

Review article

Dioxygen affinity in heme proteins investigated by computer simulation

Marcelo A. Marti, Alejandro Crespo, Luciana Capece, Leonardo Boechi,
Damián E. Bikiel, Damián A. Scherlis, Dario A. Estrin *

*Departamento de Química Inorgánica, Analítica y Química Física/Inquimae-Conicet, Facultad de Ciencias Exactas y Naturales,
Universidad de Buenos Aires, Ciudad Universitaria, Pabellón 2, Buenos Aires, C1428EHA, Argentina*

Received 29 November 2005; accepted 13 December 2005
Available online 26 January 2006

Abstract

We present an investigation of the molecular basis of the modulation of oxygen affinity in heme proteins using computer simulation. QM-MM calculations are applied to explore distal and proximal effects on O₂ binding to the heme, while classical molecular dynamics simulations are employed to investigate ligand migration across the polypeptide to the active site. Trends in binding energies and in the kinetic constants are illustrated through a number of selected examples highlighting the virtues and the limitations of the applied methodologies. These examples cover a wide range of O₂-affinities, and include: the truncated-N and truncated-O hemoglobins from *Mycobacterium tuberculosis*, the mammalian muscular O₂ storage protein: myoglobin, the hemoglobin from the parasitic nematode *Ascaris lumbricoides*, the oxygen transporter in the root of leguminous plants: leghemoglobin, the *Cerebratulus lacteus* nerve tissue hemoglobin, and the *Alcaligenes xyloxydans* cytochrome *c'*.

© 2005 Elsevier Inc. All rights reserved.

Keywords: Heme protein; DFT; QM-MM; Molecular dynamics; Oxygen binding

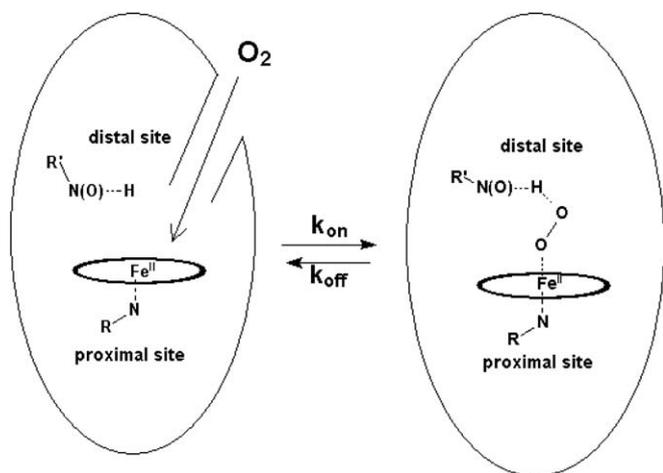
1. Introduction

Heme proteins, the family of proteins containing an iron–porphyrin complex as a prosthetic group, are found in all living organisms [1,2]. They perform a wide variety of tasks ranging from electron transport [3] to oxidation of organic compounds [4], to the sensing and transport of small molecules, namely O₂, CO and NO [5]. Out of these three ligands, O₂ is the most abundant and at the same time the one with the lowest affinity for free heme [6]. The regulation of oxygen affinity is, therefore, one of the key issues determining a heme protein's function. In most heme proteins, the active site—where the oxygen can be bound—consists of a cavity on top of the heme group known as the distal pocket. The iron atom is coordinated equatorially to the four nitrogens of the porphyrin macrocycle, and axially to a fifth or proximal ligand under the heme ring, typ-

ically a His, Cys or Tyr residue (Scheme 1). The oxygen affinity of a given heme protein, characterized by the equilibrium constant K_{O_2} , is determined by the ratio between the kinetic constants of the association and dissociation processes, k_{on} and k_{off} respectively.

Association rate constants in heme proteins depend essentially on two factors: the accessibility of O₂ to the active site, and migration of the oxygen molecule across the protein matrix [7]. Typical values span a range of three orders of magnitude, starting at about $10^5 \text{ M}^{-1} \text{ s}^{-1}$ in those systems with restrained accessibility to the iron, up to $10^8 \text{ M}^{-1} \text{ s}^{-1}$ when the association rate is mainly controlled by the diffusion in the solvent [7]. On the other hand, dissociation rate constants span a range of seven orders of magnitude, from 10^{-3} s^{-1} to 10^4 s^{-1} [7]. The dissociation rate constant is mainly determined by the nature of the interaction between the O₂ and the heme protein, which depends upon the strength of the iron–oxygen bond, and on interactions involving the protein matrix, typically the H-bond interactions of the bound O₂ ligand with the surrounding distal aminoacids (Scheme 1). A mere inspection

* Corresponding author. Tel.: +54 11 4576 3378; fax: +54 11 4576 3341.
E-mail address: dario@qi.fcen.uba.ar (D.A. Estrin).



Scheme 1. Graphical representation of the oxygen binding to heme-proteins.

of these factors, though, is not always enough to anticipate the O_2 affinity, as demonstrated by the case of two recently discovered heme proteins, neuroglobin and cytoglobin, which are able to bind oxygen despite the fact that both have a distal coordinated histidine blocking the active site [8].

