved the local structure of the exogenous, inhibitory zinc binding site of bacterial, avian and bovine bc<sub>1</sub> and, on the basis on XRD structure of the avian bc<sub>1</sub> co-crystallized with the metal, we have attempted to locate the binding cluster. We recall that the analysis strategy we used in this case can have a general interest since XAFS is normally used in combination with high-resolution crystallography, i. e. when a robust starting model is available. To the contrary, in our study, we managed to solve unknown structures by using only XAFS data and known stereochemical information.

## Appendix A

# Rigid body refinement for selected aminoacids

## A.1 Histidine

Histidine residue can be considered as a rigid unit, thus all distances and angles can be set according to Tab. 5.1, 5.2, 5.3 and treated as *set* parameters during the fitting procedure. The distance between the metal and the  $N_{\epsilon 2}$  or  $N_{\delta 1}$  atom, depending on the binding mode, is released. When considering the position of the entire ring in respect to the metal one can insert, as extra degrees of freedom, the first shell distance and the bending angle so that two degrees of freedom are sufficient to describe the system.

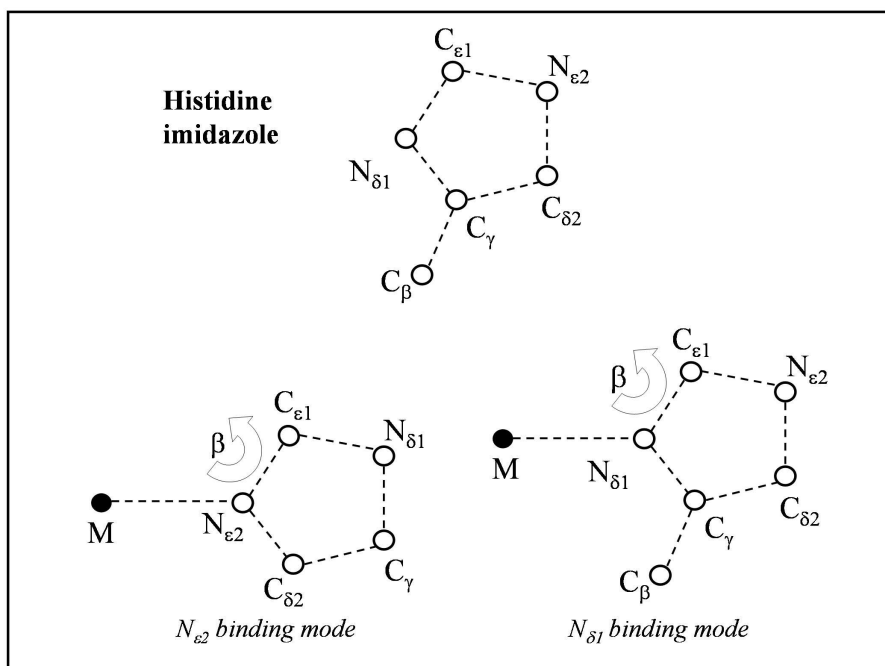


Figure A.1: Rigid body refinement of histidine residue.

## A.2 Aspartate

Aspartate residue can be considered as a structural unit with one degree of freedom: the angle  $\gamma$ . However this value cannot deviate too much (say more than  $15^\circ$ ) from a  $sp^3$  ideal geometry. The distance between the metal and the  $O_{\delta 1}$  atom is released. To describe the bending movement one can insert in principle another degree of freedom: the angle  $\gamma'$ . However, the EXAFS signal can be scarcely sensitive to this parameter when also the angle  $\gamma$  is let free, so that also for this amino acid it is sufficient to insert two degrees of freedom to perform the analysis with sufficient accuracy.

