J|A|C|S ARTICLES

Published on Web 10/04/2010

Dynamic Investigation of Protein Metal Active Sites: Interplay of XANES and Molecular Dynamics Simulations

Paola D'Angelo,*[,†] Stefano Della Longa,[‡] Alessandro Arcovito,[§] Massimiliano Anselmi,[†] Alfredo Di Nola,[†] and Giovanni Chillemi*[,||]

Department of Chemistry, University of Rome "La Sapienza", P.le Aldo Moro 5, 00185 Rome, Italy, Department of Experimental Medicine, University of L'Aquila, 67100 L'Aquila, Italy, Istituto di Biochimica e Biochimica Clinica, Università Cattolica del Sacro Cuore, L.go F. Vito 1, 00168 Rome, Italy, and CASPUR, Consortium for Supercomputing Applications, Via dei Tizii 6b, 00185 Rome, Italy

Received June 28, 2010; E-mail: p.dangelo@caspur.it; g.chillemi@caspur.it

Abstract: The effect of structural disorder on the X-ray absorption near-edge structure (XANES) spectrum of a heme protein has been investigated using the dynamical description of the system derived from molecular dynamics (MD) simulations. The XANES spectra of neuroglobin (Ngb) and carbonmonoxy-neuroglobin (NgbCO) have been quantitatively reproduced, starting from the MD geometrical configurations, without carrying out any optimization in the structural parameter space. These results provide an important experimental validation of the reliability of the potentials used in the MD simulations and accordingly corroborate the consistency of the structural dynamic information on the metal center, related to its biological function. This analysis allowed us to demonstrate that the configurational disorder associated with the distortion of the heme plane and with the different orientations of the axial ligands can affect the XANES features at very low energy. Neglecting configurational disorder in the XANES quantitative analysis of heme proteins is a source of systematic errors in the determination of Fe coordination geometry. The combined use of XANES and MD is a novel strategy to enhance the resolution and reliability of the structural information obtained on metalloproteins, making the combination of these techniques powerful for metalloprotein investigations.

Introduction

X-ray absorption fine structure (XAFS) spectroscopy is a powerful tool to investigate both local structure and dynamics on a wide class of metal-containing proteins.[1,2] This technique is very sensitive to the coordination geometry of an absorbing atom and therefore allows bond distances and angles of the surrounding atomic cluster to be measured with atomic resolution. XAFS focuses on either the extended X-ray absorption fine structure (EXAFS)[3] or the X-ray absorption near-edge structure (XANES).[4,5] While several EXAFS investigations on metalloproteins are present in the literature, the quantitative analysis of XANES spectra was rather restricted in earlier works because of the difficult theoretical treatment of the XANES features. Quantitative analysis of XAFS spectra requires knowledge of the Debye−Waller factors arising from thermal vibrations and structural disorder.[6] One of the major challenges at present is to develop a reliable theoretical scheme to include configurational disorder in the XANES calculations. In fact, while the effect of vibrations can be properly estimated using advanced theoretical approaches,[6] and becomes negligible at very low energies, the effect of structural disorder can extend down to the near edge, and it is largely dependent on the configurational space available to the specific system. Therefore, it cannot be determined easily using a general theoretical approach. A strategy to overcome this problem is to analyze the XANES spectra using a microscopic description of the system derived from molecular dynamics (MD) simulations, and this approach has been successfully used in the XANES analysis of aqueous ionic solutions.[7−10]

Today, MD simulations are a well-established tool for understanding the physical basis of the structure and function of biological macromolecules. Simulations can provide important details concerning the internal motions and resulting conformational changes of proteins that play an essential role in their

[†] University of Rome "La Sapienza".
[‡] University of L'Aquila.
[§] Università Cattolica del Sacro Cuore.
[||] CASPUR, Consortium for Supercomputing Applications.

(1) Cook, J. D.; Penner-Hahn, J. E.; Stemmler, T. L. *Methods Cell. Biol.* **2008**, *90*, 199–216.
(2) Strange, R. W.; Feiters, M. C. *Curr. Opin. Struct. Biol.* **2008**, *18*, 609–616.
(3) Lee, P. A.; Pendry, J. B. *Phys. Rev. B* **1975**, *11*, 2795–2811.
(4) Rehr, J. J.; Albers, R. C. *Rev. Mod. Phys.* **2000**, *72*, 621–654.
(5) Natoli, C. R.; Benfatto, M.; Della Longa, S.; Hatada, K. *J. Synchrotron Radiat.* **2003**, *10*, 26–42.
(6) Rehr, J. J.; Ankudinov, A. L. *Coord. Chem.* **2005**, *249*, 131–140.
(7) D'Angelo, P.; Roscioni, O. M.; Chillemi, G.; Della Longa, S.; Benfatto, M. *J. Am. Chem. Soc.* **2006**, *128*, 1853–1858.
(8) D'Angelo, P.; Zitolo, A.; Migliorati, V.; Mancini, G.; Persson, I.; Chillemi, G. *Inorg. Chem.* **2009**, *48*, 10239–10248.
(9) D'Angelo, P.; Migliorati, V.; Mancini, G.; Barone, V.; Chillemi, G. *J. Chem. Phys.* **2008**, *128*, 084502−6.
(10) D'Angelo, P.; Migliorati, V.; Mancini, G.; Chillemi, G. *J. Phys. Chem. A* **2008**, *112*, 11833–11841.

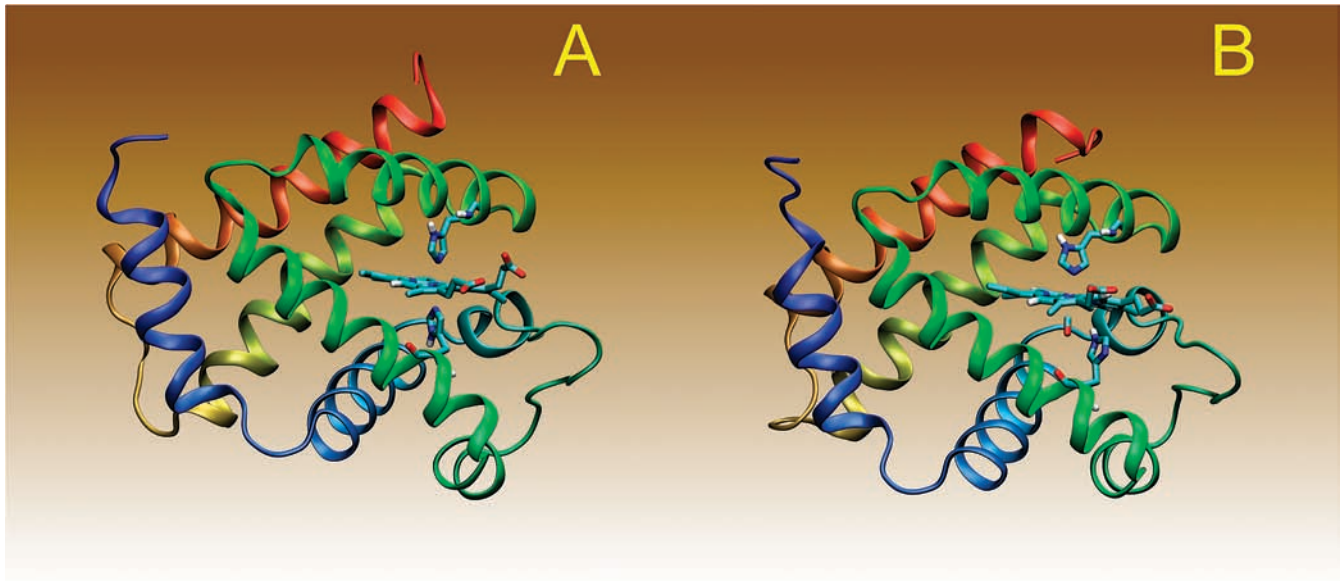

Figure 1. Ribbon representations of (A) Ngb and (B) NgbCO.

function.[11] However, the force fields used in simulations are optimized to reproduce average properties, and the possibility of validating the simulation methodology and, consequently, the reliability of the calculated results is extremely important.