As mentioned above, the function of a heme protein is intimately related to its oxygen affinity. Typically, O_2 transport proteins (such as myoglobin or hemoglobin) display moderate oxygen affinities ($K_{O_2} \approx 1 \mu M^{-1}$) and large binding rate constants ($10^7 M^{-1} s^{-1}$), ensuring the protein is easily loaded with oxygen in regions of high O_2 pressure. At the same time, moderate to high dissociation rate constants ($1-10 s^{-1}$) allow for the oxygen to be released under low O_2 pressure conditions. On the other hand, heme proteins acting as NO dioxygenases, such as *Ascaris* Hemoglobin (AscHb) [9] or the Truncated-N Hemoglobin (TrHbN) from *Mycobacterium tuberculosis* [10], display high oxygen affinities with low to very low dissociation rate constants ($0.001-0.1 s^{-1}$). This combination ensures that the protein is loaded with oxygen even at very low O_2 concentrations, providing a functional protein in anaerobic environments and thereby protecting the pathogenic microorganism from the mammalian immune response. Finally, let us mention the group of small-ligand sensing proteins, which can be divided in two. O_2 sensors, such as Fix-L [11] or HemAT [12], display a finely tuned oxygen affinity that allows a precise response to a given oxygen concentration. On the other hand, CO or NO sensors, such as the mammalian soluble guanylate cyclase (sGC) [13], have little or no oxygen affinity. These proteins must exhibit very high O_2 -dissociation rate constants to remain free in aerobic environments. From this picture, it is clear that understanding oxygen affinity and the precise factors governing it, is one of the main issues in understanding heme-proteins chemistry and biology.

Computational techniques for modeling biomolecules have emerged as an important tool to complement

experimental information [14–19]. The in silico generated models and data are essential for analyzing structural and kinetic details that are difficult to capture experimentally. Computer simulations also provide a systematic and economical tool to analyze the dependence of the property of interest on the biomolecular structure. Such studies are an invaluable tool for our understanding of biological function through the structure–function connection. As mentioned above, the study of oxygen affinity requires the examination of two factors: first, the accessibility and diffusion of the O_2 ligand to the protein's active site, and second, the evaluation of the interaction of the bound O_2 molecule with the protein. The first problem requires the assessment of the barriers for oxygen migration from the solvent to the active site. To evaluate this, extensive sampling of the protein conformational space and of the ligand molecule along the migration path is required. The method of choice to accomplish this is to perform classical molecular dynamics (MD) simulations combined with reaction-coordinate enhanced sampling methods, to establish the free energy landscape along the selected path [20–22]. The second problem requires the estimation of the iron–oxygen bond energy, plus the strength of the interaction between the bound oxygen and the distal residues. Since this is a reactive process (i.e. it involves the breaking-formation of bonds) quantum mechanical (QM) methods are needed. Hybrid quantum-mechanical molecular-mechanical schemes (QM–MM), in which the heme group and the axial ligands are studied at the QM level and the rest of the protein is treated classically, are ideal tools to tackle this problem.

In this work we present a variety of examples illustrating how computer simulation can help to understand the different aspects governing oxygen affinity of heme proteins. These examples comprise: the truncated N hemoglobin (TrHbN) and the truncated-O (TrHbO) hemoglobin from *M. tuberculosis* [10]; the *Alcaligenes xylosoxidans* cytochrome *c'* (AXCP) [23]; the oxygen transporter in the root of leguminous plants, leghemoglobin (Lba) [2]; the mammalian muscular O_2 storage protein myoglobin (Mb) [2]; the *Cerebratulus lacteus* nerve tissue Hb (CerHb) [24]; and the hemoglobin from the *Ascaris lumbricoides* (AscHb) [25]. The reported results shed light on the structural–functional relationships that control oxygen affinity and the kinetic constants, k_{on} and k_{off} . The simulation data is thoroughly contrasted with experimental evidence, while discussing the successes and limitations of the methods employed.

2. Computational methods

2.1. Initial structures and equilibration

We constructed initial structures for all our calculations from the crystal structure of the proteins involved in the present work: trHB, AXCP, Lba, Mb, CerHb and AscHb (PDB codes 1IDR [26], 1E85 [27], 1BIN [28], 2MGM

[29], 1KR7 [30] and 1ASH [31], respectively). Histidine protonation was assigned favoring H-bond formation. The systems were solvated with water molecules up to a distance of 30 Å from the heme center. All classical simulations were performed with the Amber8 package [32], using the Amber force field parametrization [33]. The systems were heated to 300 K and 500 ps molecular dynamics were performed to allow for relaxation and equilibration.

Mutant models were built in silico by replacing in the original structure the wild type residue with the corresponding mutated one. The same protocol applied to the wild type was used in the equilibration of the mutant structures.

2.2. Molecular dynamics simulations

All molecular dynamics simulations were performed using the Amber8 program [32]. For non-bonded interactions a 12 Å cut-off radius was employed. The simulations were performed at 300 K using the Berendsen thermostat [34] with a time step of 1 fs. Equilibration was followed by the corresponding MD runs. The partial charges for the oxygenated heme moiety were obtained according to the standard protocol used for the Amber force field [35]. This set of partial charges for the oxygenated heme group has been successfully employed in previous MD simulations of heme proteins [36].

If the free energy barriers are of the same magnitude as the thermal fluctuations, it is feasible to obtain the free energy profiles associated with a given process directly from classical MD simulations. In these cases an adequate sampling of the relevant configurations may be achieved in accessible simulation times, and the free energy profile can be obtained by computing the probability distribution along a selected reaction coordinate, $P(\xi)$:

$$-\beta A(\xi) \approx \ln[P(\xi)] \quad (1)$$

where $\beta^{-1} = k_B T$ is the Boltzmann constant times the temperature and A is the free energy. The free energy profiles are directly related to the experimental results, since the calculation takes into account thermal and entropic effects. However, the obtained predictions will be dependent on the employed models, specifically on the force field parametrization. This methodology was used in the calculation of the free energy difference between the two alternative conformations of *Cerebratulus lacteus* hemoglobin (CerHb).

2.3. Multiple steering molecular dynamics

The multiple steering molecular dynamics (MSMD) approach, originally proposed by Jarzynski [37], is a novel technique for computing free energy profiles when they cannot be obtained directly from MD simulations. The MSMD approach establishes the following relation between the non-equilibrium dynamics and equilibrium properties:

$$\exp[-\Delta A(\xi)/k_B T] = \langle \exp[-W(\xi)/k_B T] \rangle \quad (2)$$

where $W(\xi)$ is the external work performed on the system as it evolves from the initial to the final state through the reaction coordinate ξ . The Hamiltonian can be written as the sum of the time independent Hamiltonian of the unperturbed system plus a time-dependent external potential, usually harmonic. Thus, the free energy of a process along a reaction coordinate can be computed performing a number of finite time transformations, collecting the work done at each time step, and then properly averaging it as in Eq. (2) [38,39].