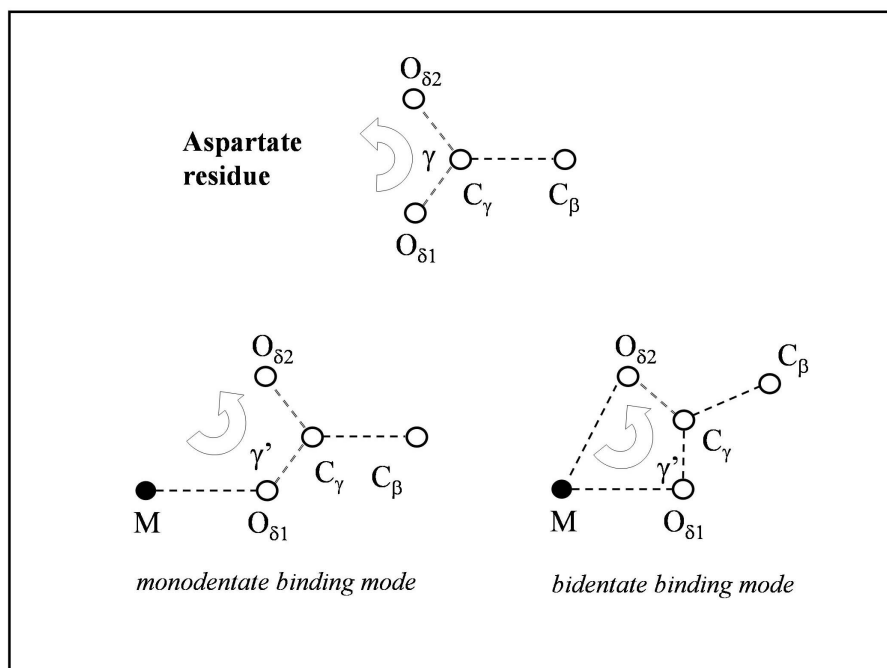


Figure A.2: Rigid body refinement of aspartate residue.

### A.3 Glutamate

For glutamate residue the same considerations made for the aspartate residue can be done.

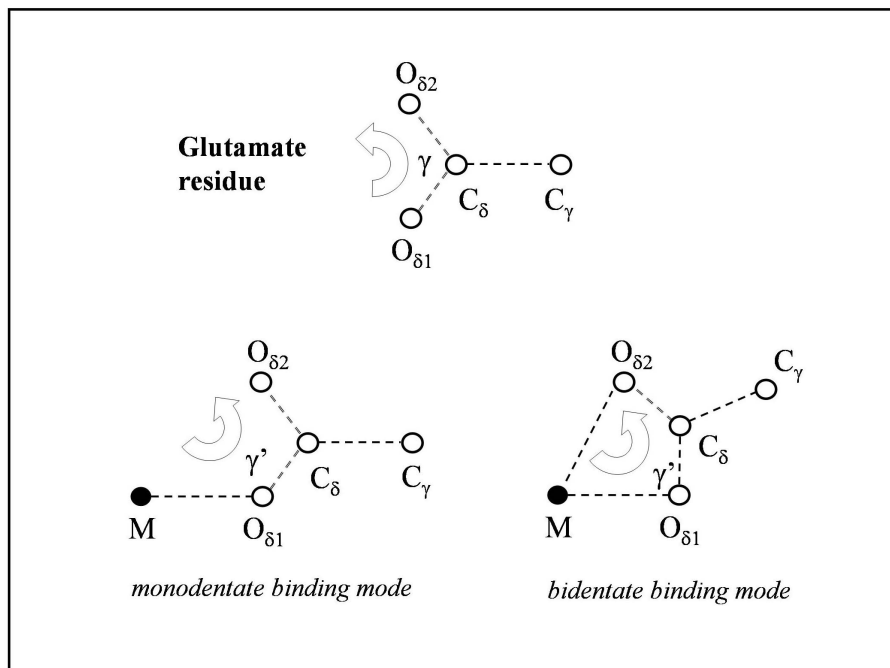


Figure A.3: Rigid body refinement of glutamate residue.

## A.4 Asparagine

For asparagine residue the same considerations made for the aspartate residue can be done.