In this paper we have extended the quantitative treatment of disorder effects to the XANES spectrum of a protein, neuroglobin (Ngb), with the aim of gaining a deeper understanding of the actual limit of the standard XANES data analysis procedures based on the multiple scattering (MS) approach. In particular, we have computed the Fe K-edge XANES spectrum of Ngb from a representative set of geometries extracted from MD simulations. This procedure allows one, on the one hand, to include disorder effects in the calculation of XANES and, on the other hand, to check the validity of the potentials used in the simulation, in particular for the metalloprotein interactions.

Ngb (Figure 1) is a globular protein composed of a single chain of 151 residues and a prosthetic group, the heme, which reversibly binds oxygen and other diatomic ligands, such as CO and NO, at the sixth coordination position.[12] Despite the small sequence similarity with other globins, Ngb displays the typical globin fold.[13,14] The ferric and the ferrous forms are both hexacoordinated to the distal His64(E7) and to the proximal His96(F8) side chains[15,16] in the absence of diatomic ligands. The physiological role of Ngb is still debated, although recent studies have proposed that Ngb is involved in neuroprotection, as the recovery from stroke in experimental animals is enhanced by overexpression of this protein.[17,18] The three-dimensional structure has been solved by X-ray diffraction (XRD) for the unliganded ferric Ngb from human and mouse[13,14] and for the CO-bound ferrous Ngb (NgbCO).[19] Crystallographic data on protein single crystals show that the binding of CO in the murine derivative is associated with structural changes involving heme sliding and a topological reorganization of the internal cavities. Starting from the coordinates of these X-ray structures, MD simulations have been carried out for ferric Ngb in solution and for the ferrous CO-bound derivative, both in solution and in the crystal.[20,21] MD simulations of NgbCO in aqueous solution have shown that the heme oscillates between two positions, similar to the ones observed in the Ngb and NgbCO crystallographic structures, although the former is more populated. Moreover, it has been shown that the ligand binding can affect the flexibility and conformation of residues near the heme pocket.[20] On the other hand, MD simulations of NgbCO in crystal environments have shown that the crystal packing affects the relative stability of the heme configurations, whereas the pattern of intermolecular contacts may reduce the mobility of specific parts of the protein surface.[21]

Recently, a combined investigation of XRD on a single crystal, XANES, and EXAFS in solution has been carried out to determine the oxidation state and the structure of the Fe-heme in both the bis-histidine and the CO-bound neuroglobin.[22] The XANES analysis in the frame of the MS approach has been carried out with the MXAN procedure,[23,24] starting from the XRD structure and obtaining, after optimization, a single representative configuration of the system. This study has shown that the Fe-heme structure of NgbCO in solution displays no

(11) van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. *Curr. Opin. Struct. Biol.* **2008**, *18*, 149–153.
(12) Burmester, T.; Weich, B.; Reinhardt, S.; Hankeln, T. *Nature* **2000**, *407*, 520–523.
(13) Vallone, B.; Nienhaus, K.; Brunori, M.; Nienhaus, G. U. *Proteins* **2004**, *56*, 85–92.
(14) Pesce, A.; Dewilde, S.; Nardini, M.; Moens, L.; Ascenzi, P.; Hankeln, T.; Burmester, T.; Bolognesi, M. *Structure* **2003**, *11*, 1087–1095.
(15) Dewilde, S.; Kiger, L.; Burmester, T.; Hankeln, T.; Baudin-Creuza, V.; Aerts, T.; Marden, M. C.; Caubergs, R.; Moens, L. *J. Biol. Chem.* **2001**, *276*, 38949–38955.
(16) Nienhaus, K.; Kriegl, J. M.; Nienhaus, G. U. *J. Biol. Chem.* **2004**, *279*, 22944–22952.
(17) Sun, Y.; Jin, K.; Mao, X.; Zhu, Y.; Greenberg, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 15306–15311.
(18) Sun, Y.; Jin, K.; Peel, A.; Mao, X. O.; Xie, L.; Greenberg, D. A. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3497–3500.
(19) Vallone, B.; Nienhaus, K.; Matthes, A.; Brunori, M.; Nienhaus, G. U. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17351–17356.
(20) Anselmi, M.; Brunori, M.; Vallone, B.; Di Nola, A. *Biophys. J.* **2007**, *93*, 434–441.
(21) Anselmi, M.; Brunori, M.; Vallone, B.; Di Nola, A. *Biophys. J.* **2008**, *95*, 4157–4162.
(22) Arcovito, A.; Moschetti, T.; D'Angelo, P.; Mancini, G.; Vallone, B.; Brunori, M.; Della Longa, S. *Arch. Biochem. Biophys.* **2008**, *475*, 7–13.
(23) Benfatto, M.; Della Longa, S. *J. Synchrotron Radiat.* **2001**, *8*, 1087–1094.
(24) Della Longa, S.; Arcovito, A.; Girasole, M.; Hazemann, J. L.; Benfatto, M. *Phys. Rev. Lett.* **2001**, *87*, 155501–04.

relevant difference with respect to the crystalline state, thereby excluding crystal-packing restraint effects, at least at the level of the active site. Moreover, elongation of the Fe−proximal histidine bond by 0.1−0.2 Å has been observed upon coordination of the CO molecule in the axial site, related to heme sliding into the large internal cavity. The EXAFS and XANES results were in an overall good agreement with each other; however, relative to EXAFS, XANES detected a 0.09 Å shorter $Fe-N_{His}$ distal distance (within the errors) in bis-histidine Ngb and a 0.14 Å shorter $Fe-N_{His}$ proximal distance in the CO-bound derivative (0.03 out of the errors). Indeed, hemes are particularly challenging for the XANES analysis due to the strong asymmetry of the Fe site, and the occasional occurrence of systematic errors in the XANES determination of heme systems in solution was previously reported[25] when using the MXAN procedure. The origin of the observed discrepancies was proposed to come mainly from the approximation used for the phenomenological broadening function that mimics the spectral damping; this effect is due to several aspects, including thermal and structural disorder. However, Fe K-edge XAS spectra of myoglobin derivatives show no direct temperature dependence,[26] and this observation suggests that the spectral damping contributions arise mainly from configurational disorder rather than thermal disorder. As a consequence, the higher the configurational disorder of the system under investigation, the bigger systematic errors affecting the XANES analysis are expected.