We use the MSMD approach to investigate the ligand migration process along the tunnel/cavity system of the CerHb and TrHbO proteins. The chosen reaction coordinate was the iron–ligand distance, with a force constant of 200 kcal/(mol Å²). We performed different sets (10 and 20 sets) of runs with pulling velocities of 0.05 Å/ps and 0.1 Å/ps respectively. In each case, the shown data corresponds to two independent free energy profile estimates using a converged set of MSMD runs [39].

2.4. QM–MM methods

The initial structures for the QM–MM calculations were taken from the MD simulations; starting from an equilibrated snapshot, the system was cooled down slowly to 0 K. This process was followed by fully hybrid QM–MM geometry optimizations using a conjugate gradient algorithm. Only residues located less than 10 Å apart from the heme reactive center were allowed to move freely in the QM–MM runs. All QM–MM computations were performed at the DFT level with the SIESTA code [40]. The SIESTA code has shown an excellent performance for medium and large systems, including biomolecules, and heme models in particular [41,42]. The use of standard norm-conserving pseudopotentials [43] avoids the computation of core electrons, smoothing at the same time the valence charge density. In our study, the nonlinear partial-core correction [44] is applied to the iron atom. Basis functions consist of localized (numerical) pseudoatomic orbitals, projected on a real space grid to compute the Hartree and the exchange-correlation potential matrix elements. For all atoms, basis sets of double zeta plus polarization quality were employed, with a pseudoatomic orbital energy shift of 25 meV, and a grid cutoff of 150 Ry [40]. Calculations were performed using the generalized gradient approximation to the exchange-correlation energy proposed by Perdew et al. [45]. Such combination of exchange-correlation functional, basis sets, and grid parameters has been already validated for heme models [36,42]. The iron porphyrinate and the relevant ligands were selected as the quantum subsystem. The rest of the protein unit, together with water molecules, was treated classically, using the Amber99 force field parametrization [35]. The interface between the QM and the MM portions of the system was treated by the scaled position link atom

method [46], adapted to our SIESTA code. Further technical details about the QM–MM implementation can be found elsewhere [47].

Finally, we have performed for all proteins QM–MM calculations of the O₂ binding energy (ΔE), which is defined as:

$$\Delta E = E_{O_2\text{-Heme}} - (E_{O_2} + E_{\text{Heme}}) \quad (3)$$

where $E_{\text{Heme-O}_2}$ is the energy of the oxy-form of the protein, E_{Heme} is the energy of the deoxy-form, and E_{O_2} is the energy of the isolated oxygen molecule. All the O₂ affinity calculations reported in this work were calculated as stated above.

It is worth mentioning at this point that DFT exhibits a bias in the description of the spin-state energetics of iron-porphyrins, in general favoring low spin configurations [48–52]. In particular, DFT energies for the free heme—which electronic structure corresponds to a high spin state—will be somehow overestimated with respect to the oxy-form—which presents a low spin ground state. As a consequence, binding energies calculated as in Eq. (3) may be flawed, leading to an apparent overbinding of O₂. However, we want to stress that even though the estimates for ΔE may be sometimes above the experimental values, the predicted trends are pretty much in line with the results of experiments. Precisely in this ability to correctly predict the trends in O₂ affinity—more than the total energies—resides the strength and value of DFT.

3. Results and discussion

3.1. The oxygen binding energy and dissociation rate constants

3.1.1. High-affinity heme proteins and H-bonding: the distal effect

Most heme proteins with very high O₂-affinity have in common the presence of several H-bond donor residues in the distal cavity. For example, the *M. tuberculosis* TrHbN, with a $k_{\text{off}} = 0.2 \text{ s}^{-1}$, has two distal H-donor residues (TyrB10 and GlnE11), as shown in Fig. 1A [53]. TrHbN is an NO dioxygenase that catalyses NO detoxifica-

tion to the innocuous nitrate, according to the reaction $\text{Fe}^{\text{II}}\text{O}_2 + \text{NO} \rightarrow \text{Fe}^{\text{III}} + \text{NO}_3^-$. As already mentioned, a high oxygen affinity ensures that the protein is loaded with oxygen even at low O₂ concentrations and therefore still functional. Mutational studies pointed to TyrB10 as the main factor responsible for the high oxygen affinity, since mutations of TyrB10 to Phe or Ala increased the k_{off} by two orders of magnitude [53].

We performed QM–MM calculations of the O₂ binding energy on this system for the native and (distal) mutant proteins in order to characterize the role of the distal residues. The starting point for our simulations was the X-ray structure of the wild type of TrHbN in the oxy-form. The TyrB10 → Ala mutant was constructed in silico, replacing the corresponding aminoacid in the wild type structure. Our results confirm the pivotal role of TyrB10 in the O₂-affinity of TrHbN, yielding for the mutant binding energies lower than for the native protein, consistently with the experimental data (see Table 1) [53].

The hemoglobin of the parasitic nematode *A. lumbricoide* (AscHb) constitutes another well known case of a high O₂-affinity heme protein [25,54]. Our calculations reveal a complex hydrogen bonding network, depicted in Fig. 1B. The hydroxyl group from TyrB10 forms the main hydrogen bond with the dioxygen, while one of the hydrogens of the GlnE7 residue is involved in multiple H-bonds, with the O₂ and the TyrB10. This rigid structure, already reported in a previous study [54], favors the oxy form of the AscHb. The magnitude of the binding energy (Table 1) is typical of high oxygen-affinity hemoglobins. It is interesting to note that although in AscHb the O₂ binding Energy is high and comparable to that of TrHbN the k_{off} value is still 2 orders of magnitude lower. Given the almost-identical distal environments (both proteins display the Tyr–Gln pair) it is expected that they have a similar oxygen binding energy. However, the lower k_{off} value in AscHb points towards an additional mechanism responsible for the unusually low value of k_{off} . In this context, a trapping cage mechanism has been proposed to account for this fact [55].