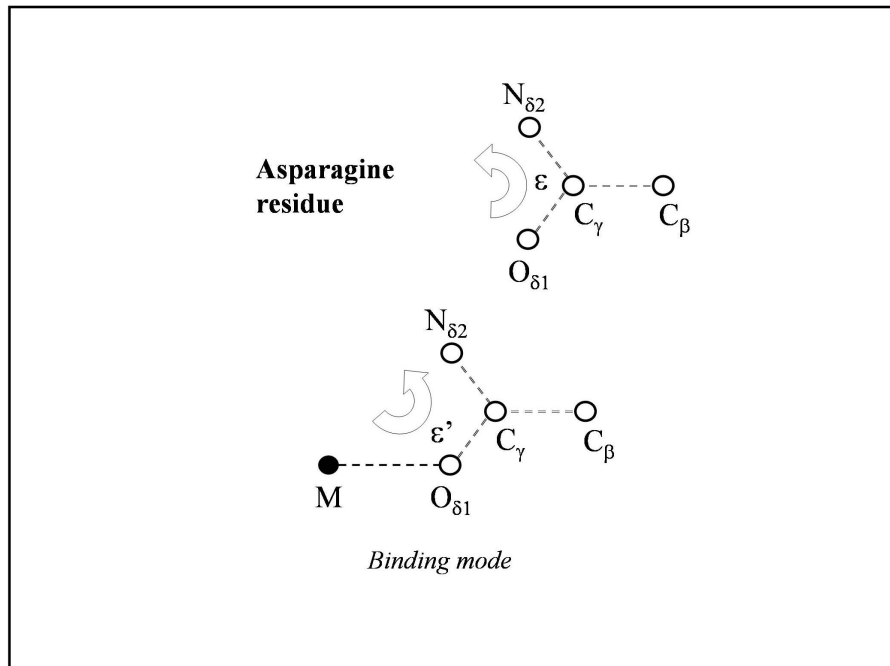


Figure A.4: Rigid body refinement of asparagine residue.

## A.5 Glutamine

For glutamine residue the same considerations made for the aspartate residue can be done.

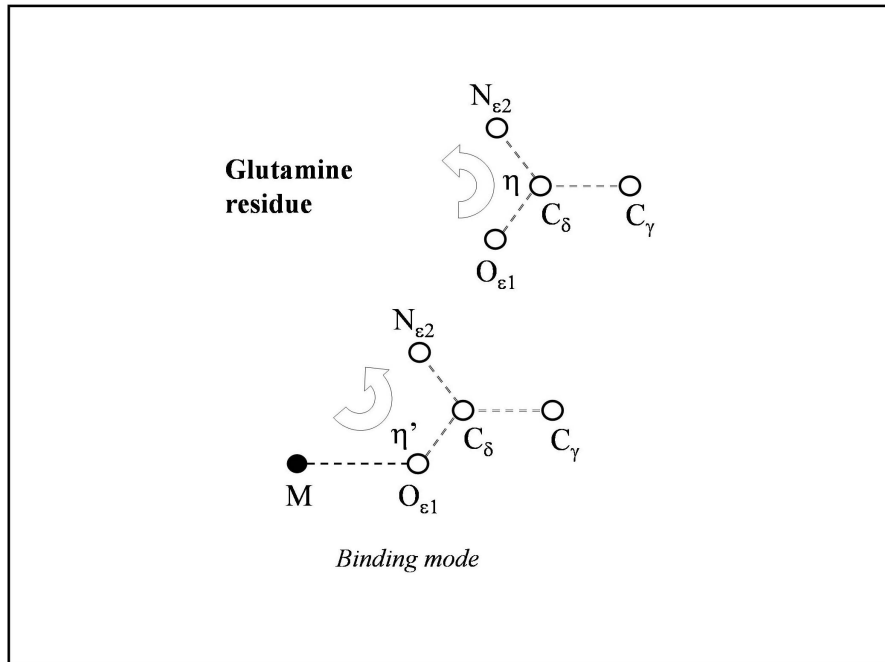


Figure A.5: Rigid body refinement of glutamine residue.

## A.6 Lysine

Lysine residue can be considered as a structural unit with one degree of freedom. Also in this case there are in principle three degrees of freedom to be considered: the first ligand distance, the angle  $\delta$ , and the bending angle  $\delta'$ . Actually three degrees of freedom are too many for this aminoacid since it gives, as a whole, a low contribution to the EXAFS signal when other residues are present. We strongly suggest to use also in this case two degrees of freedom: the first shell distance and the bending angle  $\delta'$ , since the angle delta changes only the position of the  $C_\delta$  atom which gives negligible contribution to the EXAFS signal.

