UNIVERSITÁ DEGLI STUDI DI MODENA E REGGIO EMILIA

DOTTORATO DI RICERCA IN
Multiscale Modelling, Computational Simulations and Characterization in Material and Life Sciences

XXVII Ciclo

Investigation of the effects of the environment on the dynamics/function relationship in metallo-proteins

Candidato: Licia Paltrinieri

Relatore (Tutor) : Prof. Gianantonio Battistuzzi

Correlatore (Co-Tutor) : Dott. Carlo Augusto Bortolotti

Direttore della Scuola di Dottorato: Prof.ssa Ledi Menabue

"Why is a raven like a writing desk?"

Lewis Carroll, *Alice's Adventures in Wonderland*

Contents

1 Introduction

Proteins which contain one or more metal ions are called metalloproteins, and constitute up to one third of all known proteins[1, 2]. The metals which are essential for life are relative few, compared to all the metals present in nature: iron, zinc, copper, manganese, molybdenum, cobalt, vanadium, nickel and tungsten[2, 3]. Metalloproteins utilize the peculiar chemical properties of metals to perform a wide variety of specific life processes, and they can be considered as highly elaborated metal complexes, in which the protein matrix is the ligand[1, 2, 4]. Metalloproteins are involved in a large number of vital processes, and they can be grouped according to their function[2]. Different types of metalloproteins can be recognized: those performing a structural role; storage proteins, that are involved in the uptake and the release of metal ions in a soluble form; electron transfer (ET) proteins, fundamental for the uptake, release and storage of electrons; oxygen binding proteins, required for the coordination and release of atmospheric oxygen; and catalytic proteins, which can perform enzymatic reactions[1, 2, 3, 4].

The different roles played by metalloproteins depend on the different metals and on the specific interactions between metal and the protein surrounding[5]. Redox metalloproteins are particularly relevant, since they can be involved either in electron transfer (ET) or in catalysis of oxidation-reduction reactions[6], depending on their structure and chemistry.

Usually, ET proteins feature coordination sites which undergo to minimal structural changes upon oxidation or reduction, in order to minimize the reorganization energy and to increase the electron transfer rate[7, 8, 9, 10, 11]. On the other hand, the metal site of redox metalloenzymes typically has a vacant coordination position, required for substrate binding[1, 2, 3, 4]. Thus, ET proteins and catalytic redox proteins are characterized by different structural features: the former are usually smaller and feature a rigid structure, while the latter are more flexible and their active site is more solvent exposed compared to ET proteins, allowing it to bind and react with substrates.

Proteins are crucial for all the physiological processes of every living organisms[12, 13, 14, 15, 16]. Thus, it is of fundamental importance to correctly understand the molecular details of the mechanism allowing a macromolecule to perform its biological role. The implementation of high-resolution X-ray crystallography, NMR spectroscopy and other experimental techniques, allowed to gather extraordinary relevant information concerning the relationship between the structure and the function of a number of proteins[17, 18, 19]. Nevertheless, the ”static” structure resulting from these type of investigations, does not always take into account the large variety of conformations which are accessible for a protein, as well as their possible biological roles.

The functional behavior of a protein is deeply affected by its dynamic properties. Therefore, characterization of the latter is of fundamental importance for the correct understanding of the functional role of proteins[13]. To deeply understand how the structural features of a protein are linked to its dynamic behavior it is necessary to take into account the fourth dimension, time[20]. To this end, Molecular Dynamics (MD) provides a powerful tool to investigate how the dynamics of a protein affects its functionality and to exploring the wide structural heterogeneity accessible to a protein, providing atomic resolution and achieving time scales that are compatible with many of the biological processes which deserve investigation[21]. Over the last decades, molecular dynamics simulations of biological macromolecules have been developed and optimize to an high level of reliability, becoming a common tool for proteins characterization[13, 22, 23].

It is universally known that the relationship between a protein and its environment, either water for soluble proteins or lipids for those embedded in membranes, has a crucial role in determining its functionality and dynamics[10, 11, 20, 24, 25, 26, 27]. Therefore, the aim of this work is to study, using MD, the functionality of several proteins, to understand how their behavior is affected both by their internal dynamics and by the dynamic interplay with the surrounding environment.

In particular, my research focused on four different types of proteins: cupredoxins, a multi-heme cytochrome *c*, six-coordinate globins, and a voltage-dependent membrane channel protein. These systems differ primarily for the presence or absence of a metal center in their structure. The first three systems indeed present a metal active site, whereas the membrane protein does not feature any metal in its structure. Moreover, the species featuring a metal center differ for the type of metal active site they present. Cupredoxins feature a copper metal site, multi-heme cytochrome *c* possesses two *c* type heme groups, and six-coordinate globins present a *b* type heme group, which differs from the *c* type heme mainly for its lacking of covalent bonds with the protein matrix. The thread that links the investigation of such different systems is to find how the functionality of these species is affected by the dynamic interaction between the protein scaffold and the surrounding environment, providing, thanks to MD simulations and analysis, atomic and molecular descriptors that could explain macroscopic functional properties experimentally determined.

Cupredoxins, or blue-copper proteins, are copper-containing ET proteins featuring a type 1 (T1) copper center, that display an intense blue color, due to a strong absorption around 600 nm. T1 copper proteins differ from other copper proteins for their exclusive geometry and ligand sets. Indeed, cupredoxins feature a distorted tetrahedral T1 copper site that is not in the preferred geometry for either Cu(II), which tends to be square planar, or Cu(I), which presents a more regular tetrahedral arrangement[28]. As a consequence, copper sites in cupredoxins undergo very small structural rearrangement upon redox interconversion[29, 30, 31, 32], providing the proteins optimal features for ET process. The copper atom is generally coordinated by two histidine residues and one cysteine residue in a trigonal plane, and the axial position is often occupied by a methionine residue at a relatively longer distance[6, 33, 34, 35]. The highly covalent copper-thiolate bond is responsible for the intense blue color of these proteins[36, 37]. The T1 copper centers can be found in either single- or multiple-domain proteins[38]. The former consists of the most common blue copper proteins, such as azurin, plastocyanin and amicyanin, whereas the latter includes proteins such as stellacyanin, uclacyanin and dicyanin[6]. Cupredoxins are present in archaea, bacteria and plants, where they exclusively serve as ET centers[6]. Starting from 1978, when the first crystal structure of a T1 copper protein was reported[39], several other cupredoxins crystal structures were resolved[40, 41, 42, 43]. Although the sequence identity between the various members of cupredoxins family is less than 20%[44], the overall three-dimensional structure of these proteins is highly conserved[6]. This common structure is called cupredoxins fold, and it consists of eight β-strands arranged into a Greek key β-barrel[45] and of one or two α-helices in different positions outside the core fold of the protein[6] (see figure 1.1).

An important structural feature common in all cupredoxins is a hydrophobic loop linking the two C-terminal β-strands, which contains three of the copper ligands, namely the Cysteine, one of the Histidine and the Methionine ligands. This loop region is localized on the surface of the protein, surrounding the metal atom and shielding it from direct contact with the solvent. Moreover, residues in this hydrophobic loop are involved in a number of weak interactions, such as hydrogen bonds, with other region of the protein and with the solvent, which are thought to be crucial for the

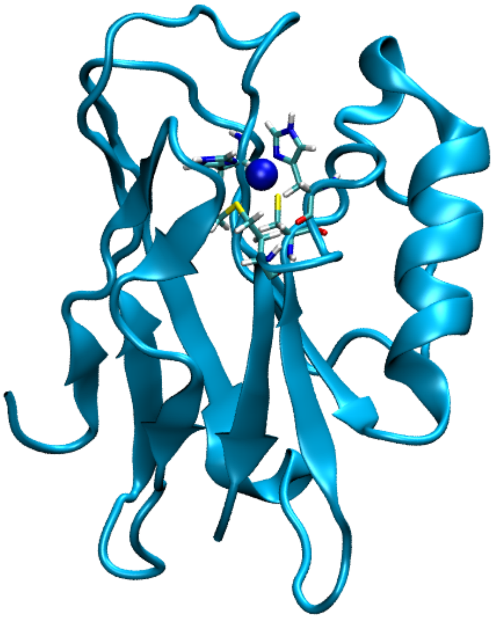


Figure 1.1: Three-dimensional cartoon structure of azurin from P. aeruginosa (PDB ID code: 4AZU[46]). The copper atom is represented as a blue sphere, the copper ligands as licorice. The image was prepared using VMD software[47].

properties and functionality of the whole protein[35, 33, 34, 48, 49, 50, 51, 52]. Thus, having this loop region hydrophobic properties and being it in direct contact with the solvent, it become very important to understand which are its interactions, both structural and dynamic, with the surrounding environment (i.e., solvent molecules) and how these interactions could have a role in controlling the whole functionality of the protein. Therefore, I performed molecular dynamics simulations on wild type azurin from *Pseudomonas aeruginosa* and four of its mutants, in which the hydrophobic loop containing the copper ligands has been replaced with that of other members of the cupredoxins family or with synthetic sequences featuring only Alanine residues[46, 48, 50, 52]. Since these proteins differed only in the composition of a small region, it was possible to focus exclusively on the contribution of this region on the electron transfer properties of this type of proteins. In particular, the determination of the ET reorganization energies[53] and the detailed study of the so-called *acid transition* (a conformational rearrangement known to occur in many reduced cupredoxins)[54], revealed the crucial role of the hydrophobic ligand-containing loop and of the effects of the dynamic interaction between proteins and solvent in controlling the functional behavior of these proteins.

The presence of multiple heme cofactors in a single protein provides to the biomoluecule the ability to rapidly exchange electrons through relatively long distances, thus enhancing the ET properties. Thus, multiheme proteins can be considered as very efficient molecular machines evolved to performed physiological reactions demanding multiple electrons, or to realize electron transport over long distances[6, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65]. The multi-heme cytochrome *c* investigated in this work is the di-heme cytochrome *c* (DHC) from *Rhodobacter sphaeroides*, one of the simplest vectorial electron transfer species, featuring two bis-histydil heme groups inserted into two distinct domain[66]. The first, N-terminal, domain, presents the typical characteristics of a class I cytochrome *c*, while the second, C-terminal, domain has no similarity to any known structures, therefore forming a novel cytochrome *c* structural motif[66] (see figure 1.2). Although the physiological function of this di-heme protein is not yet completely clear, its most plausible role is to be the electron donor for SHP protein, which is known to bind oxygen[67].

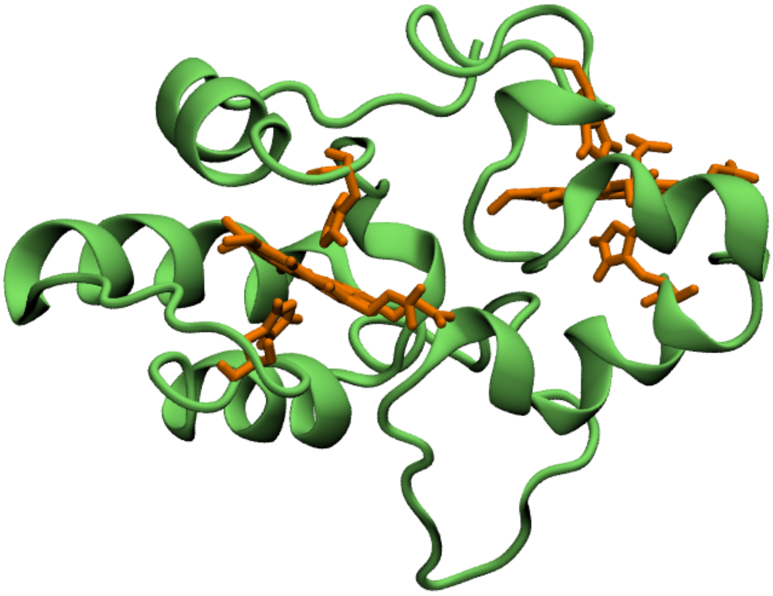
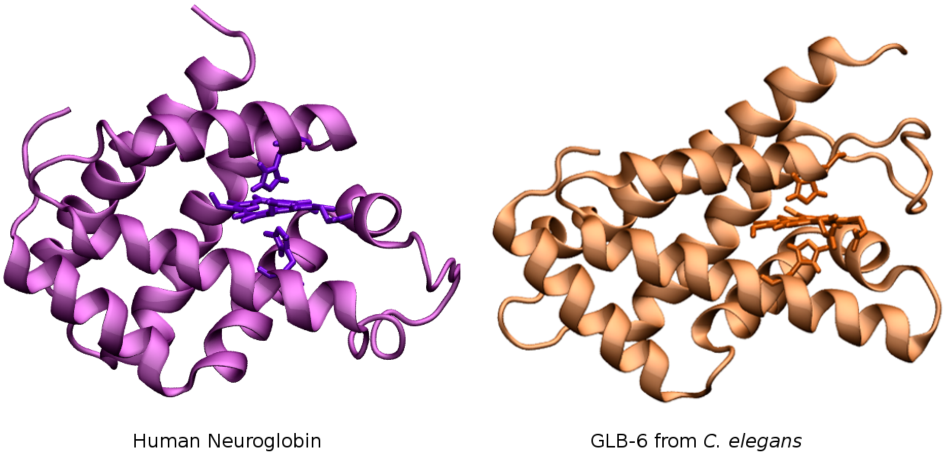


Figure 1.2: Three-dimensional cartoon structure of DHC from Rb. sphaeroides (PDB ID code: 2FWT[66]). The two heme group and their axial ligands are shown as orange licorice. The image was prepared using the VMD software[47].

In this work, performed in close collaboration with Dr. Isabella Daidone, Massimiliano Aschi (University of L'Aquila) and Andrea Amadei (University of Rome Tor Vergata), a combined QM/MM method, the Perturbed Matrix

Method (PMM)[68, 69, 70, 71, 72, 73], was used to calculate the reduction potentials of the redox centers in DHC, allowing to assign the previously reported experimental values[66] to the corresponding heme center. The very good agreement between calculated and experimental reduction potentials confirms the reliability of our theoretical approach, suggesting its applicability to more complex biosystems featuring several redox centers[74]. The investigation of this simple multi-heme protein, taken as a model system, allowed to gain important informations on the effects of the solvent exposure of the two different heme groups on the ET process, therefore helping to comprehend the molecular details beneath the relationship between the environment and the dynamics of the protein.

During the last year of my PhD, I begin to study six-coordinate globin proteins, focusing my investigation on human neuroglobin and the globin GLB-6 from the nematode *Caenorhabditis elegans*. They both feature a bis-histydil *b* heme group, surrounded by a compact α-helical fold (see figure 1.3).


Human Neuroglobin

GLB-6 from *C. elegans*

Figure 1.3: Three-dimensional structure of human neuroglobin (left) (PDB ID code: 4MPM[75]) and GLB-6 from C. elegans (right) (PDB ID code: 3MVC[76]). The heme sites and their axial ligands are represented as licorice. The image was prepared using the VMD software[47].

My aim is to verify their possible role as electron transfer (ET) proteins. Although they were deeply investigated and characterized, their physiological function(s) is still debated. Since they feature a strong axial coordination of the heme site, several studies hypothesize their possible role in redox biological processes[77, 78, 79, 80]. Therefore, I studied them from a redox metalloproteins point of view, searching for the structural and dynamic properties which can strengthen their possible function as ET proteins. In particular, I calculated the reorganization energies for both human neuroglobin and GLB-6, obtaining values in close agreement with that of typical ET proteins. Furthermore, human neuroglobin can feature an internal disulfide bridge between residues Cys46 and Cys55, depending on the redox state of the external environment[81, 82, 83], and the presence or the absence of this S-S bond is supposed to affect the overall functionality and dynamics of the protein. Thus, I simulated neuroglobin with and without this disulfide bridge, in order to discover the structural and dynamic features which are beneath the different macroscopic behavior of the two forms of the protein. Moreover, human neuroglobin is supposed to interact with human cytochrome *c* in preventing apoptosis[77, 78, 80, 84, 85, 86]. In particular, it is supposed that ferrous neuroglobin localized in the cytosol can reduce ferric cytochrome *c*, which is involved in apoptosis, while the ferrous cytochrome *c* is apparently inactive in the initiation of apoptosis. It is therefore plausible to consider an inter-molecular electron transfer reaction between human neuroglobin and human cytochrome *c*, and considering the reduction potentials featuring by the two proteins (the experimental reduction potential of neuroglobin is -115 mV vs SHE, whereas the one of cytochrome *c* is $\sim +255$ mV vs SHE) it is possible to suppose that in the ET process reduced neuroglobin transfers an electron to oxidized cytochrome *c*. Thus, I performed molecular dynamics simulations of the adduct between the proteins before (using the reduced neuroglobin structure and the oxidized cytochrome *c* one) and after (using the oxidized neuroglobin structure and the reduced cytochrome *c* one) the electron transfer process, in order to understand the effect(s) that the change in the oxidation states of the two proteins has on the stability of their complex, in particular focalizing my attention on the presence of weak interaction (such as salt bridges, hydrogen bonds and hydrophobic interactions) which can stabilize the complex, and on the possible insertion of solvent molecules between the interacting regions of the proteins in the complex. The results so far obtained indeed agree with the hypothesis that these six-coordinate globins can act as electron transfer proteins *in vivo*, and, moreover, they highlight the important role that the interplay between the protein dynamics and the external environment has in controlling the functionality of a protein.

During my PhD, I also spent six months working with the group of professor de Groot at Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, where I studied the functional behavior of the Voltage-Dependent Anion Channel (VDAC), a membrane channel protein mainly localized in the Outer Mitochondrial Membrane which regulates a number of physiological processes[87]. VDAC is a relatively small protein, functioning as a regulator of the entrance end exit of several metabolites, in this way governing the communication between mitochondria and the rest of the cell. It is composed by 19-stranded β-barrel and an N-terminal α-helix, positioned inside the channel (see figure 1.4).

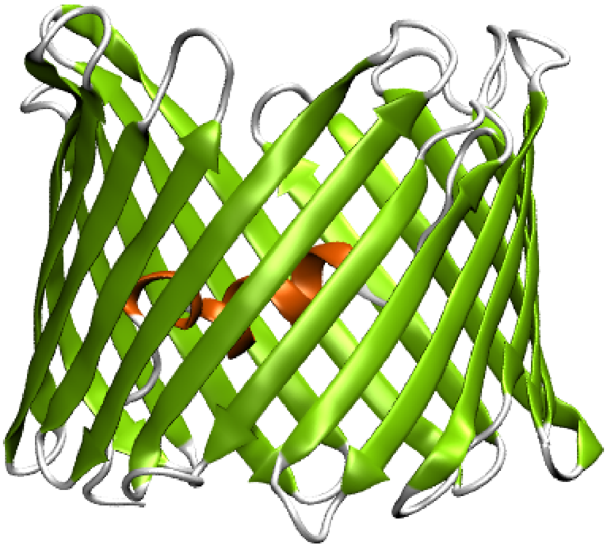


Figure 1.4: Three-dimensional structure of VDAC from Mus musculus (PDB ID code: 3EMN[88]). The β-strands are colored in green, while the N-terminal α-helix is colored in orange. The image was prepared using the VMD software[47].

Depending on several factors, first of all the value of the transmembrane potential, VDAC can undergo a structural rearrangement, in particular, at high values of both positive and negative potentials, VDAC channel tends to close, featuring several sub-states with different ionic selectivities and permeabilities[87, 89, 90]. Another factor that could affect the functional properties of VDAC is the composition of the lipid membrane in which it is embedded[91, 92, 93, 94].

Therefore, Computational Electrophysiology[95], a computational method developed by the de Groot group which simulates ion flux through membrane channels and allows to accurately predict ions conductance and selectivity, permitting the comprehension of ion conduction mechanisms[95, 96], was used to investigate the electrophysiological properties of VDAC from *Mus musculus* embedded into two different lipid membrane, namely POPC (1-

Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphoCholine) and POPE (1-Palmitoyl-2-Oleoyl-sn-3PhosphoEthanolamine), applying different transmembrane potential values, from a minimum of ~ 0.05 V to a maximum of ~ 0.4 V, in order to unravel the influence of the lipid environment and of the change in the voltage on the selectivity and permeabilities of ions and on the structural rearrangement of VDAC.

A more detailed structural and molecular description of the investigated systems is given in the corresponding chapters.

References

[1] J. A. Cowan. Wiley-VCH: New York, 1997.

[2] R. H. Holm, P. Kennepohl, and E. I. Solomon. In: *Chem. Rev.* 96 (1996), pp. 2239–2314.

[3] I. Bertini et al. University Science Book: Sausalito, CA, 1994.

[4] S. J. Lippard and J. M. Berg. University Science Books: Mill Valley, CA, 1994.

[5] F. Arnesano et al. In: *Coord. Chem. Rev.* 250 (2006), pp. 1419–1450.

[6] J. Liu et al. In: *Chem. Rev.* 114 (2014), pp. 4366–4469.

[7] C. A. Bortolotti et al. In: *J. Phys. Chem. Lett.* 2 (2011), pp. 1761–1765.

[8] V. Tipmanee et al. In: *J. Am. Chem. Soc.* 132 (2010), pp. 17032–17040.

[9] M. Cascella et al. In: *P. Natl. Acad. Sci. USA* 103 (2006), pp. 19641–19646.

[10] A. Stock. In: *Nature* 400 (1999), pp. 221–222.

[11] L. B. Sagle et al. In: *J. Am. Chem. Soc.* 128 (2006), pp. 7909–7915.

[12] K. Degtyarenko. In: *Bioinformatics* 16 (2000), pp. 851–864.

[13] L. Banci. In: *Curr. Opin. Chem. Biol.* 7 (2003), pp. 143–149.

[14] H. Feinberg, H. M. Greenblatt, and G. Shoham. In: *J. Chem, Inf. Model.* 33 (1993), pp. 501–506.

[15] H. B. Gray and J. R. Winkler. In: *Q. Rev. Biophys.* 36 (2003), pp. 341–372.

[16] A. I. Kotelnikov, E. S. Medvedev, and N. S. Goryachev. In: *Russ. Chem. B+* 60 (2012), pp. 1318–1341.

[17] C. Mapelli et al. In: *Biopolymers* 28 (1989), pp. 123–128.

[18] R. J. P. Williams. In: *Eur. J. Biochem.* 183 (1989), pp. 479–497.

[19] A. M. Gronenborn and G. M. Clore. In: *Protein Seq. Data Anal.* 2 (1989), pp. 1–8.

[20] K. Henzler-Wildman and D. Kern. In: *Nature* 450 (2007), pp. 964–72.

[21] L. Banci et al. In: *J. Biol. Inorg. Chem.* 4 (1999), pp. 824–37.

[22] M. Karplus. In: *Acc. Chem. Res.* 35 (2002), pp. 321–323.

[23] J. A. McCammon, B. R. Gelin, and M. Karplus. In: *Nature* 267 (2002), pp. 585–590.

[24] C. A. Bortolotti et al. In: *J. Am. Chem. Soc.* 134 (2012), pp. 13670–13678.

[25] S. S. Skourtis, D. H. Waldeck, and D. N. Beratan. In: *Annu. Rev. Phys. Chem.* 61 (2010), pp. 461–485.

[26] J. Zimmermann et al. In: *J. Phys. Chem. Lett.* 2.5 (2011), pp. 412–416.

[27] H. K. Ly et al. In: *ChemPhysChem* 11 (2010), pp. 1225–1235.

[28] H. B. Gray, B. G. Malmström, and R.J.P. Williams. In: *J. Biol. Inorg. Chem.* 5.5 (2000), pp. 551–559.

[29] T. Inoue et al. In: *J. Biol. Chem.* 274 (1999), pp. 17845–17852.

[30] B.R. Crane et al. In: *J. Am. Chem. Soc.* 123 (2001), pp. 11623–11631.

[31] J.M. Guss et al. In: *J. Mol. Biol.* 192 (1986), pp. 361–387.

[32] W. E. B. Shepard et al. In: *J. Am. Chem. Soc.* 112 (1990), pp. 7817–7819.

[33] C. Dennison. In: *Coordin. Chem. Rev.* 249.24 (2005), pp. 3025–3054.

[34] C. Dennison. In: *Dalton T.* 25 (2005), pp. 3436–3442.

[35] C. Dennison. In: *Nat. Prod. Rep.* 25 (2008), pp. 15–24.

[36] E. I. Solomon et al. In: *Chem. Rev.* 104 (2004), pp. 419–458.

[37] T. F. Oliveira et al. In: *J. Biol. Chem.* 283 (2008), pp. 34141–34149.

[38] Y.-C. Hsieh et al. In: *Mol. Microbiol.* 78 (2010), pp. 1101–1116.

[39] D. Aragao et al. In: *J. Biol. Inorg. Chem.* 8 (2003), pp. 540–548.

[40] M. Saraste. In: *Science* 283 (1999), pp. 1488–1493.

[41] R. Baradaran et al. In: *Nature* 494 (2013), pp. 443–448.

[42] T. Tsukihara et al. In: *Science* 272 (1996), pp. 1136–1144.

[43] R. G. Efremov and L. A. Sazanov. In: *Nature* 476 (2011), pp. 414–420.

[44] C. T. Saouma, W. Kaminsky, and J. M. Mayer. In: *J. Am. Chem. Soc.* 134 (2012), pp. 7293–7296.

[45] E. T. Adman. In: *Adv. Protein Chem.* 42 (1991), pp. 145–197.

[46] H. Nar et al. In: *J. Mol. Biol.* 221 (1991), pp. 765–772.

[47] W. Humphrey, A. Dalke, and K. Schulten. In: *J. Mol. Graphics* 14 (1996), pp. 33–38.

[48] C. Li, M. J. Banfield, and C. Dennison. In: *J. Am. Chem. Soc.* 129 (2007), pp. 709–718.

[49] C. Li et al. In: *Inorg. Chem.* 50.7 (2011), pp. 482–488.

[50] C. Li et al. In: *Proc. Natl. Acad. Sci. USA* 103 (2006), pp. 7258–7263.

[51] H. Li et al. In: *J. Am. Chem. Soc.* 36 (2004), pp. 8010–8019.

[52] K. Sato et al. In: *Proc. Natl. Acad. Sci. USA* 106 (2009), pp. 5616–5621.

[53] L. Paltrinieri et al. In: *J. Phys. Chem. Letters* 4 (2013), pp. 710–715.

[54] L. Paltrinieri et al. In: *Biochemistry* 52 (2013), pp. 7397–7404.

[55] C. M. Paquete and R. O. Louro. In: *Accounts Chem. Res.* 47.1 (2014), pp. 56–65.

[56] K. D. Bewley et al. In: *Biochim. Biophys. Acta* 1827 (2013), pp. 938–948.

[57] S. Sharma, G. Cavallaro, and A. Rosato. In: *J. Biol. Inorg. Chem.* 15 (2010), pp. 559 –571.

[58] C. M. Paquete et al. In: *Biochim. Biophys. Acta* 1767 (2007), pp. 1169–1179.

[59] C. M. Paquete and R. O. Louro. In: *Dalton T.* 39 (2010), pp. 4259–4266.

[60] B. M. Fonseca et al. In: *FEBS Lett.* 586 (2012), pp. 504–509.

[61] P. O. Quintas et al. In: *Biochim. Biophys. Acta* 1827 (2013), pp. 745–750.

[62] X. Qian et al. In: *Biochim. Biophys. Acta* 1807 (2011), pp. 404–412.

[63] M. J. Edwards et al. In: *Biochem. Soc. T.* 40 (2012), pp. 1181–1185.

[64] D. R. Lovley. In: *Biochem. Soc. T.* 40 (2012), pp. 1186–1190.

[65] J. A. Gralnick. In: *Biochem. Soc. T.* 40 (2012), pp. 1178–1180.

[66] H. R. Gibson et al. In: *Biochemistry* 45 (2006), pp. 6363–6371.

[67] B.-R. Li et al. In: *Biochem. Soc. Transact.* 36 (2008), pp. 992–995.

[68] M. Aschi et al. In: *Chem. Phys. Lett.* 344 (2001), pp. 374–380.

[69] A. Amadei, M. D'Alessandro, and M. Aschi. In: *J. Phys. Chem. B* 108 (2004), pp. 16250–16254.

[70] M. Aschi et al. In: *J. Phys. Org. Chem.* 19 (2006), pp. 518–530.

[71] C. Zazza et al. In: *J. Phys. Chem. B* 112 (2008), pp. 3184–3192.

[72] I. Daidone et al. In: *Chem. Phys. Lett.* 488 (2010), pp. 213–218.

[73] A. Amadei et al. In: *Curr. Opin. Struc. Biol.* 20 (2010), pp. 155–161.

[74] I. Daidone et al. In: *J. Phys. Chem. B* 118 (2014), pp. 7554–7560.

[75] B. G. Guimarães et al. In: *Acta Crystallogr. D* 70 (2014), pp. 1005–1014.

[76] J. Yoon et al. In: *Biochemistry* 49 (2010), pp. 5662–5670.

[77] A. Fago et al. In: *FEBS Lett.* 580 (2006), pp. 4884–4888.

[78] A. Giuffrè et al. In: *IUBMB life* 60 (2008), pp. 410–413.

[79] S. H. Bønding et al. In: *Int. J. Biol. Macromol.* 43 (2008), pp. 295–299.

[80] T. Burmester and T. Hankeln. In: *J. Exp. Biol.* 212 (2009), pp. 1423–1428.

[81] A. D. Nadra et al. In: *Proteins* 71 (2008), pp. 695–705.

[82] D. Hamdane et al. In: *J. Biol. Chem.* 278 (2003), pp. 51713–51721.

[83] A. Bocahut et al. In: *J. Phys. Chem. B* 113 (2009), pp. 16257–16267.

[84] T. Brittain et al. In: *IUBMB life* 62 (2010), pp. 878–885.

[85] L. Kiger et al. In: *PloS ONE* 6 (2011), e20478.

[86] M. Fiocchetti et al. In: *Biochim, Biophys. Acta* 1834 (2013), pp. 1744–1749.

[87] V. Shoshan-Barmatz et al. In: *Mol. Aspects Med.* 31 (2010), p. 227.285.

[88] R. Ujwal et al. In: *Proc. Natl. Acad. Sci. USA* 105 (2008), pp. 17742–17747.

[89] V. De Pinto et al. In: *FEBS Lett.* 584 (2010), pp. 1793–1799.

[90] D Gincel, S. D. Silberberg, and V. Shoshan-Barmatz. In: *J. Bioenerg. Biomembr.* 32 (2000), pp. 571–583.

[91] C. Ader et al. In: *Nat. Struct. Mol. Biol.* 15 (2008), pp. 605–612.

[92] C. Hunte and S. Richers. In: *Curr. Opin. Struct. Biol.* 18 (2008), pp. 406–411.

[93] D. Nietlispach and A. Gautier. In: *Curr. Opin. Struct. Biol.* 21 (2011), pp. 497–508.

[94] R. Phillips et al. In: *Nature* 459 (2009), pp. 379–385.

[95] C. Kutzner et al. In: *Biophys. J.* 101 (2011), pp. 809–817.

[96] U. Zachariae et al. In: *Structure* 20 (2012), pp. 1540–1549.

2 Methods

2.1 Molecular Dynamics

Molecular Dynamics (MD) is a computational chemistry technique which allows the determination of the behaviors of atomic or molecular systems at a macroscopic level, starting from interactions at the microscopic level. A molecular dynamics simulation is basically characterized by the sampling algorithm in the phase space, and the choice of the interaction potential, V(r), between the particles of the system. The way in which the phase space in sampled distinguishes the various simulative techniques. Classical molecular dynamics extracts trajectories of the atoms (i.e., a set of configurations) integrating the Newton's motion equations. Other methods utilize different algorithms to generate the set of configurations, and can be divided into two groups, one that collects the systematic search techniques, while the other includes the random search methods. At the beginning of the informatic era, Alder and Wrainwright carried out the first molecular dynamics simulations[1], studying simple fluids using models which represent atoms as hard spheres. As the potential functions used for the calculations got more complex, thus allowing to obtain simulated data directly comparable with experimental results, it became immediately evident that MD could be used as a powerful and reliable tool for the investigation of the dynamic properties of real system. Over the years, the development of more powerful computers and the appearance of high-level programming languages made possible to simulate always more complex system, from water to small protein. At the end of the 70s, the first simulations of complex biomolecule were performed. The expansion of MD in studies of biochemistry is due to the successful use of it in the reproduction of experimental data of proteins and macromolecules[2, 3, 4, 5, 6]. Nowadays, MD simulations allow to study a broad spectrum of chemical and biological processes, such as structural and functional changes as a result of mutation in proteins, thermodynamic and kinetic properties of variuos systems[7, 8], investigations and development of new drugs, etc. The exponential growth of the computers power, in conjunction with the development of parallel machines, allowed in a few years to increase the simulation time, the size of the simulated systems and the improvement of the obtained results.

2.1.1 Equation of motion

Classical MD merely describe a system of particles through the motion of their nuclei. In fact, the Born-Oppenheimer approximation allows the separation of the motion of the nuclei from the motion of the electrons, therefore treating them implicitly. Thus, the motion of the molecular system appears to be described as a system of punctual masses that move thanks to an effective potential field produced by the electrons. It is reasonable to not use a quantum treatment: if you exclude the H and He atoms, the motion of the nuclei at room temperature can be described with reasonable accuracy by classical mechanics. In classical MD, the trajectory of a molecular system (i.e., its configurations as a function of time) is generated by the simultaneous integration of the Newton's equations of motion, for all the atoms of the system:

$$\frac{d^2\mathbf{r}_i}{dt^2} = m_i^{-1}\mathbf{F}_i \tag{2.1}$$

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{r}_1, \ldots, \mathbf{r}_N)}{\partial \mathbf{r}_i} \tag{2.2}$$

The force acting on the i-th atom is $\mathbf{F}_i$, the mass is m_i, and t is time, while $V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the potential energy function, the so-called force field, which represents the forces with which nuclei interact. This function is an effective interaction, since the effect of the (omitted) electronic degrees of freedom is mediated and embedded in the effect of the atomic degrees of freedom expressed in the model.

2.1.2 The Potential Function

In MD it is possible to use different force field, mostly depending on the system under investigation. A typical effective potential for systems consisting of N atoms with molecular mass M_i (i=1,2, …, N) and positive cartesian vectors r_i, has the following form:

$$\begin{aligned} V(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N) &= \sum_{legami} \frac{1}{2} K_b (b - b_{eq})^2 + \sum_{angoli} \frac{1}{2} K_\theta (\theta - \theta_{eq})^2 \\ &+ \sum_{diedri} K_\phi [1 + cos(n\phi - \delta)] + \sum_{d.impropri} \frac{1}{2} K_\xi (\xi - \xi_{eq})^2 \\ &+ \sum_{coppie} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \sum_{coppie} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.3) \end{aligned}$$

The first term is the vibration energy of the covalent bond between two atoms linked by a totally harmonic potential, where $b_{eq}$is the minimum bond energy, and K_b is the force constant which depends on the bond type; both are derived from experimental data and/or quantum calculations. The second term (three bodies interaction) represents the energy due to the deformation of the valence angle, θ, expressed as a harmonic potential; θ_{eq} is the reference valence angle and K_θ id the force constant. The third and fourth terms represent the four bodies interactions: the third is a sine term that refers to the dihedral angles ϕ; the fourth, ξ, is of harmonic type, and it is used to describe the improper dihedral angles. The last two terms in the force field equation are the effective non-bonding interactions, expressed as van der Waals and Coulomb interactions between the i-th and the j-th atom at distance r_{ij}. The van der Waals term is expressed in different ways, depending on the force field. In the example it is expressed as classical Lennard-Jones function (constant ϵ_{ij} and σ_{ij}. The force field parameters can be determined in several ways, the most elegant method is to calibrate these parameters with the results of *ab initio* quantum calculations on small molecular aggregates. However, in this way the results are often not satisfactory, because of the many approximations which must be taken into account with this type of procedure. Alternatively, or in order to optimize a set of parameters obtained from *ab initio* calculations, experimental data can be used (crystal structures, lattice energies and dynamics, spectroscopic data, density, vaporization enthalpy, solvation free energy, NMR data, etc.)[9, 10, 11, 12]. The obtained parameters are valid only in the physical-chemical conditions in which they were determined and this implies that the force fields used in MD programs are specifically optimized for certain physical systems (organic molecules, inorganic molecules, biomolecules, etc.).