To further explore the capabilities of the method, we examined and compared the distal effect in heme proteins

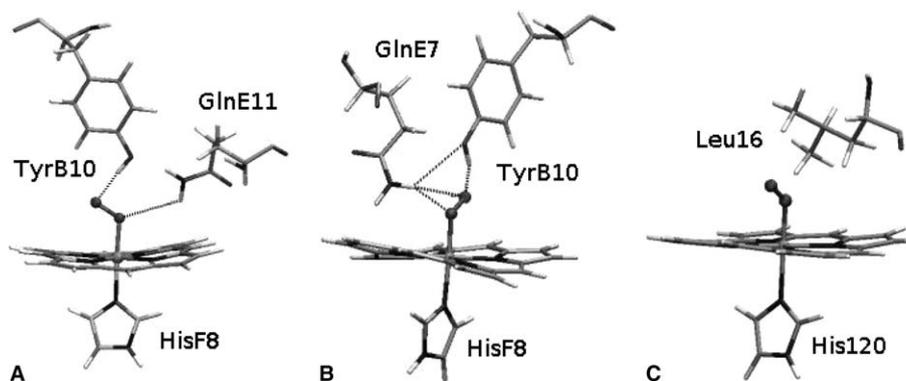


Fig. 1. Active sites of TrHbN (A), AscHb (B) and cyt *c'* (C).

Table 1
Oxygen affinity in high, moderate and low O₂ affinity heme proteins

Protein	k_{off} (s ⁻¹)	ΔE (kcal/mol)
TrHbN (wt)	0.2 ^a	37.2
TrHbN (TyrB10 → Ala)	45 ^b	29.8
Mb (wt)	12 ^a	27.0
Mb (HisE7 → Gly)	1600 ^c	22.7
Lba (wt)	5.6 ^a	29.3
Lba (HisE7 → Gly)	3.6 ^c	28.9
Lba (HisE7 → Gly, HisF8 → Gly + Im)	– ^e	24.9
Cyt <i>c'</i> (wt)	– ^f	8.4
CerHb	180 ^d	28.6
CerHb(ThrE11 → Val)	0.18 ^d	32.0
AscHb	0.004 ^a	34.3
Free Heme	– ^f	22.0

^a Ref. [10].

^b Ref. [53].

^c Ref. [60].

^d Ref. [24].

^e This mutation has not been done experimentally so the k_{off} value is not available.

^f Cyt *c'* does not bind oxygen therefore no k_{off} value is available. The free heme group only binds oxygen at very low temperatures in organic media.

displaying moderate oxygen affinity. These proteins, mainly O₂ carriers such as the mammalian muscular O₂ storage protein myoglobin, or the oxygen transporter hemoglobin (Hb), also present H-bond donor residues in the distal site (Fig. 2A and B) [2]. The number and nature of these residues, however, differ from those in TrHbN or AscHb. Carriers as Mb typically display only one His at position E7 as the H-donor residue. Our results for these proteins show that, consistently with the lower experimental affinities, the computed O₂ binding energies are

~10 kcal/mol below the one obtained for TrHbN. Also in accordance with the experimental data on Mb distal mutations, the distal HisE7 → Gly Mb mutant has an O₂ binding energy about 4 kcal/mol lower than wild type Mb (Table 1).

Finally, we investigated the O₂ binding energy in a protein with very low affinity. As mentioned above, the NO sensor sGC must remain deoxygenated in an aerobic environment in order to perform its function [13]. Since the X-ray structure of sGC is unknown, we chose instead a protein with similar behavior: the bacterial cytochrome *c'* (cyt *c'*). This heme protein is found in the periplasm of certain denitrifying, photosynthetic bacteria [56–58], which unlike most *c*-type cytochromes, contains a pentacoordinated heme. While the exact physiological role of cyt *c'* is unclear, several studies have suggested that its function is related to NO binding [59]. Additional interest in the chemistry of cyt *c'* stems from the similarities with sGC, mainly the fact that none of them binds oxygen. Cyt *c'* lacks any H-bond donor groups in the distal cavity, and in addition, it displays a sterically crowded distal pocket due to the presence of the Leu16 side chain above the iron (Fig. 1C). Our results explain this behavior, showing that O₂ binding energy for wild type cyt *c'* is only 8.4 kcal/mol (Table 1). Moreover, mutation of Leu 16 to glycine increases the binding energy to 23.3 kcal/mol, a value similar to that observed in free heme but lower than the one found in Mb (Table 1) [42].

Our calculations confirm that the distal stabilization effect (operative through the H-bond donor residues) is a key factor in determining O₂ affinity. They also show that a comparative analysis of this effect, achieved by the mutation of a particular residue in the X-ray structure, is possible by means of QM–MM simulation techniques.