Unveiling the details of the structural rearrangements of the active site upon ligand binding is fundamental for elucidating the metalloprotein biological function.[27] Structural changes to the metal coordination during substrate or ligand binding reactions are generally less than 0.1 Å and hence remain unnoticed in standard protein crystallography, as reported X-ray structures of proteins at true atomic resolution (<1.2 Å) are still scarce. The present study shows that the combination of the MS XANES theory and MD allows one to enhance the reliability of the structural information obtained on metal sites in metalloproteins from a XANES spectrum, thus making the combination of these techniques powerful for the metalloprotein investigations.

Materials and Methods

X-ray Absorption Measurements. Fe K-edge X-ray absorption spectra of Ngb were collected in fluorescence mode at the BM30B FAME beamline of the European Synchrotron Radiation Facility.[28] A 3 mM solution of Ngb in the oxidation state Fe(III) in the presence of 80 mM HEPES, pH 7.4, and 20% glycerol−water solution was used to collect the ferric form of the protein. To the same solution was added sodium dithionite to a final concentration of 50 mM under nitrogen atmosphere to collect chemically reduced species. The CO derivative was obtained by reducing the ferric sample with 50 mM sodium dithionite in a carbon monoxide atmosphere. All the spectra were collected at 15 K. The storage ring was running in the two-thirds filling mode with a typical current of 170 mA. The monochromator was equipped with a Si (111) double crystal, in which the second crystal was elastically bent to a cylindrical cross section. The energy resolution at the Fe K-edge is 0.5 eV. The X-ray photon beam was vertically focused by a Ni−Pt mirror and dynamically sagittally focused in the horizontal size. An array detector made of 30 Ge elements of very high purity was used. The spectra were calibrated by assigning the first inflection point of the Fe foil spectrum to 7111.2 eV. In the case of Ngb, the energy stability of each spectrum was carefully assessed by checking the position of a glitch in the I0 at 7220 eV. For the ferric and ferrous samples, eight spectra were recorded with a 7 s/point collection statistic and averaged. The collection time was 20 min for each XANES spectrum and 45 min for each EXAFS spectrum. No spectral changes were detected during the data collection. For the CO derivative, a clear spectral evolution was observed when the beam spot was kept fixed. Therefore, it was necessary to move the beam spot to different positions on the sample, and three equal spectra were collected and averaged for the XAS analyses of the CO derivative.

XANES Data Analysis. The XANES data analysis was carried out with the MXAN code.[23] The X-ray photoabsorption cross section is calculated using the full MS scheme in the framework of the muffin-tin approximation for the shape of the potential. All details on the potential calculations can be found in ref 23. The exchange and correlation parts of the potential are determined on the basis of the local density approximation of the self-energy. The real part of self-energy is calculated by using the Hedin−Lundqvist (HL) energy-dependent potential.

In the first step, the XANES spectrum associated with each MD configuration was calculated using only the real part of the HL potential; i.e., theoretical spectra do not account for any intrinsic and extrinsic inelastic process, while the damping associated with the experimental resolution is accounted for by convolution with a Gaussian function with full width at half-maximum (fwhm) of 0.7 eV. In the second step, to perform a comparison with the experimental data, the damping associated with the inelastic processes has to be included in the calculation. To this purpose we used a modified version of the MXAN program that reads an external theoretical spectrum (the configurational averaged calculated data) and performs a minimization in the nonstructural parameter space only. In particular, the inelastic processes are accounted for by convolution with a broadening Lorentzian function having an energy-dependent width of the form $\Gamma(E) = \Gamma_c + \Gamma_{mfp}(E)$. The constant part Γ_c accounts for the core−hole lifetime and is fixed to the tabulated value of 1.25 eV fwhm, while the energy-dependent term represents all the intrinsic and extrinsic inelastic processes.[23] The $\Gamma_{mfp}(E)$ function is zero below an energy onset E_s (which in extended systems corresponds to the plasmon excitation energy) and starts increasing from a given value A, following the universal functional form of the mean free path in solids.[29] Both the onset energy E_s and the jump A are introduced in the $\Gamma_{mfp}(E)$ function via an arctangent functional form to avoid discontinuities. E_s and A are the only two parameters that are optimized in the fitting procedure, while the E_0 value is fixed at 7121 eV. Least-squares fits of the XANES experimental data have been performed by minimizing the χ^2 function, defined as

$$\chi^2 = \frac{1}{m\varepsilon^2}\sum_{i=1}^{m}[(y_i^{th} - y_i^{exp})\varepsilon_i^{-1}]^2$$

where m is the number of experimental points, y_i^{th} and y_i^{exp} are the theoretical and experimental values of the absorption cross section, and ε_i is constant and equal to 0.5% of the experimental jump.

Molecular Dynamics Details. The XANES analyses reported herein start from the microscopic description of the deoxy-Ngb and NgbCO systems derived from MD simulations that were already successfully used to describe their global motions in solution.[20]

(25) D'Angelo, P.; Lapi, A.; Migliorati, V.; Arcovito, A.; Benfatto, M.; Roscioni, O. M.; Meyer-Klaucke, W.; Della Longa, S. *Inorg. Chem.* **2008**, *47*, 9905–9918.
(26) Scherk, C. G.; Ostermann, A.; Achterhold, K.; Iakovleva, O.; Nazikkol, C.; Krebs, B.; Knapp, E. W.; Meyer-Klaucke, W.; Parak, F. G. *Eur. Biophys. J.* **2001**, *30*, 393–403.
(27) Green, M. T.; Dawson, J. H.; Gray, H. B. *Science* **2004**, *304*, 1653–1656.
(28) Proux, O.; et al. *Phys. Scr.* **2005**, *T115*, 970–973.
(29) Muller, J. B.; Jepsen, O.; Wilkins, J. W. *Solid State Commun.* **1982**, *42*, 365–368.

The simulation parameters are as follow. The starting coordinates employed for the simulations were from the X-ray structure of CO-bound ferrous murine Ngb at 1.7 Å resolution (PDB 1w92);[19] in the case of Ngb, we used the 1.5 Å resolution structure of murine ferric bis-histidine Ngb (PDB 1q1f).[13] In order to determine the partial charges of the hexacoordinated heme in deoxy- and carboxy-Ngb, we performed quantum chemical calculations on the isolated bis(imidazole) iron(II) porphyrin [Fe(II)P(Im)$_2$] complex and the carboxy-imidazole iron(II) porphyrin [Fe(II)P(CO)(Im)] complex, respectively. For density functional calculations, Becke's three-parameter exchange[30] and the Lee−Yang−Parr correlation[31] (B3LYP) were performed. We used Ahlrichs's VTZ basis set[32] for the iron and the 6-311+G* basis set for the nitrogen atoms of the heme and the heteroatoms of the imidazole and carbon monoxide molecules. We used the 3-21G basis set for the rest of the system and for all hydrogen atoms. In the calculations, both the [Fe(II)P(Im)$_2$] and [Fe(II)P(CO)(Im)] complexes were assumed to be in the singlet, closed-shell state following previous studies.[33−37]