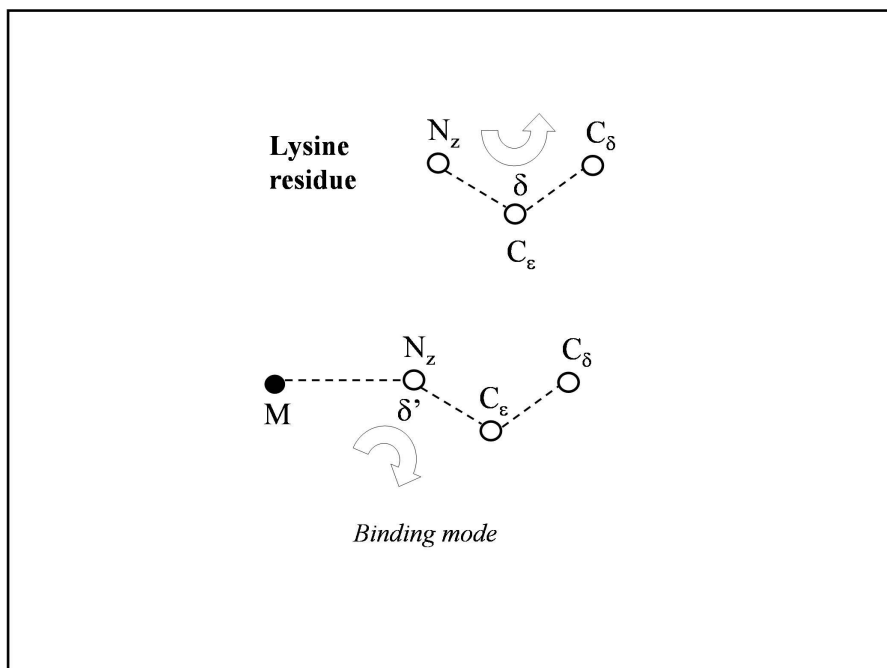


Figure A.6: Rigid body refinement of lysine residue



## A.7 Cysteine

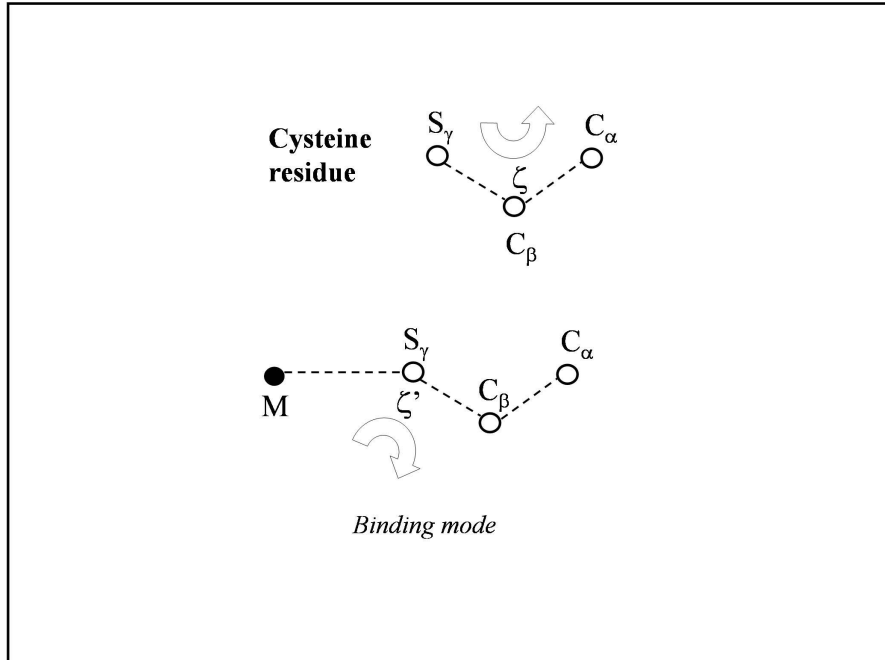


Figure A.7: Rigid body refinement of cysteine residue.