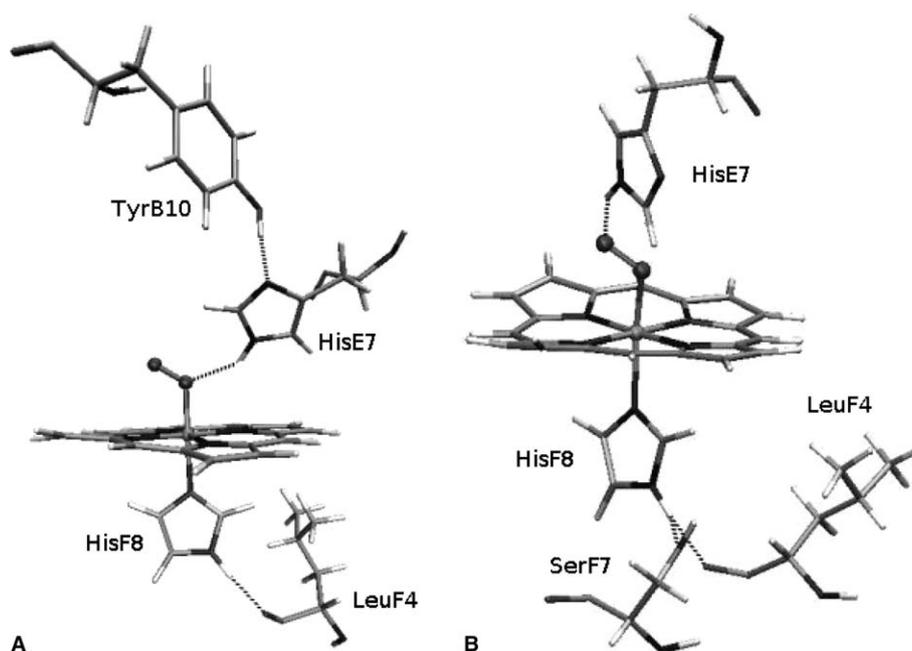


Fig. 2. Active site of Mb (A) and Lba (B).

3.1.2. Influence of the axial histidine local structure: the proximal effect

As mentioned above, the O₂ transport proteins Mb and Lba display a moderate oxygen affinity, but despite their similar distal environment (Fig. 2), they have different k_{off} values, 15 and 5.6 s⁻¹ respectively [60]. Moreover, mutational analysis revealed that distal mutations (on His E7) have a crucial impact in Mb whereas in Lba no effect is observed [2]. On the other hand, experimental evidence showed that it is possible to reduce Lba oxygen affinity by mutating the proximal histidine to glycine, and adding imidazole at the same time to occupy the sixth coordination site [60]. The opposite effect is observed on Mb in the same experiment. These experiments make manifest a significant influence of the proximal environment on oxygen affinity.

In order to investigate the role of the proximal effect we performed a comparative analysis between distal and proximal effects in Lba. First of all we obtain a slightly larger O₂ binding energy for Lba as compared to that of Mb, consistent with its somewhat lower k_{off} . Furthermore mutation of HisE7 has almost zero impact on the binding energy. On the other hand, also consistently with the experimental data, the O₂ binding energy of further mutating the proximal histidine to glycine and adding an extra imidazol, decreased by 4.0 kcal/mol (Table 1). Clearly, by constraining the histidine in a particular position, Lba is able to enhance its oxygen affinity through a proximal effect.

Visual inspection of the optimized structures for wt and mutant Lba, allowed us to understand how this enhancement is achieved. In wt Lba, the imidazole plane is oriented in a staggered conformation with respect to the pyrrolic nitrogens. In Mb it is in an eclipsed one (Fig. 3). The staggered conformation allows a better charge donation from the histidine to the iron. This is evidenced from the Mulliken charge population analysis that shows that in Lba the proximal histidine donates 0.182e to the iron whereas in Mb a value of 0.146e is observed. The larger charge density on the iron enhances the π -backdonation from the iron to the oxygen ligand, thereby increasing the Fe–O bond

strength. This is reflected in the O₂ net Mulliken charges of $-0.214e$ and $-0.289e$ for Mb and Lba, respectively.

3.2. Multiple distal site conformations: dynamical effects

The *C. lacteus* nerve tissue Hb (CerHb) is, with 109 residues, the smallest naturally occurring known Hb [30]. It has a distal cavity with the typical H-bond donors TyrB10 and GlnE7 found in the high O₂-affinity heme proteins already mentioned (Fig. 4). However, in contrast to the cases discussed above, the affinity of CerHb for O₂ is only moderate ($K_{\text{O}_2} = 1 \mu\text{M}^{-1}$), with an O₂ dissociation rate unexpectedly high ($k_{\text{off}} = 180 \text{s}^{-1}$). The molecular cause for such apparent discrepancy was recently investigated through the study of point mutations at the distal site residues TyrB10, GlnE7, and ThrE11. Strikingly, the isosteric ThrE11 → Val mutation showed a spectacular 1000-fold decrease in the rate of O₂ dissociation [24]. It was proposed that in wt CerHb, ThrE11 is able to redirect the TyrB10 phenolic OH group away from the ligand, thereby destabilizing the heme-bound O₂. The ThrE11 → Val mutation restores the availability of the OH moiety to be hydrogen bonded to the O₂ ligand, thus reducing dramatically the k_{off} value [24]. By performing MD simulations we see that two main conformations can be detected in the wt cerHb. The first conformation (conformer 1) shows TyrB10 phenolic OH group pointing towards the heme-Fe-bound O₂ molecule, whereas in the second conformation (conformer 2), the hydrogen bond between TyrB10 and the O₂ ligand is broken, and the phenolic OH group is oriented towards the OG1 atom of residue ThrE11 (Fig. 4). On the other hand, in the ThrE11 → Val mutation only conformation 1 is present.

The oxygen binding energy for wild type CerHb conformer 1 is 33 kcal/mol. A similar value of 32 kcal/mol is obtained for the ThrE11 → Val mutant. As expected by the inspection of the H-bonding patterns, the O₂ binding energy of conformer 2 is smaller (24 kcal/mol). We also performed free energy calculations to predict the relative energy and interconversion barrier for the two conformers,

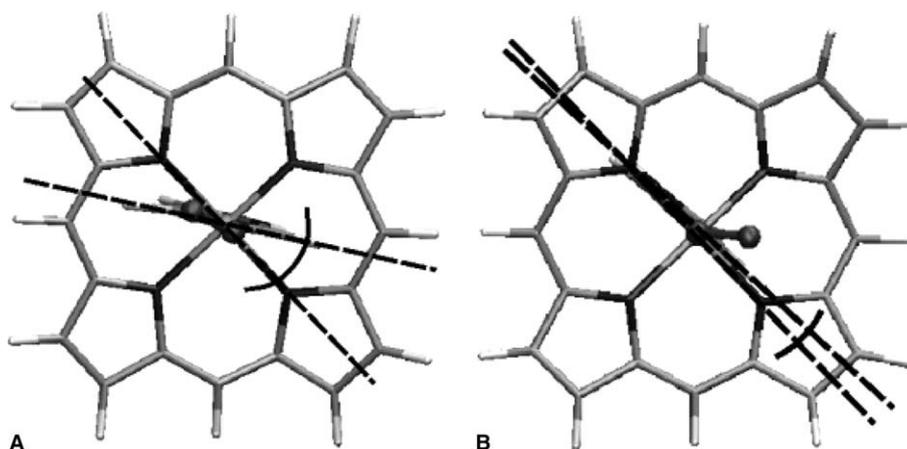


Fig. 3. Histidine rotational position in Mb and Lba.