2.1.3 Long Range Interactions

In a MD simulation, most of the computation time is occupied by the calculation of non-bonding interactions. Therefore, it is appropriate to look for a method which allows to increase the computational efficiency of the simulation of the simulation acting on the van der Waals and Coulomb components. Various techniques have been developed to reduce the computational cost. The most commonly used are the cut-off radius and the Ewald summation. The first method severs the non-bonding interactions when the distance between the interacting particles is larger than a specific value, the so called cut-off radius, r_c. Therefore, the i.th particle of the system can interact only with the ones which are contained in a sphere of r_c radius, centered in r_i. Obviously, this method introduces approximations, especially in the electrostatic calculations, which can remarkably affect dynamic and structural properties of the system. To overcome this problem, in addition to increasing the cut-off radius, a switch function can be introduced, which ensures the continuity of dynamic and structural properties beyond the cut-off distance. Despite this method suffers from the aforementioned limitations, it remains among the most common for the calculation of intermolecular forces. The more recent second method proposes the use of a periodic lattice, in which all the interactions between the central box and its infinite replicas are considered. The Coulomb energy in a periodic system formed by N particles is obtained by summing over all the pairs of atoms. One is in the central box, and the other one is in a replica:

$$E = \frac{1}{8\pi\epsilon_0} \sum_{|\mathbf{n}|=0}^{\infty} \left(\sum_{i=1}^{N} \sum_{j=1}^{N} \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \right) \tag{2.4}$$

where n are the particles and i, j are the periodic imagines. It can be shown that this summation over n, for potentials varying with r^{-1}, can not be completely convergent. Namely, its limit can vary or diverge if the order of the terms is changed. Therefore, this expression can not be used in the calculation of the electrostatic energy. To solve this problem, each charge can be considered as surrounded by a Gaussian charge distribution ρ^G of equal intensity and opposite sign[13]. This Gaussian distribution will have the effect of shielding the interactions between neighbor charge, and we could consider the interaction energy of short range type. The total charge distribution is:

$$\rho_i(\mathbf{r}) = \rho_i^q(\mathbf{r}) + \rho_i^G(\mathbf{r}) \tag{2.5}$$

where $\rho_i^q(\mathbf{r})$ is the charge distribution of the i-th particle and $\rho_i^G(\mathbf{r})$ is the corresponding Gaussian distribution:

$$\rho_i(\mathbf{r}) = q_i\delta(\mathbf{r} - \mathbf{r}_i) \tag{2.6}$$

After the calculation of the interaction energy in real space, to return to he original distribution function $\rho_i(\mathbf{r})$, a function equal to $\rho_i^G(\mathbf{r})$ must be subtracted in reciprocal space, using the Fourier transformation. The final expression of the total interaction energy will be:

$$\begin{aligned} E &= \frac{1}{8\pi\epsilon_0}\sum_{i=1}^{N}\sum_{j=1}^{N}\Bigg(\sum_{|\mathbf{n}|=0}^{\infty}\frac{q_iq_j erfc(\alpha|\mathbf{r}_{ij}+\mathbf{n}|)}{|\mathbf{r}_{ij}+\mathbf{n}|} \\ &+ \frac{1}{\pi L^3}\sum_{\mathbf{k}\neq 0}\frac{4\pi^2 q_iq_j}{k^2}\exp(-k^2/4\alpha^2)\cos(\mathbf{k}\cdot\mathbf{r}_{ij})\Bigg) \\ &- \frac{\alpha}{4\pi^{3/2}\epsilon_0}\sum_{i=1}^{N}q_i^2 + \frac{|\sum_{i=1}^{N}q_i\mathbf{r}_i|^2}{2\epsilon_0 L^3(2\epsilon'+1)} \end{aligned} \tag{2.7}$$

where $erfc(x)$ is the complementary error function, which tends to zero for $x \to \infty$. With regard to the first term, for α large enough, the sum is reduced to a single term with $\mathbf{n} = 0$; the second term is a sum over the reciprocal vectors $k = 2\pi n/L$. The last two terms are, respectively, the correction function due to the presence of a self-interaction term in $\rho_i^G(\mathbf{r})$, and the contribution of the depolarizing field to the energy, which is affect by the outer dielectric effect. Recently, other methods have been developed, such as the Particle-Particle, Particle-Mesh (PPPM) or the Particle Mesh Ewald (PME)[14], which, based on the same principles on which Ewald summations were developed, are computationally more efficient and stable. In particular, in all the simulations of this work, the PME method was used for the calculations of electrostatic forces. The PME method is substantially identical to the Ewald summations, with the only difference being that for the calculation of the energy of certain charge distribution, the charges are considered arranged on a grid. This makes considerably more efficient the calculation of the part in which the Fourier transformation is involved, which constitutes the most expensive part from the computational point of view.

2.1.4 Boundary Conditions

An important feature of the MD simulations is the way in which the boundary conditions are treated. In general, because of the computational limits, a simulated molecular system consists of 10^4-10^5 atoms, significantly less than the corresponding real system. Because of this, the molecules of a simulated system will be more frequently subject to the so-called surface effects. For example, if they are contained in a cubic box, those which are close to the surface will suffer the effects of very different forces with respect those ones present in the bulk. The study of a simulated system of limited an finite size, can lead to the creation of artifacts compared to the actual behavior, especially in the case of homogeneous liquids or solutions. Usually, the periodic boundary conditions (PBC)[13] are introduced to reduce the surface effects. In this way the central box is surrounded by identical replicas that create a condition of infinite periodicity. During a simulation, when a molecule comes out from the original box, its periodic image enters the box from the opposite side, maintaining the same velocity and direction, and allowing the reproduction of periodic systems, such as crystals or fluids. Interactions within the PBC are treated with the convention of minimum image (MI), where an atom can interact only with the first neighbors. To avoid the anisotropy due to the cubic form of MI and artifacts arising from the interaction of an atom i of the box with an atom j and its periodic image j' simultaneously, a spherical cut-off with radius r_c is used, such that the shorter side of the box is greater than twice the value of cut-off ($L \geq 2r_c$, with L is the box side). This condition is necessary, but not enough, to prevent artifacts. For example, when you want to simulate a solute in a solvent, it is also necessary that a solute atom does not interact with solute atoms belonging to periodic images. To avoid this, it is necessary that the distance of each atom of the solute from each face of the box is greater than the half of cut-off radius ($d \geq 0.5r_c$, with d the distance between solute and box face).

2.1.5 Constraints in MD

In MD simulations different types of constraints can be applied. The most common are those that block the positions of certain atoms, the so-called *Position Constraints*. The use of constraints in MD is necessary to eliminate the high frequency vibrations of bonds, which require short integration steps. In the case of biological macromolecules in solution, the vibrations involving hydrogen atoms determine the higher frequency vibrations. For the O-H bond, the stretching frequency is in the order of 10^{14} Hz, so an average period is of the order of 10 fs[15]. This limits the integration step (time step) that can be used in a simulation to about 0.5 fs, considering that to properly reproduce the trend of a periodic function is necessary to sample at least 20 times per period. With the introduction of a method which binds these bonds, or, in practice, all the covalent bonds, one can arrive at an integration step of 2 fs, that is the commonly used value, also in this work. As the binding vibrations are virtually decoupled from the other system vibrations, this constraints does not alter the rest of the dynamics, while this does not occur for the vibrations of valence angles. The most commonly used method is the so-called SHAKE[16], which was introduced in 1977. The SHAKE method consists in an iterative procedure, for which after each integration step the initial positions of the atoms $\mathbf{r'}_i(t+\Delta t)$ vary until the constraints conditions are not fulfilled in new positions $\mathbf{r}_i(t+\Delta t)$. The use of SHAKE, despite some disadvantages, allow to increase the integration step of 2-4 times.

2.1.6 Initial Conditions

Once the force field, the boundary conditions and the constraints are defined, the initial coordinates and velocities of the system must be set. If the observation time were sufficiently long, the results should be independent from the initial conditions, but since most physical systems reach equilibrium in a very longer time than those simulated, it is appropriate to perform simulations starting from a configuration which is as close as possible to the equilibrium one. The coordinates for macromolecules can be obtained from experimental data, such as X-ray and NMR. The initial velocities can be achieved with a simple Maxwell distribution, which depends on the initial temperature of the simulation:

$$\rho_{x_i} = \sqrt{\frac{m_i}{2\pi k_B T}} exp\left(-\frac{\frac{1}{2}m_i v_{x_i}^2}{k_B T}\right) \tag{2.8}$$

where ρ_{x_i} is the probability density of the component v_{x_i}.

2.1.7 Integration of the Equation of Motion

Given the complexity of the force field expression (2.3), the integration of the equations of motion (in order to obtain the trajectory of the N atoms of the system) does not have an analytical solution. For this reason, several approximate methods have been developed. They numerically extract the trajectories from an MD simulation. The characteristics of a good integration algorithm can be summarized as follows:

- it must allow the use of a long time step: it is evident that the higher the integration step, the smaller the number of integrations required in a simulation. Also, if the Δt is too large, the risk is to fail to follow certain components of the motion of the system, such as the vibrations, thus leading to misleading results. Therefore, it is appropriate to adjust the integration step on the higher frequency motions;

- it must be reversible in time and preserve constant the quantities in the system. Newton's equations are reversible in time and therefore it is desirable that the algorithm used faithfully reproduce this data;

The algorithms differ in the accuracy of calculation, the memory occupation and the type of representation. The nature of the algorithms derives from the expansion of the Taylor series of the quantities $\mathbf{r}_i(t + \Delta t)$ and $\mathbf{v}_i(t+\Delta t)$. In MD, algorithms that reach up to the third order of the Taylor expansion are generally used, as algorithms of lower order are inaccurate, while those of higher order are too expensive from a computational point of view. Tipically, the following are used:

- Verlet algorithm[17, 18]
- leap-frog algorithm

The first algorithm is fast, requires little memory and has a good conservation of energy for short times. It is obtained by additions and subtractions of Taylor's expansions of the coordinates $\mathbf{r}_i(t+\Delta t)$ and $\mathbf{r}_i(t-\Delta t)$ of particles with mass m_i:

$$\mathbf{r}_i(t + \Delta t) = -\mathbf{r}_i(t - \Delta t) + 2\mathbf{r}_i(t) + \frac{(\Delta t^2)}{m_i}\mathbf{F}_i(t) \tag{2.9}$$

where $\mathbf{F}_i(t)$ is the force exerted on the particle i at time t and where terms higher than third order ones have been neglected. This pattern of integration is therefore correct to fourth order in Δt. The velocities are given by:

$$\mathbf{v}_i(t) = \dot{\mathbf{r}}_i(t) = \frac{1}{2\Delta t}[\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)] \tag{2.10}$$

The error on velocity is on the $(\Delta t)^3$ order.

The leap-frog is a variant of the integration scheme just seen. It is used in the GROMACS simulation package[19], which is the program used in all the simulations of this work. This scheme uses integer intervals for the coordinates and semi-integer intervals for velocities:

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t + \Delta t/2)\Delta t \tag{2.11}$$

$$\mathbf{v}_i(t + \Delta t/2) = \mathbf{v}_i\left(t - \frac{1}{2}\Delta t\right) + \frac{\mathbf{F}_i(t)\Delta t}{m_i} \tag{2.12}$$

The leap-frog algorithm is very efficient from a computational point of view, being simple and requiring less memory than other, more advanced, algorithms.

2.1.8 Temperature and Pression Control Methods

Different types of statistical ensembles can be used in MD, and they are characterized according to thermodynamic changes, restricted or let free to equilibrate with the external environment:

- micro canonical ensemble (NVE), characterized by constant numbers of particles (N), volume (V) and total energy (E), i.e. an isolated system;
- canonical ensemble (NVT), in which the internal energy of the system can vary, while the temperature is held constant by heat exchange with an external thermal bath, thus an isothermal and isochore system ;
- NPT ensemble, in which the volume can vary in order to maintain constant pressure, equal to the outside pressure, thus an isothermal and isobaric system.

In order to compare the results of an MD simulation with experimental data, the simulated system must be put in the same conditions in which these data were obtained, and this often forces to run the simulation at constant pressure and temperature. Over the years, various algorithms have been developed. Usually, the methods used to generate isothermal and isobaric statistical ensembles can be divided into two groups:

- *extended system methods*[20]. These methods use feedback mechanisms to correct deviations of monitored parameters (temperature and pressure) from predetermined values. This implies that the value can fluctuate, but the amplitude of fluctuations can be controlled.
- *constraints methods.*[21, 22, 23] The equations of motion are coupled to a suitable constraint (e.g., in the case of temperature, this constraint is the constant kinetic energy).

In the case of temperature, there are two more groups, the so-called stochastic methods[24] and the weak-coupling methods[25]. The first have the effect of adjusting the temperature, in order to reproduce the effect of an external bath, or to generate a Maxwellian velocity distribution at each step, to bring the value of the total kinetic energy to the one established. In the second type of methods, which have the great advantage to be simple, the algorithm simulates the coupling of the system with an external thermal bath at the temperature T_0 and the extent of interactions between this thermostated bath and the system is determined by an arbitrary time constant τ. The coupling is achieved by changing the velocity of the atoms at every step, i.e. multiplying it by a constant λ, while the temperature T is scaled to the reference value T_0 according to the equation:

$$\frac{dT(t)}{dt} = \frac{(T_0 - T(t))}{\tau} \tag{2.13}$$

If the time constant τ introduced in the Berendsen method corresponds with the integration step, then such thermostat corresponds with the Gaussian isothermal thermostat[26].

2.2 Structural and Dynamical Analysis

2.2.1 RMSD

One of the first analysis that generally are carried out is the calculation of the Root Mean Square Deviation (RMSD) compared to a reference structure in function of time. The purpose of this analysis is to obtain a first qualitative information on the stability of the simulation and the achievement of the equilibrium. It can be expressed through the following relation:

$$RMSD(t) = \left[\frac{1}{N} \sum_{i=1}^{N} (\mathbf{r}_i(t) - \mathbf{r}_i^{ref})^2 \right]^{\frac{1}{2}} \tag{2.14}$$

where $\mathbf{r}_i(t)$ represents the position of the i-th atom at time t, $\mathbf{r}_i^{ref}$ represents its position in the reference structure and N represents the total number of the generated structures.

2.2.2 Stability of the generated structure

Through special programs you can study the behavior of secondary structure of a protein during an MD simulation, in order to understand if conformational changes occur. A very common program is the DSSP by Kabsh and

Sander[27]. It is based on the identification of hydrogen bonds in the amino acid sequence.

2.2.3 RMSF

In addition to check the stability of a system during a simulation (through RMSD and the study of secondary structures), you can determine the extent of the fluctuations through the calculation of the Root Mean Square Fluctuations (RMSF). RMSF is given by the following relation:

$$RMSF(t) = \left[\frac{1}{N} \sum_{i=1}^{T} (\mathbf{r}_i(t) - \langle \mathbf{r}_i \rangle_t)^2 \right]^{\frac{1}{2}} \tag{2.15}$$

where $\langle \mathbf{r}_i \rangle_t$ indicates the average position of each atom and N indicates the total number of the structures generated during the simulation.

2.2.4 Essential Dynamics

Essential Dynamics (ED) is a methodology that research the collective degrees of freedom that best approximate the total fluctuations of a dynamic system[28, 29], based on PCA (Principal Components Analysis) of structures generated with MD. To perform a PCA analysis on an MD of a protein, the 3D Cartesian space must be replaced with a new multi-dimensional space: a system of N atoms is described by N points in a 3D space, and each point is associated with 3 coordinates, for a total of 3N coordinates. If the system is displayed in a 3N-dimensional space, the overall structure will be represented by a single point. In this way, you can have a collection of point starting from a collection of 3D structures (trajectories) such that at each point corresponds a particular conformation in time. As illustrated in figure 2.1, which represents a 2D example, the configurations set has a preferential direction, that is, the direction of maximum variance. It is thus possible to approximate data relative only to the position along that direction, excluding positions in other directions.

If this direction is chosen as the new spatial axis, then every point may be represented by a single coordinate. Extending the two-dimensional case, the procedure is similar, with the only difference that we are not only interested in the first direction of maximum variance, but also to the other containing gradually the maximum residual variance, that is to the so-called principal components. These directions imply a space, and the subspace responsible for major fluctuations is called essential subspace. The multidimensional model has been tested on proteins configurations from MD and has been

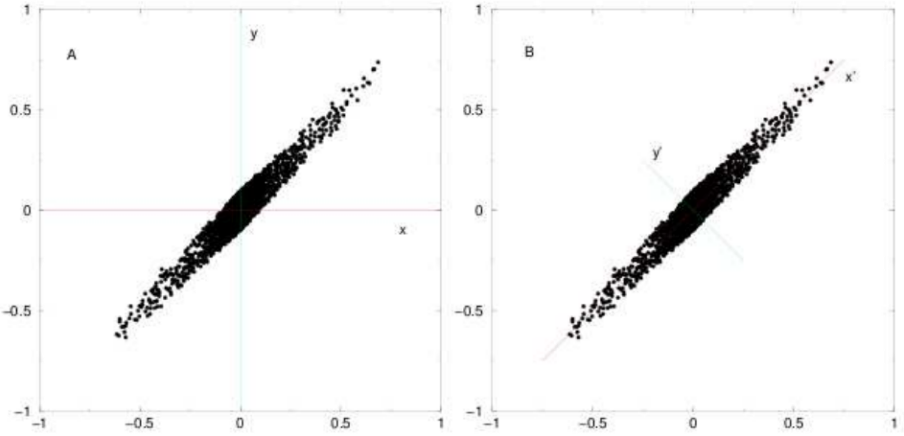


Figure 2.1: 2D example of essential dynamic. Panel A represents a collection of points after a simulation; to characterize a point you need two coordinates, x and y. Panel B represents the modification of the reference system, through the diagonalization of the covariance matrix C, such that, with an approximation, to describe a point only the x' coordinate is needing.

proven that the first 10-20 principal components are able to describe 90% of the fluctuations of a protein[28, 29, 30], i.e. a limited number of collective motions is responsible of most of the conformational fluctuations. This is possible because of the interactions, from atomic bonds to weak interactions, present in the protein structure. The first step of ED analysis involves the removal of the translational and rotational motion in order to obtain a system characterized only by internal motions, which are the object of interest of the dynamics of a protein. Then, the principal components analysis can be used, and it is based on the construction and diagonalization of the covariance matrix C of the fluctuations. To do this we define the 3N dimensional column vector $\mathbf{X}(t)$, which contains the atomic coordinates at time t, then the covariance matrix can be build, through the following equation:

$$\mathbf{C} = \langle \Delta\mathbf{X}\Delta\mathbf{X}^T \rangle \tag{2.16}$$

where $\Delta\mathbf{X} = \mathbf{X}(t) - \langle\mathbf{X}\rangle$, and the $\langle\rangle$ symbols mean an average over time. Related motions of particles corresponds to matrix elements with positive or negative high values, depending on positive or negative correlations. Independent motions correspond to matrix elements with low values. Through the orthogonal transformation $\mathbf{T}$, which contains the eigenvectors (i.e., the principal components) of C, the matrix is diagonalized, thus obtaining the corresponding eigenvalues in the diagonal matrix $\mathbf{\Lambda}$:

$$\mathbf{\Lambda} = \mathbf{T}^T\mathbf{C}\mathbf{T} \tag{2.17}$$

The eigenvalues λ_{ij} are mean square fluctuations along the corresponding eigenvectors. If these eigenvectors are arranged in descending order (relatively to the value of the corresponding eigenvalues), the ones characterized by higher eigenvalues correspond to the collective motions that best approximate the sum of the fluctuations. It is also possible to characterize each eigenvector obtained by studying its components, in order to identify the amplitude of the associated motion for each individual amino acid residue. Other types of analysis provide a projection over time of the configurations of the system generated along the trajectory on a given eigenvector. Such operation can be expressed through the following expression:

$$p_i(t) = \mathbf{X}(t) \cdot \mathbf{u}_i \tag{2.18}$$

where $\mathbf{X}(t)$ contains the coordinates of the system at time t and $\mathbf{u}_i$ is the chosen eigenvector on which the projection is performed. Orthogonally arranging in a Cartesian plane two of these projections, generally corresponding to the first two PC, you can see how the protein samples over time distinct conformational space, the so-called conformational basins. Such kind of analysis, although limited to 2D-space, allows the comparison of the dynamics of more proteins through the direct comparison of the different conformational basins. Note that the covariance matrices of different proteins should be formed by the same number of degrees of freedom. This means that for different proteins is necessary to select common parts of the amino acid sequences.

2.2.5 Functional Mode Analysis (PLS-FMA)

A recent method of analysis of the internal motions of a dynamic system has been used in this work, the so-called Functional Mode Analysis (FMA)[31], that is, the partial least square (PLS) regression applied to the analysis of functional modes (FMA). It is very useful for the investigation of collective modes internal dynamics that can match specific functional properties of interest. As previously reported, the advantage of essential dynamics consists in reducing the dimension of the conformational space of a protein, in order to highlight its principal motions[28, 29]. Although this is of great help in the functional and structural interpretation of the total motions of a protein, the PCA focuses on the *total* fluctuations, therefore the eigenvectors characterized by the highest eigenvalues will be denominated principal component only because the width of their fluctuations, thus making impossible the visualization of minor width motions, which can potentially be more interesting from a functional point of view. The FMA proves to be very useful in this regard, since it is based upon the search of the principal motion of the protein which best correlates the behavior of a particular property of interest. In order to optimize this correlation not only the advantages of the PCA methods (which reduce to the essential ones the number of motions on which to focus attention) are used, but also those of PLS regression are exploited. Data related to property that you want to study are divided into two equivalent sets; on the first (training set) will be built the regression model, and on the second (cross-validation set) will be tested the goodness of the correlation and the predictive power of the built model.

The usage of PLS regression instead of the only PCA, thus presents two important advantages:

- to correlate a particular motion of the protein (even of small entity) to a specific property of interest;
- to minimize the risk of overfitting, since the construction (and control) of a regression model allows to include the smallest possible number of coefficients (thus increasing the predictive power of the model), as it is noticeable in figure 2.2.

2.3 Perturbed Matrix Method (PMM)

As already mentioned, classical MD merely describes large molecular systems only through the motion of the nuclei, considering electrons implicitly. This type of approximation presents the great advantage of being able to perform calculations on molecular systems constituted by thousand of atoms, since relatively not expensive from a computational point of view. Nevertheless, this method presents some important limitations, especially for those chemical processes based precisely on the electrons effects. For this purpose, the *ab initio* methods, which are used to calculate the properties of a system according to the quantum mechanics principles, therefore without modulations with experimental results, appear to be the most elegant and rigorous. In fact, they are based on the solution of the Schroedinger equation, through approximations exclusively of mathematical nature. Unfortunately, the solution of this equation for systems containing many electrons turns out to be computationally too expensive, and therefore the *ab initio* methods are limited to very small systems. The combination of methods based on quantum mechanics (QM) and molecular mechanics (MM) can offer an useful compromise to utilize the advantages of both techniques. The accuracy of QM

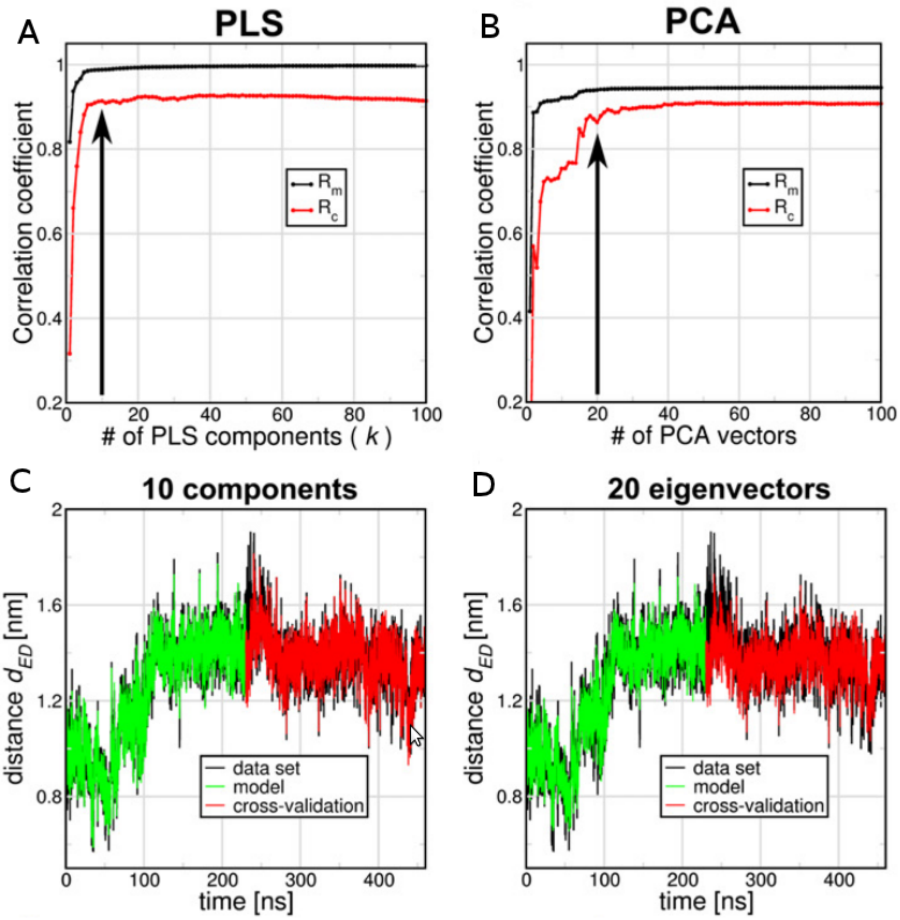


Figure 2.2: Comparison of PLS- and PCA based FMA methods used in the search of best correlation between principal motions of T4 Lysozyme (T4L) and distance (d_{ED}) between Glu11 and Asp20. In panels A and B training set (R_m, black) and cross-validation set (R_c, red) trends versus PLS components or PCA vectors are reported. In panels C and D the trends of the distance d_{ED} and of the models versus the simulation time. It is possible to note that to reach the maximum correlation, PLS-FMA model uses 10 components, whereas PCA-FMA model uses 20 vectors. The image is taken from ref. [31].

on the one hand, the relatively low computational cost and the great versatility of classical MD on the other. This is the basis concept of QM/MM hybrid models, introduced at the end of the 70s in the field of computational biology[32]. The QM/MM approach consists in the division of the system of interest into two subsets: one (QM) containing a relatively small number of atoms and described according to the principles of quantum mechanics, the other (MM) containing most of the atoms of the system, described by a classical force field where break/formation of bonds and electronic transitions are not considered explicitly. There are numerous open questions on the use of QM/MM models, for example on the choice of the QM part, on its size (invariably limited), on the description of the interactions between QM and MM parts, etc. However, the use of these hybrid methods has proved to be very useful in several fields, such as material sciences[33], the study of chemical processes in condensed phase[34], or spectroscopic properties of molecules in solution[35].

A possible application in the field of molecular dynamics is the calculation of the redox potential of electron transfer proteins. A problem of this type requires the combination of calculations of electronic structures (such as the group in which is contained the redox site) with dynamic and structural information on the protein. You could then use the classical MD to get a statistical ensemble of configurations (trajectory) of the system and QM calculations on a limited number of these configurations[36, 37]. Very accurate descriptions of the system are obtained with such type of approach, but since this procedure requires a very high number of QM calculations it becomes necessary to reduce their application filed, or, if highly refined calculations are requested, to reduce the number of configurations to consider.

It has been recently developed a theoretical method called Perturbed Matrix Method (or PMM) to obtain the perturbed molecular eigenstates[38, 39]. In PMM, instead of directly introducing the perturbation term of the Hamiltonian operator, as it is usually done in calculations based on the Hartree-Fock model, the perturbation effect is obtained diagonalizing the perturbed Hamiltonian matrix, constructed on the basis of the electronic eigenstates of the unperturbed Hamiltonian. The basic approach consists of a rigorous calculation procedure in which is admitted only one approximation, that is, the truncation of Taylor expansion of the perturbed Hamiltonian (expressed as a matrix) at a certain level. This approximation, necessary in order to reduce the computational cost, can lead to artifacts when considering also the excited states. Despite this, the first test realized with this method produced considerable results[40, 41, 42], in terms of accuracy and computational cost.

2.3.1 Mathematical Procedure

The time independent Schroedinger equation, in matrix notation, for a perturbed system is:

$$\tilde{H} c_i = U_i c_i \tag{2.19}$$

where $\tilde{H} = \tilde{H}^0 + \tilde{V}$, c_i is the i-th eigenvector of the perturbed Hamiltonian matrix $\tilde{H}$, U_i corresponds to the i-th eigenvalue, $\tilde{H}^0$ is the unperturbed Hamiltonian matrix and $\tilde{V}$ is the perturbation operator matrix.

The Hamiltonian and its eigenvalues can be expressed on a basis of electronic eigenstates of the unperturbed Hamiltonian, therefore the matrix elements of the Hamiltonian can be expressed as:

$$H_{l,l'} = \langle \Phi_l^0 \mid \hat{H} \mid \Phi_l^{0'} \rangle = U_l^0 \delta_{l,l'} + \langle \Phi_l^0 \mid \hat{V} \mid \Phi_l^{0'} \rangle \tag{2.20}$$

where Φ_l^0 is he i-th eigenvector of the unperturbed Hamiltonian, U_l^0 is the corresponding energy eigenvector, $\delta_{l,l'}$ is the Kroenecker delta and $\hat{V}$ is the perturbation operator. From these equations is clear that, to obtain eigenvectors and eigenvalues (thus, the properties of interest of the system) of the Hamiltonian eigenstates, is sufficient to diagonalize the matrix $\tilde{H}$ given by 2.20.

For a system interacting with an external field, it is possible to express on a more general basis the perturbation operator of 2.20 in terms of electric potential $\mathcal{V}$:

$$\hat{V} = \sum_j q_j \mathcal{V}(\mathbf{r}_j) \tag{2.21}$$

where $\mathbf{r}_j$ are the coordinates of the j-th particle and q_j the corresponding charge. Expanding $\mathcal{V}$ tot the second order around the equilibrium position $\mathbf{r}_0$ you have:

$$\begin{aligned} \mathcal{V}(\mathbf{r}_j) \;\; \cong \;\; & \mathcal{V}(\mathbf{r}_0) \; - \sum_{k=1}^{3} E_k (r_{j,k} - r_{0,k}) \\ - \;\; & \frac{1}{2} \sum_{k'=1}^{3} \sum_{k=1}^{3} \left(\frac{\partial E_k}{\partial r_{k'}} \right)_{\mathbf{r}=\mathbf{r}_0} (r_{j,k} - r_{0,k})(r_{j,k'} - r_{0,k'}) \end{aligned} \tag{2.22}$$

where k defines the three components of a vector in space (x, y and z). From these equations, defining q_T as the total charge, you obtain:

$$\langle \Phi_l^0 \mid \hat{V} \mid \Phi_l^{0'} \rangle \cong q_T \mathcal{V}(\mathbf{r}_0) \delta_{l,l'} - \mathbf{E} \cdot \langle \Phi_l^0 \mid \hat{\mu} \mid \Phi_l^{0'} \rangle + \frac{1}{2} T_r \left[\tilde{\Theta} \; \tilde{Q}_{l,l'} \right] \tag{2.23}$$

$$Q_{k,k'}^{l,l'} = \left[\tilde{Q}_{l,l'} \right]_{k,k'} = \sum_j q_j \langle \Phi_l^0 \mid (r_{j,k} - r0, k)(r_{j,k'} - r_{0,k'}) \mid \Phi_l^{0'} \rangle \tag{2.24}$$

where

$$\Theta_{k,k'} = -\left(\frac{\partial E_k}{\partial r_{k'}}\right)_{\mathbf{r}=\mathbf{r_0}} \tag{2.25}$$

$$\hat{\mu} = \sum_j q_j (r_j - r_0) \tag{2.26}$$

Thus, the complete perturbed Hamiltonian matrix is:

$$\tilde{H} = \tilde{H}^0 + \tilde{I} q_T \mathcal{V}(\mathbf{r}_0) + \tilde{Z}_1(\mathbf{E}) + \tilde{Z}_2(\tilde{\Theta}) \tag{2.27}$$

with

$$[\tilde{Z}_1]_{l,l'} = -\mathbf{E} \cdot \langle \Phi_l^0 \mid \hat{\mu} \mid \Phi_l^{0'} \rangle \tag{2.28}$$

$$[\tilde{Z}_2]_{l,l'} = \frac{1}{2} T_r [\tilde{\Theta}\ \tilde{Q}_{l,l'}] \tag{2.29}$$

From these last equations it can be deduced that an expansion to the second order of the electric potential requires the knowledge of the total charge, and of the unperturbed dipoles and quadrupoles. Hypothetically, it is possible to realize further expansion to greater orders, following the same criterion; that would require however additional expressions of multipoles of greater order, typically very difficult to obtain. Furthermore, for non-charged systems and homogeneous field, $\tilde{\Theta}$ becomes a null matrix, and the 2.27 results:

$$\tilde{H} = \tilde{H}^0 + \tilde{Z}_1(\mathbf{E}) \tag{2.30}$$

To be treated in an exact and rigorous way, these equations would require matrices and vectors of infinite dimensions, obviously unusable from a practical point of view. Despite this, it has been showed[39] that, for the treatment of the perturbed fundamental state and of the first excited states, it is possible to use the Hamiltonian in a matrix form of 2.19 to finite dimensions, constructing it, in 2.20, through a (finite) subspace of the unperturbed eigenstates, that goes from the fundamental state to the chosen excited one.

References

[1] B.J. Alder and T.E. Wainwright. In: *J. Chem. Phys.* 31 (1959), pp. 459–466.

[2] J.E.H. Koehler, W. Saenger, and W.F. van Gunsteren. In: *J. Mol. Biol.* 203 (1988), pp. 241–250.

[3] E. Egberts. Master's thesis, University of Groningen, 1988.

[4] U.C. Singh, S.C Weiner, and P. Kollman. In: *Procl. Natl. Acad. Sci. USA* 82 (1985), pp. 755–759.

[5] S.C. Harvey et al. In: *Science* 223 (1984), pp. 1189–1191.

[6] J. Hermans (ed.) Polycristal Bookservice P.O. Box 27, Western Springs, Ill., USA, 1985.

[7] S.N. Rao et al. In: *Nature* 328 (1987), pp. 551–554.

[8] W.F. van Gunsteren. Escom Science Publishers, Leiden, 1989, pp. 27–9.

[9] W.L. Jorgensen et al. In: *J. Chem. Phys.* 79 (1983), p. 926.

[10] S. Lifson, C.E. Felder, and A. Shanzer. In: *J. Am. Chem. Soc.* 105 (1983), p. 3866.

[11] E. Giglio, A.M. Liquori, and L. Mazzarella. In: *Nuovo Cim.* 1 (1969), p. 135.

[12] A. Di Nola and E. Giglio. In: *Acta Crystallogr.* A 26 (1970), p. 144.

[13] M.P. Allen and D.J. Tildesly. Oxford University Press, Oxford, 1989.

[14] U. Essmann et al. In: *J. Chem. Phys.* 103 (1995), pp. 8577–93.

[15] T. Schlick, E. Barth, and M. Mandziuk. In: *Annu. Rev. Biomol.* 26 (1997), pp. 181–222.

[16] J.P. Ryckaert, G. Ciccotti, and H.J.C. Berendsen. In: *J. Comp. Phys.* 23 (1977), pp. 327–41.

[17] L. Verlet. In: *Phys. Rev.* 159 (1967), p. 98.

[18] L. Verlet. In: *Phys. Rev.* 165 (1968), pp. 201–23.

[19] H. J. C. Berendsen, D. van der Spoel, and D. van Drunen. In: *Comp. Phys. Comm.* 95 (1995), pp. 43–56.

[20] S. Nose. In: *J. Chem. Phys.* 81 (1984), pp. 511–519.

[21] W.G. Hoover et al. In: *Pyhs. Rev. A* 22 (1980), p. 1960.

[22] D.J. Evans et al. In: *Phys. Rev. A* 28 (1983), p. 1016.

[23] J.M. Haile and S. Gupta. In: *J. Chem. Phys.* 79 (1983), p. 3067.

[24] A. Andersen. In: *J. Chem. Phys.* 72 (1980), p. 2384.

[25] H.J.C. Berendsen et al. In: *J. Chem. Phys.* 81 (1984), p. 3684.

[26] A. Amadei. Groningen, 1998.

[27] W. Kabsch and C. Sander. In: *Biopolymers* 22 (1983), pp. 2577–2637.

[28] A. Amadei, A.B.M. Linssen, and H.J.C. Berendsen. In: *Proteins: Struct. Funct. Genetic.* 17 (1993), pp. 412–25.