# Bibliography

- [1] A. Levina, R. S. Armstrong, and P. A. Lay. *Coord. Chem. Rev.*, 249:141–160, 2005.
- [2] J. Rehr and R. C. Albers. *Rev. Mod. Physics*, 72(3):621–654, 2000.
- [3] R. H. Holm, P. Kennepohl, and E. I. Solomon. *Chem. Rev.*, 96:2239–2314, 1996.
- [4] R. Engh and R. Huber. *Acta Cryst.*, A47:392–400, 1991.
- [5] D. E. Sayers and E. A. Stern. *Phys. Rev. Letters*, 27(18):1204–1207, 1971.
- [6] P. A. Lee, P. Eisenberger, and B. M. Kincaid. *Rev. Mod. Phys.*, 53(4):769–806, 1981.
- [7] N. Binsted and S. S. Hasnain. *Physica B*, 208 and 209:117–120, 1995.
- [8] A. Mijovilovich and W. Meyer-Klaube. *J. Synchrotron Rad.*, 10:64–68, 2003.
- [9] W. L. Schaich. *Phys. Rev. B*, 8(8):4028–4032, 1973.
- [10] M. Benfatto, S. Della Longa, and C. R. Natoli. *J. of Synchrotron Radiation*, 10(1):51–57, Jan 2003.

- [11] M. Girasole J. L. Hazemann S. Della Longa, A. Arcovito and M. Benfatto. *Phys. Rev. Lett.*, 87:155501, 2001.
- [12] M. Benfatto and S. Della Longa. *J. Synchrotron Rad.*, 8:1087–1094, 2001.
- [13] G. Beni and P.M. Platzman. *Phys. Rev. B*, 14:1514–1518, 1976.
- [14] R. Kronig. *Z Phys.*, 70:317, 1931.
- [15] R. Kronig. *Z Phys.*, 75:468, 1931.
- [16] J. E. Muller, O. Jepsen, O. K. Andersen, and J. W. Wilkins. *Phys. Rev. Letters*, 40:720–722, 1978.
- [17] Joly Y. *Phys. Rev. B*, 63(12), 2001.
- [18] S. S. Hasnain and R. W. Strange. *JSR*, 10:9–15, 2003.
- [19] H. F. Frauenfelder, F. Parak, and R. D. Young. *Annu. Rev. Biophys. Biophys. Chem.*, 17:451–479, 1988.
- [20] F. Schotte, M. Lim, T. A. Jackson, A. V. Smirnov, J. Soman, J. S. Olson, G. N. Phillips, M. Wulff, Jr, and P. A. Anfinrud. *Science*, 300(5627):1944–1947, 2003.
- [21] C. Bressler and M. Chergui. *Chem. Rev.*, 104:1781–1812, 2004.
- [22] S. S. Hasnain and K. O. Hodgson. *J. Synchrotron Rad.*, 6:852–864, 1999.
- [23] R. W. Strange and S. S. Hasnain. *J. of Synchrotron Radiation*, 10:9–15, 2003.
- [24] J. Jaklevic, J. A. Kirby, M. P. Klein, A. S. Robertson, G. S. Brown, and Eisenberger P. *Solid State Commun.*, 23:679–682, 1977.

- [25] N. J. Blackburn, R. W. Strange, J. Reedijk, A. Volbeda, A. Farooq, K. D. Karlin, and J. Zubietta. *Inorg. Chem.*, 28:1349–1357, 1989.
- [26] R. W. Strange, N. J. Blackburn, P. F. Knowles, and S. Samar Hasnain. *J. Am. Chem. Soc.*, 109:7157–7162, 1987.
- [27] A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson. *Phys. Rev. B*, 58(12):7565–7576, 1998.
- [28] N. Binsted, S. J. Gurman, J. W. Campbell, and P. Stephenson. *EX-CURVE. Daresbury Laboratory Program*. CCLRC Daresbury Laboratory, Warrington WA4 4AD, Cheshire, UK, 1982.
- [29] A. Filipponi, A. Di Cicco, T. A. Tyson, and C. R. Natoli. *Solid State Commun.*, 78:265–268, 1991.
- [30] M. F. Perutz, S. S. Hasnain, P. J. Duke, J. L. Sessler, and J. E. Hahn. *Nature*, 295:535 – 538, 1982.
- [31] N. J. Blackburn, S. S. Hasnain, N. Binsted, G. P. Diakun, and C. D. Garner. *Biochem. J.*, 219:985–990, 1982.
- [32] Scherk C.G., A. Ostermann, K. Achterhold, O. Iakovleva, C. Nazikoll, B. Krebs, E.V. Knapp, W. Meyer-Klaucke, and F.G. Parak. *Eur. Biophys. J.*, 30:393–304, 2001.
- [33] N. Binsted, R. W. Strange, and S. Samar Hasnain. *Biochem.*, 31:12117–12125, 1992.
- [34] A. V. Poiarkova and J. J. Rehr. *Phys. Rev. B*, 59:948–957, 1999.
- [35] N. Dimakis and G. Bunker. *J. of Synchrotron Radiation*, 8:297–299, 2001.