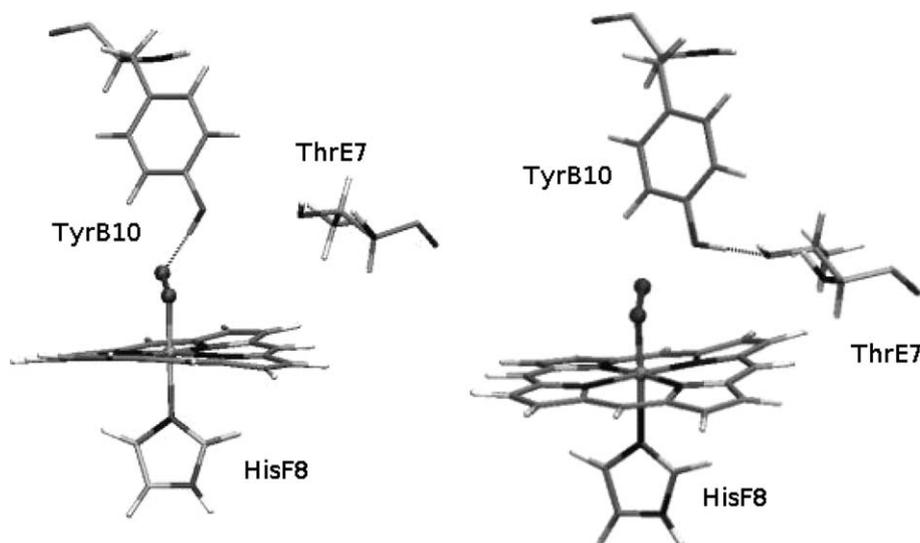


Fig. 4. Alternative Conformers in CerHb. Conformers 1 and 2 (right and left panels, respectively).

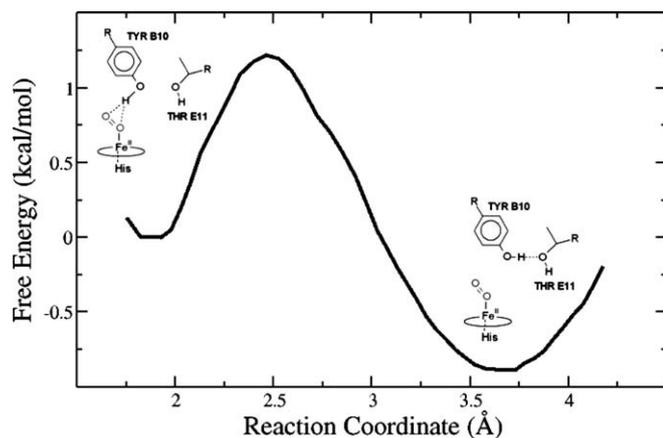


Fig. 5. Free energy profile for the interconversion of the CerHb conformers.

(Fig. 5) finding a low barrier (2 kcal/mol) with both conformations almost isoenergetical. The O_2 affinity value observed experimentally for CerHb [24] is thus expected to correspond to the average of both values, weighted according to the relative populations of the two conformers. If, based on Fig. 5, we assume a relative weight of 50% for each conformer, an estimative value of 28.6 kcal/mol is obtained for the O_2 binding energy. The difference in O_2 binding energies between the wt protein and the ThrE11 \rightarrow Val mutant is also consistent with the experimentally observed difference in the k_{off} values. Finally, it is interesting that this dynamic equilibrium between two distinctive distal states predicted by the MD simulations was unnoticed in the experimental results, and points to a novel mechanism to control ligand affinity [61].

3.3. Small ligand migration in heme proteins

As mentioned in the introduction, a central step determining the ligand association rate in heme proteins is the

diffusive process from the solvent to the active site. Interestingly, the internal structure of heme proteins usually reveals a number of hydrophobic tunnel-connected cavities with enough volume to host small gaseous ligands (about 100 \AA^3) [10]. Since these packing defects are associated with decreased thermodynamic stability, there is a wide interest in understanding their role and the reason why they are preserved throughout evolution. Myoglobin X-ray structures of photolyzed MbCO have shown intermediate states in which the CO is transiently located in the different cavities [62]. Moreover, kinetic analysis of different mutations affecting the size of the cavities indicate that they influence the rate of ligand entry/escape [7]. In the TrHb group of proteins, for example, this tunnel-cavity system is topologically conserved and can be thought as a pathway connecting the solvent with the protein's active site [63]. These channels linking the solvent and the active center are also evident in CerHb. X-ray data of samples with Xenon atoms, and Fourier transform infrared temperature derivative spectroscopy (FTIR-TDS), have revealed that these cavities can trap small ligands and that protein motions affect the dynamics of ligand migration along the tunnels [10,24].

To analyze the connection between the tunnel cavity system and the oxygen migration process, we computed the free energy profile along the tunnel for trHbO and CerH. These two proteins display significantly different oxygen association rate constants: the values of k_{on} are $1.1 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ for TrHbO and $2.4 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$ for CerHb.