[29] A.E. Garcia. In: *Phys. Rev. Lett.* 66 (1992), pp. 2696–9.

[30] A. Kitao, F. Hirata, and N. Go. In: *J. Chem. Phys.* 158 (1991), pp. 447–72.

[31] T. Krivobokova et al. In: *Biophys. J.* 103 (2012), pp. 786–96.

[32] A. Warshel and M. Levitt. In: *J. Mol. Biol.* 103 (1976), pp. 227–49.

[33] F.F. Abraham et al. In: *Europhys. Lett.* 44 (1998), pp. 783–92.

[34] J. Gao and M. Freindorf. In: *J. Phys. Chem. A.* 101 (1997), pp. 3182–94.

[35] F. Aquilante et al. In: *Mol. Phys.* 101 (2003), pp. 1945–53.

[36] B.G. Schulze and J.D. Evanseck. In: *J. Am. Chem. Soc.* 121 (1999), pp. 6444–54.

[37] E.S. Manas, J.M. Vanderkooi, and K.A. Sharp. In: *J. Phys. Chem. B.* 103 (1999), pp. 6334–48.

[38] M. Aschi et al. In: *Chem. Phys. Lett.* 344 (2001), pp. 374–380.

[39] R. Spezia et al. In: *Chem. Phys. Lett.* 365 (2002), pp. 450–456.

[40] A. Amadei et al. In: *J. Chem. Phys.* 122 (2005), p. 124506.

[41] M. Anselmi et al. In: *Biophys. J.* 92 (2007), pp. 3442–3447.

[42] M. Aschi and A. Amadei. In: *IJLSS* 2 (2009), pp. 29–38.

3 The active site loop modulates the reorganization energy of blue copper proteins by controlling the dynamic interplay with solvent

3.1 Cupredoxins

Cupredoxins are copper-containing proteins involved in electron transfer (ET) processes in both plants and bacteria [1]. Their structure features a rigid β-barrel structure, whose C-terminal strands are linked by a loop that shields the copper atom from direct contact with the solvent. The length and sequence of the loop region can vary for different species, apart from the conserved presence of three copper ligands (His, Cys and Met). Recent studies proved the central role played by the loop in determining the structure and the physiological properties of cupredoxins [2, 3, 4], but the molecular details of how it tunes the redox functionality were not yet fully unveiled. The elucidation of the role of the hydrophobic, ligand-containing loop in the ET processes is central not only for a deeper understanding of the structure/function relationships throughout the whole family of cupredoxins, but also for tailoring the redox properties of biomolecules to be used in a number of applications such as bioelectronic devices, nano-biosensors or artificial photosynthetic centres[5, 6, 7]. Among the most deeply investigated issues in biological ET is the complex relationship that links the rate of the process to the structural features of the biomolecule[8, 9, 10] and to their dynamic interplay with the solvent. One of the key factors in determining the ET rate is the reorganization energy λ, which represents the free-energy required for the reactants to move from the equilibrium reactant structural ensemble to the equilibrium product ensemble, without transferring electrons [11]. The reorganization energy can be conveniently divided in two terms: for a metal-containing protein, the first (λ_{in}) is mainly determined by the geometrical rearrangement of the first coordination sphere of the active site, while the contributions from the protein matrix and from the solvent are included in the so called outer-sphere reorganization energy λ_{out}.

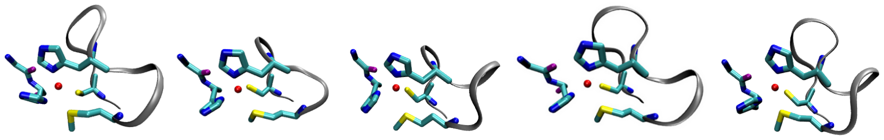
For heme proteins like cytochrome *c*, that were thouroughly investigated as paradigmatic systems for ET processes, one of the determining factors to the λ_{out} was proved to be the accessibility of the active site to the solvent [12, 13]. A similar result was obtained for azurin, the most studied member of cupredoxins, for which the interface between protein and water in the hydrophobic region surrounding copper-binding His117 was found to be the main contribution to λ_{out} [14].

Here, we performed classical molecular dynamics (MD) simulations to investigate native azurin and four of its mutants, in which the ligand-containing loop has been replaced either with that of other members of the blue copper family (amicyanin and plastocyanin) or with synthetic sequences featuring only Ala residues (see figure 3.1). The comparison of five distinct proteins that differ only in the loop composition allowed us to focus exclusively on the contribution of this region on the ET properties. Moreover, through MD simulation, we could take into account the contribution of thermal fluctuations and in general of the dynamic relationship between protein and solvent, which were suggested to be crucial for the redox functionality of a protein [15, 16, 17, 18, 19, 20, 21].

In particular, we calculated the outer sphere reorganization energy for all the investigated proteins and we monitored changes in the loop region over time and upon reduction. The relatively small but significant changes in λ_{out} for the five loop mutants were found to be related to differences in the extent of both short- and long-range solvent reorganization and to structural rearrangement of the whole protein scaffold following reduction/oxidation. We believe that our most important finding is that the nature and length of the relatively small loop region can tune the ET functionality by inducing sizeable effects on the dynamics/function relationship of the cupredoxin as a whole.

3.2 Molecular Dynamic Simulations

The MD simulations of Azurin and its loop-mutants were performed in water using the GROMACS software package [25] and the AMBER99 force field [26]. As the starting point of the simulations, oxidized structures of Azurin and loop mutants AZAMI, AZPC, 4A3A and 4A4A (PDB id codes: 4AZU, 2FTA, 2HX7, 3FSW, 3FSZ) were taken as starting conformations. Water was represented with the TIP3P model [27]. For each starting structure, one protein molecule was solvated in a periodic rhombic dodecahedral box. The simulations were performed in the NVT ensemble at the experimental temperature of 300 K. The temperature was kept constant by the



AZU
CTFPGHSALM

AZAMI
CTPHPFM

AZPC
CSPHQGAGM

4A3A
CAAAAHAAAM

4A4A
CAAAAHAAAAM

Figure 3.1: Active site structures of P. aeruginosa azurin (AZU) and of its four copper-binding loop mutants AZAMI, AZPC, 4A3A and 4A4A. The copper atom is represented as a red sphere, the metal ligands as licorice and the hydrophobic loop is depicted as a gray ribbon. The sequence of the loops for all the species is also shown. The image was prepared with VMD software [22] using the crystallographic structures of the species (PDB ID codes: 4AZU[23], 2FTA[4], 2HX7[2], 3FSW[24], 3FSZ[24]).

isokinetic temperature coupling [28]. MD simulations were performed for both reduced and oxidized forms of all the variants; the two oxidation states were obtained in our simulations by changing the atomic charges of the protein active site (i.e. the Cu metal atom and the atoms belonging to the five copper-coordinating residues). The atomic charges calculated at the DFT level of the theory were taken form literature [29]. No other changes at the force field have been made. Periodic boundary conditions were applied to the simulation box and the long-range electrostatic interactions were treated with the Particle Mesh Ewald method [30] using a grid spacing of 0.12 nm combined with a fourth-order B-spline interpolation to compute the potential and forces. The real space cut-off distance and the van der Waals cut-off distance were set to 0.9 nm. A time step of 2 fs was used. Bond lenghts were constrained with LINCS algorithm. The N-terminal and C-terminal ends of the proteins were modeled as NH_3^+ and CO_2^-, respectively. Positive counter ions were added (Na^+) to make the simulation box neutral. The starting structures were subjected to an energy minimization in vacuum (3000 steps) using the steepest descent method, followed by a two-step minimization protocol in solvent, using the conjugate gradient method. The first minimization was performed with the coordinates of the protein held fixed, allowing only the water and the counter ions to move, and the second was performed on the atoms of both the protein and the solvent molecules. The temperature of the systems was then increased form 50 to 300 K in 100 ps MD before the production simulations were performed for 100 ns.

To evaluate λ_{out} from molecular dynamics simulations, we used the following relation,[31] that connects it to the fluctuations of the energy gap η, where η=E_{red}-E_{ox}:

$$\lambda out = \frac{< \eta >_{ox} - < \eta >_{red}}{2} = \frac{< E_{red} - E_{ox} >_{ox} - < E_{red} - E_{ox} >_{red}}{2} \quad (3.1)$$

3.3 Determination of the reorganization energy

The loop region of native *P. aeruginosa* azurin (AZU hereafter) and that of four loop mutants is shown in Figure 1. The protein scaffold of AZU was conserved for all the mutants, but different copper-binding loop sequences were introduced to yield the variants named AZAMI (azurin scaffold and amycianin loop region, CTPHPFM), AZPC (featuring the plastocyanin loop, CSPHQGAGM), 4A3A and 4A4A mutants, whose metal-binding sequences are CAAAAHAAAM and CAAAAHAAAAM, respectively. The crystal structures of all these cupredoxins are available, [2, 4, 23, 24] and they were taken as starting points to perform 100ns-long MD simulations for both the reduced and oxidized state, from which the reorganization energy for all the 5 cupredoxins was obtained. Several methods were suggested in literature for the calculation of reorganization free energies at different levels of the theory [13, 14, 32, 33, 34, 35, 36, 37]. Here, we are comparing proteins where the ligands of the copper atom are conserved: therefore, it is resonable to infer that λ_{in} does not undergo dramatic changes between the different loop mutants. Moreover, for native azurin the contribution of the λ_{in} term to the total reorganization free energy was found to be significantly smaller than the λ_{out} one [14]. We therefore decided to focus only on λ_{out}, that was calculated with the energy gap fluctuations method [31]. The values of λ_{out} are listed in table 3.1.

Experimental estimates for total reorganization energy are available only for azurin, whose λ has been fixed between 0.6 and 0.8 eV [37, 38]. Our calculated value for azurin is higher than the experimental ones: this is most likely due to the use of a non-polarizable force field for the calculations, which is known to introduce some overestimation (which accounts to roughly 30%) of the reorganization energies values [13]. Taking this effect into account would lead to a calculated λ_{out} value for azurin of approximately 0.73 eV, which would therefore be in line with the aformentioned range, also including the previously estimated value for λ_{in} of 0.15 eV [39]. Calculated λ closer to experimental results could have been obtained by choosing more refined the-

	λ_{out} (eV)	<SASA> Active site	$\vert\Delta\mu_{water}\vert$	RMSD (ox. vs red.)
AZU	1.05 (0.03)	0.198 (0.048)	12.12 (0.91)	0.842
AZAMI	1.13 (0.02)	0.565 (0.110)	13.10 (1.32)	0.962
AZPC	1.26 (0.01)	0.442 (0.084)	8.32 (0.01)	1.724
4A3A	1.18 (0.03)	0.476 (0.035)	19.52 (0.65)	1.110
4A4A	0.97 (0.02)	0.261 (0.015)	3.26 (0.02)	0.456

Table 3.1: Calculated outer sphere reorganization energy λ_{out}, solvent accessible surface area of the active site averaged over both reduced and oxidized ensemble, changes (reduced - oxidized state) in the total dipole of the 260 water molecules closest to the Cu atom and root mean square deviation between the average structures obtained from MD in the oxidized and reduced ensemble. Estimated errors are given in brackets.

oretical approaches; nevertheless, since our main goal is the comparison of the outer sphere reorganization free energies for different species, we chose a classical method over a computationally more demanding, quantum mechanical approach. Moreover, as our goal is to investigate the effects of the dynamics of the mutants on their ET kinetic properties, MD allows for a much broader sampling of conformations than QM-based methods presently do.

AZPC displays the highest value for λ_{out} (1.26 eV), followed by 4A3A and AZAMI (1.18 and 1.13 eV, respectively). AZU (1.05 eV) and 4A4A (0.97 eV) feature the smallest reorganization energies. The difference displayed by 4A3A and 4A4A is quite surprising, as there is just a slight difference in the loop composition between the two variants, with 4A4A featuring one additional alanine residue in the hydrophobic loop. Mere comparison of the crystal structures of the two proteins could not provide an explanation for this discrepancy.

3.4 Comparison with experimental data

The reliability of our calculations could be assessed by a direct comparison with the values of activation enthalpies ($\Delta H^{\#}$) for heterogeneous protein-electrode ET, recently obtained by electrochemical measurements performed on the same set of proteins, immobilized on a gold electrode functionalized with a self-assembled monolayer (SAM) [40]. Since the binding orientation on the SAM is the same for the five cupredoxins [40], it is reasonable to assume a constant contribution of the adsorption process to the activation enthalpy of the five species with respect to what could be measured for freely-diffusing proteins: therefore, the experimental data collected for the surface-immobilized loop mutants can serve as a test for our calculations. Figure 3.2 displays the plot of calculated λ_{out} versus the experimental $\Delta H^{\#}$ values: a very good correlation can be observed, suggesting that our simulations can be confidently used to calculate observables related to the molecular determinants to the ET kinetics.

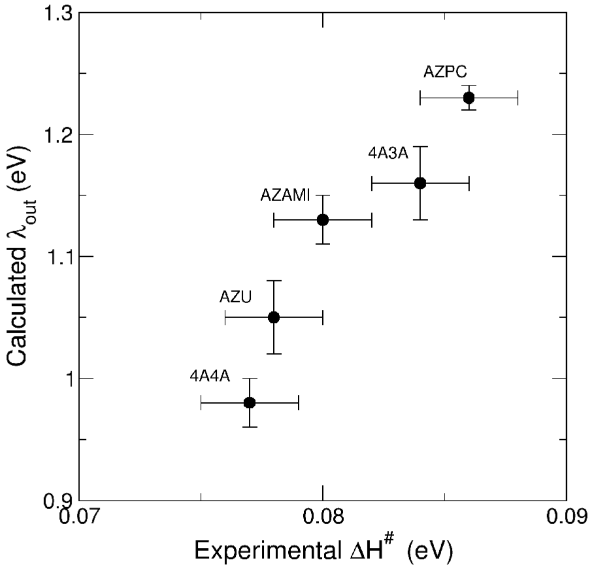


Figure 3.2: Plot of the calculated values for outer-sphere reorganization energy versus experimental activation enthalpy values obtained by electrochemical means. The experimental were taken from ref. [40]

.

3.5 Dependence of the reorganization energy on solvent accessible surface of the active site

As mentioned above, rearrangements of the hydrophobic region around the copper atom upon reduction/oxidation were assigned a major role in determining the total reorganization energy for azurin[14]. We first monitored the extent of exposure to water of the protein active site along the whole MD by calculating the solvent accessible surface area (SASA) of the active site, defined as the Cu atom and the five metal-coordinating residues.

The calculated SASA values, averaged over both Cu(I) and Cu(II) oxidation states, are listed in Table 3.1. The three cupredoxin mutants displaying the highest SASA values, AZAMI, 4A3A and AZPC, are those for which the biggest λ_{out} was also calculated, while the SASA are significantly lower for the mutants with the smallest reorganization energy values, namely AZU and 4A4A. Similarly to what was found for cytochrome *c*, our results show that the solvent accessibility to the protein active site is clearly (and unsurprisingly) a major contribution to λ_{out} also for cupredoxins; nevertheless, this structural parameter does not completely justify smaller but meaningful differences in reorganization energy between AZU and 4A4A, or between AZAMI, 4A3A and AZPC. For example, although the solvent accessible surface area of the active site for 4A4A is bigger than the corresponding azurin value, the latter protein displays a significantly higher reorganization energy. In the case of the blue-copper species under investigation, the finer regulation of λ_{out} seems to be achieved through the combined contributions of additional factors. We therefore monitored in deeper detail the interaction between protein and solvent, with a particular focus on the relationship between λ_{out} and the reorganization of the solvent upon reduction.

3.6 Dependence of the reorganization energy on solvent rearrangement around the active site

It was demonstrated both through experiments[41] and calculations[14, 37] that the water rearrangement caused by a change in the copper oxidation state is the major determinant (approximately up to 80%) to the total reorganization energy for azurin. In order to check whether this counclusion could be extended to the whole set of cupredoxins, we investigated the relationship between λ_{out} and the reorganization of the solvent upon reduction.

We quantified the reduction-driven modification of the hydration of the active site by determining N_W in both oxidation states (see figure 3.3): N_W

can be defined as the number of water molecules in a sphere centered on the copper atom and featuring a 0.9 nm radius. Figure 3.3 displays the probability that N_W takes a particular value along the whole simulation, calculated for each protein and for both the reduced and oxidized ensembles (please note that while figure 3.3 depicts the changes in N_W between the two oxidation states, the values for the SASA in Table 3.1 were averaged between the two ensembles.) The species featuring the most dramatic rearrangements as a function of the Cu oxidation states is AZPC, followed by AZAMI and 4A3A, as indicated by the shifts of the maxima and average values of the normalized distribution of configurations with a given N_W. Changes in N_W are smaller for AZU and almost absent for 4A4A, thus strengthening the hypothesis that the hydration shell has dramatic effects on the overall outer-sphere reorganization energy.

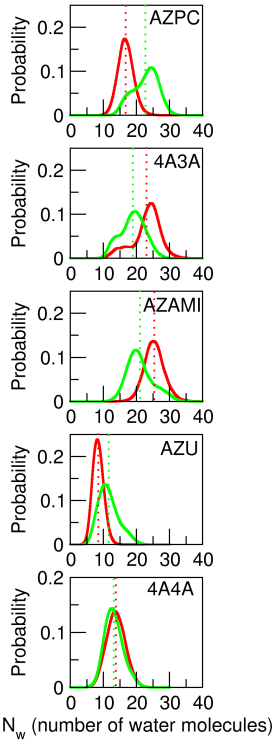


Figure 3.3: Normalized distribution of configurations with a given number of water molecules N_W within 0.9 nm from the Cu atom in the reduced (black) and oxidized (green) ensemble, respectively. Dashed lines correspond to average N_W values for the two ensembles.

The N_W distribution plots only reflects changes in the number of solvent molecules in close proximity to the active site, but the reorganization energy can be affected by the water molecules within up to roughly 1.6 nm from the metal ion, as previously suggested[14] and as further proved by the radial distribution function $g(r)$ of the water oxygen atoms with respect to Cu (see figure 3.7). To further investigate this long-range effect, we decided to follow reduction-induced changes of the dipole moment of the solvent $|\Delta\mu_{water}|$, which is directly connected to the energy of the system and therefore is a promising candidate for monitoring the outer sphere reorganization energy of our proteins. In particular, we calculated the module of total dipole moment μ_{water} of the 260 water molecules closest to the copper atom, i.e. the number of solvent molecules which are comprised within a shell of 1.6 nm centered on Cu, which was previously chosen [14] as representative of the solvent region perturbed by the electric field of the metal ion embedded in the protein matrix. We then monitored changes in μ_{water}, induced by the reduction of Cu(II) to Cu(I), obtaining the module of the difference of the solvent dipole between the two oxidation states, $|\Delta\mu_{water}|$. The calculated values are listed in Table 3.1: large $|\Delta\mu_{water}|$ values indicate significant reorientation of the solvent molecules upon copper reduction, while for species displaying smaller $|\Delta\mu_{water}|$ it is reasonable to infer a less extended reorganization of hydration water molecules. Figure 3.4 shows the plot of the outer-sphere reorganization energy versus the module of the difference in the dipole moment of the hydration water molecules: it is interesting to see how an increase in the differences between the solvent dipole moment going from the oxidized ensemble to the reduced one $|\Delta\mu_{water}|$ corresponds to larger reorganization energy values. The only species which does not follow this trend is AZPC. Although it is hard to provide an explanation for its behavior, it is possible that for this mutant λ_{out} is dominated by the solvent rearrangements in the proximity of the Cu, which are high enough to account for the biggest reorganization energy value among the loop mutants.

These results provide evidences for the correlation, valid for all the investigated cupredoxins, between changes in the structure of the water molecules within the hydration shell of the active site and the outer-sphere reorganization energy. Most importantly, it is evident how the solvent reorganization is not limited to a re-orientation of the total dipole moment, but is also composed by significant variations of the number of solvent molecules within a given shell.

3.7 Dependence of reorganization energy on RMSD

The reorganization energy is defined as the free energy accompanying structural rearrangements upon reduction/oxidation: therefore, the reorganization of the protein scaffold is a natural quantity to be analyzed. We extracted

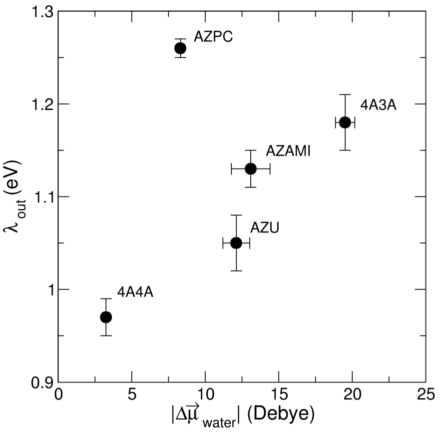


Figure 3.4: Calculated values for outer sphere reorganization energy plotted against the module of the difference between the dipole moment of the 260 water molecules closest to the Cu atom calculated in the reduced and oxidized state.

the average structures from the MD simulations for both ensembles, and we calculated the root mean square deviation RMSD between the backbones atoms of the average oxidized structure and the average reduced one. The results are listed in Table 1. Higher values for RMSD were found for the species displaying the largest λ_{out} values; on the contrary, the two loop mutants with the smallest RMSD (0.842 Å and 0.456 Å for AZU and 4A4A, respectively) are also those with the lowest reorganization energy. The high correlation between λ_{out} and the RMSD between the average structures obtained in the two ensembles is plotted in figure 3.5. This finding is consistent with the predictions of the Marcus theory, according to which structurally similar reduced and oxidized forms would display small reorganization energy values.

In order to determine which regions of the proteins provided the most significant contribution to the overall structural rearrangement, we calculated RMSD between oxidized and reduced form for each secondary structure element of the cupredoxins. The results are reported in figure 3.6.

As expected, β-sheets regions for all the mutants are only slightly affected by change in oxidation states, while the loop regions generally feature higher RMSD values. Interestingly, for all the five proteins, the loop spanning residues 9 to 18, the 36-to-44 one and the region encompassing the α helix and the loop which follows display RMSD values among the highest. Although far from each other in the primary structure, these structural elements carry the residues that form the extended hydrophobic area surround-

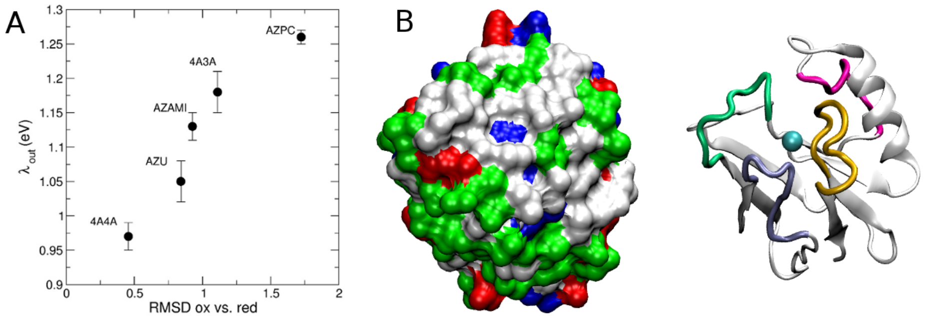


Figure 3.5: A. Plot of the calculated values for outer-sphere reorganization energy versus RMSD between the average structures of oxidized and reduced ensemble. B. Surface (left) and cartoon (right) representations of native azurin, prepared starting from PDB file 4AZU.pdb. Left: hydrophobic residues are colored in white, polar residues in green, basic residues in blue and acidic residues in red. Right: the loops displaying the highest RMSD values for all the mutants are colored in blue (loop spanning residues 9 to 18), green (36-44), pink (72-79) and yellow (112-121).

ing the copper atom, together with the loop that was mutated in the present work (see figure 3.5). This hydrophobic patch was previously assigned a central role in determining the reorganization energy for native azurin[14], based on the assumption that this portion of the protein would form less hydrogen bonds with the surrounding solvent molecules, that would then be greatly affected by changes in the copper oxidation state. The present results allow to extend this hypothesys to the whole family of cupredoxins, since the residues forming the hydrophobic region close to the Cu atom are invariably among the most significant contributors to the overall RMSD, which was proved to be directly correlated to λ_{out}.

3.8 Radial distribution function

We calculated the radial distribution function $g(r)$ of the water oxygen atoms with respect to Cu. The Cu-O $g(r)$ was computed for both Cu(II) and Cu(I), and the difference between the radial distribution functions (Δg(r)) for the two oxidation states was then obtained for all the mutants (see Figure S1). The water molecule which is closest to the Cu atom is invariantly hydrogen-bonded to His117, and its presence generates the first peak of the $g(r)$, placed at a distance of approximately 0.5-0.6 nm from the metal (see figure 3.7).

The variants feature different $\Delta g(r)$ plots, proving that changes in the

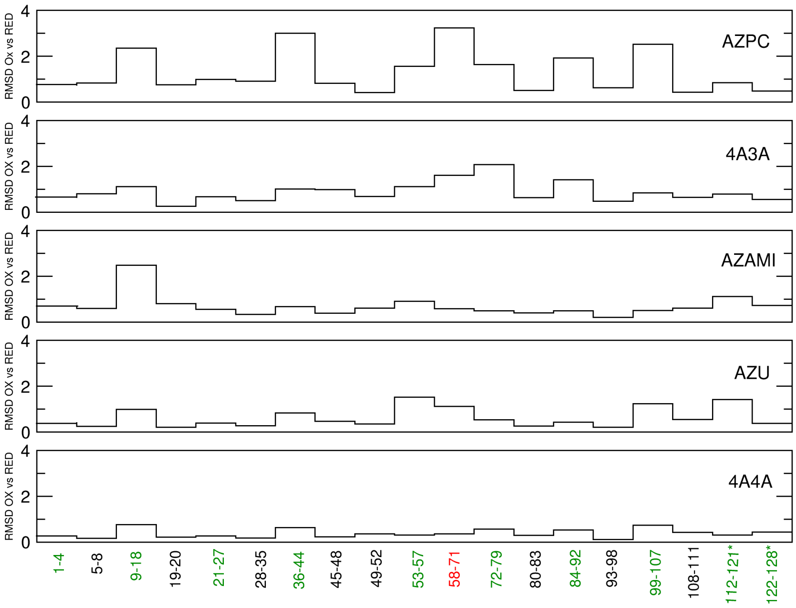


Figure 3.6: RMSD of each secondary structure element of the investigated proteins, calculated between the average oxidized and reduced structures obtained from our MD simulations. The structural elements are labelled by the residues spanned by each of them (i.e. 1-4 stands for "structural element composed by residues 1 to 4", and so on). Black colour indicates β sheets, green stands for loop regions and red for α helix. The numbering refers to native azurin sequence. The symbol $*$ indicates that the numbering of the last two structural elements varies according to the different loop length and composition.

loop region have significant effects on the structure of the water molecules around the active site. Since the plot was calculated as the difference between the Cu-O radial distribution functions in the two oxidation states, more pronounced peaks within 0.9-1 nm from the metal indicate more significant changes in the structure of the water molecules within the first hydration shell following a change in the oxidation state of Cu. It is interesting to see how the peaks of the $g(r)$ differential plot in this region are markedly higher and/or broader for AZPC, 4A3A and AZAMI (i.e. the variants featuring the highest λ_{out} values) than the corresponding plots for AZU and 4A4A. Although on a mostly qualitative basis, this plot provides a hint of the connection between solvent rearrangement and the reorganization energy.

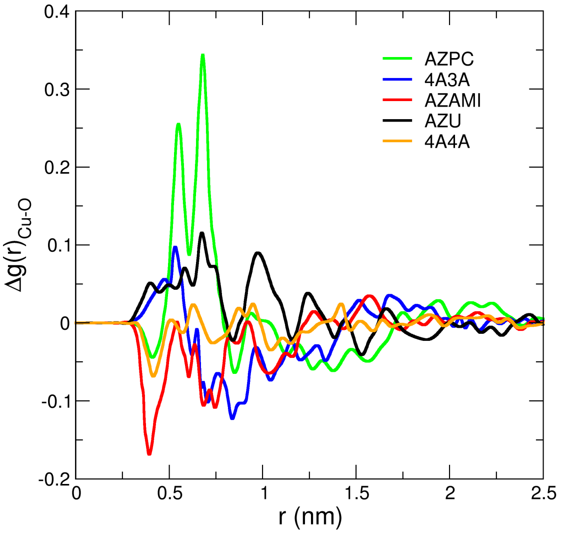


Figure 3.7: Difference (oxidized - reduced ensemble) of Cu-O radial distribution functions for AZU (black), AZAMI (red), AZPC (green), 4A3A (blue) and 4A4A (orange).

In conclusion, the present work unveils the role of the hydrophobic loop region in determining the outer sphere reorganization energy in cupredoxins. Similarly to what was reported for heme-containing ET proteins, λ_{out} turns out to be dependent on the accessibility to the solvent of the active site. Nevertheless, the active site SASA is not the only factor which controls the outer-sphere reorganization energy. Indeed, changes in length and composition of the relatively small hydrophobic loop region result in different dynamics of the whole molecule, leading to reduction-dependent displacements in the protein scaffold, especially in the loop-containing portion of the molecule close to the copper. These structural variations affect the extent of reorganization of the water molecules close to the active site, as well as long-range solvent re-orientation. The combined effects of these molecular determinants contribute to the protein/solvent interaction, leading to a fine regulation of λ_{out}.

References

[1] C. Dennison. In: *Coordin. Chem. Rev.* 249.24 (2005), pp. 3025–3054.

[2] C. Li, M. J. Banfield, and C. Dennison. In: *J. Am. Chem. Soc.* 129 (2007), pp. 709–718.

[3] G. Battistuzzi et al. In: *Biochim. Biophys. Acta* 1794 (2009), pp. 995–1000.

[4] C. Li et al. In: *Proc. Natl. Acad. Sci. USA* 103 (2006), pp. 7258–7263.

[5] N. M. Marshall et al. In: *Nature* 462.7269 (2009), pp. 113–116.

[6] W. Li et al. In: *ACS Nano* 6.12 (2012), pp. 10816–10824.

[7] I. Ron et al. In: *J. Am. Chem. Soc.* 132.12 (2010), pp. 4131–4140.

[8] C. C. Page et al. In: *Nature* 402 (1999), pp. 47–52.

[9] H. B. Gray and J. R. Winkler. In: *Q. Rev. Biophys.* 36 (2003), pp. 341–372.

[10] C. A. Bortolotti et al. In: *J. Phys. Chem. C* 111.32 (2007), pp. 12100–12105.

[11] D. S. Bendall. BIOS Scientific Publisher Ltd:Oxford,UK, 1996.

[12] C. A. Bortolotti et al. In: *J. Phys. Chem. Lett.* 2 (2011), pp. 1761–1765.

[13] V. Tipmanee et al. In: *J. Am. Chem. Soc.* 132 (2010), pp. 17032–17040.

[14] M. Cascella et al. In: *P. Natl. Acad. Sci. USA* 103 (2006), pp. 19641–19646.

[15] C. A. Bortolotti et al. In: *J. Am. Chem. Soc.* 134.33 (2012), pp. 13670–13678.

[16] A. Stock. In: *Nature* 400 (1999), pp. 221–222.

[17] L. B. Sagle et al. In: *J. Am. Chem. Soc.* 128 (2006), pp. 7909–7915.

[18] K. Henzler-Wildman and D. D. Kern. In: *Nature* 450 (2007), pp. 964–972.

[19] S.S. Skourtis, D.H. Waldeck, and D.N. Beratan. In: *Annu. Rev. Phys. Chem.* 61 (2010), pp. 461–485.

[20] J. Zimmermann et al. In: *J. Phys. Chem. Lett.* 2.5 (2011), pp. 412–416.

[21] H. K. Ly et al. In: *ChemPhysChem* 11 (2010), pp. 1225–1235.

[22] W. Humphrey, A. Dalke, and K. Schulten. In: *J. Mol. Graphics* 14 (1996), pp. 33–38.

[23] H. Nar et al. In: *J. Mol. Biol.* 221 (1991), pp. 765–772.

[24] K. Sato et al. In: *Proc. Natl. Acad. Sci. USA* 106 (2009), pp. 5616–5621.

[25] H.J.C. Berendsen, D. van der Spoel, and R. van Drunen. In: *Comput. Phys. Commun.* 91.1-3 (1995), pp. 43–56.

[26] J. Wang, P. Cieplak, and P. A. Kollman. In: *J. Comp. Chem.* 21 (2000), pp. 1049–1074.

[27] W.L. Jorgensen et al. In: *J. Chem. Phys.* 79 (1983), p. 926.

[28] D. Brown and J.H.R. Clarke. In: *Mol. Phys.* 51 (1984), pp. 1243–1252.

[29] M. van den Bosch et al. In: *Chembiochem* 6.4 (2005), pp. 738–746.

[30] T. Darden, D. York, and L. Pedersen. In: *J. Chem. Phys.* 98 (1993), pp. 10089–92.

[31] T. Simonson. In: *Rep. Prog. Phys.* 66 (2003), pp. 737–787.

[32] A. Heck et al. In: *J. Phys. Chem. B* 116.7 (2012), pp. 2284–2293.

[33] J. Blumberger. In: *Phys. Chem. Chem. Phys.* 10 (2008), pp. 5651–5667.

[34] T. Simonson. In: *P. Natl. Acad. Sci.* 99 (2002), pp. 6544–6549.

[35] I. Muegge et al. In: *J. Phys. Chem. B* 101 (1997), pp. 825–836.

[36] J. Blumberger and M. L. Klein. In: *J. Am. Chem. Soc.* 128 (2006), pp. 13854–13867.

[37] S. Corni. In: *J. Phys. Chem. B* 109 (2005), pp. 3423–3430.

[38] H. B. Gray, B. G. Malmström, and R.J.P. Williams. In: *J. Biol. Inorg. Chem.* 5.5 (2000), pp. 551–559.

[39] U. L. F. Ryde and M. H. M. Olsson. In: *Int. J. Quantum Chem.* 81.2001 (2000), pp. 335–347.

[40] S. Monari et al. In: *J. Am. Chem. Soc.* 134.29 (2012), pp. 11848–11851.

[41] B.R. Crane et al. In: *J. Am. Chem. Soc.* 123 (2001), pp. 11623–11631.

4 How the dynamics of the metal-binding loop region controls the acid transition in cupredoxins

4.1 Acid Transition in Cupredoxins

Ionization of amino acid side chains is crucial to structural and functional properties of biomolecules, such as stability, solubility, binding and enzymatic activity [1, 2, 3]. For this reason, the activity of some proteins can be dramatically modulated by intracellular pH changes, which have been shown to be involved in the regulation of physiological processes like apoptosis and proliferation, migration and transport [4]. In this respect, an interesting case study is provided by cupredoxins, which are copper-containing redox proteins that act as electron carriers in several crucial cellular processes in both plants and bacteria [5, 6, 7]. The strong metal equatorial ligands at their T1 copper center are provided by the thiolate sulfur of a Cys and the nitrogen atoms of two His imidazoles. A Met usually act as a fourth, weaker axial ligand, imposing a distorted tetrahedral geometry [6]. A fifth weaker axial ligand is present in some species, as in the case of azurin (Gly45). Most of the blue copper proteins, such as plastocyanin (PC), amycianin (AMI), pseudoazurin and stellacyanin, undergo a pH-dependent structural rearrangement, known as the acid transition, which consists of protonation of the C-terminal His ligand in the Cu(I) protein, with pK_a values ranging from ∼7 to 5 (see ref. [6] and references therein). As a consequence of these structural changes, the reduction potential E^0 of the protein increases dramatically, rendering the species biologically inactive. Several hypotheses have been formulated concerning the possibility that this pH-induced structural rearrangement serves a physiological purpose [8, 9, 10]. In the case of PC, for example, it could down-regulate photosynthetic activity, as a consequence of significant lowering of the pH in the thylakoid lumen under extreme exposure to light [8, 11, 10]. It is possible that the acid transition of cupredoxins endows them with the ability to act as “pH-sensors”, as proteins whose activities are sensitive to small, physiologically relevant, changes in pH [4]. Therefore, understanding the pH-dependence of functional properties of biomolecules is central to the elucidation of their physiological activity. Moreover, the capability of a protein to undergo reversible conformational changes upon an external stimulus, such as alteration of pH, with subsequent changes in one or more functional property like E^0 could provide a biomolecular switch, with potential applications in biomolecular electronics, biosensing or molecular machines [12, 13].