- [36] N. Dimakis and G. Bunker. *Phys. Rev. B*, 70:195114, 2004.
- [37] N. Dimakis and G. Bunker. *Phys. Rev. B*, 65:201103(R), 2004.
- [38] N. Dimakis and G. Bunker. *Biophys. J.*, 91:L87–89, 2006.
- [39] A. V. Poiarkova. *X-ray absorption fine structure Debye-Waller factors*. PhD thesis, Physics, 1999.
- [40] A. Marcelli S. Pascarelli, F. Boscherini and S. Mobilio. *Rev. Sci. Instrum.*, 63:927–930, 1992.
- [41] S. Pascarelli, F. F. Boscherini, F. D’Acapito, F. Hardy, C. Meneghini, and S. Mobilio. *J. Synchrotron Rad.*, 3:147–155, 1996.
- [42] T. Matsushida and H. Hashizume. *X-ray Monochromators*, volume 1. North Holland, 1983.
- [43] G. Ciatto, F. d’Acapito, F. Boscherini, and S. Mobilio. *J. Synchrotron Rad.*, 11:278–283, 2004.
- [44] F. Francia, L. Giachini, G. Palazzo, A. Mallardi, F. Boscherini, and G. Venturoli. *Bioelectrochemistry*, 63:73–77, 2004.
- [45] M. Weik, R. BG Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and Sussman L. *PNAS*, 97:623–628, 2000.
- [46] M. S. Weiss, S. Panjikar, C. Mueller-Dieckmann, and P. A. Tucker. *J. Synchrotron Rad.*, 12:304–309, 2005.
- [47] J. Yano, J. Kern, K. Irrgang, M. J. Latimer, U. Bergmann, P. Glatzel, Y. Pushkar, J. Biesiadka, B. Loll, K. Sauer, J. Messinger, A. Zouni, and V. K. Yachandra. *PNAS*, 102(34):12047–12052, 2005.

- [48] E. A. Stern, M. Newville, B. Ravel, Y. Yacoby, and D. Haskel. *Physica B*, 208 and 209:117–120, 1995.
- [49] M. Newville, P. Livins, Y. Yacobi, E. A. Stern, and J. J. Rehr. *Phys. Rev. B*, 47:14126–14131, 1993.
- [50] B. Ravel and M. Newville. *J. Synchrotron Rad.*, 12:537–541, 2005.
- [51] A. L. Ankudinov and J. J. Rehr. *Phys. Rev. B*, 56:R1712, 1997.
- [52] L. Mattheiss. *Phys. Rev.*, 133:A1399, 1964.
- [53] U. von Barth and Hedin L. *J. Phys. C.*, 5:1629, 1972.
- [54] L. Hedin and Lundqvist S. *Solid State Phys.*, 23:1, 1969.
- [55] D. D. Koelling and B N Harmon. 10(16):3107–3114, 1977.
- [56] J. Mustre de Leon, J. Yacoby, E. A. Stern, and J. J. Rehr. *Phys. Rev. B*, 42:10843–10851, 1990.
- [57] M. Newville. *FEFFIT*. University of Chicago. GSE-CARS, Bldg 434A. APS, Argonne National Lab, Argonne, IL60439, 1998.
- [58] S. D. Kelly, K. M. Kenner, G. E. Fryxell, J. Liu, S. V. Mattigod, and K. F. Ferris. *J. Phys. Chem. B*, 105:6337–6346, 2001.
- [59] K. O. Hodgson and S. S. Hasnain. *J. of Synchrotron Radiation*, 6:852–864, 1999.
- [60] M.Y. Okamura, M. L. Paddock, Graige M.S., and Feher G. *Biochim. Biophys. Acta*, 1458:148–163, 1991.
- [61] A.R. Crofts and C.A. Wraight. *Biochim. Biophys. Acta*, 726:149–185, 1991.