From Fig. 6 it is clear that the tunnel is quite different in both proteins. The main difference seems to be due to the presence of three residues in TrHbO: Trp88, Met92 and Leu48. This aminoacids are occluding the channel and thus generating a barrier to ligand transport. On the other hand, in CerHb no discontinuities are observed in the tunnel connecting the iron with the solvent.

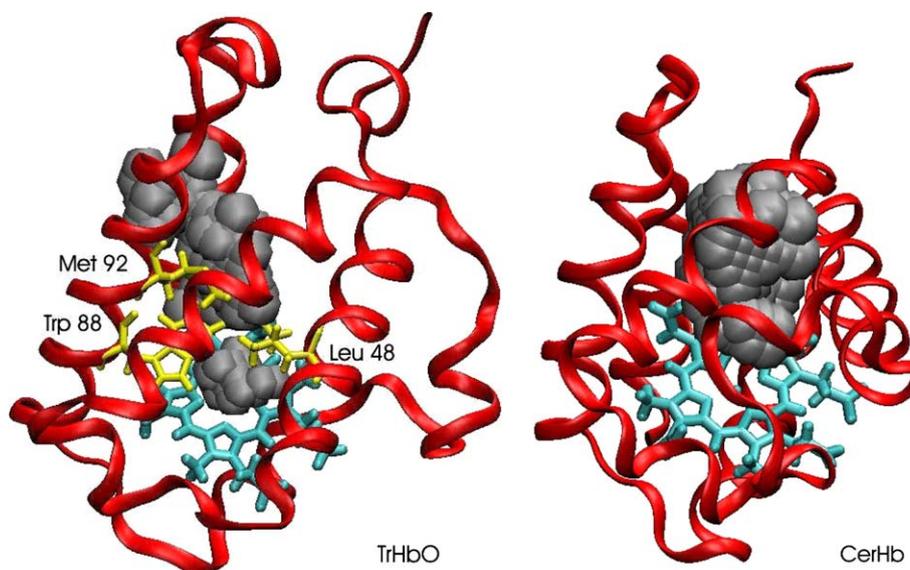


Fig. 6. Cavity system for *M. tuberculosis* TrHbO and for CerHb calculated from the corresponding X-ray structures using a probe of 1.4 Å.

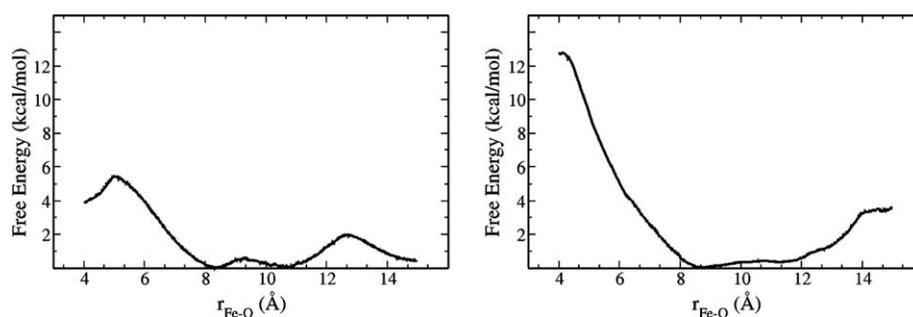


Fig. 7. Free energy profile (kcal/mol) for O₂ migration along the tunnel in CerHb (left panel) and trHbO (right panel).

In CerHb the profile shows that there is a small barrier (about 2 kcal/mol) that separates the ligand in water from two secondary docking sites, separated by a very small barrier. The largest barrier for oxygen binding to the iron is about 5 kcal/mol, and corresponds to a rearrangement of the distal cavity residues located at about 5–7 Å of the iron. It is noteworthy that the presence of secondary docking sites identified in the profile for CerHb has also been evidenced through FTIR-TDS [24]. In TrHbO the picture shows that there is a 3 kcal/mol free energy stabilization for the ligand inside the protein respect to the solvent, without any appreciable entry barrier. There is only one big tunnel-cavity of around 5 Å long (7.5–12.5 Å) in the reaction-coordinate free energy profile. The most striking difference between these two proteins is the fact that in TrHbO the tunnel is not continuously connected with the heme active site, presenting a barrier of more than 10 kcal/mol at about 6 Å in the Fe–O distance. The nature of this barrier can be ascribed to the presence of the TrpG8 side chain, which lies in close Van der Waals contact with LeuE15 and MetG12, blocking the tunnel. Interestingly, TrpG8 is conserved in all the TrHbO proteins sequenced so far [10]. The predicted barriers in TrHbO and CerHb,

10 and 5 kcal/mol respectively, are consistent with the experimental trends in the association rate constants (see Fig. 7).

4. Conclusions

The present results make apparent the contribution of computational classical and QM–MM techniques to the exploration of the molecular basis of oxygen affinity of heme proteins. In particular, ligand migration paths can be investigated with classical MD simulations in combination with advanced sampling tools, such as MSMD, to yield information about free energy barriers and possible secondary docking sites. On the other hand, QM–MM tools are specially suited to investigate the protein influence (proximal and distal effects) on oxygen binding energies to the heme. These two frameworks, by covering different time and space scales, complement each other in providing a complete picture of the mechanism of O₂ binding to the protein. The agreement with the experimental data, when available, constitutes a stringent test to the reliability of these approaches. Computational simulation is a precious tool even if it is often used to validate and contrast exper-

imental observations, because it offers a microscopic view, sometimes resolved in real time, as no other method is capable to provide. With the refinement of the methodology and the hardware, classical and QM–MM simulation will be more and more at the center of the stage in structural and molecular biology, likely to become independent tools with the same hierarchy as any experimental technique.