The acid transition is not observed for azurin (AZ), possibly the most studied cupredoxin, and a pK_a value < 2 has been estimated [14]. What are the molecular determinants that hinder protonation of the His117 ligand in AZ? As three of the copper binding residues (Cys112, His117 and Met121, AZ numbering), including the His whose protonation triggers the acid transition, are situated on the C-terminal hydrophobic loop, previous work has been devoted to the elucidation of how the structural features of this secondary structure element could affect the pK_a of the His ligand. To this end, four AZ mutants, in which this C-terminal loop has been replaced by those of other cupredoxins (AMI and PC, thus generating the AZAMI and AZPC loop mutants) or with non native polyalanine loops (AZ4A3A and AZ4A4A) were designed and produced [15, 16, 17]. These loop mutants were then characterized by solving their crystallographic structures and functionally, via spectroscopic and electrochemical studies [15, 16, 18, 17, 19].

The loop length, which was found to correlate with the pK_a of His protonation, was suggested as one of the factors affecting the thermodynamics of the acid transition, together with second coordination sphere effects such as hydrogen bonding and π interactions [15, 16, 18, 17, 20] and differences in the solvent accessibility of the C-terminal His ligand [21, 22, 17]. Nevertheless, it has been difficult to state which specific molecular features relate the length of the loop and/or its composition to changes in functionality. For example, the hydrophobic loop of AZ is composed of the same number of residues as that of AZ4A3A, but only the latter species undergoes the acid transition. Therefore, additional investigations are required to elucidate the determinants of the thermodynamics of the acid transition in cupredoxins. Keeping in mind that the functionality of a protein is ultimately controlled by its dynamics [23, 24, 25], we decided to go beyond the structure/function relationship and add the time dimension to the investigation of AZ and of its loop mutants. We therefore performed molecular dynamics simulations of AZ, AZAMI, AZPC, AZ4A3A and AZ4A4A (whose active site structures are displayed in figure 4.1), looking for dynamic aspects of cupredoxins that could play a key role in influencing the acid transition.

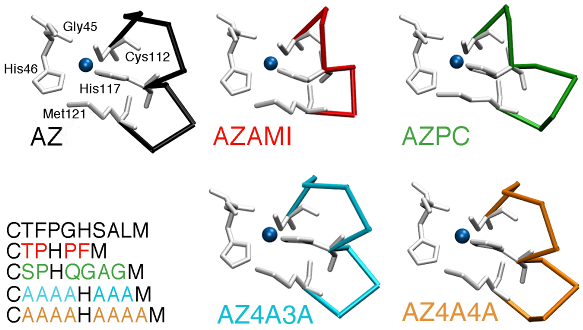


Figure 4.1: Active site structures of the five proteins investigated in the present work. The five metal-coordinating residues are displayed as white licorice and, for sake of clarity, are labelled only in AZ. The copper atom is displayed as a blue sphere. The hydrophobic loop is displayed as a C_α trace and coloured differently for AZ and its mutants; the colour code is mantained throughout the chapter. On the lower left corner, the amino acid sequences of the different loops are also shown. The image was prepared with VMD software [26] using the crystal structures 4AZU[27], 2FTA[15], 2HX7[16], 3FSW[18], 3FSZ[18].

4.2 Molecular Dynamics Simulations

Since the acid transition is observed only for the reduced species [9], we used the MD simulations performed for the study of the reorganization energy for the reduced ensemble. Molecular dynamics does not take into account variations of the bonds between copper and its ligands nor does it allow observation of the breaking/formation of chemical bonds. Therefore, our work did not address the reasons of the different behaviour of the two oxidation states of cupredoxins with respect to the acid transition. The atomic charges of the active site, calculated at the DFT level of the theory, were taken form the literature [28]. The distances and the angles between Cu and the five coordinating atoms were constrained to the values in the corresponding crystal structures, in order to avoid unrealistic distortion of the copper geometry that could be caused by the treatment of the Cu-ligand interactions at a classical level of the theory [28]. It is reasonable not to simulate the whole mechanism of His protonation and dissociation, since our goal is to determine whether the characteristics of the system before the transition takes place affect the propensity to undergo this process. Three independent production runs (replicas) were carried out for each protein, in order to improve the conformational sampling [29]. The length of the first simulation was 100 ns, the length of each additional replica was 60 ns. Values and errors were obtained averaging over all replica simulations. Functional Mode Analysis was implemented as previously described [30, 31]. Before analysis, trajectories in the GROMACS XTC format were least-square fitted to a reference frame to filter out overall translation and rotation [31]. The first half of the trajectory was used for model building, while the remaining part served for the cross-validation of the model. The number of PLS components to be used was evaluated by calculating the correlation coefficients for the model training subset (R_m) and for the cross-validation subset (R_c) as a function of the number of components, and by choosing the lowest number of latent vectors which would ensure convergence of the R_c value. Controls to assess the significance of the FMA-derived modes were carried out as follows: the trajectories for two different proteins (generically labelled as A and B) were merged into one, saving only the coordinates of a common subset of atoms (residues 1 to 112). The merged trajectory was least-square fitted to a reference frame to filter out overall translation and rotation. The time dependence of SASA of the active site for the two proteins A and B was also merged into a single file. Again, the first half of the trajectory, belonging to protein A, was used for model building, while the second part, obtained by MD of protein B, served for the cross-validation of the mode. A very low R_c value is expected if the differences between A and B are indeed significant. The Y72A AZ, P114F-AZAMI and P114F-AZPC mutants were prepared *in silico* by homology modelling with SWISS-MODEL 8.05 [32, 33, 34, 35, 36] using as input the FASTA sequence of the mutants, and as the template the PDB structure of AZ (PDB-ID 4AZU, chain A[27]),AZAMI (PDB-ID 2FTA, chain A[15] and AZPC (PDB-ID 2HX7, chain A[16]), respectively. The simulation setup and the analysis tools are the same used for other proteins.

4.3 Results

4.3.1 Active site solvent accessible surface

The proton triggering the acid transition must be provided by the solvent, and we have therefore focused our efforts on the interplay between the protein, and in particular the region around its copper active site, and surrounding water molecules. The first quantity that we calculated was the solvent accessible surface area (SASA) of the active site, where the active site is defined as the metal atom plus its five ligands (Gly45, His46, Cys112, His 117 and Met121, AZ numbering). In fact, changes in the SASA of the active site can quantitatively describe the extent of the interface protein/solution that can affect, through electrostatic and dipolar interactions, the apparent

Protein	Exp. $pK_a{}^{[a,b]}$	Calc. $\langle SASA \rangle$ (nm^2)$^{[a]}$	SASA X-ray (nm^2)$^{[a]}$
AZ	<2	0.242 (0.003)	0.109 (0.022)
AZAMI	5.5 (0.2)	1.024 (0.057)	0.979 (0.249)
AZPC	4.4 (0.2)	0.504 (0.023)	0.443 (0.044)
AZ4A3A	4.5 (0.1)	0.882 (0.071)	0.260 (0.018)
AZ4A4A	4.4 (0.1)	0.506 (0.050)	0.241 (0.035)

Table 4.1: Experimental pK_a values for protonation of the C-terminal His ligand. Calculated Solvent Accessible Surface Area (SASA) values of the protein active site, averaged over all the replicas performed in the reduced ensemble, for the five proteins under investigation. Mean of the active site SASA for the different chains of the crystal structure. [a] Associated standard error is given in brackets. [b] Taken from refs. [15, 16, 18].

dielectric constant and (de)stabilize the protonated His117 [22, 37]. The calculated SASA values, averaged over 3 independent simulations (replicas hereafter), are listed in table 4.1, and the plot of the experimental pK_a values versus the calculated active site SASAs is displayed in figure 4.2. It can be readily seen that the protein featuring the highest pK_a value, AZAMI, is also the one for which the highest SASA was obtained. On the contrary, the accessibility of the active site of AZ to the solvent is much smaller than that of the other species, in accordance with its lack of acid transition at pH >2. AZPC, AZ4A3A and AZ4A4A all feature intermediate pK_a values, and their SASA values lie between the two extremes represented by AZAMI and AZ. Therefore, a general trend can be observed: the more the protein active site is exposed to the surrounding solvent, the higher the corresponding pK_a of His protonation.

The SASA values displayed in figure 4.2 are average quantities resulting from the whole sampling provided by our simulations, therefore indicated as $\langle SASA \rangle$. It is also interesting to take into account the fluctuations of the SASA values with time. Thus, we calculated the normalized probability distributions of the active site SASA along the 100 ns-long MD simulations and we plot them in figure 4.3.

The distribution for AZ is much narrower than that of the loop mutants, and never exceeds 0.5 nm^2. All the other proteins span a significantly wider range of SASA values. This finding is related to different dynamics of the five proteins along the simulation time. The conformations accessed by AZ along our simulations all feature very close active site SASA values, while the fluctuations of the other species, and especially those of AZAMI and AZ4A3A, lead to a structural sampling displaying a set of conforma-

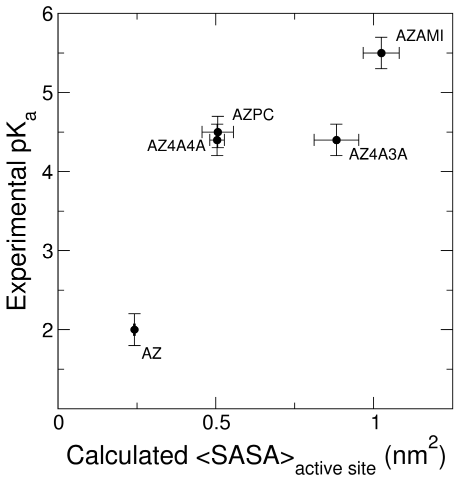


Figure 4.2: Experimental pK$_a$ values for the protonation of the C-terminal His ligand versus the calculated SASA of the active site. The pK$_a$ value for AZ could not be determined experimentally, and its corresponding data point on the plot was drawn assuming a pK$_a$ value of 2.

tions characterized by significant changes in the solvent accessibility of the active site. Therefore, the average active site $\langle$SASA$\rangle$ for AZ is not only the smallest, but also the one which features significantly lower fluctuations with respect to its four mutants. The differences concerning the active site SASA between the five proteins are not simply structural but stem from different dynamics of the molecules, as suggested by the better correlation between experimental pK$_a$ and calculated SASA, with respect to that with SASA values obtained from crystal structures (see table 4.1). Differences between calculated SASA and the corresponding values from crystal structures are most likely due to the low temperature at which X-ray diffraction data were collected and the close packing of molecules in the crystal. In general, proteins in solution and under physiological conditions can display a larger structural heterogeneity than that shown in a crystal structure [38, 39]. Nevertheless, as previously suggested [17], a correlation can also be observed when plotting pK$_a$ vs. SASA from crystal structures, see figure 4.4, thus supporting our MD results.

Further insight into the relationship between the protein active site and the solvent may be obtained by studying the structure of the hydration layers around the metal center. This can be achieved by calculating the radial distribution function $g(r)$ between the copper and the oxygen atoms of the solvent. This is displayed in 4.5, which provides the probability of

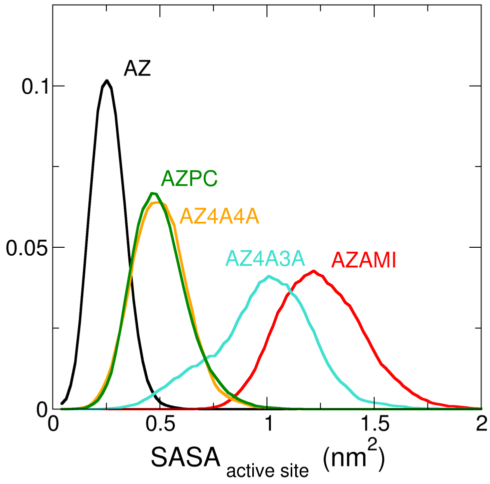


Figure 4.3: Normalized distribution of the calculated SASA for the active site of the five investigated proteins.

finding the oxygen atom of a water molecule at a given distance r from the copper atom. The first sharp peak in the Cu-O radial distribution function for AZ falls slightly below 0.7 nm, as previously reported [40], and corresponds to a solvent molecule hydrogen-bonded to His117. On the contrary, all the other proteins, although invariably showing a peak at the same r value, also feature a non-zero probability of finding a water molecule at shorter distances. This is particularly evident in the case of AZAMI, whose first $g(r)$ peak falls around 0.4 nm. This difference in the solvent structure around the copper atom, which likely stems from the substitution of Phe with a much less bulky side chain [16], as observed for the Phe114Pro mutant of AZ [41], allows us to hypothesize that water molecules can get significantly closer to the protein active site in the loop mutants than in AZ, thereby favouring the probability of the acid transition.

4.3.2 Structural motions involved in the dynamic changes of SASA

All the results collected so far provide a picture of an active site in AZ which is significantly less exposed to the solvent and less prone to undergoing structural fluctuations that could allow water molecules to approach the copper, probably as a consequence of significant differences between the dynamics of AZ and the four mutants. We therefore turned our attention to searching for the structural motions of the loop-mutants which are involved in the

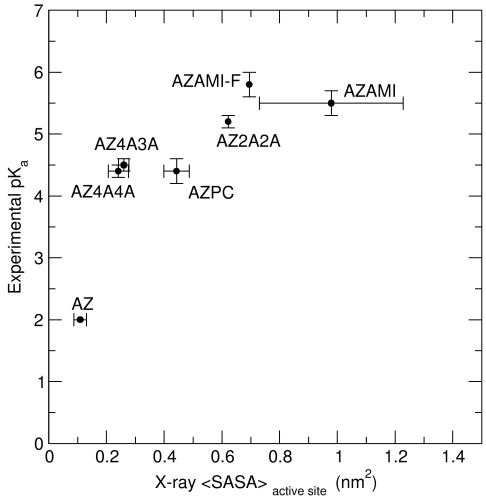


Figure 4.4: Experimental pK$_a$ values for the protonation of the C-terminal His ligand versus the SASA of the active site, obtained from the corresponding crystal structures used as a starting point for the simulations. The pK$_a$ value for AZ could not be determined experimentally, and its corresponding data point on the plot was drawn assuming a pK$_a$ value of 2. The errors associated with the SASA values are the standard error between the SASA for the different chains in the crystal structures of the same protein. Please note the very large variation for the four chains of AZAMI. The PDB codes for crystal structures that were used are the following: AZU: 4AZU.pdb, AZAMI: 2FTA.pdb, AZPC: 2HX7.pdb, AZ4A3A: 3FSW.pdb, AZ4A4A: 3FSZ.pdb, AZAMI-F: 2FT6.pdb, AZ2A2A: 2XV0.pdb. Experimental pK$_a$ values for AZAMI-F and AZ2A2A were taken from Li et al., Proc. Natl. Acad. Sci. USA, 2006, 103, 7258-7263, Li et al., Inorg. Chem., 2011, 50, 482-488 and Battistuzzi et al., Biochim. Biophys. Acta, 2009, 1794, 995-1000, respectively.

dynamic changes of the active site SASA over time. This was achieved using Functional Mode Analysis (FMA) [30, 31], which enables identification of the collective atomic modes of a protein that maximally correlate to a functional quantity of interest, in our case the time dependent SASA of the active site. By applying Partial Least-Squares (PLS) based FMA [31] to our simulations, the protein dynamics underlying the SASA fluctuations could be detected and compared. The quality of the models calculated by FMA was evaluated by cross-validation. We obtained convincing cross-validation correlation coefficients (R$_c$) for the loop mutants (ranging from ≈0.72 to ≈0.85), indicating that a reliable collective model underlying the SASA fluctuations could be built. On the contrary, in the case of AZ, PLS-based FMA could not provide an acceptable model correlating the time dependence of active site SASA to collective protein motions. Indirectly, this again points

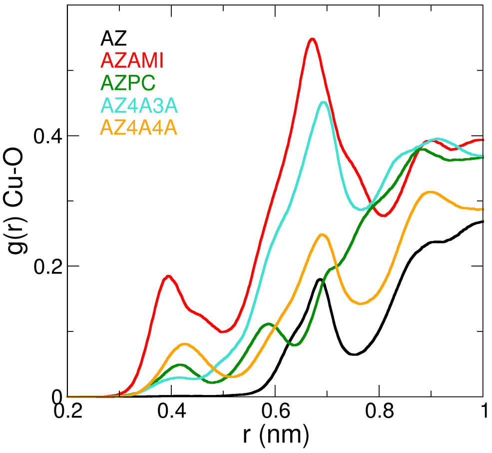


Figure 4.5: Cu-O radial distribution function $g(r)$. The O atoms taken into account for the calculations are only those belonging to solvent.

to different behavior of AZ with respect to the loop mutants. In order to test whether the differences between the FMA-derived motions were significant, controls were carried out (see Methods section) by building an FMA model for one trajectory and cross-validation by another. Extremely low R_C values (≈ 0.15) show that, indeed, the collective dynamics underlying SASA fluctuations differ significantly for the different cases. The three dimensional representation of the maximally correlated modes contributing to the changes in the active site SASA are depicted in figure 4.6.

The motions contributing mainly to SASA fluctuations involve concerted movements of several portions of the protein. Moreover, although the mutants all have the same protein scaffold and the only structural difference lies in the length and composition of the rather short hydrophobic loop, the detected functional modes are significantly different. This finding implies that the C-terminal loop is crucial to overall protein dynamics and, consequently, functional properties, and that modifying a few amino acids in this region can lead to dramatic changes. It is possible to discern which residues contribute most to each SASA-correlated internal mode, by calculating the root mean square fluctuations (RMSF) per residue within the collective movement obtained by FMA. In general, loops are the secondary structure elements most involved in the collective movements, and none of the modes describe dramatic structural rearrangements, as expected for an electron transfer protein, which in general can be considered rather rigid systems. In the case of AZAMI, the major contribution to the FMA-derived

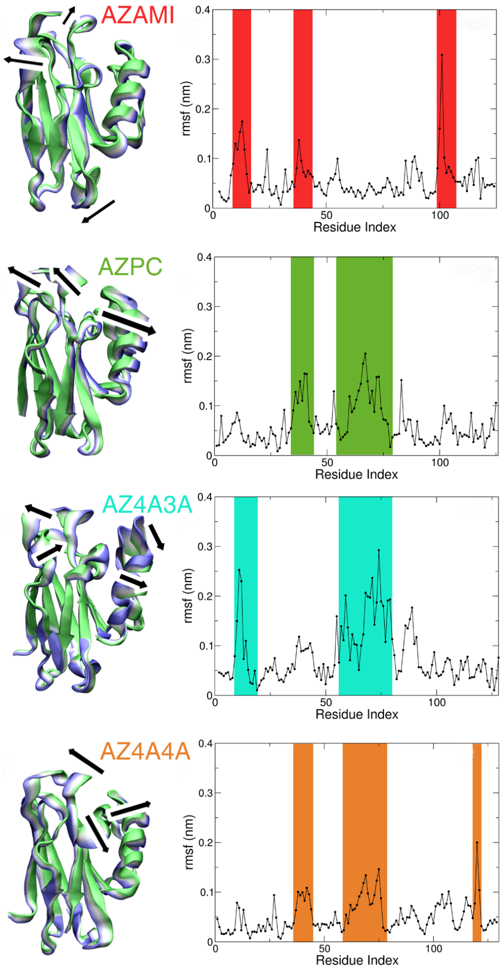


Figure 4.6: Backbone representation of the collective motions determining to the fluctuations in the active site SASA values. No representation could be shown for AZ, as PLS-based FMA could not provide a reliable model.

modes come from residues belonging to the 9-18, 36-44 and 99-107 loops. The collective mode of AZPC is dominated by the motion of the only α-helix, spanning residues 58 to 71, and by the following loop, but a significant contribution also derives from the 36-44 loop. Similar behaviour was observed for native plastocyanin, whose loops surrounding the copper atom display the most significant structural rearrangements upon protonation of the C-terminal His92 ligand [42]. For AZ4A3A, the highest RMSF values are displayed again by the residues forming the α-helix and by the following loop, together with the 9-18 loop. The ensemble-weighted maximally correlated mode for AZ4A4A arises from the concerted movements of the α-helix, the 36-44 loop and of the final part of the introduced polyalanine loop. Therefore, with the exception of the involvement of the 99-107 loop for AZAMI, the collective modes that cause fluctuations of the active site SASA values involve the concerted movements of the loops surrounding the copper atom and of the α-helix. The mutated hydrophobic loop does not move significantly, except in AZ4A4A, as it contains three of the copper ligand whose distance to the metal atom is kept constant throughout the whole simulation time.

So far, we have shown that the ability of AZ and the loop mutants to undergo the acid transition most likely depends on the extent and fluctuations in the exposure of the active site to solvent due to collective protein motions. The final questions that needs to be addressed is why the region around the copper site in AZ is so significantly less flexible, thus hindering His protonation and detachment even at low pH values. In AZ, and, to a lesser extent, in the loop mutants, the copper atom is shielded from direct contact with the solvent by the presence of a large hydrophobic area, formed by the close packing between the 9-18, 36-44, 72-79 loops as well as by the ligand-containing 112-121 loop. The relative position of these loop regions is kept constant by a complex network of hydrogen bonds, all falling within the so-called “second-coordination sphere” [43, 44, 41]. The importance of some key residues which do not bind copper to the spectroscopic and functional properties of AZ (and other cupredoxins) is well documented [43, 44, 45, 41, 20].

The hydrogen bonds accepted by the thiolate group of Cys112 from the backbone amides of Asn47 and Phe114 are among the most important to the functionality of AZ [43, 44, 45, 41, 20]. The former is present in all the investigated species whilst the Cys112-Phe114 interaction is absent in both AZAMI and AZPC [15, 16, 18]. This is because both variants feature a Pro in place of the Phe [16]. In both AZ4A3A and AZ4A4A, the backbone NH of Ala114 does act as a hydrogen bond donor to the Cys112 thiolate, but fluctuations in this bond length in the two mutants are much higher than in AZ, often exceeding 3.5 Å, particularly for AZ4A4A (data not shown).

Another potentially important second-coordination sphere interaction in AZ is the hydrogen bond between the imidazole ring of His46 and the backbone carbonyl of Asn10, which is mantained in both AZPC and AZ4A4A. This bond is also present in the AZAMI and AZ4A3A crystal structures, but is quickly and permanently lost during our simulations for both species. As a consequence, the loop spanning residues 9 to 18 is more free to move away from the active site, thus signifcantly contributing to the collective motions underlying SASA fluctuations for AZAMI and AZ4A3A (see figure 4.6).

The enhanced freedom of movement of the α-helix and the loop which follows it in the four loop mutants is more difficult to justify. One tentative explanation could lie in the absence of Phe114 in the mutants with respect to AZ. The phenyl ring of this residue is located very close to the aromatic side chain of Tyr72, a residue that belongs to the loop located after the helix and a T-shaped $\pi - \pi$ interaction could exist between Phe114 and Tyr72 (see figure 4.7, panel A). To test this hypothesis, we monitored both the distance r between the centers of mass of the two aromatic rings and the angle θ between the normals to the ring planes along our simulations of AZ, obtaining r=0.50±0.03 nm and θ=63.0±0.5 °. Therefore, the relative orientation of the two lateral chains is rather constant, and also the distance between the two rings does not change markedly. Moreover, these values are in agreement with those for the most stable T-shaped structure calculated in vacuo between tyrosine and phenylalanine side chains (r=0.51 and θ=74.2) [46]. Although the stacked conformation is known to be the most stable for a Tyr-Phe $\pi - \pi$ interaction, it has also been stated that distal interacting Phe-Tyr are prevalently T-shaped [46], as in our case, where the distance between the C_α atoms of the two residues is 0.59±0.02. To prove whether the Tyr72-Phe114 interaction is important for the structural stability of AZ, we created *in silico* and simulated the Y72A mutant of azurin, which cannot have a $\pi - \pi$ interaction with Phe114. Removal of the aromatic side chain of Tyr72 leads to a higher freedom of movement of the α-helix region, as witnessed by the comparison between the RMSF per residue of WT AZ and Y72A. In fact, while for AZ the region encompassing residues 50 to 80 does not undergo significant fluctuation throughout the whole MD sampling, the same protein region for the Y72A variant displays the highest RMSF values (figure 4.7).

Further insights into the role played by Phe114 were provided by the construction and modelling *in silico* of the P114F mutants of AZAMI and AZPC. In the case of P114F-AZAMI, reverting Pro back to Phe leads to

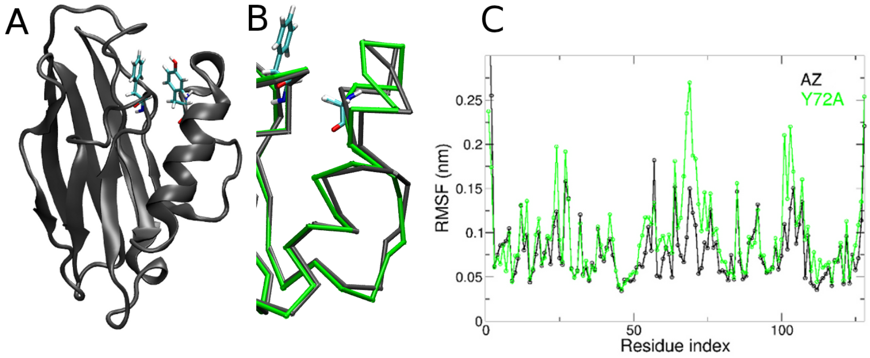


Figure 4.7: A: Cartoon representation of AZ. Tyr72 and Phe114 are represented in licorice sticks. B: Superimposition of the average structure along the MD simulation obtained for WT azurin (black) and for the Y72A mutant (green) with Phe114 and the mutated Ala72 displayed as licorice sticks. C: Comparison between the RMSF per residue for AZ (black) and Y72A (green).

a dramatic decrease in the active site SASA, whose value for the P114F-AZAMI is 0.684±0.019 nm^2, significantly lower than that for AZAMI (see table 4.2). The $g(r)$ plot for the AZAMI-P114F is also remarkably different from that of AZAMI, due to the removal of the 0.4 nm peak.

Protein	$\langle$SASA$\rangle$ (nm^2)[a]	r (nm)[a]	θ (°)[a]
AZAMI	1.024 (0.057)	-	-
P114F-AZAMI	0.684 (0.019)	0.51 (0.03)	89.55 (16.11)
AZPC	0.504 (0.023)	-	-
P114F-AZPC	0.485 (0.071)	0.51 (0.06)	65.66 (31.74)

Table 4.2: Active site average SASA, distances between the centers of mass ot the two aromatic rings of Tyr72 and Phe114 (r) and angles between the normals to the ring planes (θ) for AZAMI, AZPC and their correspondin P114F mutants. [a] Associated standard errors are given in brackets.

The differences between AZAMI and its mutant can be ascribed to restoration of the $\pi - \pi$ interaction between the aromatic rings of Phe114 and Tyr72 (see figure 4.8, panel B), as witnessed by the r and θ values for the AZAMI-P114F variant (see table 4.2). On the contrary, the P114F mutation in AZPC does not affect the active site SASA and the changes in the g(r) plot are opposite to those observed for P114F-AZAMI. It is important to stress that it in AZAMI and AZPC the newly introduced Phe is adjacent to the His ligand, whereas in AZ there are two intervening residues: there-

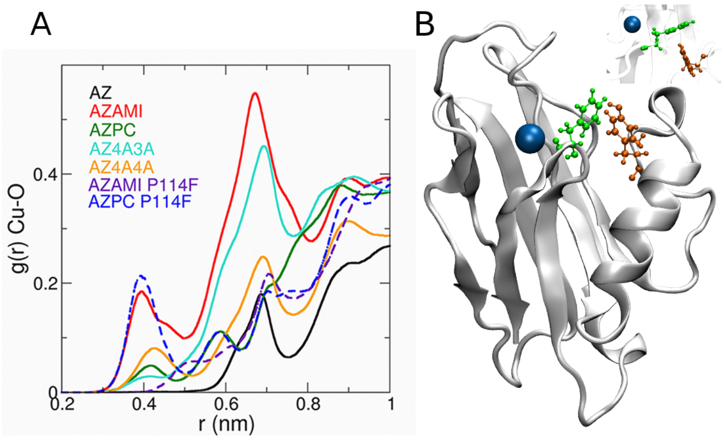


Figure 4.8: A: Cu-O radial distribution function $g(r)$ for the five investigated species and for the P114F-AZAMI and P114F-AZPC variants. The purple dashed line and the blue dashed line correspond to the P114F-AZAMI mutant and to the P114F-AZPC mutant, respectively. B: Cartoon representation of a representative structure of the P114F-AZAMI mutant along the simulation. Tyr72 (orange) and Phe114 (green) are represented with balls and sticks. The inset displays the protein active site from the "northern" side of the molecule, showing the T-shaped conformation of the rings of Tyr72 and Phe114.

fore, care should be taken when directly comparing the P114F mutants to AZ, and only indirect inferences can be obtained by these additional simulations. However, these results suggest that Phe114 is one of the structural attributes contributing to the differences between the species but is not the only factor affecting the protein/solvent interplay, which is determined by a number of structural and dynamical features.

Overall, our results are in agreement with previous works investigating the internal mobility of cupredoxins, which demonstrate that these proteins, even under native conditions, can access multiple conformations involving low-populated states which are essential for their functionality [47, 48]. Similar findings were also obtained for other systems [38], and in particular cytochrome c [49, 25], another ET protein which is usually considered as being rigid.

We have investigated the molecular effectors of the acid transition in cupredoxins by comparing the dynamic behavior of WT AZ and of four variants in which the C-terminal hydrophobic loop was mutated. We find that the SASA of the protein active site largely affects the pK$_a$ of the C-terminal His ligand on this loop, which triggers the whole structural transition. Notably, not only the average SASA value seems to be crucial to the pH-dependent rearrangement, but also the extent of its flucutations over time. Indeed, the effect of loop mutations is not only structural but also dynamic. Another significant difference between AZ and its loop mutants concerns the structure of the water layer surrounding the protein active site, as witnessed by the copper-oxygen radial distribution function $g(r)$. The absence of peaks at distances shorter than 0.7 nm for AZ, in contrast with all the other investigated species, suggests that for the loop mutants solvent molecules can approach the active site thus enhancing the probability of the acid transition. In order to identify the protein regions mostly contributing to the SASA fluctuations we performed FMA-based investigations. These allowed us to ascribe the different behavior of the loop mutants to enhanced freedom of movement of the loops surrounding the active site with respect to AZ, which is a consequence of the loss of some crucial second-coordination sphere interactions. Our results converge to strengthen the growing awareness that the functional properties of ET proteins are ultimately determined by the dynamic interplay between the protein scaffold and its environment. Similar conclusions have already been made for several biomolecules, very different from ET species in both structure and function, including membrane voltage channels [50, 51].

References

[1] N.V. Di Russo et al. In: *PLoS Comput. Biol.* 8 (2012), e1002761.

[2] G.R. Grimsley, Scholtz J.M., and C.N. Pace. In: *Protein Sci.* 18 (2009), pp. 247–251.

[3] A. Narayanan and M. P. Jacobson. In: *Curr. Opin. Struc. Biol.* 19.2 (2009), pp. 156–63.

[4] J. Srivastava, D. L. Barber, and M. P. Jacobson. In: *Physiology* 22 (2007), pp. 30–9.

[5] H. B. Gray, B. G. Malmström, and R.J.P. Williams. In: *J. Biol. Inorg. Chem.* 5.5 (2000), pp. 551–559.

[6] C. Dennison. In: *Coordin. Chem. Rev.* 249.24 (2005), pp. 3025–3054.

[7] M. G. Savelieff et al. In: *Proc. Natl. Acad. Sci. USA* 105.23 (2008), pp. 7919–24.

[8] J.M. Guss et al. In: *J. Mol. Biol.* 192 (1986), pp. 361–387.

[9] G. Battistuzzi et al. In: *Biochemistry* 41 (2002), pp. 14293–14298.

[10] M. A. S. Hass et al. In: *Biochemistry* 46.50 (2007), pp. 14619–28.

[11] K. Sato, T. Kohzuma, and C. Dennison. In: *J. Am. Chem. Soc.* 125 (2003), pp. 2101–2112.

[12] C. A. Bortolotti et al. In: *Chem. Sci.* 3.3 (2012), pp. 807–810.

[13] M. M. Boyle et al. In: *Chem. Sci.* 2 (2011), pp. 204–210.

[14] L. J. C. Jeuken et al. In: *J. Am. Chem. Soc.* 122 (2000), pp. 12186–12194.

[15] C. Li et al. In: *Proc. Natl. Acad. Sci. USA* 103 (2006), pp. 7258–7263.

[16] C. Li, M. J. Banfield, and C. Dennison. In: *J. Am. Chem. Soc.* 129 (2007), pp. 709–718.

[17] C. Li et al. In: *Inorg. Chem.* 50.7 (2011), pp. 482–488.

[18] K. Sato et al. In: *Proc. Natl. Acad. Sci. USA* 106 (2009), pp. 5616–5621.

[19] G. Battistuzzi et al. In: *Biochim. Biophys. Acta* 1794 (2009), pp. 995–1000.

[20] S. Yanagisawa et al. In: *J. Am. Chem. Soc.* 130 (2008), pp. 15420–15428.

[21] G. Battistuzzi et al. In: *Biochemistry* 36 (1997), pp. 16247–16258.

[22] D. M. Hunter et al. In: *Inorg. Chem.* 40 (2001), pp. 354–360.

[23] M Karplus and J Kuriyan. In: *Proc. Natl. Acad. Sci. USA* 102.19 (2005), pp. 6679–85.

[24] K. Henzler-Wildman and D. D. Kern. In: *Nature* 450 (2007), pp. 964–972.

[25] C. A. Bortolotti et al. In: *J. Am. Chem. Soc.* 134 (2012), pp. 13670–13678.

[26] W. Humphrey, A. Dalke, and K. Schulten. In: *J. Mol. Graphics* 14 (1996), pp. 33–38.

[27] H. Nar et al. In: *J. Mol. Biol.* 221 (1991), pp. 765–772.

[28] M. van den Bosch et al. In: *Chembiochem* 6.4 (2005), pp. 738–746.

[29] E. Papaleo et al. In: *PLoS comput. biol.* 7.5 (2011), e1002056.

[30] J. S. Hub and B. L. de Groot. In: *PLoS Comput. Biol.* 5 (2009), e1000480.

[31] T. Krivobokova et al. In: *Biophys. J.* 103 (2012), pp. 786–796.

[32] K. Arnold et al. In: *Bioinformatics* 22 (2006), pp. 195–201.

[33] F. Kiefer et al. In: *Nucleic Acids Res.* 37 (2009), pp. D387–D392.

[34] T. Schwede et al. In: *Nucleic Acids Res.* 31 (2003), pp. 3381–3385.

[35] N. Guex and M. C. Peitsch. In: *Electrophoresis* 18 (1997), pp. 2714–2723.

[36] M. C. Peitsch. In: *Bio-technol* 13 (1995), pp. 658–660.

[37] D. G. Isom et al. In: *Proc. Natl. Acad. Sci. USA* 108 (2011), pp. 5260–5265.

[38] K. Henzler-Wildman and D. Kern. In: *Nature* 450 (2007), pp. 964–972.

[39] P. R. Blake et al. In: *Prot. Sci.* 1 (1992), pp. 1522–1525.

[40] M. Cascella et al. In: *P. Natl. Acad. Sci. USA* 103 (2006), pp. 19641–19646.

[41] S. Yanagisawa, M. J. Banfield, and C. Dennison. In: *Biochemistry* 45 (2006), pp. 8812–8822.

[42] M. A. S. Hass et al. In: *J. Am. Chem. Soc.* 126 (2004), pp. 753–765.

[43] J. J. Warren et al. In: *J. Inorg. Biochem.* 115 (2012), pp. 119–26.

[44] T. D. Wilson, Y. Yu, and Y. Lu. In: *Coord. Chem. Rev.* 257.1 (2013), pp. 260–276.

[45] N. M. Marshall et al. In: *Nature* 462.7269 (2009), pp. 113–116.

[46] R. Chelli et al. In: *J. Am. Chem. Soc.* 124.21 (2002), pp. 6133–43.

[47] D. M. Korzhnev et al. In: *Prot. Sci.* 12 (2003), pp. 56–65.

[48] L. Ma et al. In: *Biochemistry* 42 (2003), pp. 320–330.

[49] M.M.G. Krishna et al. In: *J. Mol. Biol.* 331 (2003), pp. 29–36.

[50] A. Alessandrini and P. Facci. In: *Comm. Int. Biol.* 4 (2011), pp. 346–348.