All our quantum chemical calculations were carried out using the Gamess US package.[38] The partial charges were obtained from the CHELPG algorithm,[39] and the fitted charges were constrained to exactly reproduce the total charge and the calculated dipole moment of the system. We chose the following set of partial charges: for deoxy-Ngb, 0.3 e for the iron and −0.075 e for the pyrrolic nitrogen atoms; for NgbCO, 0.6 e for the iron and −0.15 e for the heme nitrogen atoms; for the CO bound to the hexacoordinated ferrous heme, 0.17 e for the carbon and −0.17 e for the oxygen atom. Each protein was solvated in a box with explicit single point charge (SPC) water molecules,[40] large enough to contain the protein and 0.8 nm of solvent on all sides. The total number of atoms for the systems was about 21 000. MD simulations were performed with the Gromacs software package[41] using the GROMOS96 force field.[42] The additional parameters for hexacoordinated heme and bound CO were taken from the GROMOS force field parameter sets 53A6.[43] The bond length between the heme iron and the N_{His} atom of proximal histidine was set at 2.11 Å, corresponding to the crystallographic distance. An additional simulation of NgbCO, 5 ns long, was performed, setting the Fe−N_{His} distance at 1.96 Å. Simulations were carried out at a constant temperature of 300 K within a fixed-volume rectangular box using periodic boundary conditions. The initial velocities were taken randomly from a Maxwellian distribution at 300 K, and the temperature was held constant by the isothermal algorithm.[44] The Particle Mesh Ewald method[45] was used for calculation of the long-range interactions. For all systems, the solvent was relaxed by energy minimization followed by 100 ps of MD at 300 K while restraining protein atomic positions with a harmonic potential. The systems were then minimized without restraints, and their temperature was brought to 300 K in a stepwise manner: 50 ps MD runs were carried out at 50, 100, 150, 200, 250, and 300 K, before starting the production runs at 300 K.

Computational Procedure. In the first step of the analysis, trajectories containing only the iron atom, the porphyrin macrocycle, the imidazole ring of the axial histidine(s), and the CO molecule for NgbCO were extracted from the total MD trajectory. The cluster size and the l_{max} value (i.e., the maximum *l*-value of the spherical harmonic expansion of the scattering path operators) were chosen on the basis of a convergence criterion. In the case of deoxy-Ngb, the trajectory was chosen in the simulation window from 60 to 90 ns, and we extracted 100 snapshots, saved every 300 ps. For NgbCO, a more careful sampling of the MD simulation was performed because a previous study[20] had singled out three different regions along the MD trajectory, characterized by different heme displacements along the first eigenvector, as obtained by means of essential dynamics analysis, a consolidated technique to characterize the dynamical behavior of proteins.[46] An analogous heme displacement was not observed in the deoxy-Ngb system, where the heme visits a unique conformational basin for the whole simulation. In particular, the first region in the NgbCO simulation resembles the NgbCO crystal structure, and it is identified by a first essential dynamics eigenvector lower than −1 (named SL−); the second resembles the deoxy-Ngb crystal structure, and it shows a first eigenvector bigger than 0.7 (named SL+); finally, the third region samples the average position assumed by the heme plane according to the MD simulation (named SL0). From each of these regions, we extracted 300 snapshots that were used in the analysis. Each snapshot was used to generate the XANES associated with the corresponding instantaneous geometry, and the averaged theoretical spectrum was obtained by summing all the spectra and dividing by the total number of MD snapshots used. As previously mentioned, at this stage only the real part of the HL potential has been used; i.e., theoretical spectra do not account for any intrinsic and extrinsic inelastic process.

An important question when dealing with the computation of spectra from MD simulations is the total sampling length that is necessary to have a statistically significant average. To this end, we carried out a statistical treatment of the data. In particular, we calculated a residual function, defined as

$$\mathrm{rms} = \sqrt{\sum_i [\alpha^N(E_i) - \alpha^{N-1}(E_i)]^2}$$

where $\alpha^N(E_i)$ is the theoretical spectrum averaged over *N* snapshots, and the sum is extended over all the energy points E_i. A residual value of 10^{-5} was chosen to establish the number of spectra necessary to have a statistically significant average.

Results

Deoxy-Ngb. Figure 2 shows 100 XANES theoretical spectra (without broadening terms) of ferric Ngb, associated with each instantaneous configuration (black lines) obtained from 100 frames extracted from the MD simulation, together with the XANES average theoretical spectrum (red line). The calculated XANES spectra present noticeable differences up to 100 eV after the edge, showing the sensitivity of XANES to the geometrical changes visited by the MD simulation. The low-energy

(30) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(31) Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.
(32) Schaefer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.
(33) Loew, G. H.; Harris, D. L. *Chem. Rev.* **2000**, *100*, 407–420.
(34) Smith, D. M. A.; Dupuis, M.; Straatsma, T. P. *Mol. Phys.* **2005**, *103*, 273–278.
(35) Harvey, J. N. *J. Am. Chem. Soc.* **2000**, *122*, 12401–12402.
(36) McMahon, B. H.; Stojkovic, B. P.; Hay, P. J.; Martin, R. L.; Garcia, A. E. *J. Chem. Phys.* **2000**, *113*, 6831–6850.
(37) Li, J.; Ai, Y. J.; Xie, Z. Z.; Fang, W. H. *J. Phys. Chem. B* **2008**, *112*, 8715–8723.
(38) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; et al. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
(39) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361–373.
(40) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Co.: Dordrecht, The Netherlands, 1981; pp 331−342.
(41) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
(42) van Gunsteren, W. F.; Billeter, S.; Eising, A.; Hunenberger, P.; Kruger, P.; Mark, A. E.; Scott, W.; Tironi, I. *Biomolecular Simulations: The GROMOS96 Manual and User Guide*; BIOMOS BV: Zurich/Groningen, 1996.
(43) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.
(44) Brown, D.; Clarke, J. H. R. *Mol. Phys.* **1984**, *51*, 1243–1252.
(45) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(46) Amadei, A.; Linssen, A. B.; Berendsen, H. J. *Proteins* **1993**, *17*, 412–425.

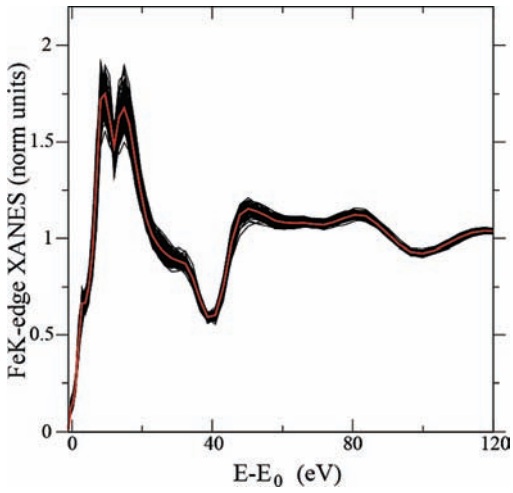

Figure 2. Comparison of the theoretical XANES spectrum of Ngb obtained from the MD average (red line) and several spectra associated with individual MD configurations (black lines).

part, up to 20 eV, is actually affected by configurational disorder, and the importance of making a proper sampling of the configurational space is evident, while the EXAFS region is less sensitive to configurational changes. This result is in contrast to what was found in the case of ionic aqueous solutions, where the structural disorder was found to have the greatest influence in the energy range above 25 eV.[7] This difference is explained by the opposite nature of the structural disorder present in the two kinds of systems: In aqueous solutions, the structural disorder is due to the high mobility of the water molecules in the first and second hydration shells that causes a strong broadening of the ion−water pair distribution functions. As a result, the main effect of the coexistence of several geometrical configurations around the ion is the damping of the high-energy (EXAFS) signal. Conversely, in the case of Ngb, the disorder is mainly associated with heme plane distortions and different positioning of the two histidine ligands with respect to heme iron. It gives rise to a complex configurational landscape in which relevant differences in the multiple scattering paths are detected in the low-energy region of the XAS spectrum. A pictorial description of the configurational landscape visited by the MD simulation and its effect on the corresponding XANES theoretical spectrum is shown in the Supporting Information, Movie 1.