- [62] G. Feher, J.P. Allen, M.Y. Okamura, and Rees D.C. *Nature*, 33:111–116, 1989.
- [63] L. A. Sazanov and P. Hinchliffe. *Science*, 311:1430–1436, 2006.
- [64] E.A. Berry, L.-S. Huang, L.K. Saechao, N.G. Pon, M. Valkova-Valchanova, and F. Daldal. *Photosynth. Res.*, 81:251–275, 2004.
- [65] A.R. Crofts, S.W. Meinhardt, K.R. Jones, and M. Snozzi. *Biochim. Biophys. Acta*, 723:202–218, 1983.
- [66] Z. Zhang, L. Huang, V.M. Shulmeister, Y.I. Chi, K.K. Kim, L.W. Hung, and A.R. Crofts. *Nature*, 392:677–684, 1998.
- [67] A.R. Crofts. *Annu. Rev. Physiol.*, 66:689–733, 2004.
- [68] E.A. Berry and L.-S. Huang. *FEBS Lett.*, 555:13–20, 2003.
- [69] A.Y. Mulikidjanian. *Biochim. Biophys. Acta*, 1709:5–34, 2005.
- [70] P. Brezinski. *Trends in Biochem. Sci.*, 29:380–387, 2004.
- [71] M.L. Paddock, M.S. Graige, G. Feher, and M.Y. Okamura. *Proc. Natl. Acad. Sci.*, 96:6183–6188, 1999.
- [72] L. Giachini, F. Francia, A. Mallardi, G. Palazzo, Carpena E., Federico Boscherini, and Giovanni Venturoli. *Biophys. J.*, 88:2038–2046, 2005.
- [73] M.L. Paddock, L. Sagle, A. Tehrani, J.T. Beatty, G. Feher, and M.Y. Okamura. *Biochemistry*, 42:9626–9632, 2003.
- [74] P. Adelroth, M.L. Paddock, L.B. Sagle, G. Feher, and M.Y. Okamura. *Proc. Natl. Acad. Sci.*, 97:13086–13091, 2000.
- [75] M. Lorusso, T. Cocco, A.M. Sardanelli, M. Minuto, F. Bonomi, and S. Papa. *Eur. J. Biochem.*, 197:555–561, 1991.

- [76] S.S. Klishin, W. Junge, and A.Y. Mulkidjanian. *Biochim. Biophys. Acta*, 1553:177–182, 2002.
- [77] E.A. Berry, M. Guergova-Kuras, L. Huang, and A. R. Crofts. *Ann. Rev. Biochem.*, 69:1005–1075, 2000.
- [78] P. Nicholls and A.P. Singh. *Life Sci. Adv.*, 7:321–326, 1988.
- [79] A. Kannt, T. Ostemann, Muller H., and Ruitenberg M. *FEBS Lett.*, 7:142–146, 2001.
- [80] A. Aagaard, Namslauer A., and P. Brzezinski. *Biochim. Biophys. Acta*, 1555:133–139, 2002.
- [81] S.S. Kuznetsova, N.V. Azarkina, T.V. Vygodina, S.A. Siletsky, and Konstantinov A.A. *Biochemistry*, 70:128136, 2005.
- [82] D.A. Mills, B. Schmidt, C. Hiser, E. Westley, and S. Fergusson-Miller. *J. Biol. Chem.*, 277:1489414901, 2002.
- [83] K. Faxen, L. Salomonsson, P. Adelroth, and P. Brzezinski. *Biochim. Biophys. Acta*, 1757:388394, 2006.
- [84] Mark S. Sharpley and Judy Hirst. *J. Biol. Chem.*, 281:34803 – 34809, 2006.
- [85] K. Cheung, R. W. Strange, and S. S. Hasnain. *Acta Cryst.*, D56:697–704, 2000.
- [86] I. L. Alberts, K. Nadassy, and S. J. Wodak. *Protein science*, 7:1700–1717, 1998.
- [87] H. D. Bellamy E. A. Berry, Z. Zhang and L. Huang. *Biochim. Biophys. Acta*, 1459:440–448, 2000.