[51] A. Alessandrini and P. Facci. In: *J. Mol. Recognit.* 24 (2011), pp. 387–396.

5 Unambiguous assignment of reduction potentials in diheme cytochromes

5.1 Di-heme cytochrome *c*

Multicenter metalloproteins and particularly multiheme proteins are among the most fascinating molecular machineries evolved by Nature to carry out reactions requiring multiple electrons and accomplish electron transport over long distances [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12]. They have attracted an increasing interest in these last years, not only in the effort of elucidating the fundamental aspects of the physiological processes they take part to, but also for their potential use as novel players in the field of nanobiotechnology (e.g. as molecular wires) [13, 14, 15, 16, 17, 18]. Although, particularly for the soluble species, it is possible to experimentally determine the reduction potentials of the heme centers,[1, 13] assigning each E^0 value to its corresponding center in a sequence-specific manner is not straightforward, although thorough investigations combining NMR and visible spectroscopy have been succesfully employed [19, 1]. Detailed thermodynamic characterization of multicenter redox proteins is of fundamental importance for a deeper understanding of vectorial electron transfer (ET) in multiheme systems. Moreover, a detailed picture of the free energy landscape across these systems is mandatory for understanding and controlling their ET properties at molecular level. It follows that the availability of a reliable predictive theoretical tool for the estimation of the E^0 values in such systems would constitute an important achievement in this context, to be used either as a predictive tool or in synergy with the aforementioned experimental approaches to unveil the dynamics/function relationship determining their functionality. This is even more important for multiheme proteins for which only structural data are available and experimental determination of E^0 is difficult, as for membrane embedded species. Here, we study a small ET system featuring more than one heme group using a combined Perturbed Matrix Method (PMM)/Molecular Dynamics (MD) approach. As this study serves as a benchmark for assessing the capability of the PMM/MD approach to unambiguously assign the reduction potentials to the metal centers in a multiheme protein, we focus on one of the simplest vectorial ET species, namely the DHC from *Rhodobacter sphaeroides.* Different theoretical approaches can be employed to estimate the reduction potential of a metalloprotein[20, 21, 22, 14, 23, 24, 25, 26, 27]. We calculated the E^0 of the redox centers in DHC using the PMM/MD approach [28, 29, 30, 31, 32, 33], which allowed us to assign the previously reported E^0 values to the corresponding heme center. The extremely good agreement between calculated and experimental E^0 values testifies the solidity of our theoretical approach, suggesting its applicability to more complex biosystems featuring several redox centers. Low-redox potential diheme cytochromes *c*, previously known as cytochromes *c*551.5 [34], are soluble 16 kDa proteins. The DHC from *Rb. sphaeroides* accepts electrons from a membrane-embedded cytochrome b/membrane diheme cytochrome chimera, and passes them to a single-heme species known as sphaeroides heme protein (SHP)[35], which is at present the only c-type heme protein known to bind molecular dioxygen. Although the physiological function of this small gene cluster is still unknown, it is much likely that it serves as a conserved ET pathway, as the same cluster was found also in several other protobacteria [35, 36]. DHC folds into two distinct domains, each containing a single heme *c*: the N-terminal domain spans the first 72 residues and closely resembles the folding of class I cytochromes *c*, while the C-terminal one is composed by residues 94 to 139 and is different from any available structure of c-types cytochromes[37]. The two domains are connected by a 17 residue-long loop[37]. The reduction potentials for the hemes of DHC (-310 and -240 mV vs SHE) and SHP (-105 mV) were measured by spectroelectrochemical experiments [37]. The sign and magnitude of the E^0 for the two species reinforced the hypothesis of the role of DHC as electron carrier between its putative partners, but it was not possible to discern which of the two DHC hemes would feature the lower and higher E^0.

5.2 Methods

5.2.1 Calculation of the reduction potential

To estimate the reduction potential E^0 associated to the $Fe^{+3} + e^- \rightarrow Fe^{+2}$ semireaction occurring at a heme site of the DHC protein, MD simulations of the protein in solution were firstly performed in both the reduced and oxidized ensemble, i.e. with the considered heme in its reduced or oxidized state, respectively (details of the MD simulations are given below). E^0 was, then, calculated by applying the mixed quantum mechanics/molecular mechanics approach named PMM [28, 29, 30, 31, 32, 33] to the MD simulations, closely paralleling the procedure used in our previous works[38, 20, 21]. In what follows, the use of PMM/MD approach to the calculation of reduction potential is briefly outlined.

The (Helmholtz) free energy (ΔA^0) change associated to the $Fe^{+3} + e^- \rightarrow Fe^{+2}$ semireaction was calculated using the following equation (ΔA^0 is related to the reduction potential via $E^0 = -\Delta A^0/nF$, with F the Faraday constant and n is the number of electrons involved in the process):

$$\Delta A^0 \cong -k_B T \ln \left\langle e^{-\beta \Delta \mathcal{U}} \right\rangle_{ox} = k_B T \ln \left\langle e^{\beta \Delta \mathcal{U}} \right\rangle_{red} \tag{5.1}$$

In the above equation $\Delta \mathcal{U} = \mathcal{U}_{red} - \mathcal{U}_{ox}$, with $\mathcal{U}_{red}$ and $\mathcal{U}_{ox}$ the perturbed electronic ground-state energy of the reduced (Fe^{+2}) and oxidized (Fe^{+3}) state, respectively, of the quantum center (QC), i.e. the portion of the system involved in the reduction transition (see below for the definition of the QC). $\mathcal{U}_{red}$ and $\mathcal{U}_{ox}$ are evaluated at each frame of a classical MD simulation via the PMM approach, i.e. by diagonalizing at each MD frame the perturbed Hamiltonian matrices of the reduced and oxidized QC, respectively (details of the PMM procedure can be found in Refs. [28, 29, 30, 31, 32, 33, 38, 20, 21]), and the averaging is performed in the reduced or oxidized ensemble as indicated by the angle brackets subscript. Although 5.1 is based on in principle an exact relation, given the sampling problems of finite-time simulations, the best estimate of the reduction free energy is obtained by averaging the values provided by the reduced and oxidized ensembles[39]:

$$\Delta A^0 \cong \frac{-kT \ln \left\langle e^{-\beta \Delta \mathcal{U}} \right\rangle_{ox} + kT \ln \left\langle e^{\beta \Delta \mathcal{U}} \right\rangle_{red}}{2} \tag{5.2}$$

5.2.2 MD simulations

The crystal structure of the diheme cytochrome *c* (DHC) (PDB id code: 2FWT [37]) lacks of the first 12 and the last 3 residues. These residues were modelled with the EasyModeller software (version 4.0) [40] using as input sequence the FASTA sequence of the protein and as a template the PDB structure (PDB id code: 2FWT [37]). The resulting structure was used as the starting point for the MD simulations.

All MD simulations were performed using the GROMACS software package [41] version 4.5. The protein was put at the center of a dodecahedral box filled with single point charge (SPC) water molecules [42]. The ionic strength of the simulated protein solution was chosen in order to mimic typical experimental ionic conditions. To do so, we added to the simulation box the number of counter-ions neutralizing the protein net charge (9 to 11 sodium ions added, according to the different reduction states under investigation). This corresponds roughly to a $[\mathrm{Na}^+] \cong 0.15\text{-}0.20$ M. The protonation state of ionizable amino acid side chains was chosen according to the protonation state at pH = 7. The Gromos96 (53a6 version) force field parameters [43] were adopted for the protein and the heme in its reduced form. The atomic partial charges for the oxidized heme and the missing parameters describing axial and covalent bonds between the protein and the heme were taken from Ref.[38].

A standard protocol was adopted for initiating the simulations. Firstly, a two-step minimization protocol in solvent was performed using the conjugate gradient method. The first minimization involved only the water and the counter ions, with the coordinates of the protein held fixed, while the second minimization was performed on both the protein and the solvent molecules. The temperature of the systems was then gradually increased from 50 to 300 K in a single MD run of 100 ps. The trajectories were then propagated for 100 ns in the NVT ensemble at 300 K. The temperature was kept constant by the isokinetic temperature coupling [44]. All bond lengths were constrained using the LINCS algorithm [45] and an integration step of 2.0 fs was used. Long-range electrostatics was computed by the Particle Mesh Ewald (PME) method [46], with a grid spacing of 0.12 nm combined with a fourth-order cubic interpolation and a real space cutoff of 0.9 nm.

As the DHC protein has two heme sites (N-terminal heme and C-terminal heme, with His28- His51, and His106-His128 as axial ligands, respectively), four different redox states of the hemes were considered: N-terminal heme reduced and C-terminal heme oxidized (Cter:OX-Nter:RED), N-terminal heme oxidized and C-terminal heme reduced (Cter:RED-Nter:OX), both hemes reduced (Cter:RED-Nter:RED) and both hemes oxidized (Cter:OX-Nter:OX).

In order to calculate the reduction potential associated to each of the four transitions (steps A1, A2, B1 and B2, see results and Discussion section), the following strategy was used. For each OX → RED transition, four simulations were performed: two with a suitable number of counter ions neutralizing the starting state (one simulation with the considered heme in its reduced state and the other in its oxidized state) and the other two with a suitable number of counter ions neutralizing the end state (one simulation with the considered heme in its reduced state and the other in its oxidized state). The reduction potential was, then, evaluated by averaging the reduction potential calculated with 5.2 on each of the two different ionic conditions. A summary of the different MD simulations performed is presented in Table 5.1.

	Cter:OX-Nter:OX (total charge=-9)	Cter:RED-Nter:OX (total charge=-10)	Cter:OX-Nter:RED (total charge=-10)	Cter:RED-Nter:RED (total charge=-11)
simulation with 9 Na^+	√	√	√	
simulation with 10 Na^+	√	√	√	√
simulation with 11 Na^+		√	√	√

Table 5.1: Overview of the MD simulations performed in the present work.

5.2.3 Quantum chemical calculations

In order to apply the PMM/MD procedure for the calculation of the redox properties it is necessary to preselect the QC, i.e. the subportion of the simulated system to be treated explicitly at the electronic level. In the present case the atoms of the prosthetic group and of the sidechains of the axial ligands, i.e. the sidechains of the two histidines, were selected as QC. Then, quantum chemical calculations are performed on the isolated QC in order to obtain unperturbed energies and related properties for the ground state and a few excited states (see below) which are necessary to build the Hamiltonian matrix to be diagonalized at each MD frame and yielding the instantaneous perturbed electronic ground-state energy (as mentioned above).

For obtaining the unperturbed ground-state geometry, energy and related properties of the QC for both its redox states, i.e. with Fe^{+2} and Fe^{+3}, quantum chemical calculations were performed on the isolated QC at the Time Dependent Density Functional Theory (TD-DFT) level with Becke's three parameters exchange and Lee, Yang, and Parr correlation functionals (B3LYP). The atomic basis sets were as follows: (i) for the iron atom we used the LANL2DZ effective core potential for the inner electrons and a double Gaussian basis set of (5S,5P,5D)/[3S,3P,2D] quality for the valence electrons [47]; (ii) for the hydrogen, carbon, nitrogen and oxygen atoms we used a standard 6-31+G(d) [48] Gaussian basis set. The first twelve unperturbed excited electronic (vertical) states were obtained on the ground state geometry using TD-DFT calculations for both redox states. Although in general this level of theory might not provide a fully correct description of electronic excited states, in the case of the heme group it has already proved to represent a good compromise between computational costs and chemical accuracy [49]. All quantum chemical calculations were carried out using the Gaussian03 package [50].

5.2.4 Electrostatic Potential Calculations

The electrostatic properties of DHC were investigated using the APBS software [51]. The .pqr input file required to run APBS were prepared using PDB2PQR [52, 53] on the DHC structure from the 2FWT Protein Data Bank entry [37]. The electrostatic potential was obtained by solving the Linearized Poisson Boltzmann Equation (LPBE) at 298.5 K, using dielectric constant values of 2 and 78.54 for protein (solute) and solvent, respectively.

5.2.5 Docking

Prediction of the interaction between DHC and SHP was performed using the freely available ZDOCK server [54, 55]. As input proteins, the crystal structures of DHC and SHP (2FWT.pdb [37] and 1DW0.pdb, chain A [56], respectively). The ZDOCK 3.0.2 was chosen, whose scoring function is composed by IFACE Statistical Potential, Shape Complementarity, and Electrostatics. Prediction of the most favorable DHC-SHP adduct was fully unbiased, i.e. no likely interaction patches on either partner was suggested during the job submission.

5.3 Results and Discussion

In the spectroelectrochemical experiment which led to the aforementioned -310 and -240 mV E^0 values, the two hemes are sequentially reduced: in the starting conditions, both C- and N-terminal hemes are oxidized (we will refer to this form as to Cter:OX-Nter:OX). When the redox potential of the solution becomes increasingly more negative, one of the two hemes is reduced: this could lead either to the Cter:OX-Nter:RED or to Cter:RED-Nter:OX, according to which of the two redox centers features the most positive E^0. The spectroelectrochemical titration is then carried on until the second heme is also reduced, yielding the Cter:RED-Nter:RED state. A scheme depicting the two possible scenarios describing the experimental process is provided in 5.1. The overall reduction process could either occur via PATH A (light yellow), through steps A1 and A2, or following alternative PATH B (light blue). In order to assign each experimentally obtained E^0 value to the corresponding heme reduction, we performed 100 ns long MD simulations in the four redox-state ensembles, and applied the PMM theoretical framework to calculate the (Helmoltz) free-energy change upon reduction ΔA^0 for all the possible one-electron reduction processes (steps A1, A2, B1 and B2, respectively), as depicted in figure 5.1. The reduction potential E^0 was then calculated via 5.2.

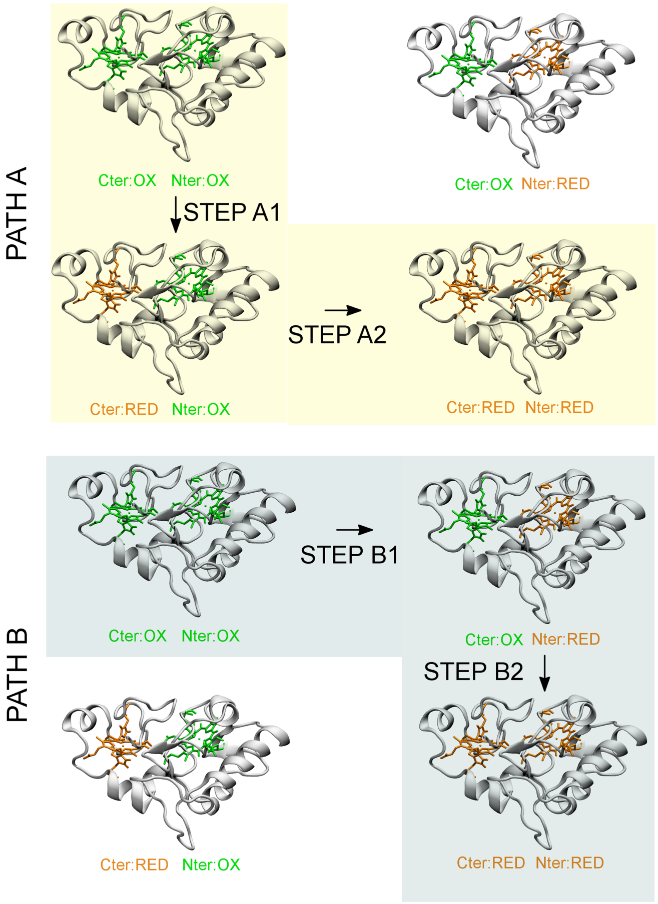


Figure 5.1: The two possible two-steps-one-electron paths for the reduction of Rb. sphaeroides DHC in the spectroelectrochemical experiment described in ref. [37]. The image was prepared using the VMD software [57].

The calculated ΔA^0, along with their corresponding E^0, are shown in figure 5.2. Please note that calculations provide absolute E^0 values, which were subsequently reported as relative to SHE, by substracting the IUPAC recommended 4.420 V value for the hydrogen semireaction[58, 59]. The comparison between the ΔA^0 for the Cter:OX-Nter:OX → Cter:OX-Nter:RED (step B1) and Cter:OX-Nter:OX → Cter:RED-Nter:OX (step A1) processes (-395 kJ/mol vs -401 kJ/mol, respectively) indicates that the reduction of the C-terminal heme is thermodynamically favored. We then estimated ΔA^0 for the second reduction, corresponding to the Cter:RED-Nter:OX → Cter:RED-Nter:RED process (step A2), to be -394 kJ/mol. Nevertheless, an additional factor suggests that path A is favored over path B. The ΔA^0 for step B2 is significantly more negative than that for B1, indicating that the Cter:OX-Nter:RED → Cter:RED-Nter:RED reduction (step B2) is thermodynamically favored over B1, and that the reduction potential for B2 should be even more positive than that for step B1. If that were the case, this would result experimentally in a simultaneous reduction of both hemes occurring at the same E value, corresponding to the E^0 of the Cter:OX-Nter:OX → Cter:OX-Nter:RED step. Such a process would therefore involve the simultaneous ET of two electrons, one for each heme, in disagreement with the two-times-one-electron Nernst process observed experimentally [37] and with cyclic voltammetry studies performed on a homologous diheme bacterial cytochrome [36], showing the occurrence of two distinct one-electron sequential reduction processes. Our results therefore suggest that path A is the most likely description of the spectroelectrochemical experiment, so that the -240 and -310 mV E^0 values can be unambiguously assigned to the subsequent reduction of C- and N-terminal heme iron, respectively.

According to our calculations, the difference in E^0 between the two hemes is 72 mV, in strikingly good agreement with the 70 mV difference obtained experimentally. But even most importantly, this approach yields calculated absolute E^0 values (-0.264 and -0.336 V see figure 5.2) very similar to the experimental values.

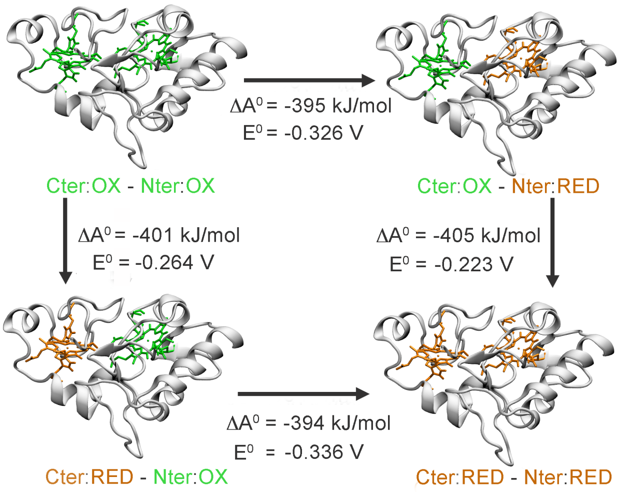


Figure 5.2: Calculated ΔA^0 and E^0 for the four possible investigated reduction processes. The standard error associated to the calculated ΔA^0 and to the ΔE^0 (± 2 kJ/mol and ± 0.020 V, respectively) was obtained by averaging the estimated standard errors of the reduction potentials for all possible investigated reduction processes.

It is worth stressing that the design of the computational experiment is a faithful reproduction of the electrochemical event, as reduction of the second heme center is modelled considering the other heme in its reduced state.

Once the agreement between predicted and experimental E^0 values has been assessed, we wanted to check whether the assignment of the high- and low-potential heme to the C-terminal and N-terminal center, respectively, is supported by the molecular features of the two heme environments. To this end, we first mapped the presence of charged residues in the surroundings (within 0.7 nm of any atom of the heme group) of N- and C-terminal centers. The results are shown in figure 5.3, panel A. Positively and negatively charged residues prevail in the environment of the C- and N-terminal heme center, respectively. The spatial distribution of charged residues clearly affects the surface electrostatic potential of DHC, whose distribution is such that a wide region of negative potential surrounds the N-terminal lobe, while the opposite is observed for the C-terminal lobe (see figure 5.3, panel B). In the latter case, the positive electrostatic potential is expected to stabilize the less positively charged ferrous heme over the ferric state. This effect is thereby consistent with the higher E^0 value of the C-terminal heme.

Besides electrostatics, another key factor affecting the redox thermodynamics of heme proteins is the exposure of the heme center to the solvent[60, 61, 12]. A larger solvent accessibility is known to stabilize the oxidized over the reduced form[62, 63, 64], whereas the effect of water surrounding the negatively charged propionates is opposite [38, 65, 66, 67]. We calculated the SASA values of the porphyrin ring ($<\text{SASA}>_{porph}$) and of the propionate groups only ($<\text{SASA}>_{prop}$) for both N- and C-terminal hemes, in the Cter:OX-Nter:OX ensemble. While the $<\text{SASA}>_{porph}$ values do not differ significantly for the two redox centres (0.955 ± 0.149 vs 0.837 ± 0.017 nm^2 for C- and N-terminal hemes, respectively), the propionate groups of the C-terminal heme are significantly more exposed than those of the N-terminal one ($<\text{SASA}>_{prop/Cter} = 1.257 \pm 0.113$ nm^2, $<\text{SASA}>_{prop/Nter} = 0.947 \pm 0.040$ nm^2). Comparison of the calculated SASA values therefore further supports our finding that the C-terminal heme features a more positive E^0 value.

The assignment of the reduction potential values to each individual heme allows hypotheses to be put forward concerning the ET processes involving the DHC from *Rb. sphaeroides*. Diheme cytochromes most probably work *in vivo* as electron shuttles carrying one electron at a time [68, 69, 70]. The most reasonable mechanism for ET in the small system under investigation is the following: the membrane embedded chimera complex transfers one

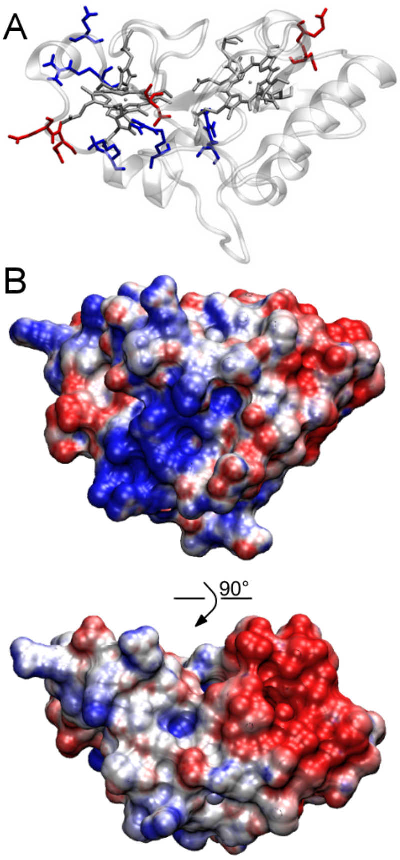


Figure 5.3: A) The distribution of positively (blue licorice) and negatively (red) charged residues within 0.7 nm from each heme of Rb. sphaeroides DHC. B) Electrostatic potential mapped on the molecular surface of Rb. sphaeroides DHC. Potentials less than -5 kT/e are colored in red, and those greater than + 5 kT/e are depicted in blue.

electron to the DHC heme featuring the most negative E^0 (the N-terminal one), through a kinetically favored step. The electron is then transferred to C-ter heme by internal ET, which is thermodynamically favored due to the -6 kJ/mol Δ E^0 between the Cter:OX-Nter:RED and the Cter:RED-Nter:OX states, and finally accepted by SHP, which features the more positive reduction potential. We tested this hypothesis by performing blind docking predictions of interaction between DHC and SHP, using the Fast Fourier Transform based ZDOCK server [55, 54]. In the predicted structure of the DHC/SHP adduct featuring the highest score (see figure 5.4), the relative orientation of the two partners is such that the C-terminal lobe of DHC is in direct contact with SHP. The same holds for four out of the top five predictions performed by the docking software. Although this is a preliminary result on the interaction between the two proteins, it strengthens the assignment of the more positive E^0 value to the C-terminal heme of DHC. The relative orientation of DHC and SHP, as predicted by the docking calculations, is such that the contact region in DHC is mainly composed by residues 85 to 88 and 100to 109. For SHP, the residue involved in the interaction are 56 to 64 and those forming the α-helix spanning residues 77 to 91. The binding regions of both DHC and SHP feature the presence of several ionizable amino acid side chains, in agreement with the experimental finding that SHP and DHC bind effectively only at low ionic strength, suggesting that their interaction is electrostatic in nature[37].

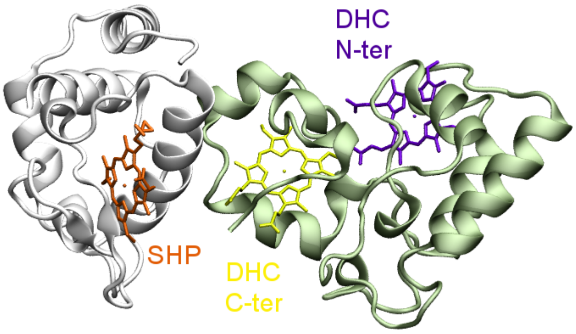


Figure 5.4: The structure of the adduct between DHC and SHP, as predicted by the ZDOCK web server using as input the corresponding crystal structures (2FWT.pdb [37] and 1DW0.pdb[56], respectively.)

We have presented a computational approach which proves to be a robust tool for assigning reduction potentials in multiheme proteins, and to investigate the thermodynamics of the electron flow. This approach was employed to successfully assign the reduction potentials previously obtained by spectroelectrochemical means to each heme group in the diheme cytochrome *c* from *Rb. sphaeroides*. Although we benchmarked our method against a system featuring two hemes, this approach can be extended to more complex systems bearing multiple redox centres. We also show that using PMM/MD calculations to estimate the E^0 of each redox centre in multiheme species allows insight to be gained into the intermolecular interactions within complex ET systems.

References

[1] C. M. Paquete and R. O. Louro. In: *Accounts Chem. Res.* 47.1 (2014), pp. 56–65.

[2] K. D. Bewley et al. In: *Biochim. Biophys. Acta* 1827 (2013), pp. 938–948.

[3] S. Sharma, G. Cavallaro, and A. Rosato. In: *J. Biol. Inorg. Chem.* 15 (2010), pp. 559 –571.

[4] C. M. Paquete et al. In: *Biochim. Biophys. Acta* 1767 (2007), pp. 1169–1179.

[5] C. M. Paquete and R. O. Louro. In: *Dalton T.* 39 (2010), pp. 4259–4266.

[6] B. M. Fonseca et al. In: *FEBS Lett.* 586 (2012), pp. 504–509.

[7] P. O. Quintas et al. In: *Biochim. Biophys. Acta* 1827 (2013), pp. 745–750.

[8] X. Qian et al. In: *Biochim. Biophys. Acta* 1807 (2011), pp. 404–412.

[9] M. J. Edwards et al. In: *Biochem. Soc. T.* 40 (2012), pp. 1181–1185.

[10] D. R. Lovley. In: *Biochem. Soc. T.* 40 (2012), pp. 1186–1190.

[11] J. A. Gralnick. In: *Biochem. Soc. T.* 40 (2012), pp. 1178–1180.

[12] J. Liu et al. In: *Chem. Rev.* 114 (2014), pp. 4366–4469.

[13] S. J. Marritt et al. In: *Biochem. J.* 444 (2012), pp. 465–474.

[14] M. Breuer et al. In: *J. Am. Chem. Soc.* 134 (2012), pp. 9868–9871.

[15] M. Breuer, K. M. Rosso, and J. Blumberger. In: *P. Natl. Acad. Sci. U.S.A.* 111.2 (2014), pp. 611–616.

[16] N. F. Polizzi, S. S. Skourtis, and D. N. Beratan. In: *Faraday Discuss.* 155 (2012), pp. 43–61.

[17] N. S. Malvankar and D. R. Lovley. In: *Chemsuschem* 5 (2012), pp. 1039–1046.

[18] L. Morgado et al. In: *Biochem. Soc. T.* 40 (2012), pp. 1295–1301.

[19] B. M. Fonseca et al. In: *J. Biol. Inorg. Chem.* 14 (2009), pp. 375–385.

[20] A. Amadei, I. Daidone, and C. A. Bortolotti. In: *RSC Advances* 3.42 (2013), pp. 19657–19665.

[21] I. Daidone et al. In: *J. Phys. Chem. Lett.* 5 (2014), pp. 1534–1540.

[22] J. Blumberger. In: *Phys. Chem. Chem. Phys.* 10 (2008), pp. 5651–5667.

[23] B. S. Perrin Jr., S. Niu, and T. Ichiye. In: *J. Comp. Chem.* 34 (2013), pp. 576–582.

[24] J. Mao, K. Hauser, and M. R. Gunner. In: *Biochemistry* 42 (2003), pp. 9829–9840.

[25] P. Voigt and E. W. Knapp. In: *J. Biol. Chem.* 278 (2003), pp. 1993–2001.

[26] A. Galstyan, A. Robertazzi, and E. W. Knapp. In: *J. Am. Chem. Soc.* 134.17 (2012), pp. 7442–7449.

[27] A. P. Gámiz-Hernández et al. In: *ChemPhysChem* 11 (2010), pp. 1196–1206.

[28] M. Aschi et al. In: *Chem. Phys. Lett.* 344 (2001), pp. 374–380.

[29] A. Amadei, M. D'Alessandro, and M. Aschi. In: *J. Phys. Chem. B* 108 (2004), pp. 16250–16254.

[30] M. Aschi et al. In: *J. Phys. Org. Chem.* 19 (2006), pp. 518–530.

[31] C. Zazza et al. In: *J. Phys. Chem. B* 112 (2008), pp. 3184–3192.

[32] I. Daidone et al. In: *Chem. Phys. Lett.* 488 (2010), pp. 213–218.

[33] A. Amadei et al. In: *Curr. Opin. Struc. Biol.* 20 (2010), pp. 155–161.

[34] I. Vandenberghe et al. In: *Biochemistry* 37 (1998), pp. 13075–13081.

[35] T. E. Meyer, J. A. Kyndt, and M. A. Cusanovich. In: *BMC Biochem.* 11 (2010), p. 24.

[36] G. Di Rocco et al. In: *J. Biol. Inorg. Chem.* 16 (2011), pp. 461–471.

[37] H. R. Gibson et al. In: *Biochemistry* 45 (2006), pp. 6363–6371.

[38] C. A. Bortolotti et al. In: *J. Am. Chem. Soc.* 134 (2012), pp. 13670–13678.

[39] I. Muegge et al. In: *J. Phys. Chem. B* 101 (1997), pp. 825–836.

[40] B. K. Kuntal, P. Aparoy, and P. Reddanna. In: *BMC Res. Notes* 3 (2010), pp. 226–230.

[41] H. J. C. Berendsen, D. van der Spoel, and D. van Drunen. In: *Comp. Phys. Comm.* 95 (1995), pp. 43–56.

[42] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. In: *J. Phys. Chem.* 91 (1987), pp. 6269–6271.

[43] W. F. van Gunsteren et al. Zürich, Switzerland: Hochschuleverlag AG an der ETH Zürich, 1996.

[44] D. Brown and J. H. R. Clarke. In: *Mol. Phys.* 51 (1984), pp. 1243–1252.

[45] B. Hess et al. In: *J. Comput. Chem.* 18 (1997), pp. 1463–1472.

[46] T. Darden, D. York, and L. Pedersen. In: ().

[47] P. J. Hay and W. R. Wadt. In: *J. Chem. Phys.* 82 (1985), pp. 299–310.

[48] R. Krishnan et al. In: *J. Chem. Phys.* 72 (1980), pp. 650–654.

[49] M. Aschi et al. In: *J. Comput. Chem.* 25 (2004), pp. 974–984.

[50] M. J. Frisch *et al.* Gaussian, Inc., Wallingford CT, 2004.

[51] N. A. Baker et al. In: *P. Natl. Acad. Sci. U.S.A.* 98 (2001), pp. 10037–10041.

[52] T. J. Dolinsky et al. In: *Nucleic Acids Res.* 32 (2004), W665–W667.

[53] T. J. Dolinsky et al. In: *Nucleic Acids Res.* 35 (2007), W522–W525.

[54] B. G. Pierce et al. In: *Bioinformatics* 30 (2014), pp. 1771–1773.

[55] B.G. Pierce, Y. Hourai, and Z. Weng. In: *PLoS ONE* 6 (2011), e24657.

[56] D. Leys et al. In: *J. Biol. Chem.* 275 (2000), pp. 16050–16056.

[57] W. Humphrey, A. Dalke, and K. Schulten. In: *J. Mol. Graphics* 14 (1996), pp. 33–38.

[58] W. R. Fawcett. In: *Langmuir* 24 (2008), pp. 9868–9875.

[59] A.A. Isse and A. Gennaro. In: *J. Phys. Chem. B* 114 (2010), pp. 7894–7899.

[60] S. Casalini et al. In: *J. Phys. Chem. B* 114 (2010), pp. 1698–1706.

[61] L. Paltrinieri et al. In: *J. Phys. Chem. Letters* 4 (2013), pp. 710–715.

[62] F. A. Tezcan, J. R. Winkler, and H. B. Gray. In: *J. Am. Chem. Soc.* 120.51 (1998), pp. 13383–13388.

[63] A. Fantuzzi et al. In: *Biochemistry* 41.27 (2002), pp. 8718–8724.

[64] S. Monari et al. In: *Electrochim. Acta* 56 (2011), pp. 6925–6931.

[65] A. Dolla et al. In: *Biochimie* 76 (1994), pp. 471–479.

[66] M. Unno et al. In: *J. Biol. Chem.* 271 (1996), pp. 17869–17874.

[67] G. R. Moore et al. In: *Biochim. Biophys. Acta* 829 (1985), pp. 83–96.

[68] H. Chang et al. In: *Biochemistry* 49 (2010), pp. 7494–7503.

[69] Q. Chi et al. In: *J. Phys.: Condens. Matter* 20 (2008), p. 374124.

[70] D. Heitmann and O. Einsle. In: *Biochemistry* 44 (2005), pp. 12411–12419.

6 Six-coordinate globins: are they electron transfer proteins?

6.1 Globins

6.1.1 Neuroglobin

Globins are relatively small heme proteins, characterized by a common fold formed by 5 to 8 α-helices and containing a single heme *b*, that exert several biological functions[1, 2, 3]. The most famous members of the globin family are the well known and thoroughly studied hemoglobin and myoglobin, whose biological functions are related to the transport and storage of molecular oxygen. Both these proteins can indeed bind exogenous ligands as O_2 at the sixth axial position of the ferrous heme, whereas in the absence of an external ligand they feature a five-coordinate heme[2]. The iron atom in heme *b* can be found in both the oxidation states, ferrous (Fe^{2+}) and ferric (Fe^{3+}). Its four equatorial binding sites are occupied by the heme pyrrole nitrogens, and one of the axial ligand is invariably one histidine residue, known as proximal histidine[2].

About fifteen years ago, new globin proteins different from the known hemoglobin and myoglobin and present in almost all kingdoms of life, from vertebrate to plants and bacteria, were discovered[4, 1, 2, 5]. Among these new proteins, of particular interest is a class of globins containing a six-coordinate heme group[1, 6, 7]. Neuroglobin (NGB), a six-coordinate globin which is mostly localized in the nervous tissues[1, 8, 2], is one of the most interesting members of this class. In the last years, neuroglobin orthologues were found in different organisms, from mammalians, to avians, fishes and amphibians[9].

The structure of NGB is similar to that of myoglobin, although the two proteins share less than 30% identity[10]. Neuroglobin is a monomer of about 150 residues, depending from the species, with a molecular mass of 17 kDa[1]. It features the typical globin fold, with 8 α-helices arranged in a $“three-over-three''$ α-helical sandwich[11, 3] (see figure 6.1).