An important question when dealing with the computation of spectra from MD simulations is the total sampling length necessary to have a statistically significant average. To this end, we have carried out a statistical treatment of the data, and the results of this analysis are shown in Figure 3, where the residual function (rms) used to test the convergence (see Methods) is plotted against the number of averaged spectra. As evident from the plot, a statistically significant average is obtained when 100 individual spectra are used.

In a recent work, MXAN has been used to carry out a quantitative analysis of the XANES spectrum of deoxy-Ngb and NgbCO.[22] This analysis was performed starting from the crystallographic structure (XANES static analysis), and selected structural parameters were optimized to gain the best agreement between the experimental and theoretical data. In particular, minimization of the χ^2 function for deoxy-Ngb was performed in the space of three structural parameters: the heme-core size, leading to the Fe−N_{pyrrol} distance, and the two Fe−N_{His} distances associated with the proximal and distal histidines. In addition, five nonstructural parameters, Γ_c (the core−hole width), E_s^{normal}, E_s^{heme}, A_s^{normal}, and A_s^{heme} (associated with the phenomenological broadening function), were used in the fitting procedure. The results of this analysis are shown in the upper panel of Figure 4, where the Ngb XANES experimental spectrum is compared to the best-fit XANES static theoretical curve. The structural parameters obtained from this analysis were found to be in good agreement with the XRD and EXAFS determinations, with the exception of the Fe−N_{His} distal distance, which was found to be slightly shorter (2.03 ± 0.07 Å) as compared to EXAFS (2.12 ± 0.03 Å) and XRD (2.12 ± 0.05 Å). The χ^2 value obtained from this analysis was 2.36.

As previously mentioned, inelastic losses have not been taken into account in the averaged theoretical XANES spectrum calculated from the MD configurations (see Figure 2). In order to carry out a comparison with the experimental data, all inelastic

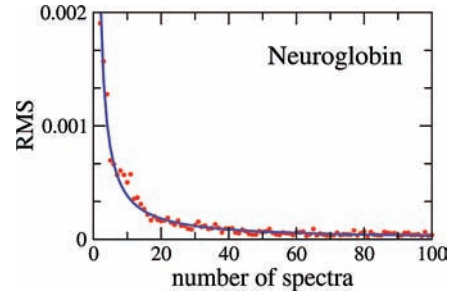

Figure 3. Residual function (rms) of the XANES averaged spectra of Ngb as a function of the number of MD snapshots used in the analysis.

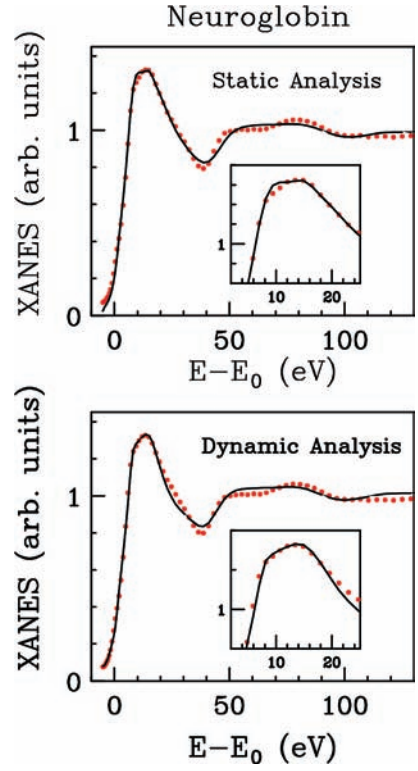

Figure 4. (Top) Comparison between the Fe K-edge XANES experimental spectrum of Ngb (red circles) and the best-fit theoretical spectrum obtained from the static analysis (black line). (Bottom) Comparison between the XANES experimental spectrum of Ngb (red circles) and the MD-averaged theoretical spectrum obtained from the dynamic analysis (black line). (Insets) Zoom-in of the edge region of each XANES spectrum.

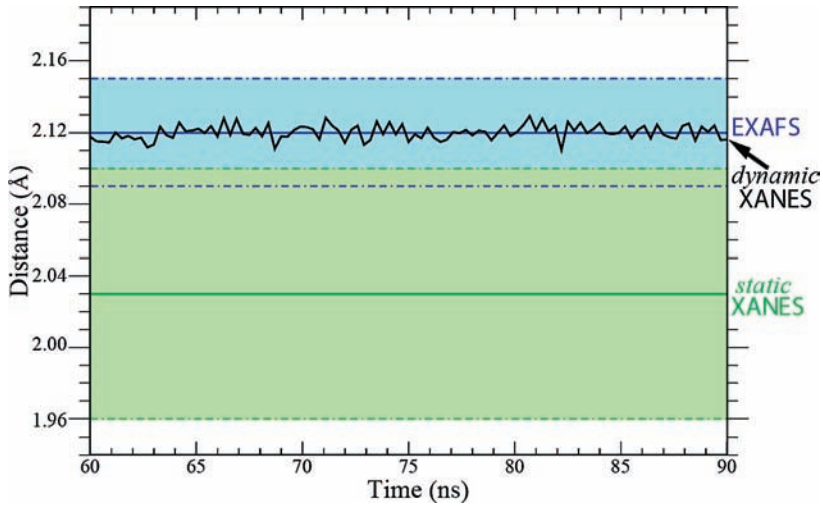

Figure 5. Fe−N_{His} distal distance as sampled by the MD simulation of Ngb in the time window from 60 to 90 ns, used to carry out the "dynamic XANES" analysis (full black line). For comparison, the Fe−N_{His} distal distances obtained by EXAFS and "static XANES" analysis (see ref 22) are shown as blue and green lines, respectively, together with the corresponding error bars (dash-dotted lines).

processes have been accounted for by convoluting the theoretical averaged spectrum with a broadening Lorentzian function, and the corresponding E_s and A_s nonstructural parameters have been optimized. In the bottom panel of Figure 4, the experimental XANES spectrum of deoxy-Ngb is compared with the averaged theoretical spectrum coming from MD. The agreement between the experimental and theoretical curves is quite good ($\chi^2 = 2.54$) and comparable to the one obtained from the "static" analysis, where three structural and five nonstructural parameters were optimized. Moreover, despite the lower number of parameters used to mimic the inelastic losses used in the dynamic analysis, some features of the edge (see insets in Figure 4) are better reproduced as compared to the "static" fit.