- [88] T. A. Link and G von Jagow. *J. Biol. Chem.*, 270:25001–25006, 1995.
- [89] N. Capitanio E. De Nitto G. Capitanio, P. L. Martino and S. Papa. *Biochemistry*, 45:1930–1937, 2006.
- [90] R. A. Schott. *Annu. Rev. Biophys. Biophys. Chem.*, 18:137–158, 1989.
- [91] J. Hatefi. *Methods in Enzymology*, 53:11–14, 1978.
- [92] I. M. Fearnley L. A. Sazanov, S. Y. Peak-Chew and J. Walker. *Biochemistry*, 39:7229–7235, 2000.
- [93] K. Clark-Baldwin, D. L. Tierney, N. Govindaswamy, E. S. Gruff, C. Kim, J. Berg, S. A. Koch, and J. E. Penner-Hahn. *J. Am. Chem. Soc.*, 120:8401–8409, 1998.
- [94] S.B. Leslie, E. Israeli, B. Lighthaert, J.H. Crowe, and L.M. Crowe. *Appl. Environ. Microbiol.*, 91:3592–3597, 1995.
- [95] M. Uritani, M. Takai, and K. Yoshinaga. *J. Biochem.*, 117:774–779, 1995.
- [96] W.Q. Sun and P. Davidson. *Biochim. Biophys. Acta*, 1425:235–244, 1998.
- [97] J.H. Crowe, J.F. Carpenter, and L.M. Crowe. *Annu. Rev. Physiol.*, 60:73–103, 1998.
- [98] L.M. Crowe. *Comp. Biochem. Physiol.*, 131:503–513, 2002.
- [99] J.F. Carpenter and J.H. Crowe. *Biochemistry*, 28:3916–3922, 1989.
- [100] P.S. Belton and A.M. Gil. *Biopolymers*, 34:957–961, 1994.
- [101] J.G. Sampedro and S. Uribe. *Mol. Cell. Biochem*, 256:319–327, 2004.

- [102] J.L. Green and C.A. Angell. *J. Phys. Chem.*, 93:2880–2882, 1989.
- [103] L. Cordone, P. Galajada, E. Vitrano, A. Gassmann, A. Ostermann, and F. Parak. *Eur. Biophys. J.*, 27:173–176, 1998.
- [104] F. Parak, E.W. Knapp, and D. Kucheida. *J. Mol. Biol.*, 161:177–194, 1982.
- [105] Cordone L., M. Ferrand, E. Vitrano, and G. Zaccai. *Biophys. J.*, 76:1043–1047, 1999.
- [106] W. Doster, S. Cusack, and W. Petry. *Nature*, 337:754–756, 1989.
- [107] I. Köper, M.C. Bellissent-Funel, and W. Petry. *Eur. Biophys. J.*, 29:263, 2000.
- [108] G. Cottone, L. Cordone, and G. Ciccotti. *Biophys. J.*, 80:931–938, 2001.
- [109] G. Cottone, G. Ciccotti, and L. Cordone. *J. Chem. Phys.*, 117:9862–9866, 2002.
- [110] S. Giuffrida, G. Cottone, F. Librizzi, and L. Cordone. *J. Phys. Chem. B*, 107:13211–13217, 2003.
- [111] V.V. Ponkratov, J. Friedrich, and J.M. Vanderkooi. *J. Chem. Phys.*, 117:4594–4601, 2002.
- [112] L. Cordone, G. Cottone, S. Giuffrida, G. Palazzo, Giovanni Venturoli, and C. Viappiani. *Biochim. Biophys. Acta*, 1749:252–281, 2005.
- [113] G. Palazzo, A. Mallardi, A. Hochkoepler, L. Cordone, and Giovanni Venturoli. *Biophys. J.*, 82:558–568, 2002.

- [114] F. Francia, G. Palazzo, A. Mallardi, L. Cordone, and G. Venturoli. *Biophys. J.*, 85:2760–2775, 2003.
- [115] D.C. Koningsberger and R. Prins. *X-ray absorption: Principles, Applications, Techniques of EXAFS, SEXAFS and XANES*. Wiley, New York.
- [116] M. Cheng, A.M. Rich, R.S. Armstrong, P.J. Ellis, and P.A. Lay. *Inorg. Chem.*, 38:5703–5708, 1999.
- [117] A.M. Rich, R.S. Armstrong, P.J. Ellis, H.C. Freeman, and A. Lay. *Inorg. Chem.*, 37:5743–5753, 1998.
- [118] F. Librizzi, C. Viappiani, S. Abbruzzetti, and L. Cordone. *J. Chem. Phys.*, 116:1193–1200, 2002.