As customary for six-coordinate globins, a histidine serves as sitxh axial

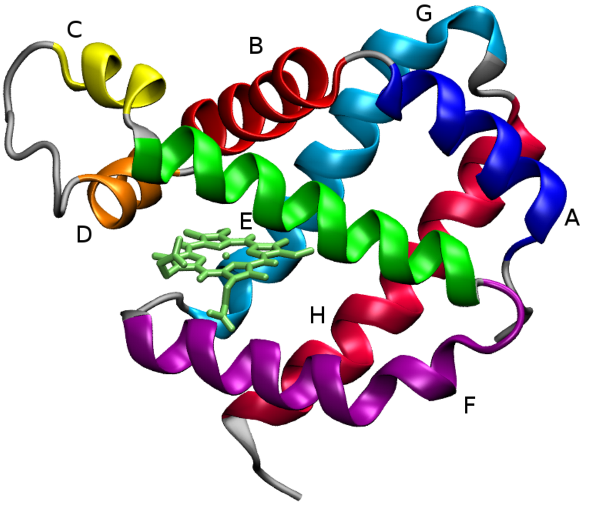


Figure 6.1: Three-dimensional structure of human neuroglobin (PBD ID code: 4MPM[12]) with the 8 α-helices labeled from A to H, in cartoon representation. The heme site is represented as licorice. The image was prepared using the VMD software[13].

ligand, known as "distal" histidine[8]. As it is common in globin proteins to design the α-helices with progressive letters, the distal histidine, whose position is largerly conserved, is referred as HisE7, where E7 indicates the seventh residue in the E-helix, whereas the proximal histidine is labeled HisF8 (as the eighth residue in the F-helix). In the absence of hexogenous ligands, the distal histidine is bound to the heme iron atom both in the ferrous and in the ferric protein, thus, to bind to neuroglobin exogenous ligands must compete with the distal histidine.[14, 8, 2, 15] Despite being one of the most deeply investigated six-coordinate globins so far, the physiological role of NGB is not completely clear yet. Various functions for this protein have been proposed so far[16, 17]: some studies assign to NGB a myoglobin-like role, suggesting that it could act to supply O_2 to the nervous cells, protecting them in case of hypoxia and ischemia[10, 18, 19]. Other potential physiological roles suggested for NGB include the detoxification of damaging surplus of NO to NO_3 or the production of NO from NO_2 for signalling functions[20, 21, 22, 23, 24]. Moreover, NGB may be involved in scavenging dangerous reactive oxygen or nitrogen species (ROS and RNS, respectively)[18, 25], or exerts a intracellular signalling role involving the binding of NGB to the α subunit of heterotrimeric G proteins, inhibiting the dissociation of GDP from Gα and triggering the release of G$\beta\gamma$, thus protecting the cell from apoptosis[26, 27, 28], or can partecipate to redox processes, interacting with cytochrome *c* to limit apoptosis in cytosol[29, 30, 31, 32, 33, 34].

The ferrous neuroglobin is able to bind O_2, CO and NO[8, 2, 15, 35, 36], whereas the ferric protein is able to bind only CN^-[37]. Although neuroglobin displays an high rate of binding and a low rate of dissociation for O_2, CO, and NO, thus indicating an high intrinsic affinity for these ligands, the rate limiting step in the binding reaction invloves the dissociation of the distal histidine[8, 15, 35], suggesting that *in vivo* the binding of the above ligands is rather low. These results, together with the high rate of auto-oxidation[8] suggest that neuroglobin would not be very efficient in the storage and transport of oxygen or other ligands, which is the physiological role of standard globins (as hemoglobin or myoglobin).

Several neuroglobins feature three conserved cysteine residues (in human neuroglobin CYS46, CYS55 and CYS120)[38, 39, 19, 40, 41]. In human neuroglobin, CYS46 and CYS55 form an intramolecular disulfide bridge[38, 19, 40]. The breaking of this S-S bond or the mutation of the cysteine with other residues decrease the affinity for O_2 by an order of magnitude and, in the absence of the disulfide bridge, the dissociation rate of the distal histidine decreases of a factor of 10[38]. These results suggest that the affinity for O_2 is directly related to the redox state of the cell[38, 39]. Experimental and computational studies hypothsize that the intramolecular disulfide bridge can modulate the functionality of human neuroglobin[42, 39, 19, 40, 41, 24, 43, 44], suggesting that the presence of the S-S bond favors the exogenous ligand binding, increasing the dissociation rate of the distal histidine.

Among all the putative functions of neuroglobin, one of the most intriguing is its interaction with cytochrome *c*, which would assign neuroglobin a role in preventing apoptosis. According to several studies[29, 31, 32, 17, 33, 45, 34] it is reasonable to consider neuroglobin as an electron transfer protein, which can react with cytochrome *c* released in the cytosol during apoptosis. Ferrous neuroglobin can reduce ferric cytochrome *c*, which is involved in the apoptotic process, while the ferrous cytochrome *c* is apparently inactive in the initiation of apoptosis[46, 47]. One of the evidences for this possible role for neuroglobin is the very high rate of electron transfer between NGB and cytochrome *c* ($k = 2 \times 10^7\ M^{-1}\ s^{-1}$)[29], which is comparable to those of the fastest inter-protein redox reactions, such as that between cytochrome *c* and cytochrome *c* oxidase, ($k = 4 \times 10^6$ - $1 \times 10^7\ M^{-1}\ s^{-1}$)[48].

6.1.2 GLB-6 from *C. elegans*

Recently, other six-coordinate globin proteins have been found in the nematode *Caenorhabditis elegans*[49, 50, 51]. The function of these proteins is still discussed, but it was recently demonstrated that some of these globins may play a key role in a gas sensory signaling pathway in *C. elegans*. For example, the neural globin GLB-5 serves as an O_2 sensor in the gas sensory neurons and controls neuronal responses of the nematode to changing O_2 concentration[52, 53, 54, 55]. Furthermore, GLB-5 rapidly oxidizes upon reversible binding of molecular oxygen, which is known to be an important feature for a signailing protein[53]. Besides these gas sensory globins, *C. elegans* possesses a neural globin, GLB-6, which features unusual ligand binding and redox properties compared to classical globins[56]. In particular, ligand binding experiments showed that GLB-6 does not bind any of the typical exogenous ligands for globins, such as CO, NO and CN^-[56], which are widely known to form highly stable complexes with the heme. Furthermore, the reduction potential of GLB-6, measured using potentiometric redox titration[56], is -193 x± 2 mV vs SHE, which is sensibly lower compared to those of other six-coordinate bis-histidyl globins, such as human neuroglobin or human cytoglobin, which feature reduction potential values of -115 mV vs SHE and -28 mV vs SHE, respectively[6]. The structural features of GLB-6 can, at least in part, help to explain the different behavior of this protein. GLB-6 is formed by 161 amino acid residues, arranged in seven helices, corresponding to the A, B, C, E, F, G and H helices of the globin-like domain (see figure 6.2).

The heme group is axially coordinated by two histidine residues, as in human and murine neuroglobin[8, 1], human cytoglobin[57], nonsymbiotic rice hemoglobin[58, 59], and a bacterial globin-coupled sensor from *Geobacter sulfurreducens*[60]. Although this similarity, the GLB-6 heme domain differs from those of other bis-histidyl ligated globins, and this could help to explain the unusual properties of this protein. In particular, GLB-6 features an elongated E helix, and totally lacks a D helix. The E helix presents an extra helix loop of about four amino acid residues at the N-terminal side, and therefore, the distal histidine (His60) is positioned at the E11 position instead of E7 as in other globins. It is to note that the His60 is called "distal" only on the basis of sequence alignment with other globins. Due to the complete absence of D helix, GLB-6 presents a direct connection between C and E helices. It is possible that these structural features lead to a more rigid structure, imposing a severe restriction on the helical movement of the heme domain, thus partially inhibiting exogenous ligand binding. In addi-

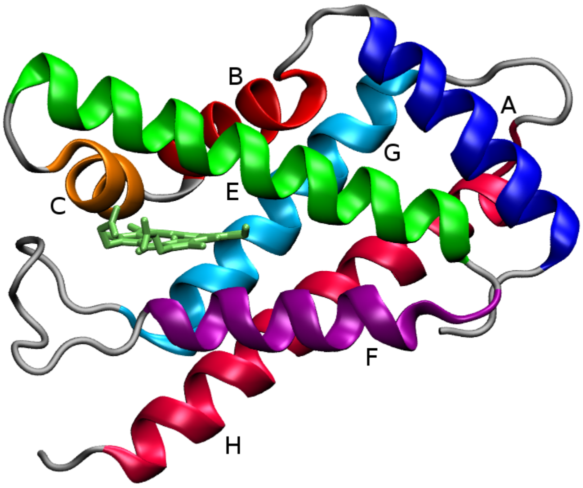


Figure 6.2: Three-dimensional structure of GLB-6 from C. elegans (PDB ID code: 3MVC[56]) with the 7 α-helix labeled from A to H, excluding the missing D helix, represented as cartoon. The heme site is represented as licorice. The image was prepared using the VMD software[13].

tion, GLB-6 has a shorter F helix compared to neuroglobin, and a longer F-G loop. The consequence is that the proximal histidine (His92) is found at the end of the helix, and it is more exposed to the solvent (i.e., to an hydrophilic environment) than in neuroglobin, whose heme is embedded in a hydrophobic pocket. This feature could explain the lower reduction potential of GLB-6 with respect to other globins, since the exposure to a more hydrophilic, aqueous environment thermodynamically favors the ferric state over the ferrous state of the heme[61]. Furthermore, the reduction potential of GLB-6 is lower than that of the $O_2/O_2{}^{\bullet}-$ redox couple, suggesting that GLB-6 may directly interact with molecular oxygen to generate superoxide, which is known to have an important role in several physiological responses[62, 63]. The above characteristics lead to the hypothesis that GLB-6 could function as a redox sensor[56] or electron transfer protein[45], rather than a gas storage or transporter protein, contrary to the most of globins[1, 2].

The aim of my investigation by means of Molecular Dynamics of human neuroglobin and GLB-6 from *C. elegans* is to verify their putative role as electron transfer proteins. Both proteins feature a strong bond of the distal histidine residue with the heme iron atom, which makes their active site structure rigid and less prone to undergo significant conformational changes upon reduction/oxidation processes. The latter features are typical of electron transfer proteins, such as cytochromes and cupredoxins[64, 65, 66, 67, 68, 69, 70, 71]. In particular, the alignment of three-dimensional structures of neuroglobin and GLB-6, despite the differences in the three-dimensional structures reported above, is quite good[56] (see figure 6.3). This feature allows to compare the structural and dynamic analysis performed on the two proteins, in order to verify the hypothesis of the existence of a family of six-coordinated globins functioning as electron transfer proteins[45].

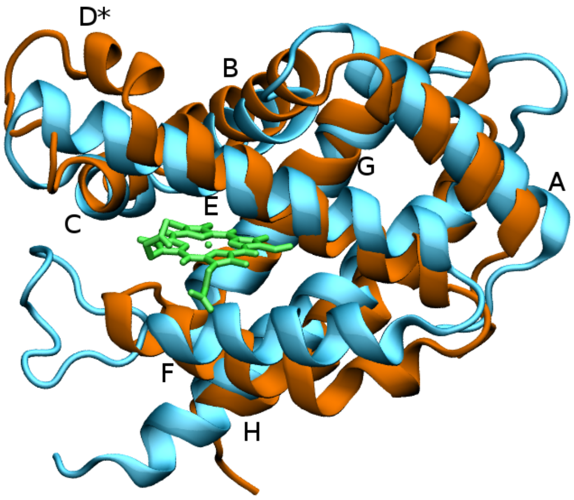


Figure 6.3: Overall alignment of three-dimensional structure of human neuroglobin and GLB-6 from C. elegans (PDB ID code: 4MPM[12] and 3MVC[56], respectively), with α-helices labeled from A to H. The D helix is marked with an asterisk to indicate that it is present only in the structure of neuroglobin. The image was prepared using the VMD software[13].

6.2 Molecular Dynamics simulations

The crystal structure of human neuroglobin (PDB ID code: 4MPM[12], chain B) and GLB-6 from *C. elegans* (PDB ID code: 3MVC[56], chain A) were used as the starting point for the MD simulations. The crystal structure of human neuroglobin lacks the first and the last two residues, while the crystal structure of GLB-6 lacks residues from 124 to 127. Both proteins are then modelled using the EasyModeller software (version 4.0)[72], using as input the corresponding FASTA sequence and as template the corresponding PDB structure, in order to obtain the two complete proteins. The GROMACS software package[73], version 4.5.5, was used for all the MD simulations. The proteins were put at the center of a dodecahedral box, that then was filled with single point charge (SPC) water molecules[74]. The appropriate number of counter-ions were added in order to neutralize the systems; in particular, six Na^+ ions were added to the box of neuroglobin, and 3 Cl^- ions were added to the box of GLB-6. The parameters for the proteins and the hemes in their reduced form were taken from the Gromos96 (53a6 version) force field, whereas the atomic partial charges and the missing parameters for the heme group and its axial ligands in their oxidized form were taken from ref.[75]. For all the simulations, a standard protocol was used: firstly, the systems energy was minimized in vacuum, using the steepest descent method; then, after the addition of the solvent and counter-ions, a two-step minimization protocol was performed with the conjugate gradient method. The first minimization held fixed the coordinates of the proteins, allowing only the water and counter-ions to move. The second minimization involves both the protein and solvent molecules. Then the temperature of the systems was gradually increased from 50 to 300 K in a 100 ps MD run, and was kept constant for all the simulation time by the isokinetic temperature coupling[76]. The simulation runs were performed in the NVT ensemble. The bond lengths were constrained with the LINCS algorithm[77] using an integration step of 2.0 fs. The Particle Mesh Ewald (PME) method[78] was used to compute the long-range electrostatics, with a grid spacing of 0.12 nm, a fourth-order cubic interpolation and a real space cut-off of 0.9 nm.

Depending on the redox state of the external environment, neuroglobin can feature an internal disulfide bridge between Cys46 and Cys55, which affects the dynamics of the protein, in particular the ability of the distal histidine to move away from the active site. Thus, I simulated both the ferrous and ferric forms of neuroglobin with and without the S-S bond(see figure 6.4).

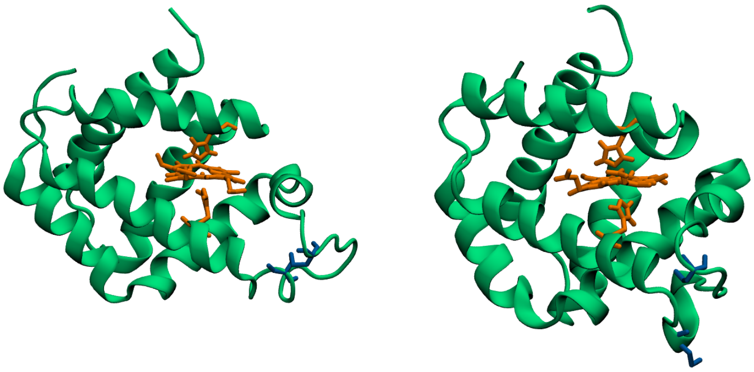


Figure 6.4: Three-dimensional cartoon structure of human neuroglobin (PDB ID code: 4MPM[12]) with disulfide bridge (left) and without disulfide bridge (right) between Cys46 and Cys55. The heme site, its axial ligands and the two cysteine residues involved in the disulfide bridge are represented as licorice. The image was prepared using the VMD software[13].

6.2.1 Stability of the simulations

In order to verify the stability of the MD simulations, some preliminary analysis have been performed. First of all, the root mean square deviation (RMSD) over time of the backbone atoms of the proteins was calculated, both for the reduced and oxidized state, using as a reference the starting structures. The RMSD trends over time are reported in figures 6.5, 6.6, 6.7 for neuroglobin and GLB-6; for the former it is showed the RMSD in the presence and in the absence of the disulfide bridge. The RMSD values have the typical trend of this kind of analysis: the systems tend to rapidly separate from their reference structures in the first ns of simulation, and then they reach an equilibrium condition over time. These results, together with their low mean values (see table 6.1), allow to assert that the obtained simulations are adequately stable, therefore they can be utilized to obtain reliable informations about the studied systems, with the expedient of removing the first ns of simulation before performing the analysis.

GLB-6 from *C. elegans*			
Property	X-ray	MD_{red}	MD_{ox}
RMSD (nm)	-	0.28 (0.03)	0.35 (0.04)
% α-helix	73.91	68.12 (6.49)	64.40 (5.98)
Human Neuroglobin with disulfide bridge			
Property	X-ray	MD_{red}	MD_{ox}
RMSD (nm)	-	0.38 (0.07)	0.30 (0.03)
% α-helix	82.12	70.83 (4.15)	67.28 (4.15)
Human Neuroglobin without disulfide bridge			
Property	X-ray	MD_{red}	MD_{ox}
RMSD (nm)	-	0.29 (0.03)	0.34 (0.03)
% α-helix	82.12	67.85 (3.85)	63.50 (4.65)

Table 6.1: Structural properties of the investigated proteins, for crystal structures and MD simulations in the reduced and oxidized ensembles. The standard deviations are given in brackets.

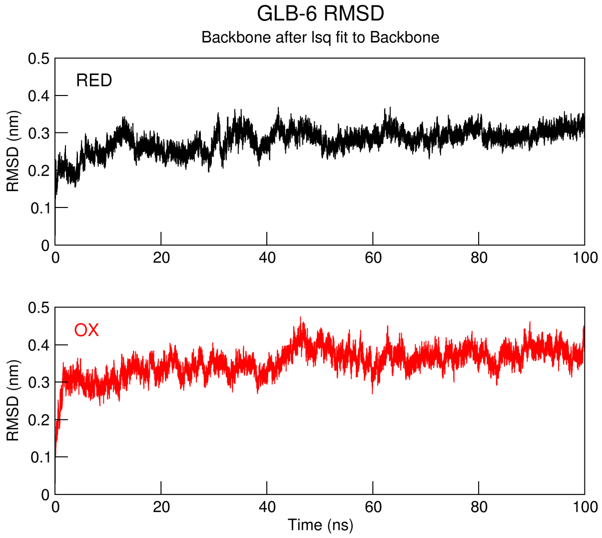


Figure 6.5: Calculated RMSD for GLB-6 from C. elegans in the reduced (top) and oxidized (bottom) ensembles.

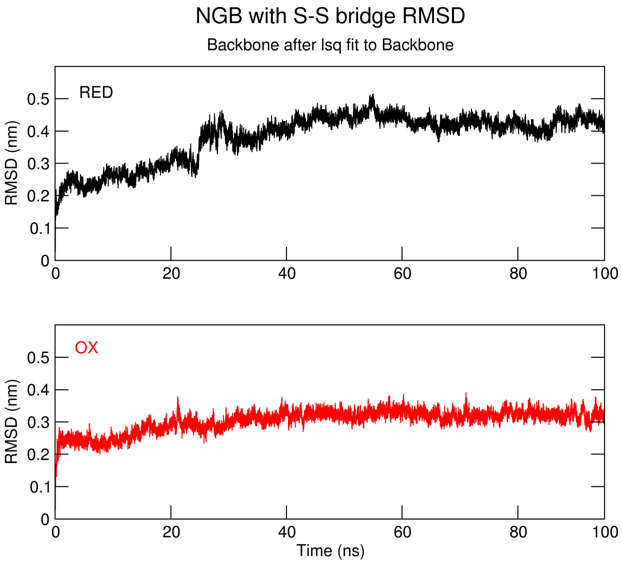


Figure 6.6: Calculated RMSD for human neuroglobin with disulfide bridge in the reduced (top) and oxidized (bottom) ensembles.

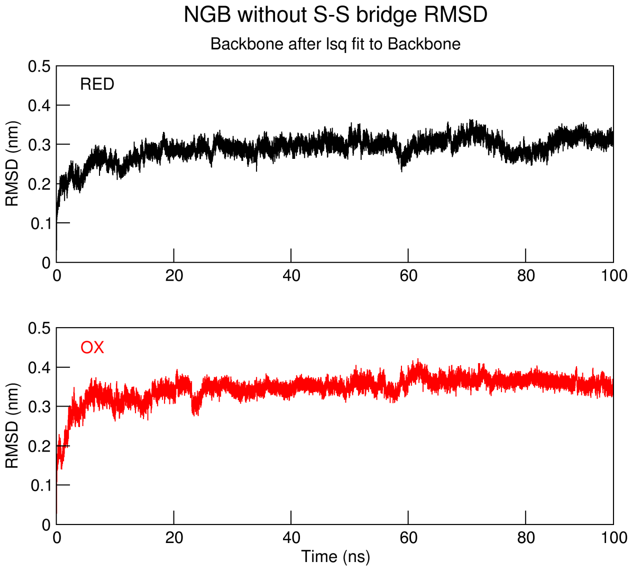


Figure 6.7: Calculated RMSD for human neuroglobin without disulfide bridge in the reduced (top) and oxidized (bottom) ensembles.

Subsequently, the root mean square fluctuations (RMSF) per residue were calculated for all the investigated systems, reported in figures 6.8, 6.9, 6.10. In general, the wider fluctuations involve the residues at the extremities of the amino acid sequence, and the residues forming the loops, whereas the α-helix regions are definitely less flexible. In particular, the regions featuring higher flexibility correspond to the loops surrounding the heme site, both in neuroglobin and in GLB-6. In neuroglobin they include residues from 45 to 55 for both the reduced and oxidized form with disulfide bridge, whereas in the absence of the disulfide bridge this region is a little less flexible. This is a rather unpredictable result, since disulfide bridges typically increase the rigidity of proteins, while their dissociation should enhance the flexibility, at least of the region containing the cysteine residues forming the bond. However, in the disulfide bridge-containing neuroglobin the distal histidine shows a lower affinity for the heme than in form without the disulfide bridge[38, 19, 40]. As a consequence, the detachment of the distal histidine and the binding to the heme iron atom of exogenous ligands, as O_2, is favored in the former species[38, 19, 40]. Our result agree with these findings, since a higher flexibility of the region containing the two cysteine residues involved in the formation of the disulfide bridge, which is positioned in the heme surrounding, can help the access of an external ligand to the heme distal cavity.

The larger fluctuations in GLB-6 corresponds to the regions spanning residues 45 to 52, 90 to 103 and 120 to 135.

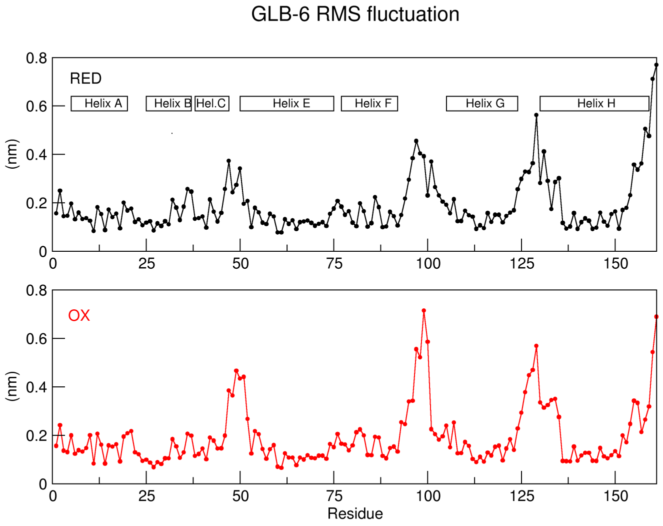


Figure 6.8: Calculated RMSF for GLB-6 from C. elegans in the reduced (top) and oxidized (bottom) ensembles.

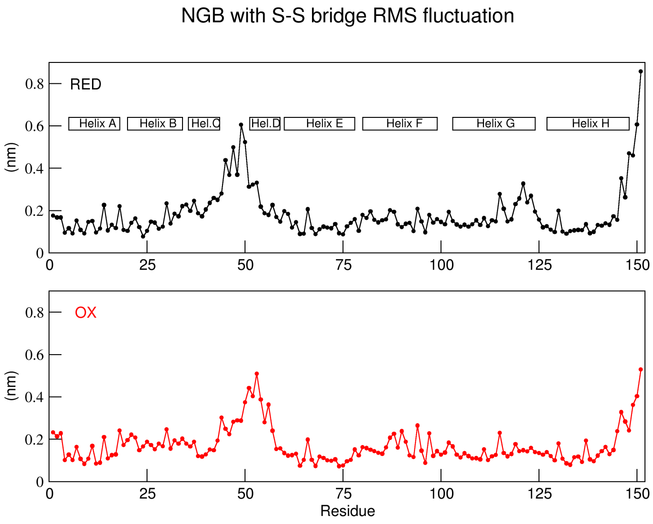


Figure 6.9: Calculated RMSF for human neuroglobin with disulfide bridge in the reduced (top) and oxidized (bottom) ensembles.

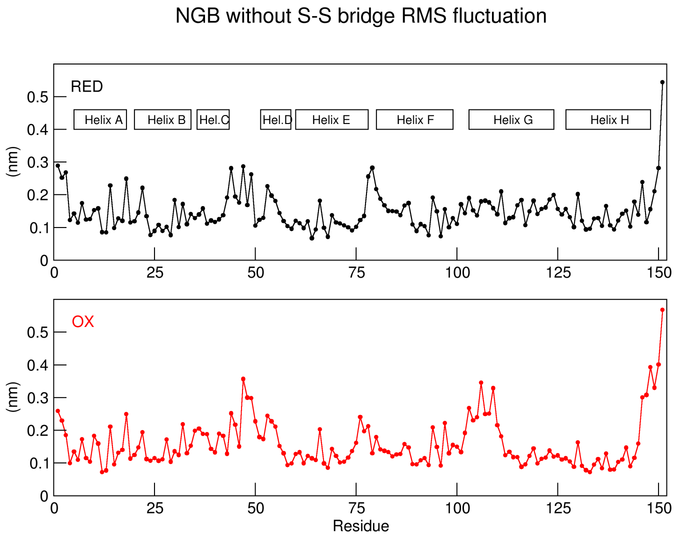


Figure 6.10: Calculated RMSF for human neuroglobin without disulfide bridge in the reduced (top) and oxidized (bottom) ensembles.

Finally, the α-helix contents, expressed as percentage, during the simulations has been calculated, and it has been compared with the values obtained for the crystal structures. This calculation was performed using the DSSP program[79] implemented in GROMACS. The results are reported in table 6.1, and they show how all the α-helices are mostly conserved during the simulations, further proving their stability.

6.3 Calculation of the reduction potential E^0

Since the aim of this work is to verify if the investigated globins can be considered as electron transfer proteins, the reliability of the calculation methods must be validate. In particular, the reduction potential of all the simulated species were calculated and compared with the corresponding experimental values, in order to test the reliability of the Perturbed Matrix Method (PMM)[80, 81, 82, 83, 84, 85] for the calculation and prevision of other redox properties of globins. The calculation of the reduction potential (E^0) for neuroglobin and GLB-6 was thus performed using the PMM approach on the MD simulations of the proteins in their reduced and oxidized state, following the procedure reported in ref.[75, 86, 87, 88]. The reduction potential derives from the calculated value of the Helmholtz free energy change (ΔA^0) related to the semi reaction occurring at the heme site ($Fe^{3+} + e^- \rightarrow Fe^{2+}$), which is associated with E^0 via $E^0 = -\Delta A^0/nF$, where F is the Faraday constant and n is the number of electrons involved in the process. ΔA^0 is calculated using the following expression:

$$\Delta A^0 \cong \frac{-kT \ln \left\langle e^{-\beta \Delta \mathcal{U}} \right\rangle_{ox} + kT \ln \left\langle e^{\beta \Delta \mathcal{U}} \right\rangle_{red}}{2} \tag{6.1}$$

In the above equation, $\Delta\mathcal{U} = \mathcal{U}_{red} - \mathcal{U}_{ox}$, where $\mathcal{U}_{red}$ and $\mathcal{U}_{ox}$ are the perturbed electronic ground-state energies of the reduced and oxidized state, respectively, of the quantum center (i.e., the portion of the system involved in the redox process, in this case defined as the heme group and the side chain of the axial ligands) and it is averaged in the reduced and oxidized ensemble as indicated by the angle brackets[75, 87, 88, 89].

The obtained E^0 values for single ensemble (reduced and oxidized) and as the average on the two ensembles, are reported in table 6.2. It is interesting to note that the difference between the estimation of the reduction potential in the two ensemble is very low compared to the absolute E^0 value. However, this result indicates that each simulation ensemble is affected by a systematic error, arising from the fact that not all the crucial regions of the phase space have been accessible. Nevertheless, if the two values obtained on the two ensemble are considered as the maximum and minimum limit of the estimation of the reduction potential, and if we consider that the standard errors on the two ensembles are similar, we can assume that averaging the values of E^0 on the two ensembles, the systematic errors are erased. Thus, we can consider the calculated value of E^0 as a value not affected by systematic errors.

The calculated E^0 values have been compared to the corresponding ex-

	$\langle E^0_{calc}\rangle_{red}$ (V)	$\langle E^0_{calc}\rangle_{ox}$ (V)	E^0_{calc} (V)	$E^0_{exp}{}^b$ (V)
GLB-6	4.208	4.056	4.132	4.227
NGB with S-S bridge	4.263	4.210	4.236	4.305*
NGB without S-S bridge	4.375	4.192	4.283	4.305*

Table 6.2: Comparison between the calculated and experimental E^0 values. The associated standard error for $\langle E^0_{calc}\rangle_{red}$ and $\langle E^0_{calc}\rangle_{ox}$ is 0.027 V and for E^0_{calc} is 0.040 V, obtained by averaging the estimated standard errors of the reduction potentials for all the simulations. ^{b} Experimental values are taken from ref[56]. Note that it is not reported if the experimental value of E^0 for human neuroglobin is referred to the form with or without the disulfide bridge, thus the same value is reported for the two forms

.

perimental values, as reported in table 6.2. It turns out that there is a good correspondence between calculated and experimental values, thus proving the reliability of the PMM method in the calculation and prevision of redox properties of metalloproteins. The slightly higher value obtained for neuroglobin without disulfide bridge compared to the one with it, although not impressive, could be in part justified considering the RMSF of the two proteins. Neuroglobin without the S-S bridge is slighlty less flexible with respect to the protein with the bridge, therefore it can be supposed that its active site is less exposed to the solvent, providing a higher E^0 value being exposed to a less hydrophilic environment.

6.4 Calculation of reorganization energy

The reorganization energy, λ, is the energy needed for the reactants to move from the equilibrium reactant geometry to the equilibrium product geometry, without the transfer of electrons[67]. It is therefore one of the most important parameters from which the rate constant of an electron transfer (ET) process depends. In fact, according to the semiclassical Marcus theory[90], the rate of a nonadiabatic ET reaction can be expressed as:

$$k_{ET} = \frac{2\pi}{\hbar} \frac{H_{DA}^2}{\sqrt{4\pi\lambda RT}} \, e^{\frac{-(\Delta G^0+\lambda)^2}{4\lambda RT}} \tag{6.2}$$

It is therefore obvious that the lower the reorganization energy, the higher the k_{ET}. For a metalloprotein involved in electron transfer reaction, typically the total reorganization energy has low values, and it can be conveniently divided in two terms, the inner-sphere reorganization energy, λ_{in} and the outer-sphere reorganization energy, λ_{out}[90]The first term is due to contributions from the active site, and it is related to the geometrical rearrangement of the active site. The second one takes into account the effects of the changes in the polarization of the surrounding protein matrix and the solvent upon reduction/oxidation processes[90, 64].Nevertheless, it is to note that the dominant factor in the reorganization energy values for metalloproteins is thought to be the outer-sphere contribution[91, 71, 64].

Since the reliability of MD simulations has been proved by their stability (see related paragraph) and the Perturbed Matrix Method has proved to be trustworthy for the calculation of redox properties of metalloproteins, thanks to the good results obtained for the calculation of the reduction potentials of the investigated globins, they were used for the calculation of the the reorganization energy for neuroglobin and GLB-6, following the procedure in ref.[86]. In this method, the reorganization energy is provided by the following equation:

$$\lambda = \frac{\lambda_{red} + \lambda_{ox}}{2} \tag{6.3}$$

where λ_{red} and λ_{ox} are defined as:

$$\lambda_{red} = \langle \Delta \mathcal{U} \rangle_{ox} - \Delta A^0 - kTln\frac{\sigma_{red}}{\sigma_{ox}} \tag{6.4}$$

$$\lambda_{ox} = -\langle \Delta \mathcal{U} \rangle_{red} + \Delta A^0 + kTln\frac{\sigma_{red}}{\sigma_{ox}} \tag{6.5}$$

where $\langle \Delta \mathcal{U} \rangle_{red}$ and $\langle \Delta \mathcal{U} \rangle_{ox}$ are the averaged perturbed electronic ground-state energies for the reduced and oxidized ensemble, respectively, and σ_{red}

and σ_{ox} are the standard deviation associated to the $\Delta\mathcal{U}_{red}$ and $\Delta\mathcal{U}_{ox}$ respectively.

The reorganization energy of the two proteins was also calculated with another procedure, the so-called *Energy Gap Fluctuations* method[92], that evaluates λ_{out} (i.e., the outer sphere reorganization energy) with a relation that connect it to the fluctuations of the energy gap η, where η=E$_{red}$ - E$_{ox}$:

$$\lambda_{out} = \frac{<\eta>_{ox} - <\eta>_{red}}{2} = \frac{<E_{red} - E_{ox}>_{ox} - <E_{red} - E_{ox}>_{red}}{2} \tag{6.6}$$

The calculated values for reorganization energies for both neuroglobin (with and without disulfide bridge) and GLB-6, obtained with both the PMM and the *Energy Gap Fluctuations* approaches, are reported in table 6.3. Two observations have to be made regarding these values. First of all, it can be noted that the values obtained with the PMM approach are invariably slightly higher than the ones obtained with the *Energy Gap Fluctuations* method, but this is due to the fact that the first method calculates the *total* reorganization energy, i.e., involving both λ_{in} and λ_{out}, whereas the second one provides *only* the λ_{out} value. Since the λ_{in} contribution is thought to be relatively small in electron transfer metalloproteins[93, 64] (for example, in the order of $\sim$ 0.15 eV for azurin[93], which is a typical model system for ET processes, thus obtaining a total λ value of $\sim$ 0.8 eV), our values are therefore in good agreement. Furthermore, the values obtained either with the PMM approach or the *Energy Gap Fluctuations* method are invariably slightly higher than the experimental total λ values reported for the most common electron transfer proteins (between 0.5 and 1.5 eV for cytochrome *c*[64, 68, 94, 95, 96] and between 0.6 and 0.8 eV for azurin[70, 97]), but this is most probably due to the use of a nonpolarizable force field, which introduce some overestimation (in the amount of $\sim$ 30%) of the reorganization energy values. In fact, the calculated λ_{out} values, obtained using nonpolarizable force fields, for cytochrome *c* is between 0.60 eV and 0.9 eV[64, 69, 89], and the one for azurin lie in the range between 0.8 eV to 1.0 eV[98, 71]. Therefore, the obtained values for neuroglobin and GLB-6 lie in the range of the typical calculated λ values for electron transfer proteins. These results, together with the fact that structural and dynamical analysis of both the investigated globins proved that they possesses a quite rigid and stable structures, which show only minimal changes in the oxidized form compared to the reduced one, can be used as evidence for their plausible electron transfer function.

	$\langle\Delta U\rangle_{red}$ (kJ/mol)^a	$\langle\Delta U\rangle_{ox}$ (kJ/mol)	ΔA^0 (kJ/mol)	λ_{PMM} (eV)	λ_{outEGF} (eV)
GLB-6	-497.66	-318.82	-398.73	0.92	0.82
NGB with S-S bridge	-487.27	-349.56	-408.76	0.70	0.65
NGB without S-S bridge	-496.95	-330.64	-413.28	0.88	0.75

Table 6.3: Redox thermodynamic properties obtained with the PMM method and reorganization energies obtained both from PMM method and Energy gap fluctuations (EGF) method for the investigated systems. The associated standard errors are 2.06 kJ/mol for $\langle\Delta U\rangle_{red}$ and $\langle\Delta U\rangle_{ox}$ and 2.63 kJ/mol for ΔA^0, and 0.02 eV for λ_{PMM} and λ_{outEGF}, obtained by averaging the associated standard errors for these parameters for all the simulations.

6.5 Effects of the disulfide bridge on human neuroglobin

To gain further informations on the effect(s) of the disulfide bridge on the dynamics of human neuroglobin, the investigation of the collective motions of the protein, by means of Principal Components Analysis (PCA), was performed. This technique allows to establish which are the principal motions controlling the dynamics of a protein, and can provides important correlations with some of the previous cited results. The PCA was performed on the MD simulations of the reduced and oxidized neuroglobin, either with and without the disulfide bridge, considering only the $C\alpha$ atoms. The residues positioned at the both ends of the amino acid sequence were not considered in the analysis, since their fluctuations result to be prevalent compared to those involving the central residues, which however are more interesting. Therefore, only residues between 4 and 148 were examined. In order to obtain the eigenvectors related to the motions of the proteins and the corresponding eigenvalues, the covariance matrix of the fluctuations of the atomic positions was constructed and diagonalized. Figure 6.11 shows the eigenvalues for the reduced and oxidized forms of neuroglobin with and without disulfide bridge arranged in descending order. As anticipated in the Methods section, from figure 6.11 we can observe that the majority of the collective motions is described by a relative low number of eigenvector. In particular, for all the considered simulations, the sum of the first ten eigenvalues (on a total number of over 400) corresponds to $\sim 70\%$ of the total value of the covariance matrix.