Figure 5 shows the variation of the Fe−N_{His} distal distance as sampled by the MD simulation in the time window from 60 to 90 ns, which has been used for the dynamic XANES analysis as explained in the Materials and Methods. The average distance imposed in the MD simulation matches the starting crystallographic value, in perfect agreement with the EXAFS determination, and it is 0.09 Å longer than the XANES determination obtained from the static fit. Note that the use of a single configuration in the XANES analysis may give rise to systematic errors in the structural determination due to the lack of a proper sampling of the configurational space visited by the heme site. The same behavior is not observed for the other structural parameters of the ferrous deoxy-Ngb system, i.e., Fe−N_{His} proximal and Fe−N_{pyrol} distances, which show a better agreement between the optimized values of the static XANES procedure and the EXAFS and XRD results (see Figures S1 and S2, Supporting Information). This discrepancy between EXAFS and static XANES is also not present in those systems, such as ions in solution, having a more closely packed symmetry and where the contribution of multiple scattering terms is less pronounced.[47−50]

Carbonmonoxy-Ngb. In the MD simulations of deoxy-Ngb, the heme position has been found not to move far away from the one observed in the Ngb crystal structure.[13,14] On the other hand, the MD simulation of NgbCO has shown that the heme rapidly fluctuates between two positions that roughly correspond to the Ngb and the NgbCO crystal structures.[20] This behavior has been highlighted by essential dynamics analysis. Figure 6A shows the trajectory of the heme displacement along its first eigenvector in NgbCO. The position of the heme in the crystal structure of NgbCO corresponds to the negative (about −0.5) values at $t = 0$ and $t \approx 65$ ns. In the remaining part of the simulation, the heme group position roughly corresponds to its position in the Ngb crystal structure. Note that the latter heme position is not allowed in the NgbCO crystal structure, due to close contacts that would occur between bound CO and the distal His64 (E7). However, the simulation indicates that small displacements of the His64 (E7) side chain drastically reduce atomic hindrance. Therefore, the fairly large repositioning of the whole heme associated with CO binding to deoxy-Ngb is a surprisingly fast process, occurring within the time window of the 90 ns MD simulation. In order to determine the sensitivity of XANES toward the different arrangements that the heme pocket assumes during the simulation, we have extracted 300 frames from three different regions, corresponding to the lowest (SL−), highest (SL+), and intermediate (SL0) values of the projection along the first essential dynamics eigenvector. In this case, due to the higher mobility of the heme pocket detected by the MD simulation, 300 frames have been used to calculate the averaged theoretical XANES spectra, and the attainment of convergence has been checked. The SL− and SL+ regions are represented by the green and blue bands in Figure 6A, respectively, while the red line represents the intermediate state SL0, which corresponds to a projection value of −0.25. The averaged theoretical XANES spectra obtained from the three analyses are compared in Figure 6B. While the energy region above 30 eV is not sensitive to the heme rearrangements sampled by the simulation, noticeable differences among the spectra are present in the edge region. This result reinforces the sensitivity of this energy range toward heme plane distortions and ligand orientations with respect to the heme plane.

(47) D'Angelo, P.; Zitolo, A.; Migliorati, V.; Persson, I. *Chem.—Eur. J.* **2010**, *16*, 684–692.
(48) Chillemi, G.; Mancini, G.; Sanna, N.; Barone, V.; Della Longa, S.; Benfatto, M.; Pavel, N. V.; D'Angelo, P. *J. Am. Chem. Soc.* **2007**, *129*, 5430–5436.
(49) Hayakawa, K.; Hatada, K.; D'Angelo, P.; Della Longa, S.; Natoli, C. R.; Benfatto, M. *J. Am. Chem. Soc.* **2004**, *126*, 15618–15623.
(50) D'Angelo, P.; Benfatto, M.; Della Longa, S.; Pavel, N. V. *Phys. Rev. B* **2002**, *66*, 064209.

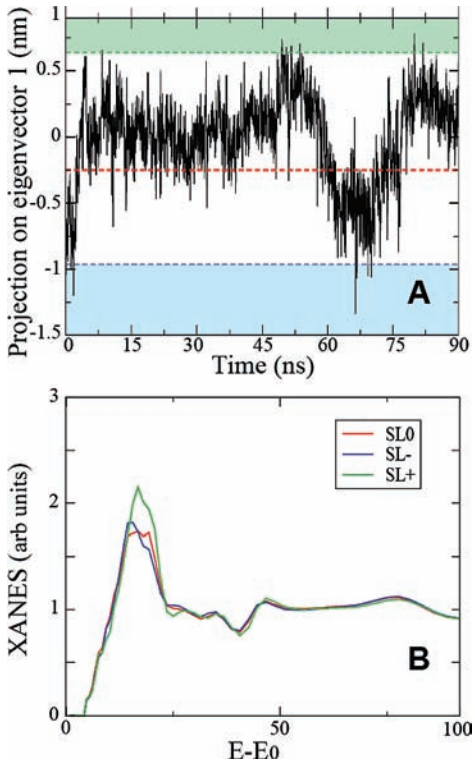

Figure 6. (A) Trajectory of the heme displacement along its first essential eigenvector in NgbCO (see ref 20). The regions corresponding to the lowest (SL−) and highest (SL+) values of the projection along the first essential dynamics eigenvector are highlighted by green and blue bands, respectively. The intermediate state SL0, which corresponds to a projection value of −0.25, is shown with a red dashed line. (B) Averaged theoretical XANES spectra obtained from the analyses carried out with frames corresponding to the three regions SL0, SL−, and SL+ are shown with red, blue, and green lines, respectively.

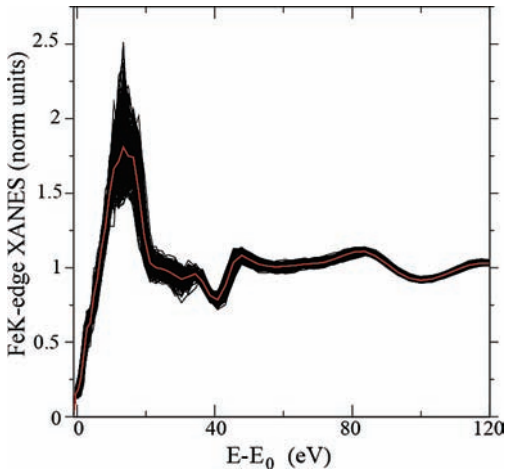

Figure 7. Comparison of the theoretical XANES spectrum of NgbCO obtained from the MD average (red line) and several spectra associated with individual MD configurations (black lines).