Therefore, the first ten eigenvectors for neuroglobin both with and without disulfide bridge, in the reduced and the oxidized ensemble have been studied. For each eigenvector it is possible to reveal the amino acid residues that mainly contribute to the corresponding eigenvalue. As an example, in

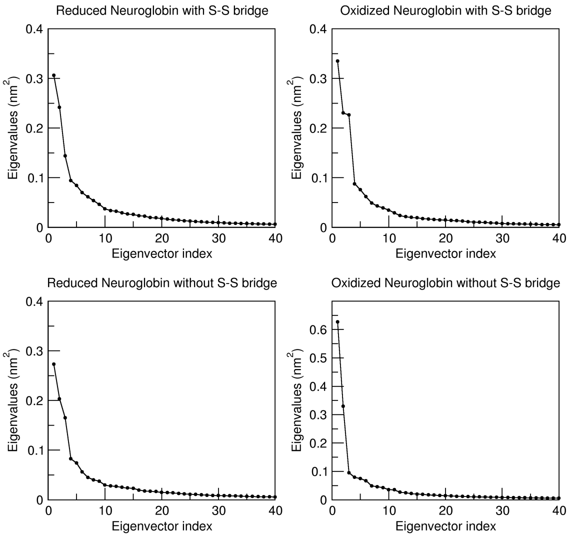


Figure 6.11: Graphical representation of eigenvalues obtained from PCA analysis on human neuroglobin with and without disulfide bridge in the reduced and oxidized ensemble.

figure 6.12 the amino acid residues mainly involved in the first mode for all the investigated species are reported. It appears that, the residues involved in the definition of the first mode of neuroglobin with the disulfide bridge are roughly the same in the reduced and the oxidized ensemble, whereas the analysis on neuroglobin without the disulfide bridge shows that the first mode involves different regions of the structure. This result, together with the values obtained from the calculation of the difference between the trace of the covariance matrix for the reduced and oxidized protein, with and without the disulfide bridge (table 6.4), can help to explain the lower calculated reorganization energy for neuroglobin with disulfide bridge compared to that for the protein without disulfide bridge. The small difference between the traces of covariance matrices for the same protein in the reduced and oxidized ensemble indicate that there is a low conformational rearrangement of the protein as a consequence of the electron transfer, and therefore, a lower reorganization energy.

	Trace MD_{red} (nm^2)	Trace MD_{ox} (nm^2)	Δ traces (red-ox) (nm^2)
Ngb with S-S bridge	1.829	1.761	0.068
Ngb without S-S bridge	1.625	2.056	0.431

Table 6.4: Traces of covariance matrices for human neuroglobin with and without disulfide bridge calculated in the reduced and oxidized ensembles, and their difference.

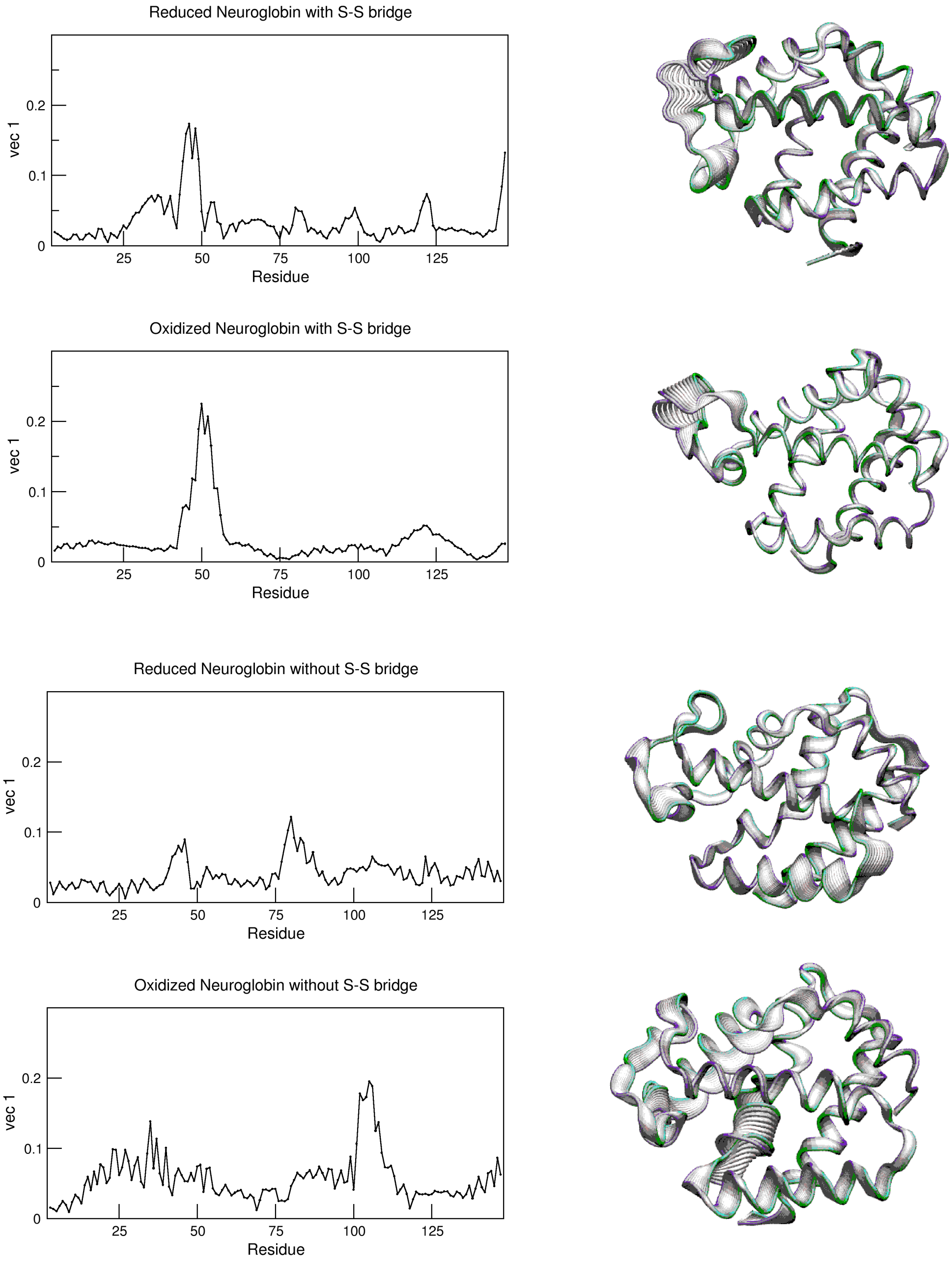


Figure 6.12: RMSF and three-dimensional representation of human neuroglobin with (top) and without (bottom) disulfide birdge in the reduced and oxidized ensembles, obtained for the first mode using PCA.

If we consider the extent of the fluctuations in the region of the disulfide bridge for neuroglobin with and without the disulfide bridge for the first mode, we can observe that the protein with the disulfide bridge displayed wider fluctuations, as in the total RMSF reported previously. This result is in agreement with several studies reporting that the access of exogenous ligands to the heme distal site and the replacement of the distal histidine is easier in neuroglobin with the disulfide bridge than in the protein without it[38, 19, 40]. Indeed, a higher flexibility of the protein can help this process.

6.6 Interaction between human neuroglobin and human cytochrome *c*

Considering the possible role of human neuroglobin as an electron tranfer protein, the MD simulation of its complex with one of its putative biological partner, i.e., the human cytochrome *c*[29, 31, 32, 17, 33, 45, 34] has been performed. Indeed, it is supposed that ferrous neuroglobin can reduce ferric cytochrome *c*, which is released in cytosol during apoptosis, while the reduced form of cytochrome *c* is apparently inactive in the initiation of the apoptotic process[46, 47]. An inter-molecular electron transfer (ET) process is constituted by three main step: the formation of a redox complex, the electron tunneling across a biomolecular interface, and the dissociation of the reduced and oxidized products[99]. In an inter-molecular ET process between two metalloproteins, the two redox partners must possess specific features in order for the electron transfer to occur. In order to maximize the efficiency in the ET chain, redox complexes must present short lifetimes, i.e., after the electron transfer, the proteins involved in the process have to rapidly separate[99, 100]. Furthermore, their reduction potentials must be different, and their values provide information on the direction of the ET process. In the case of human neuroglobin and human cytochrome *c* this condition is verified, since the experimental reduction potential of human neuroglobin is -115 mV vs SHE and the one of cytochrome *c* is $\sim$ 255 mV vs SHE[101, 48]; these values indicate that neuroglobin transfer an electron to cytochrome *c*, reducing it. Another important parameter which affects the possibility of an electron transfer process to occur is the value of the electron transfer rate constant. An experimental work by Fago *et al.*[29] determined the electron transfer rate of the ET process between neuroglobin and cytochrome *c*, and it is equal to 2.0 x 10^7 M^{-1} s^{-1}, a value very close to the one of the reaction between cytochrome *c* and cytochrome *c* oxidase, which lies in the range of 4 x 10^6 - 1 x 10^7 M^{-1} s^{-1}[48]. The extent of the difference between the reduction potential values of neuroglobin and cytochrome *c* ($\sim$ 370 mV) compared to the one between typical redox partners as cytochrome *c* and cytochrome *c* oxidase ($\sim$ 255 mV [102]) and the value of the electron transfer rate, support the hypothesis that neuroglobin could be involved in ET processes and that cytochrome *c* could be considered as one of its redox partners. It is therefore reasonable to assess that the inter-molecular electron transfer reaction between reduced neuroglobin and oxidized cytochrome *c* is possible.

The structure of the complex between neuroglobin and cytochrome *c* was obtained using the freely available ZDOCK server[103, 104], starting from the equilibrated structures of human neuroglobin (PBD ID code: 4MPM[12], chain B) and of cytochrome *c* (PDB ID code: 1J3S[105]). In particular, we utilized ZDOCK 3.0.2, which features a scoring function composed by IFACE statistical potential, shape complementarity and electrostatics. Ten possible structures of the complex were generated, the first one being the most probable, and therefore, the one we used for the simulations. The prediction of these structures was fully unbiased, i.e., no likely interaction patches on either partner were suggested during the job submission. For further details on the docking process see ref.[104, 103]. The complex was simulated for 50 ns (following the same procedure reported above for the simulations og GLB-6 and neuroglobin) in two different oxidation states: the first one has the neuroglobin in its reduced form and the cytochrome *c* in its oxidized form, thus simulating the conditions before the electron transfer occurs (pre-ET form). The second one has the neuroglobin in its oxidized state, and the cytochrome *c* in its reduced state, therefore mimicking the complex after the electron transfer(post-ET form) (see figure 6.13). In these simulations, neuroglobin is considered in its form *with* the disulfide bridge between Cys46 and Cys55, mainly because it represents its most probable conformation (the crystal structure is indeed reported with the S-S bond). Nevertheless, since the environment in which the supposed function of preventing apoptosis is the cytosol, thus a reductant environment, future works on the interaction between neuroglobin and cytochrome *c* aim to simulate the adduct considering neuroglobin without the disulfide bridge. However, since the previously reported results do not show large differences between the two forms of neuroglobin, it is plausible to suppose that also for this complex the behavior of the protein with and without S-S bridge will not be extensively affected.

It is important to highlight that I have **not** simulated the electron transfer process, but only the redox condition before and after it, studying the effect(s) that the change of the charge of the two proteins involved has on the behavior, and in particular on the stability, of the complex between human neuroglobin and human cytochrome *c*.

The structure of the adduct between neuroglobin and cytochrome *c* obtained from the docking analysis shows that the two proteins approach in the corresponding regions of the heme sites, establishing an average distance between the two heme sites of 11.15 ± 0.09 Å, a value in agreement with inter-molecular ET processes, set between 10 and 15 Å[106]. All the regions in the vicinity of the two heme sites contribute to the formation of the complex, establishing several weak interactions (salt bridges, hydrogen

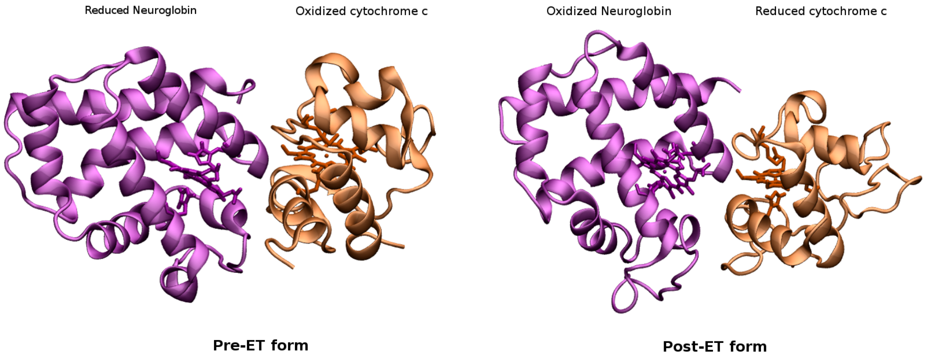


Figure 6.13: Three-dimensional cartoon structure of the complex between human neuroglobin (PDB ID code: 4MPM[12]) and human cytochrome c (PBD ID code: 1J3S[105]) before the electron transfer (left) and after the electron transfer (right) occurs. The heme groups and their axial ligands are represented as licorice. The image was prepared using the VMD software[13].

bonds and hydrophobic interactions) (reported in table 6.5), which probably contribute to the stability of the complex, avoiding the two proteins moving away. Some of these weak interactions are highlighted in figure 6.14. Salt bridges and hydrogen bonds were determined using the Gromacs tools g_saltbr and g_hbond, whereas the hydrophobic interactions were determined using the online software PLATINUM[107], which allows to find out molecular hydrophobic and hydrophilic properties using the concept of *"Molecular Hydrophobicity Potential"* (MHP): each atom is assigned a hydrophobic or hydrophilic constant, derived from experimental data, then the hydrophobic/hydrophilic properties of a molecule can be determined on its surface, and by comparison with the properties of neighboring molecules the hydrophobic effect can be evaluated[107] In particular, hydrophobic interactions were determined uploading the equilibrated structures of human neuroglobin and human cytochrome c as the receptor and ligand. To evaluate the hydrophobic interactions between a ligand and its receptor, PLATINUM calculates hydrophobic/hydrophilic properties of both the partners, determining which are the most favorable.

	NGB	Cytochrome c
Salt Bridges	COO Asp37	$N\epsilon2$ His26
	COO Asp149	$N\zeta$ Lys79
	COO Glu151	$N\zeta$ Lys79
	COO Glu60	$N\zeta$ Lys8
Hydrogen Bonds	O Val99	$N\epsilon2H\epsilon1$ Gln16
	$O\delta2$ Asp37	$N\epsilon2H\epsilon2$ His26
	$N\zeta H\zeta$ Lys102	O His26
	O Phe42	$N\epsilon2H\epsilon2$ Gln16
Hydrophobic Interactions	$C\beta$ Ala98	$C\gamma2$ Ile81
	$C\beta$ Ala98	$C\gamma2$ Val83
	$C\gamma$ Leu41	$C\delta2$ His26
	$C\delta2$ Tyr44	$C\epsilon$ Met12
	$C\beta$ Arg94	$C\gamma1$ Val83
	$C\gamma$ Lys95	$C\gamma2$ Val83
	$C\beta$ Ala98	CBC heme
	$C\gamma2$ Val99	$C\beta$ Gln16

Table 6.5: List of atoms involving in weak interactions for the complex between human neuroglobin and human cytochrome c.

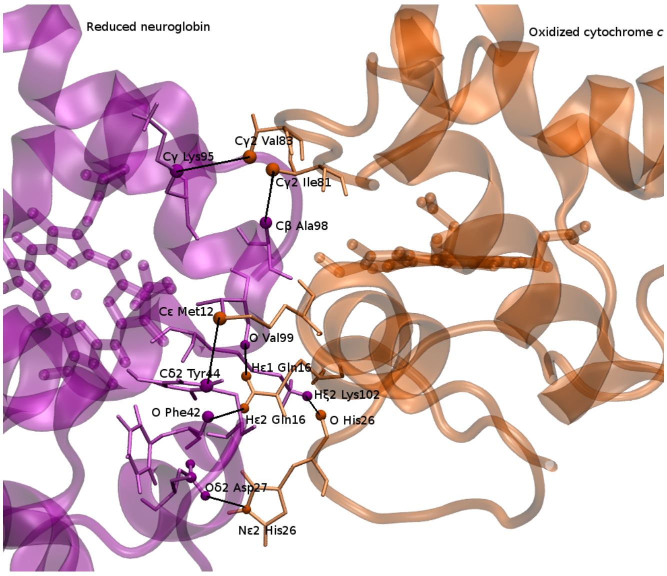


Figure 6.14: Zoom on the interacting regions of human neuroglobin and human cytochrome c in their complex in the pre-ET form. Some of the weak interactions reported in table 6.5 are highlighted. The image was prepared using the VMD software[13].

Setting cut-off values of 0.4 nm for salt bridges and hydrogen bonds[108, 109, 110], and 0.5 nm for hydrophobic interactions (since they are stronger than salt bridges and hydrogen bonds)[108, 109, 110] most of these interactions disappear in the simulation of the complex after the electron transfer, as showed in the normalized probability distribution of the distances between the atoms involved in these interactions(see figures 6.15, 6.16, 6.17 and table 6.6), indicating that, once the neuroglobin has transferred one electron to cytochrome *c*, reducing it, the complex between the two proteins is no longer stable as in the pre-ET form.

NGB - Cyt *c*	pre-ET form	post-ET form
Salt Bridges	$\langle distance \rangle$ (nm)	$\langle distance \rangle$ (nm)
COO Asp37 - Nϵ2 His26	0.652 (0.363)	1.724 (0.133)
COO Asp149 - Nζ Lys79	0.866 (0.209)	1.664 (0.339)
COO Glu151 - Nζ Lys79	0.764 (0.231)	1.783 (0.533)
COO Glu60 - Nζ Lys8	0.836 (0.209)	1.351 (0.238)
Hydrogen Bonds	$\langle distance \rangle$ (nm)	$\langle distance \rangle$ (nm)
O Val 99 - Nϵ2Hϵ1 Gln16	0.309 (0.063)	0.471 (0.160)
Oδ2 Asp37 - Nϵ2Hϵ2 His26	0.659 (0.365)	1.718 (0.153)
Nζ Lys102 - O His26	0.560 (0.202)	1.310 (0.168)
O Phe42 - Nϵ2Hϵ2 Gln16	0.504 (0.084)	0.882 (0.122)
Hydrophobic Interactions	$\langle distance \rangle$ (nm)	$\langle distance \rangle$ (nm)
Cβ Ala98 - Cγ2 Ile81	0.458 (0.075)	0.913 (0.255)
Cβ Ala98 - Cγ2 Val83	0.443 (0.089)	0.598 (0.162)
Cγ Leu41 - Cδ2 His26	0.582 (0.196)	1.768 (0.082)
Cδ2 Tyr44 - Cϵ Met12	0.510 (0.128)	0.676 (0.190)
Cβ Arg94 - Cγ1 Val83	0.563 (0.114)	0.922 (0.234)
Cγ Lys95 - Cγ2 Val83	0.454 (0.069)	1.058 (0.301)
Cβ Ala98 - CBC heme	0.521 (0.037)	0.969 (0.203)
Cγ2 Val99 - Cβ Gln16	0.555 (0.077)	0.689 (0.093)

Table 6.6: Average values of the distances between the atoms forming the weak interactions, for both the pre-ET and post-ET form of the complex between neuroglobin and cytochrome c. Standard deviations are given in brackets.

The average values of the distances between the atoms involving in the formation of weak interactions are invariably higher for the post-ET complex compared with the pre-ET form, often exceeding to a larger extent the typical values for the realistic existence of such kind of interactions in proteins[108, 109, 110], with values ranging from ~ 0.8 to ~ 1.8 nm.

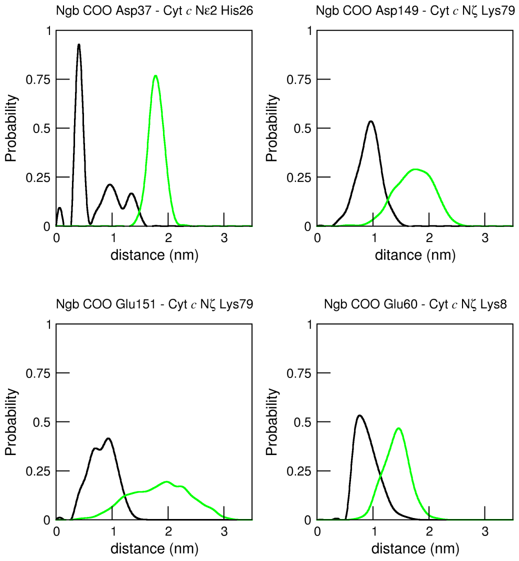


Figure 6.15: Normalized probability distribution of the distances between the atoms forming salt bridges in the complex between neuroglobin and cytochrome c, in the pre-ET (black) and post-ET (green) simulation, respectively.

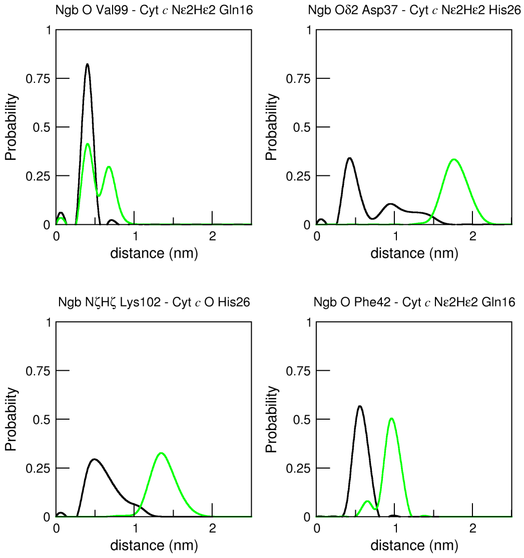


Figure 6.16: Normalized probability distribution of the distances between the atoms forming hydrogen bonds in the complex between neuroglobin and cytochrome c, in the pre-ET (black) and post-ET (green) simulation, respectively.

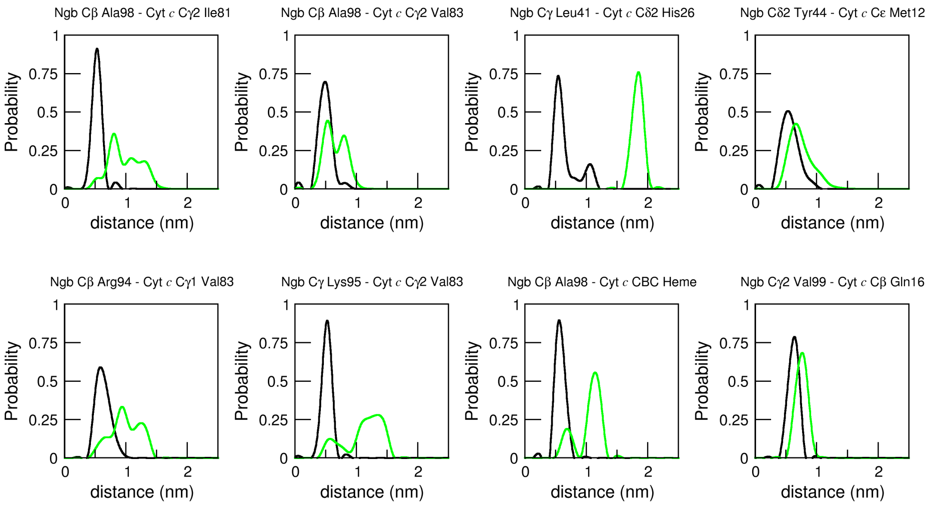


Figure 6.17: Normalized probability distribution of the distances between the atoms forming hydrophobic interactions in the complex between neuroglobin and cytochrome c, in the pre-ET (black) and post-ET(green) simulation, respectively.

Together with their normalized probability distributions (see figures 6.15, 6.16 and 6.17), which show how during the simulation time most of these distances lies in a range compatible with the presence of these weak interactions (0.3-0-5 nm) in the pre-ET form with respect to the post-ET complex, these results suggest that the pre-ET complex is probably more stable compared to the post-ET one, strengthen the hypothesis that, *in vivo*, the electron transfer process involves the transfer of an electron from reduced neuroglobin to oxidized cytochrome *c*.

Another evidence that verify the hypothesis of a different behavior of the complex before and after the electron transfer was obtained calculating the number of solvent molecules which enter between the contact surfaces of the two proteins, obtained using a sphere of radius 1.2 nm from the Cα atom of residue Val99 of neuroglobin, which is localized more or less at the center of the contact surface of the two protein forming the complex (see figure 6.18).

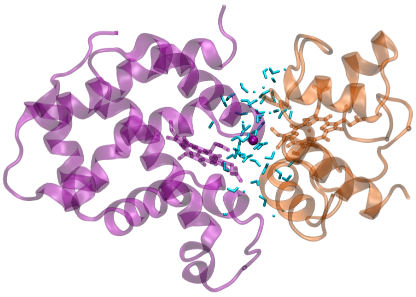


Figure 6.18: Water molecules (as light blue licorice) entering the surface between neuroglobin and cytochrome c complex in a radius of 1.2 nm from the Cα of residue Val99 of neuroglobin (purple sphere). Neuroglobin structure is represented as purple cartoon, while cytochrome c structure is showed as orange cartoon.

As can be seen from figure 6.19, in the first ns of simulation the number of water molecules between the two proteins is almost the same, whereas after about 10 ns of simulation it diverges, lowering for the pre-ET form and increasing for the post-ET form, thus indicating that in the pre-ET complex the two proteins tend to become closer, while in the post-ET complex they tend to diverge.

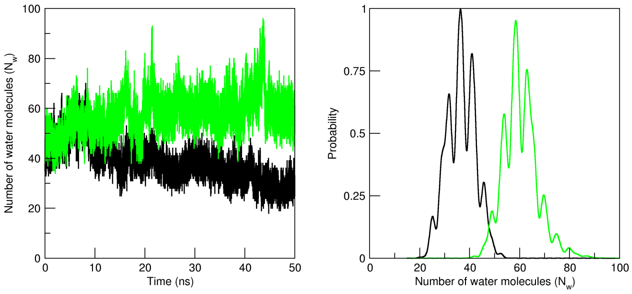


Figure 6.19: Number of water molecules entering the contact surface between neuroglobin and cytochrome c. On the left the trend over the simulation time is reported, while on the right the normalized probability ditribution is showed (black for the pre-ET simulation, green for the post-ET simulation, respectively).

The center of mass of the complex in the two oxidation states has also been calculated. As can be seen from figure 6.20, this is a further evidence that the complex before the electron transfer is more stable and tight compared to the one after ET.

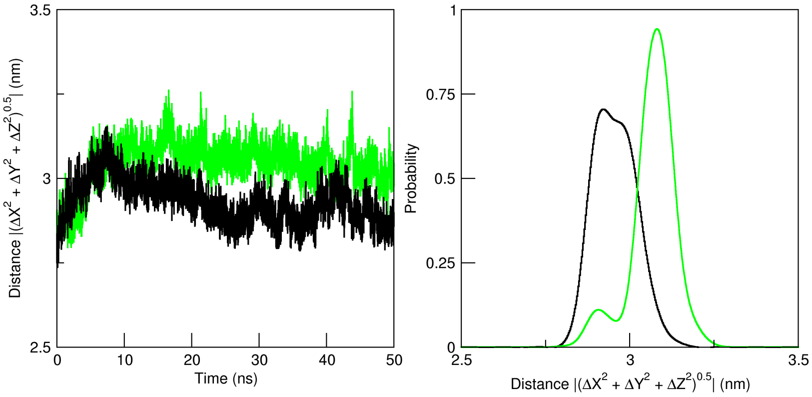


Figure 6.20: Center of mass of the complex between neuroglobin and cytochrome c over simualtion time (left) and as a normalized probability distribution (right) for the complex before ET (black) and after ET (green).

Using the *Energy Gap Fluctuations* method the reorganization energy of the complex has been calculated, resulting in the amount of 1.40 ± 0.02 eV, and since the cytochrome c has a value of 0.73 ± 0.05 eV[64], it corresponds to the sum of the reorganization energy of the single proteins (being the λ value of neuroglobin 0.70 ± 0.02 eV). Moreover, the calculated reorganization energy for this complex lies in the range of the typical values for ET processes, namely between 0.5-1.5 eV[106]. All these results lead to the conclusion that probably, *in vivo*, the complex between human neuroglobin and human cytochrome c is favored when neuroglobin is in its reduced form and the cytochrome c is in its oxidized form, with respect to the one with the inverted charges, thus confirming the possible role of neuroglobin in preventing apoptosis.

References

[1] T. Burmester et al. In: *Nature* 407 (2000), pp. 520–523.

[2] J. T. Trent, R. A. Watts, and M. S. Hargrove. In: *J. Biol. Chem.* 276 (2001), pp. 30106–30110.

[3] A. Pesce et al. In: *EMBO Rep.* 3 (2002), pp. 1146–1151.

[4] M. Couture et al. In: *J. Biol. Chem.* 274 (1999), pp. 6898–6910.

[5] T. R. Weiland et al. In: *J. Am. Chem. Soc.* 126 (2004), pp. 11930–11935.

[6] P. Halder, J. T. III Trent, and M. S. Hargrove. In: *Proteins* 182 (2007), pp. 172–182.

[7] S. Kakar et al. In: *Biophys. Chem.* 152 (2010), pp. 1–14.

[8] S. Dewilde et al. In: *J. Biol. Chem.* 276 (2001), pp. 38949–38955.

[9] T. Burmester et al. In: *IUBMB life* 56 (2004), pp. 703–707.

[10] A. Pesce et al. In: *Structure* 11 (2003), pp. 1087–1095.

[11] L. Holm and C. Sander. In: *FEBS Lett.* 315 (1993), pp. 301–306.

[12] B. G. Guimarães et al. In: *Acta Crystallogr. D* 70 (2014), pp. 1005–1014.

[13] W. Humphrey, A. Dalke, and K. Schulten. In: *J. Mol. Graphics* 14 (1996), pp. 33–38.

[14] M. Couture et al. In: *J. Biol. Chem.* 276 (2001), pp. 36377–36382.

[15] J. M. Kriegl et al. In: *P. Natl. Acad. Sci. U.S.A.* 99 (2002), pp. 7992–7997.

[16] M. Brunori and B. Vallone. In: *Cell. Mol. Life Sci* 64 (2007), pp. 1259–1268.

[17] T. Burmester and T. Hankeln. In: *J. Exp. Biol.* 212 (2009), pp. 1423–1428.

[18] P. Ascenzi et al. In: *Biochem. Mol. Biol. Edu.* 32 (2004), pp. 305–313.

[19] A. D. Nadra et al. In: *Proteins* 71 (2008), pp. 695–705.

[20] S. Herold et al. In: *J. Biol. Chem.* 279 (2004), pp. 22841–22847.

[21] M. Brunori et al. In: *Proc. Natl. Acad. Sci. U.S.A.* 102 (2005), pp. 8483–8488.

[22] M. G. Petersen, S. Dewilde, and A. Fago. In: *J. Inorg. Biochem.* 102 (2008), pp. 1777–1782.

[23] K. Jin et al. In: *Neurosci. Lett.* 430 (2008), pp. 135–137.

[24] M. Tiso et al. In: *J. Biol. Chem.* 286 (2011).

[25] A. Fago et al. In: *IUBMB life* 56 (2004), pp. 689–696.

[26] K. Wakasugi and I. Morishima. In: *Biochem. Biophys. Res.* 330 (2005), pp. 591–597.

[27] K. Wakasugi, T. Nakano, and I. Morishima. In: *J. Biol. Chem.* 278 (2003), pp. 36505–36512.

[28] S. Watanabe and K. Wakasugi. In: *Biochem. Biophys. Res.* 369 (2008), pp. 695–700.

[29] A. Fago et al. In: *FEBS Lett.* 580 (2006), pp. 4884–4888.

[30] A. Giuffrè et al. In: *Biochem. Bioph. Res. Co.* 367 (2008), pp. 893–898.

[31] A. Giuffrè et al. In: *IUBMB life* 60 (2008), pp. 410–413.

[32] S. H. Bønding et al. In: *Int. J. Biol. Macromol.* 43 (2008), pp. 295–299.

[33] T. Brittain et al. In: *IUBMB life* 62 (2010), pp. 878–885.

[34] M. Fiocchetti et al. In: *Biochim, Biophys. Acta* 1834 (2013), pp. 1744–1749.

[35] A. Pesce et al. In: *IUBMB life* 56 (2004), pp. 657–664.

[36] S. Orlowski and W. Nowak. In: *Biosystems* 94 (2008), pp. 263–266.

[37] L. Kiger et al. In: *IUBMB life* 56 (2004), pp. 709–719.

[38] D. Hamdane et al. In: *J. Biol. Chem.* 278 (2003), pp. 51713–51721.

[39] H. Ishikawa et al. In: *Proc. Natl. Acad. Sci. U.S.A.* 104 (2007), pp. 19309–19314.

[40] A. Bocahut et al. In: *J. Phys. Chem. B* 113 (2009), pp. 16257–16267.

[41] L. Astudillo et al. In: *Biophys. J.* 99 (2010), pp. L16–L18.

[42] E. Vinck et al. In: *J. Am. Chem. Soc.* 126 (2004), pp. 4516–4517.

[43] M. Ezhevskaya et al. In: *J. Inorg. Biochem.* 105 (2011), pp. 1131–1137.

[44] A. N. Morozov et al. In: *J. Chem. Inf. Mod.* 54 (2014), pp. 1997–2003.

[45] L. Kiger et al. In: *PloS ONE* 6 (2011), e20478.

[46] D. Suto et al. In: *Biochem. J.* 392 (2005), pp. 399–406.

[47] Z. Pan, D. W. Voehringer, and R. E. Meyn. In: *Cell Death Differ.* 6 (1999), pp. 683–688.

[48] M. T. Wilson et al. In: *Biochem. J.* 147 (1975), pp. 145–153.

[49] D. Hoogewijs et al. In: *IUBMB Life* 56 (2004), pp. 697–702.

[50] D. Hoogewijs et al. In: *BMC Genomics* 8 (2007), p. 356.

[51] E. Geuens et al. In: *BMC Biochem.* 11 (2010), p. 17.

[52] P. T. McGrath et al. In: *Neuron* 61 (2009), pp. 692–699.

[53] A. Persson et al. In: *Nature* 458 (2009), pp. 1030–1033.

[54] J. M. Gray et al. In: *Nature* 430 (2004), pp. 317–322.

[55] A. J. Chang et al. In: *PLoS Biol.* 4 (2006), e274.

[56] J. Yoon et al. In: *Biochemistry* 49 (2010), pp. 5662–5670.

[57] H. Sawai et al. In: *Biochemistry* 42 (2003), pp. 5133–5142.

[58] S. M. G. Duff, B. A. Wittenberg, and R. D. Hill. In: *J. Biol. Chem.* 272 (1997), pp. 16746–16752.

[59] R. Arredondo-Peter et al. In: *Plant Physiol.* 115 (1997), pp. 1259–1266.

[60] A. Pesce et al. In: *J. Mol. Biol.* 386 (2009), pp. 246–260.

[61] A. G. Mauk and G. R. Moore. In: *J. Biol. Inorg. Chem.* 2 (1997), pp. 119–125.

[62] D. I. Brown and K. K. Griendlin. In: *Free Radical Biol. Med.* 47 (2009), pp. 1239–1253.

[63] M. P. Murphy. In: *Biochem. J.* 417 (2009), pp. 1–13.

[64] C. A. Bortolotti et al. In: *J. Phys. Chem. Lett.* 2 (2011), pp. 1761–1765.

[65] T. B. Karpishin et al. In: *Structure* 2 (1994), pp. 415–422.

[66] A. Alessandrini, S. Corni, and P. Facci. In: *Phys Chem Chem Phys* 8 (2006), pp. 4383–4397.

[67] D.S. Bendall. BIOS Scientific Publisher Ltd:Oxford,UK, 1996.

[68] D. H. Murgida and P. Hildebrandt. In: *J. Phys. Chem. B* 106 (2002), pp. 12814–12819.

[69] T. Simonson. In: *P. Natl. Acad. Sci.* 99 (2002), pp. 6544–6549.

[70] S. Corni. In: *J. Phys. Chem. B* 109 (2005), pp. 3423–3430.