Figure 7 shows, as an example, the 300 XANES theoretical spectra of the intermediate state SL0 associated with each instantaneous configuration (black lines), together with the XANES average theoretical spectrum (red line) of NgbCO. In this case the differences among the instantaneous theoretical spectra are much larger as compared to deoxy-Ngb in the low-energy region (see Figure 2), as a consequence of the higher mobility of the heme pocket. This behavior is highlighted in the Supporting Information, Movie 2, where some configurations sampled in the SL0 state are shown together with the corresponding calculated XANES spectra. The effect of the structural disorder associated with the different orientations of the histidine and CO molecule, and with the distortions of the heme plane on the XANES and EXAFS regions, is shown in Figure 8. The two spectra associated with different heme configurations show marked differences in the energy region up to about 60 eV after the edge, while they are identical for higher energy values. This result is useful to understand the different approach that has to be adopted to include the Debye−Waller factor in the XANES and EXAFS regions of the absorption spectra of heme proteins.

A comparison of the SL0, SL+, and SL− averaged theoretical spectra with the experimental data is shown in Figure 9, together with the "static" analysis reported in ref 22. Also in this case, only three nonstructural parameters have been used to mimic the inelastic losses, and the agreement between the theoretical and experimental spectra is comparable to that obtained in the "static" fit ($\chi^2 = 1.71$ and 1.85 for the "static" and dynamic analysis, respectively) in the case of the SL0 configurations. As far as the SL− state is concerned, the quality of the fit is still good ($\chi^2 = 2.0$), even if the edge region is not reproduced as accurately as in the SL0 case, and the double-peak feature is lost (see the insets of Figure 9). Conversely, a poorer agreement has been obtained in the SL+ case ($\chi^2 = 2.5$), and the mismatch between the theoretical and experimental spectra is detected mainly in the energy region above 50 eV. These results are in line with the MD simulation, in which the SL+ conformations are rarely probed.

An interesting observation concerns the structural parameters sampled by the MD simulation. As previously mentioned, use of a single configuration in the XANES analysis gives rise to a systematic error in the structural determination of the Fe−N_{His} proximal distance in the CO-bound derivative.[22] Figure 10 shows the trend of the Fe−N_{His} proximal distance for the 300 SL0 frames used in the dynamic XANES analysis. Note that the Fe−N_{His} proximal distance fluctuates around a value that is in very good agreement with the EXAFS determination. The oscillations of the Fe−N_{pyrrol} and Fe−CO distances probed by the MD simulation are shown in the Supporting Information, Figures S3 and S4.

To get a deeper understanding of the effect of the modification of the heme structure, we extracted 300 frames from a simulation carried out by setting the Fe−N_{His} distance at 1.96 Å, which corresponds to the value obtained from the static fit. The comparison between the experimental data and theoretical spectra is shown in the Supporting Information, Figure S5. The agreement between the experimental and theoretical XANES spectra becomes poorer when the Fe−N_{His} distance obtained from the static XANES analysis is used in the MD simulations ($\chi^2 = 3.2$), confirming that the origin of the systematic error obtained from the static XANES fit stems from the neglect of the structural disorder.

A last remark concerns the CO ligand geometry obtained from our analysis. The Fe−C−O angle determined from X-ray crystallography is 157° (PDB 1w92),[19] while the combined XANES and EXAFS analysis determined an almost linear configuration (170−180°).[22] As far as the MD simulation is concerned, an interesting result is that, even if the simulation started from the crystallographic geometry, the Fe−C−O angle goes immediately to a value of about 176°, and this linear geometry is kept during the whole trajectory (see Supporting Information, Figure S6).[20] The resolution of the crystal structure

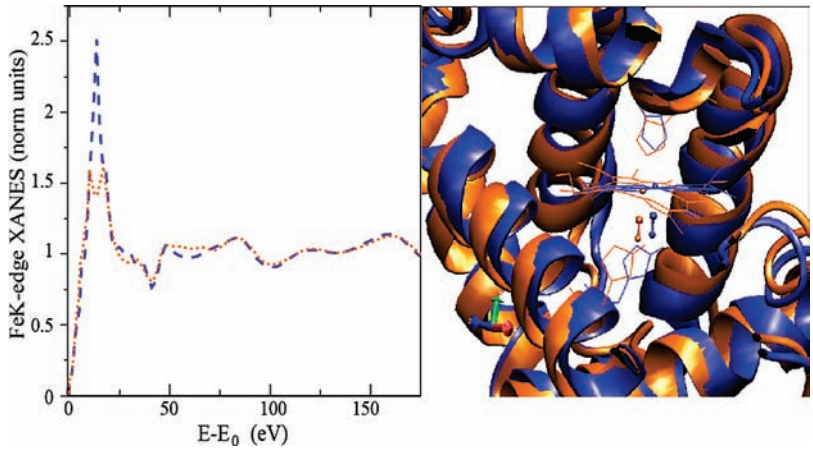

Figure 8. Ribbon representations of two configurations extracted from the MD simulation for NgbCO and corresponding theoretical XANES spectra.

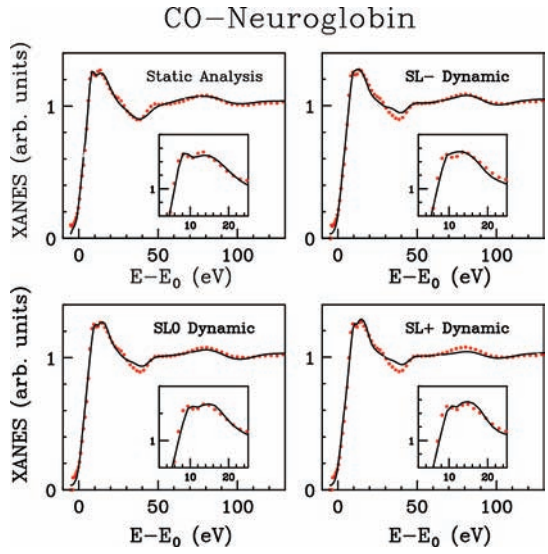

Figure 9. (Top left) Comparison between the Fe K-edge XANES experimental spectrum of NgbCO (red circles) and the best-fit theoretical spectrum obtained from the static analysis (black line). (Top right and bottom) Comparison between the XANES experimental spectrum of NgbCO (red circles) and MD-averaged theoretical spectra obtained from frames corresponding to the three regions SL0, SL−, and SL+ (black lines). (Insets) Zoom-in of the edge region of each XANES spectrum.

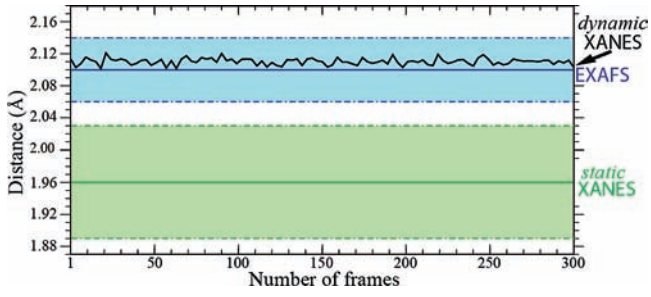

Figure 10. $Fe-N_{His}$ proximal distance as sampled by the MD simulation of NgbCO in the 300 frames of the SL0 region and used to carry out the "dynamic XANES" analysis (solid black line). For comparison, the $Fe-N_{His}$ proximal distances obtained by EXAFS and "static XANES" analysis (see ref 22) are shown in blue and green lines, respectively, together with the corresponding error bars (dash-dotted lines).