[71] M. Cascella et al. In: *P. Natl. Acad. Sci. USA* 103 (2006), pp. 19641–19646.

[72] B. K. Kuntal, P. Aparoy, and P. Reddanna. In: *BMC Res. Notes* 3 (2010), pp. 226–230.

[73] H. J. C. Berendsen, D. van der Spoel, and D. van Drunen. In: *Comp. Phys. Comm.* 95 (1995), pp. 43–56.

[74] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. In: *J. Phys. Chem.* 91 (1987), pp. 6269–6271.

[75] C. A. Bortolotti et al. In: *J. Am. Chem. Soc.* 134 (2012), pp. 13670–13678.

[76] D. Brown and J. H. R. Clarke. In: *Mol. Phys.* 51 (1984), pp. 1243–1252.

[77] B. Hess et al. In: *J. Comput. Chem.* 18 (1997), pp. 1463–1472.

[78] T. Darden, D. York, and L. Pedersen. In: ().

[79] W. Kabsch and C. Sander. In: *Biopolymers* 22 (1983), pp. 2577–2637.

[80] M. Aschi et al. In: *Chem. Phys. Lett.* 344 (2001), pp. 374–380.

[81] A. Amadei, M. D'Alessandro, and M. Aschi. In: *J. Phys. Chem. B* 108 (2004), pp. 16250–16254.

[82] M. Aschi et al. In: *J. Phys. Org. Chem.* 19 (2006), pp. 518–530.

[83] C. Zazza et al. In: *J. Phys. Chem. B* 112 (2008), pp. 3184–3192.

[84] I. Daidone et al. In: *Chem. Phys. Lett.* 488 (2010), pp. 213–218.

[85] A. Amadei et al. In: *Curr. Opin. Struc. Biol.* 20 (2010), pp. 155–161.

[86] A. Amadei, I. Daidone, and C. A. Bortolotti. In: *RSC Advances* 3.42 (2013), pp. 19657–19665.

[87] I. Daidone et al. In: *J. Phys. Chem. Lett.* 5 (2014), pp. 1534–1540.

[88] I. Daidone et al. In: *J. Phys. Chem. B* 118 (2014), pp. 7554–7560.

[89] I. Muegge et al. In: *J. Phys. Chem. B* 101 (1997), pp. 825–836.

[90] R. A. Marcus and N. Sutin. In: *BBA-Bioenergetics* 811 (1985), pp. 265–322.

[91] H. B. Gray and J. R. Winkler. In: *Q. Rev. Biophys.* 36 (2003), pp. 341–372.

[92] T. Simonson. In: *Rep. Prog. Phys.* 66 (2003), pp. 737–787.

[93] U. L. F. Ryde and M. H. M. Olsson. In: *Int. J. Quantum Chem.* 81.2001 (2000), pp. 335–347.

[94] H. B. Gray and J. R. Winkler. In: *Annu. Rev. Biochem.* 256 (1996), pp. 537–561.

[95] R. A. Scott and A. G. Mauk. University Science Book: Sausalito, CA, 1995.

[96] J. Cheng et al. In: *Isr. J. Chem.* 37 (1997), pp. 259–266.

[97] H. B. Gray, B. G. Malmström, and R.J.P. Williams. In: *J. Biol. Inorg. Chem.* 5.5 (2000), pp. 551–559.

[98] L. Paltrinieri et al. In: *J. Phys. Chem. Letters* 4 (2013), pp. 710–715.

[99] A. N. Volkov and N. A. J. van Nuland. In: *PLos Comput. Biol.* 8 (2012), e1002807.

[100] P. B. Crowley and M. Ubbnik. In: *Acc. Chem. Res.* 36 (2003), pp. 723–730.

[101] M. J. Eddowes and H. A. O. Hill. In: *J. Am. Chem. Soc.* 101 (1979), pp. 4461–4464.

[102] R. W. Hendler, G. S. Sidhu, and K. Pardhasaradhi. In: *Biophys. J.* 58 (1990), pp. 957–967.

[103] B. G. Pierce et al. In: *Bioinformatics* 30 (2014), pp. 1771–1773.

[104] B.G. Pierce, Y. Hourai, and Z. Weng. In: *PLoS ONE* 6 (2011), e24657.

[105] W. Y. Jeng et al. In: *to be published* ().

[106] G. W. Canters and C. Dennison. In: *Biochimie* 77 (1995), pp. 506–515.

[107] T. V. Pyrkov et al. In: *Bioinformatics* 25 (2009), pp. 1201–1202.

[108] S. Kumar and R. Nussinov. In: *ChemBioChem* 37 (1997), pp. 604–617.

[109] P. Atkins and J. de Paula. Oxford University Press: Oxford, UK, 2006.

[110] R. Chang. Edwards Brothers Inc: Sausalito, CA, 2005.

7 Computational Electrophysiology of Voltage-Dependent Anion Channel, VDAC

7.1 Voltage-Dependent Anion Channel, VDAC

The central role played by mitochondria in a number of fundamental physiological processes, from cell signaling events, inter-organellar communication, aging, cell proiferation, diseases and cell death[1] is nowadays universally recognized. One of the most important mitochondrial proteins that controls most of these biological processes is the Voltage-Dependent Anion Channel, VDAC, a relatively small protein (~ 30 kDa) which is highly conserved in all eukaryotics[2, 3, 4, 5, 6, 7, 8]. VDAC is predominantly localized in the Outer Mitochondrial Membrane (OMM), and its main function is to regulate the entrance and exit of several metabolites, therefore governing the communication between mitochondria and the rest of the cell. VDAC is composed by 283 amino acid residues, forming a 19-stranded β-barrel, of which each β strand is connected to another by flexible loops. and an N-terminal α-helix which is positioned inside the channel, but is not part of it (see figure 7.1).

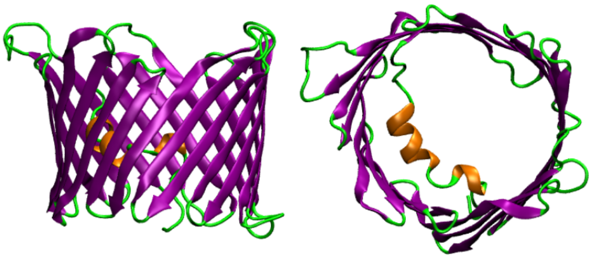


Figure 7.1: Three-dimensional cartoon representation of VDAC structure from Mus musculus (PDB ID code: 3EMN[9]).

It is noteworthy that VDAC is formed by an uneven number of β-strands, in contrast with the even number featured by the bacterial outer membrane proteins[10], thus requiring a parallel interaction of two adjacent β-strands,

$\beta 1$ and $\beta 19$. VDAC is defined as a voltage-dependent anion channel since its electrophysiological characterization shows that *in vitro* this protein possesses the ability to sense the transmembrane imposed electrical potential and to react to it exhibiting a partial closure[11, 12, 13]. In particular, at low values of transmembrane potential (~ 10 mV) VDAC exhibits a long-lived open state[1], whereas at higher potentials, both positive and negative (> 40 mV), several sub-states are present, featuring different ionic selectivities and permeabilities[14, 12]. The closure of the channel is more frequent as the voltage increases, leading to a decrease of the conductance, which is maximum in the completely open state. The extent of this decrease in conductance depends on the considered species from which VDAC is taken[15, 16]. In the open, main conductance, state, VDAC allows the permeation of small ions (Cl^-, K^+, Na^+) as well as large anions, such as glutamate[14] and ATP[13] and large cations, for example acetylcholine, dopamine[14] and Tris[17]. Despite the protein's name,in the open state of VDAC the anion selectivity is indeed only slightly higher compared to the cation selectivity for species composed by equally mobile cations and anions, such as KCl[14]. The slight anion selectivity can be explained considering the structural data, which show a modest prevalence of positive charged residues positioned in the pore wall[9]. It is quite obvious that the closed state of the VDAC channel can prevent the passage of larger molecules, and limit the permeation of small ions. In order to clarify the mechanisms of VADC gating process, different models have been proposed, based on the 3D structure of the protein. Some of these models stress the role of N-terminal α-helix, which might block the channel moving from the barrel wall to the center of the pore[9, 18]. Other models, based on electron microscopy and electrophysiological studies[19, 20], suggest larger conformational rearrangements which also involve the β-barrel region, supported by the conformational instability of the channel due to the parallel disposition of strands $\beta 1$ and $\beta 19$, and by the large number of charged residues in strands 1-8[1].

Another element that could affect the behavior of VDAC, in particular with regard of ions selectivity and conductance, is the effect of the membrane environment, which has been explicitly taken into account in several studies[21, 22, 23, 24].

During the period I spent working at Max Planck Insitute for Biophysical Chemistry in Göttingen, Germany, under the supervsion of prof. Bert de Groot, I studied the behavior of VDAC in two different lipid membranes (1-Palmitoyl-2-Oleoyl-sn-glycero-3-PhosphoCholine - POPC - and 1-Palmitoyl-2-Oleoyl-sn-3-PhosphoEthanolamine - POPE - (see figure 7.2)), by means of Molecular Dynamics and Computational Electrophysiology[25], in order to verify the effect of the lipid environment on the electrophysiological properties of the channel.

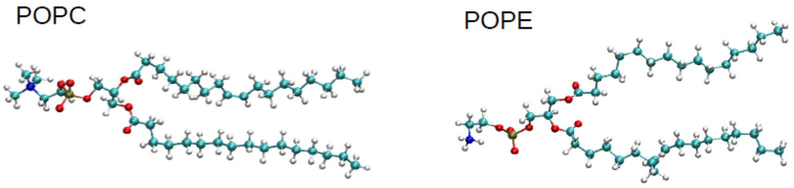


Figure 7.2: Balls and sticks representation of POPC (left) and POPE (right) phospholipids.

7.2 Molecular Dynamics and Computational Electrophysiology of VDAC

7.2.1 Molecular Dynamics Simulations

The MD simulations of VDAC incorporated in pure POPC and pure POPE lipid membrane were performed in water using the GROMACS software package[26], version 4.5.5, utilizing the CHARMM36 force field[27], a quite recently developed force field which, in contrast with the generally previously used CHARMM27 and CHARMM27c force fields, that systematically provided values of the surface area per lipid smaller than the experimental ones, allows to obtain a better representation of the characteristics of pure lipid systems and heterogeneous systems involving membrane proteins. In the simulations, water was represented with TIPS3P model[28]. As a first step, the protein (PDB ID code: 3EMN[9]) was embedded inside the lipid membrane, using the g_membed tool[29], which is a Gromacs implemented tool, developed by the Grubm*ü*ller group at Max Planck Institute for Biophysical Chemistry, which, in a very efficient way, includes a protein into an equilibrated and hydrated lipid membrane with a minimal perturbation. Then the system was energy minimized (10000 steps) using the steepest descent method, and neutralized adding the suitable amount of counter ions. A final concentration of 0.5 M NaCl was obtained. After energy minimization, the system was subjected to an equilibration MD run, with position restraints on protein atoms, in order to equilibrate the lipid membrane and the solvent around the protein. Finally, the simulation runs were performed in the NPT ensemble, at 303 K for POPC membrane and 310 K for POPE membrane, and at a pressure of 1.0 bar. The temperature was kept constant by v-rescale coupling and the pressure was kept constant by the Berendsen coupling[30]. Periodic boundary conditions were applied to the simulation box, and the long range electrostatic interactions were treated with the Particle Mesh Ewald method[31], using a grid space of 0.15 nm combined with a fourth-order B-spline interpolation to compute the potential and forces. The real space cut-off distance was set to 1.2 nm, and the van der Waals cut-off distance was set to 1.0 nm. A time step of 2 fs was used. Bond lengths were constrained with LINCS algorithm. The N-terminal and C-terminal ends of the protein were invariably modeled as NH_3^+ and CO_2^-, respectively.

7.2.2 Stability of the simulations

In order to verify the reliability of the simulated systems, first of all I calculated the values of the area per lipid (APL), for pure POPC and pure POPE membranes. To obtain these values, I simulated for 80 ns two systems formed only by a fully hydrated double layer of POPC and POPE respectively, in presence of a NaCl concentration of 0.5 M (to recreate the conditions of the simulations containing the VDAC). The area per lipid was calculated using the following expression:

$$APL = \frac{(L_{boxX})(L_{boxY})}{N_{lipid}/2} \tag{7.1}$$

where L_{boxX} and L_{boxY} are the size of of the simulated box along the X and the Y coordinate, respectively; N_{lipid} is the total number of lipids forming the bilayer, which has to be divided by two in order to obtain the APL value for a single lipid molecule.

The results, compared with the experimental values, are reported in table 7.1.

	N_{lipid}	APL_{calc} ($\text{Å}^2/lipid$)	APL_{exp} ($\text{Å}^2/lipid$)
POPC	236	65.9 ± 0.2	68.3 ± 1.5 [a]
POPE	239	59.4 ± 0.3	60.25 ± 0.5 [b]

Table 7.1: Calculated and experimental values and associated errors of area per lipid for POPC and POPE membranes. [a] *taken from ref.[32],* [b] *taken from ref.[33].*

The very good agreement between the calculated and experimental values indicates that the simulations are trustworthy, and that the properties of the lipid bilayer are well represented.

Moreover, to further prove the stability and reliability of the simulations, I performed MD runs of 80 ns on two systems, one containing VDAC embedded in pure POPC membrane, and the other with VDAC in pure POPE (see figure 7.3), and I have calculated the RMSD and RMSF on the protein atoms (see figure 7.4 and 7.5). The trends of the RMSD and their very low average values (0.175 ± 0.020 nm for VDAC-POPC and 0.167 ± 0.018 nm for VDAC-POPE), together with the relative low values of the fluctuations of amino acid residues (the higher fluctuations corresponds to the loops connecting the β-sheet of the protein), are a strong evidence that the simulated systems are stable and reliable, and that they can be used to further analysis.

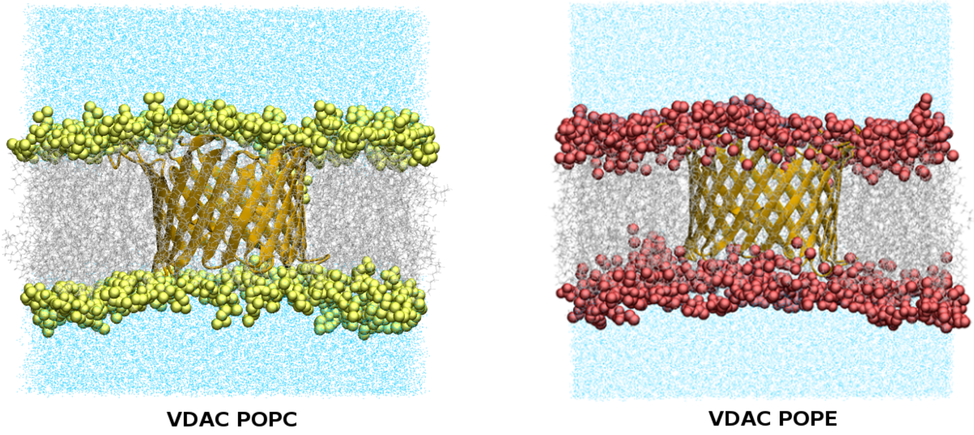


Figure 7.3: 3-D representation of the simulated systems: VDAC-POPC on the right and VDAC-POPE on the left.

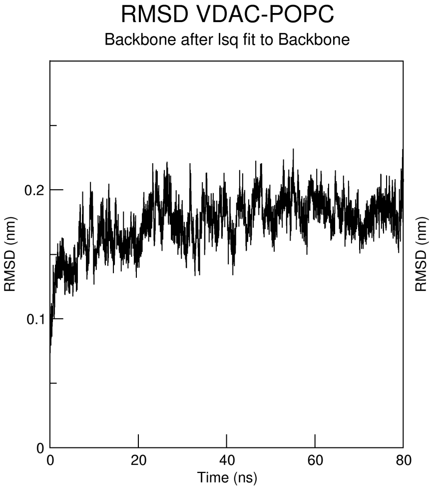


Figure 7.4: Calculated RMSD for VDAC in POPC (left) and POPE (right) lipid membrane.

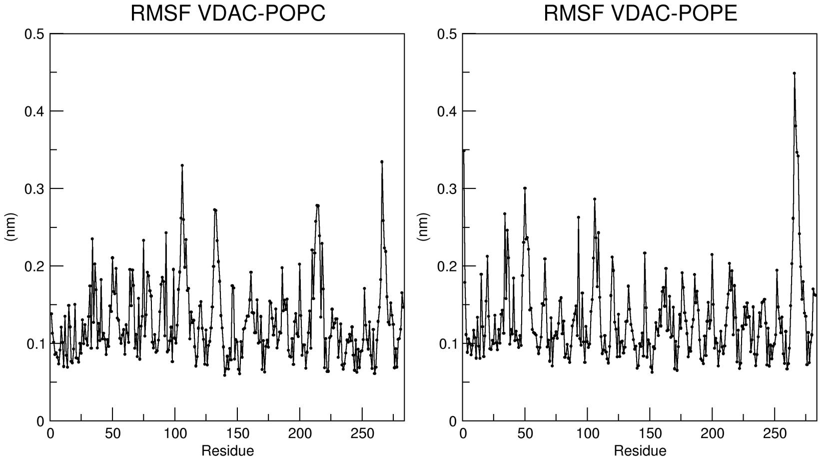


Figure 7.5: Calculated RMSF for VDAC in POPC (left) and POPE (right) lipid membrane.

7.2.3 Computational Electrophysiology on VDAC

Recently, the research group of professors de Groot and Grubm*ü*ller at Max Planck Institute for Biophysical Chemistry, where I spent 6 months, developed a new computational method which simulates ion flux through membrane channels founded on biologically realistic electrochemical gradients, the so-called *Computational Electrophysiology* (CompEl)[25], which is usable in Gromacs. This techinique allows to accurately predict ions conductance and selectivity, and permits the clarification of ion conduction mechanisms[25, 34]. In particular, CompEl allows to control both an ionic concentration gradient and a potential difference across the lipid membrane during MD simulations[25]. In order to utilize CompEl, the investigated system has to be prepared in a specific manner; i.e., a single box containing the protein embedded in the lipid membrane and the solvent must be doubled, thus obtaining two channel in the same simulation box (see figure 7.6). This simulation setup allows monitoring both positive and negative potential differences simultaneously. Then, a charge imbalance has to be set between the compartment α and β, thus earning two compartments of unequal ionic strength, and thereby creating a difference in the transmembrane potential, which depends precisely on the entity of this difference in charge. Changing the charge imbalance allows to obtain different transmembrane potential, and thus to study the electrophysiological behavior of a channel protein in different conditions. The transfer of water molecules and ions from α to β and vice versa can only occur through protein channels, therefore, it is possible to measure the cationic and anionic current flows, dependent on the transmembrane potential applied. The ionic concentration (and hence the charge imbalance) and the potential difference between the two compartments must be kept constant during the overall simulation time, and this is achieved by means of exchanges (swaps) of ions with solvent molecules from one compartment to another[25].

To prepare the systems for CompEl simulations, I extracted two well equilibrated structures of VDAC embedded either in POPC or POPE from the 80 ns MD runs described previously, and I doubled the two boxes, thus obtaining two fully hydrated systems each one containing two VDAC channel, embedded in their corresponding lipid membrane, with NaCl concentration of ~ 0.5 M (see figure 7.6).

Then, I imposed seven different charge imbalance between the two compartments, simply removing the same number of Na^+ atoms from a compartment and Cl^- atoms from the other, in order to maintain the neutrality of the system, naming the obtained systems 2q, 4q, 6q, 8q, 10q, 12q and

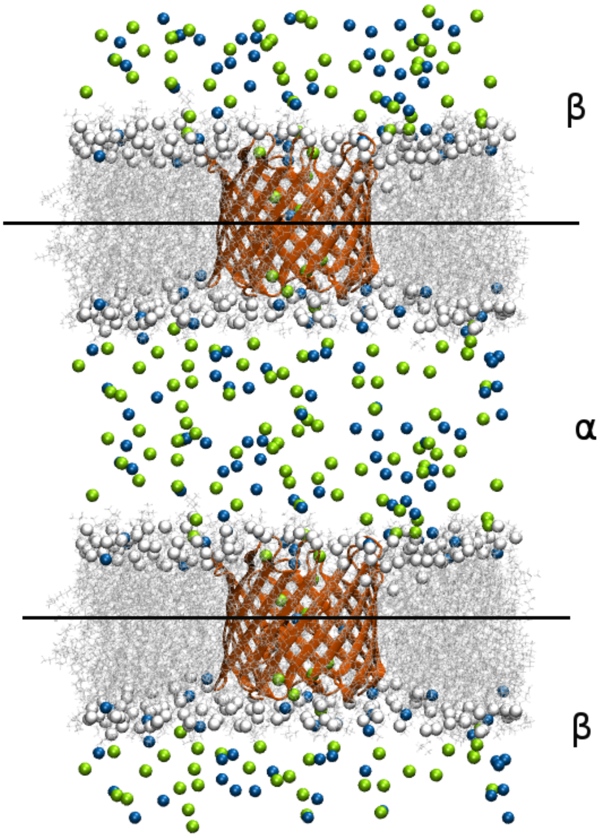


Figure 7.6: Three-dimensional representation of CompEl simulation box for VDAC-POPC. VDAC is colored in orange, POPC in gray, Na^+ ions in blue and Cl^- ions in green. Water molecules are omitted for clarity.

14q (where the number represents the total number of ions I removed, and q stands for an indication of charge), thus obtaining six different values of transmebrane potential. I performed CompEl simulations of 60 ns for 10 replicas for each system, in order to obtain a good sampling, changing the intial velocity for each replica. A summary of the performed simulations is reported in table 7.2.

For each replica, the values of transmembrane potential, anionic current, anionic conductance, cationic current and cationic conductance were calculated, and then they were averaged to obtain mean values for each investigated system. The results are reported in table 7.3 and a plot of total current vs potential can be seen in figure 7.7, from which it can inferred that

	Na^+ removed from comp. α	Cl^- removed from comp. β	Number of replicas
2q	1	1	10
4q	2	2	10
6q	3	3	10
8q	4	4	10
10q	5	5	10
12q	6	6	10
14q	7	7	10

Table 7.2: Summary of the CompEl simulations performed on VDAC both in POPC and POPE membrane.

current is directly proportional to the potential.

VDAC POPC

System	Potential (V)	Anionic Current (nA)	Anionic Conductance (nS)	Cationic Current (nA)	Cationic Conductance (nS)
2q	0.055 (0.006)	0.177 (0.018)	3.509 (0.475)	0.019 (0.004)	0.364 (0.060)
4q	0.103 (0.006)	0.380 (0.023)	3.696 (0.133)	0.037 (0.004)	0.356 (0.034)
6q	0.162 (0.008)	0.568 (0.038)	3.477 (0.147)	0.051 (0.006)	0.325 (0.045)
8q	0.221 (0.009)	0.762 (0.045)	3.439 (0.141)	0.072 (0.008)	0.331 (0.034)
10q	0.274 (0.007)	1.019 (0.030)	3.723 (0.089)	0.094 (0.004)	0.344 (0.013)
12q	0.324 (0.007)	1.159 (0.032)	3.578 (0.045)	0.132 (0.007)	0.405 (0.017)
14q	0.382 (0.008)	1.443 (0.034)	3.787 (0.087)	0.168 (0.010)	0.439 (0.024)

VDAC POPE

System	Potential (V)	Anionic Current (nA)	Anionic Conductance (nS)	Cationic Current (nA)	Cationic Conductance (nS)
2q	0.062 (0.005)	0.220 (0.033)	3.498 (0.505)	0.016 (0.003)	0.288 (0.056)
4q	0.121 (0.008)	0.497 (0.030)	4.187 (0.190)	0.026 (0.005)	0.212 (0.044)
6q	0.184 (0.006)	0.722 (0.032)	3.925 (0.139)	0.048 (0.004)	0.263 (0.022)
8q	0.259 (0.005)	1.018 (0.041)	3.921 (0.125)	0.073 (0.008)	0.281 (0.027)
10q	0.293 (0.008)	1.141 (0.030)	3.911 (0.093)	0.098 (0.007)	0.334 (0.021)
12q	0.353 (0.006)	1.334 (0.054)	3.772 (0.110)	0.143 (0.009)	0.403 (0.024)
14q	0.398 (0.006)	1.515 (0.044)	3.797 (0.070)	0.180 (0.010)	0.451 (0.024)

Table 7.3: Average values and their associated errors - in brackets - of transmembrane potentials and anionic and cationic currents and conductances for VDAC-POPC and VDAC-POPE.

The calculated values of conductance fit very well with the experimental ones, reported to be between 3-4 nS[35, 36, 37, 16, 38, 39, 15]. This result not only proves the goodness of the Computational Electrophysiology technique, but also strengthens the reliability of the performed simulations. From these data, it can be inferred that the membrane composition affects the ion conductance through the embedded VDAC. In fact, when VDAC is surrounded by POPE phospholipids, it displays a larger total mean value for the anionic conductance (3.859 ± 0.073 nS) and a smaller one for cation (0.318 ± 0.029 nS) compared to VDAC in POPC (3.601 ± 0.047 nS and 0.366 ± 0.014 nS for anionic and cationic conductance, respectively). Unsurprisingly, for both VDAC-POPC and VDAC-POPE model systems, anionic conductance is preferred with respect to the cationic one, as could be expected for an anion channel.

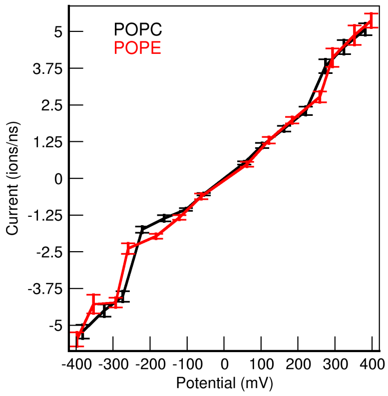


Figure 7.7: Total current expressed as the number of ions which pass through the VDAC channels per ns versus applied potential in mV for VDAC-POPC (black) and VDAC-POPE (red).

Since it is known that the value of transmembrane potential affects the structural properties of VDAC[1, 11, 12, 13, 14], in particular high values of potential lead to a closure of the VDAC channel, therefore to a conformational rearrangement, I evaluated the extent of this conformational change in my CompEl simulations calculating the degree of ellipticity of the VDAC pore for all the simulated transmebrane potentials (see figure 7.8). From these results it can be inferred that the closure of the VDAC channel is favored by high negative potentials, whereas the higher positive ones seems not to alter very much the structure pf the protein, for both POPC and POPE lipid membrane.

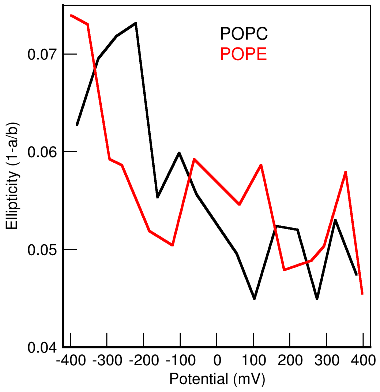


Figure 7.8: Ellipticity values of VDAC-POPC (black) and VDAC-POPE (red) in function of applied potential, obtained in CompEl simulations.

All these results provide important information about the effect that the lipid environment and transmembrane potential have on the behavior of the channel protein VDAC. Further work on this system is in progress, and it relates in particular on the permeation of ATP through the protein, and on the study of VDAC embedded in a lipid membrane whose composition recreate the one of the real outer mitochondrial membrane.

References

[1] V. Shoshan-Barmatz et al. In: *Mol. Aspects Med.* 31 (2010), p. 227.285.

[2] M. Colombini. In: *Biochim. Biophys. Acta* 1818 (2012), pp. 1457–1465.

[3] M. Colombini. In: *Nature* 279 (1979), pp. 643–645.

[4] S. J. Schein, M. Colombini, and A. Finkelstein. In: *J. Membr. Biol.* 30 (1976), pp. 99–120.

[5] M. Colombini. In: *J. Membr. Biol.* 111 (1989), pp. 103–111.

[6] K. Mihara and R. Sato. In: *EMBO J.* 4 (1985), pp. 769–774.

[7] H. Kayser et al. In: *Biol. Chem. Hoppe Seyler* 370 (1989), pp. 1265–1278.

[8] R. Kleene et al. In: *EMBO J.* 6 (1987), pp. 2627–2633.

[9] R. Ujwal et al. In: *Proc. Natl. Acad. Sci. USA* 105 (2008), pp. 17742–17747.

[10] G. E. Schulz. In: *Bioinformatics* 1565 (2002), Biochim. Biophys. Acta.

[11] V. De Pinto et al. In: *FEBS Lett.* 584 (2010), pp. 1793–1799.

[12] T. Hodge and M. Colombini. In: *J. Membr. Biol.* 157 (1997), pp. 271–279.

[13] T. K. Rostovtseva and M. Colombini. In: *Biophys. J.* 72 (1997), pp. 1954–1962.

[14] D Gincel, S. D. Silberberg, and V. Shoshan-Barmatz. In: *J. Bioenerg. Biomembr.* 32 (2000), pp. 571–583.

[15] O. Ludwig et al. In: *Eur. Biophys. J.* 15 (1988), pp. 269–276.

[16] O. Ludwig et al. In: *Biochim. Biophys. Acta* 860 (1986), pp. 268–276.

[17] R. Benz, M. Kottke, and D. Brdiczka. In: *Biochim. Biophys. Acta* 1022 (1990), pp. 311–318.

[18] S. Hiller et al. In: *Science* 321 (2009), pp. 1206–1210.

[19] S. Peng et al. In: *Biophys. J.* 62 (1992), pp. 123–131.

[20] J. Zimmerberg and V. A. Parsegian. In: *Nature* 323 (1986), pp. 36–39.

[21] C. Ader et al. In: *Nat. Struct. Mol. Biol.* 15 (2008), pp. 605–612.

[22] C. Hunte and S. Richers. In: *Curr. Opin. Struct. Biol.* 18 (2008), pp. 406–411.

[23] D. Nietlispach and A. Gautier. In: *Curr. Opin. Struct. Biol.* 21 (2011), pp. 497–508.

[24] R. Phillips et al. In: *Nature* 459 (2009), pp. 379–385.

[25] C. Kutzner et al. In: *Biophys. J.* 101 (2011), pp. 809–817.

[26] H. J. C. Berendsen, D. van der Spoel, and D. van Drunen. In: *Comp. Phys. Comm.* 95 (1995), pp. 43–56.

[27] J. B. Klauda et al. In: *J. Phys. Chem. B* 114 (2010), pp. 7830–7843.

[28] A. D. MacKerell et al. In: *J. Phys. Chem.* 102 (1998), pp. 3586–3616.

[29] M. G. Wolf et al. In: *J. Comput. Chem.* 31 (2011), pp. 2169–2174.

[30] H. J. C. Berendsen et al. In: *J. Chem. Phys.* 81 (1984), pp. 3684–3690.

[31] T. Darden, D. York, and L. Pedersen. In: ().

[32] N. Kucerka, S. Tristram-Nagle, and J. F. Nagle. In: *J. Membr. Biol.* 208 (2006), pp. 193–202.

[33] M. Rappolt et al. In: *Eur. Biophys. J.* 31 (2003), pp. 575–585.

[34] U. Zachariae et al. In: *Structure* 20 (2012), pp. 1540–1549.

[35] E.-M. Krammer, F. Homble, and M. Prevost. In: *Biochim. Biophys. Acta* 1828 (2012), pp. 1284–1292.

[36] N. Roos, R. Benz, and D. Brdiczka. In: *Biochim. Biophys. Acta* 686 (1982), pp. 204–214.

[37] H. Troll et al. In: *J. Biol. Chem.* 267 (1992), pp. 21072–21079.

[38] T. K. Rostovtseva and S. M. Bezrukov. In: *Biophys. J.* 74 (1998), pp. 2365–2373.

[39] H. Freitag, W. Neupert, and R. Benz. In: *Eur. J. Biochem.* 123 (1982), pp. 629–636.

8 Conclusions

8.1 Conclusions

My PhD work aimed at unveiling the effects of the environment on the dynamics and functionality of four different types of proteins, using a Molecular Dynamics approach. The first group of investigated proteins was a set of cupredoxins, namely native azurin and four of its mutants (AZAMI, AZPC, AZ4A3A and AZ4A4A), in which the hydrophobic ligand-containing loop has been replaced with that of other cupredoxins (amicyanin and plastocyanin) or with synthethic sequences formed only by alanines. The results revealed that the outer-sphere reorganization energy in cupredoxins, does not depend only on the solvent accessible surface of the active site, but also on the lenght and composition of the small ligand-containing loop region, which influence the dynamics of the whole molecule as well as long-range solvent reorientation. The above molecular features are responsible of the protein/solvent interplay, regulating the extent of reorganization energy. Moreover, we unravelled the molecular effectors of the acid transition in cupredoxins, showing that the solvent accessible surface (SASA) of the active site largely affects the pK_a of the C-terminal His ligand on the ligand-containing loop, which is known to trigger the whole conformational transition.

The water layer surrounding the protein active site is another factor that determines the ability of the loop mutants to undergo the acid transition, in contrast with native azurin. Indeed, water molecules can approach the active site closer in the loop mutants than in native azurin, thus increasing the probability of the acid transition. The FMA investigation allowed us to identify the protein regions which mainly contribute to the SASA fluctuations, resulting in a enhanced freedom of movement of the protein loops surrounding the active site in the loop mutants compared to azurin, as a consequence of the loss of some crucial second-coordination sphere interactions.

We have presented a computational approach, PMM, which proved to be a robust tool to assign reduction potentials in multiheme proteins, and to investigate the thermodynamics of the electron flow. This approach was succesfully employed to assign the experimental reduction potentials to each heme group in the diheme cytochrome *c* from *Rb. sphaeroides.* Although we benchmarked our method against a system featuring two hemes, this approach can be extended to more complex systems featuring multiple redox centres. We also show that PMM/MD approach to estimate the E^0 of each redox center in multiheme species allows insight to be gained into the intermolecular interactions within complex ET systems.

The reliability of the PMM method was further confirmed by calculating the reduction potentials for human neuroglobin, with and without internal disulfide bridge between Cys46 and Cys55, and GLB-6 from *Caenorhabditis elegans*, which turn out to be in very good agreement with the experimental values.

The MD simulations and analysis performed on these globins agreed with the hypothesis that human neuroglobin and GLB-6 act as electron transfer proteins, because of their stable and rigid structures (indicated by RMSF calculations) and of their small ET reorganization energies (comparable with that of cytochromes or cupredoxins). The above hypothesis is further strengthened by the results concerning the adduct between human neuroglobin and its putative partner cytochrome *c*, obtained performing MD simulations on the complex before (pre-ET; i.e., reduced neuroglobin and oxidized cytochrome *c*) and after (post-ET; i.e., oxidized neuroglobin and reduced cytochrome *C*) electron transfer. These simulations investigated the electron transfer process only from a thermodynamic point of view, studying the effect(s) that the change of the charge of the two proteins has on the behavior, and in particular on the stability, of the adduct. Indeed, we found that a large set of weak interactions is present in the pre-ET complex, most of which disappear in the post-ET form. Moreover, the number of water molecules which can interpose between the two proteins in pre-ET form is definitely smaller than the post-ET form, indicating that the change of the charges of the two proteins largerly affects the stability of the complex. These results suggest that *in vivo*, electron transfer process is possible only between the reduced form of human neuroglobin and the oxidized human cytochrome *c*.

The above results concerning six-coordinate globins provide molecular evidences that strenghten the hypothesis that these proteins can act as electron transfer proteins *in vivo.* Moreover, they provide informations on the dynamic effects of the disulfide bridge on human neuroglobin, opening the road for future studies.

The key role of the environment in determining the function and dynamics of proteins was further analyzed by investigating the electrophysiological behavior of the Voltage-Dependent Anion Channel (VDAC), an important membrane protein localized in the outer mitochondrial membrane, embedded into two different lipid bilayers, POPC and POPE. The obtained results suggest that the lipid membrane compostion and the extent of the transmembrane potential affect the ion conductance and selectivity through the VDAC and its ability to undergo conformational rearrangement. In particular, VDAC in POPE lipids shows higher anionic and smaller cationic conductances compared to POPC bilayer, although in both VDAC-POPE and VDAC-POPC anionic conductance is higher than the cationic one, as could be expected for an anion channel.

All these results help to better understand how the functionality of a protein is influenced by its dynamics and by its interaction with the surrounding environment.