is 1.7 Å; therefore, the error on the angle determination is quite large, and the values determined by XANES and EXAFS are expected to be more accurate. The final agreement between the MD and XAFS results on the determination of the Fe−C−O angle suggests that the force field used in the MD simulation is reliable at the atomic resolution required for the XAFS analysis. Moreover, from the comparison between our results and the crystallographic determination of the Fe−C−O angle, some interesting biological issues can be addressed. According to the MD simulations by Di Nola et al.[20,21] at room temperature, the heme sliding distributions along the first essential eigenvector are different in the crystal and in solution. In the former case, the conformational distribution of NgbCO peaks around a position close to the crystallographic structure of deoxy-Ngb (i.e., with no heme sliding and a major steric hindrance between the CO molecule and the distal histidine). In the crystal, the MD simulation at room temperature shows a double-peak distribution. A first, higher maximum is found around a position close to the NgbCO crystallographic structure, with sliding of the heme and no steric hindrance between the CO molecule and the distal histidine. A second, lower maximum is located near the structure of deoxy-Ngb. However, at low temperature in the cryo-trapped crystal, only the first conformation is observed by XRD as significantly populated. According to the MD picture, our XAS experiment performed at cryogenic temperature in solution probes the entire conformational ensemble centered around the deoxy-like conformation (with no heme sliding), whereas only the more populated conformation (with sliding of the heme) is visualized by XRD in the crystal. XANES calculations averaged on MD trajectories in the crystalline state are the aim of a future work. However, according to recent "static" XAS measurements carried out on NgbCO single crystals,[51] the local structure around the Fe ion in NgbCO, including the Fe−C−O bending angle, seems unaltered between the crystal and the solution sample. As a consequence, no residual energy is stored in this site due to crystal packing, to be delivered with cracking at room temperature (that actually occurs in time at room temperature), in agreement with the MD-simulated rearrangement described above. This result reinforces the model proposed to describe the functional transition of the protein when an exogenous ligand replaces the distal histidine in the sixth position. The residual energy involved in the protein relaxation responsible for cracking the crystal does not reside in the heme pocket but is rapidly transferred at peripheral sites of the protein (more likely the

(51) Arcovito, A.; Ardiccioni, C.; Cianci, M.; D'Angelo, P.; Vallone, B.; Della Longa, S. *J. Phys. Chem. B* [Online early access]. DOI: 10.1021/jp104395g. Published Online: Sept 29, 2010.

CD corner), so that the final tertiary rearrangement of the protein in solution can be reached.

Discussion

In this work we have performed a quantitative treatment of structural disorder effects on the XANES spectrum of a heme protein using the dynamical description of the system derived from MD simulations without carrying out any optimization in the structural parameter space. This analysis allowed us to demonstrate that the structural disorder around the Fe atom in heme proteins is mainly associated with the distortion of the heme plane and with the different orientations of the axial ligands, giving rise to relevant differences in the multiple scattering paths associated with the second and third heme shells. Inspection of Movies 1 and 2 (Supporting Information) reveals a more complex dynamics of the protein and a corresponding higher mobility of the heme site for NgbCO as compared to Ngb. In particular, in Ngb, the proximal and distal histidines anchor the heme to the whole protein, thus reducing the mobility of the prosthetic group. Conversely, in the NgbCO derivative, much larger translational and rotational motions are observable both at level of the heme plane and for the proximal histidine. The differences in the MS paths are mainly visible in the low-energy region, and the individual XANES spectra associated with the distinct configurations show the presence of intense peaks near the edge (see Figures 2 and 7). As a result of the configurational disorder, all these pronounced features are smeared out in the average theoretical spectra, which show the characteristic peaks at about 8 and 16 eV after the edge (see insets of Figures 4 and 9) present in the experimental data. It is interesting to note that, in the case of deoxy-Ngb, a clear double peak is detected in the theoretical average spectrum, while a much broader and less intense feature is visible for NgbCO due to the larger deviation among instantaneous spectra. The stronger damping of the edge features observed for NgbCO is clearly associated with the higher structural disorder of the heme pocket probed by the MD simulation. It is important to stress that additional damping associated with a finite core−hole lifetime, experimental resolution, and inelastic processes (the latter being negligible below about 20 eV) has to be applied to the average theoretical spectra before they can be compared with the experimental data. This additional damping further reduces the intensity of the edge features, but the presence of a double-peak structure is still visible in the experimental spectra. When a single geometrical configuration is used, the edge features maintain their full intensity, and, as a consequence, the XANES static analysis is affected by a systematic error due to the neglect of structural disorder. This effect interferes with the electronic damping and the broadening terms used to account for inelastic losses. A direct consequence is the low accuracy in the determination of the axial parameters (i.e., the Fe−N_{His} and Fe−sixth ligand) for heme proteins in solution. In this case, the XANES signal is dominated by heme-plane scattering, and the errors affecting the axial parameters are bigger than in the reported angle-resolved experiments, where the strong heme scattering contribution can be depressed by the polarization dependence.[24]

One outstanding result of this study is that, for the first time, it has been possible to directly check the validity of the force fields used in MD simulations of a biological macromolecule at the precision of XAS (0.05 Å). The very good agreement between the XANES experimental data and the average theoretical spectrum confirms the quality of the potentials used in the simulations and reinforces the structural and dynamical description of the protein obtained from the MD results.

The force-field generation for metalloproteins is a key aspect that has limited the application of MD to this important class of biomolecules.[52] The transferability of the parameters associated with the atoms of the metallo groups among different proteins, in fact, is not straightforward, and it could lead to important errors in the model.[53] The application of our procedure to other metalloproteins will allow one us, on the one hand, to gain structural and dynamic information on the metallo center at a precision of hundredths of angstroms and, on the other hand, to validate the model, thus giving the possibility to investigate the long-range motions of the protein, often linked to its biological function.

In conclusion, we have explored the structural and dynamic properties of Ngb in the deoxy and the CO-bound forms, combining MD simulations and MS XANES theory. We unveiled the origin of a discrepancy observed between the XANES and EXAFS results, occurring when the XANES spectra are analyzed using a static model cluster. We have found that the analysis of the XANES experimental data of heme proteins in solution can be successfully improved by using a reliable description of the dynamical behavior of the heme pocket. We have developed a general procedure to analyze the XANES in combination with MD simulations that can be applied to the study of other biological macromolecules.

Acknowledgment. A.A. was supported by funds assigned in the research program D.1 "Hemeproteins of biomedical interest" of the Catholic University. This work was also supported by CASPUR with the Standard HPC Grant 2010 entitled "A combined X-ray absorption spectroscopy, Molecular Dynamics simulations and Quantum Mechanics calculation procedure for the structural characterization of ill-defined systems". We acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities and the staff of BM30B for the helpful support.

Supporting Information Available: Figures S1−S6, movies 1 and 2, and complete ref 28. This material is available free of charge via the Internet at http://pubs.acs.org.

JA1056533

(52) Banci, L. *Curr. Opin. Chem. Biol.* **2003**, *7*, 143–149.
(53) Shi, W.; Chance, M. R. *Cell. Mol. Life Sci.* **2008**, *65*, 3040–3048.