5. Abbreviations

TrHbN	truncated-N hemoglobin
TrHbO	truncated-O hemoglobin
Mb	myoglobin
Hb	hemoglobin
AscHb	<i>Ascaris lumbricoides</i> hemoglobin
Lba	leghemoglobin
CerHb	<i>Cerebratulus lacteus</i> hemoglobin
Cyt <i>c'</i>	cytochrome <i>c'</i>
sGC	soluble guanylate cyclase
AXCP	<i>Alcalignes xylooxidans</i> cytochrome <i>c'</i>
DFT	density functional theory
QM–MM	quantum mechanics–molecular mechanics
MD	molecular dynamics
MSMD	multiple steering molecular dynamics
FTIR–TDS	Fourier transform infrared temperature derivative spectroscopy

Acknowledgements

This work was partially supported by grants from Fundación Antorchas, Universidad de Buenos Aires (Grant X038), ANPCYT (Grant 06-08447).

References

- [1] K.M. Kadish, K.M. Smith, R. Guilard (Eds.), *The Porphyrin Handbook*, 4, Academic Press, San Diego, 2000.
- [2] S. Kundu, J.T. Trent III, M.S. Hargrove, *Trends Plant Sci.* 8 (2003) 387–393.
- [3] G. Simonneaux, A. Bondon, *Chem. Rev.* 105 (2005) 2627–2646.
- [4] S. Shaik, D. Kumar, S.P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* 105 (2005) 2279–2328.
- [5] R. Jain, M.K. Chan, *J. Biol. Inorg. Chem.* 8 (2003) 1–11.
- [6] J.S. Olson, G.N. Phillips Jr., *J. Biol. Inorg. Chem.* 2 (1997) 544–552.
- [7] E.E. Scott, Q.H. Gibson, J.S. Olson, *J. Biol. Chem.* 276 (2001) 5177–5188.
- [8] H. Sawai, M. Makino, Y. Mizutani, T. Ohta, H. Sugimoto, T. Uno, N. Kawada, K. Yoshizato, T. Kitagawa, Y. Shiro, *Biochemistry* 44 (2005) 13257–13265.
- [9] Q.H. Gibson, M.H. Smith, *Proc. R. Soc. Lond. B* 163 (1965) 206–214.
- [10] M. Milani, A. Pesce, M. Nardini, H. Ouellet, Y. Ouellet, S. Dewilde, A. Bocedi, P. Ascenzi, M. Guertin, L. Moens, J.M. Friedman, J.B. Wittenberg, M. Bolognesi, *J. Inorg. Biochem.* 99 (2005) 97–109.
- [11] M.A. Gilles-Gonzalez, G.S. Ditta, D.R. Helinski, *Nature* 350 (1991) 170–172.
- [12] S. Hou, R.W. Larsen, D. Boudko, C.W. Riley, E. Karatan, M. Zimmer, G.W. Ordal, M. Alam, *Nature* 403 (2000) 540–544.
- [13] S. Moncada, R.M. Palmer, E.A. Higgs, *Pharmacol. Rev.* 43 (1991) 109–142.
- [14] C. Simmerling, B. Strockbine, A. Roitberg, *J. Am. Chem. Soc.* 124 (2002) 11258–11259.
- [15] S.A. Hassan, L. Gracia, G. Vasudevan, P.J. Steinbach, *Meth. Mol. Biol.* 305 (2005) 451–492.
- [16] C. Domene, D.A. Doyle, C. Venice-Bryan, *Biophys. J.* 89 (2005) 1–3.
- [17] G. Monard, K.M. Merz Jr., *Acc. Chem. Res.* 32 (1999) 904–911.
- [18] U. Ryde, *Curr. Opt. Chem. Biol.* 7 (2003) 136–142.
- [19] R.A. Friesner, V. Guallar, *Ann. Rev. Phys. Chem.* 56 (2005) 389–427.
- [20] D.R. Nutt, M. Mewly, *Proc. Natl. Acad. Sci. USA* 101 (2004) 5998–6002.
- [21] C. Bossa, M. Anselmi, D. Roccatano, A. Amadei, B. Vallone, M. Brunori, M.A. Di Nola, *Biophys. J.* 86 (2004) 3855–3862.
- [22] H. Xiong, A. Crespo, M.A. Marti, D.A. Estrin, A.E. Roitberg, *Theor. Chem. Acc.*, in press.
- [23] T.E. Meyer, M.D. Kamen, *Adv. Protein. Chem.* 35 (1982) 105–212.
- [24] A. Pesce, M. Nardini, P. Ascenzi, E. Geuens, S. Dewilde, L. Moens, M. Bolognesi, A.F. Riggs, A. Hale, P. Deng, G.U. Nienhaus, J.S. Olson, K. Nienhaus, *J. Biol. Chem.* 279 (2004) 33662–33672.
- [25] D. Goldberg, *Chem. Rev.* 99 (1999) 3371–3378.
- [26] M. Milani, A. Pesce, Y. Ouellet, P. Ascenzi, M. Guertin, M. Bolognesi, *EMBO J.* 20 (2001) 3902–3909.
- [27] D.M. Lawson, C.E. Stevenson, C.R. Andrew, R.R. Eady, *EMBO J.* 19 (2000) 5661–5671.
- [28] M.S. Hargrove, J.K. Barry, E.A. Brucker, M.B. Berry, G.N. Phillips Jr., J.S. Olson, R. Arredondo-Peter, J.M. Dean, R.V. Klucas, G. Sarath, *J. Mol. Biol.* 266 (1997) 1032–1042.
- [29] M.L. Quillin, R.M. Arduini, J.S. Olson, G.N. Phillips Jr., *J. Mol. Biol.* 234 (1993) 140–155.
- [30] A. Pesce, M. Nardini, S. Dewilde, E. Geuens, K. Yamauchi, P. Ascenzi, A.F. Riggs, L. Moens, M. Bolognesi, *Structure (Camb)* 10 (2002) 725–735.
- [31] J. Yang, A.P. Kloek, D.E. Goldberg, F.S. Mathews, *Proc. Natl. Acad. Sci. USA* 92 (1995) 4224–4228.
- [32] D.A. Pearlman, D.A. Case, J.W. Caldwell, W.R. Ross, T.E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, P. Kollman, *Comp. Phys. Commun.* 91 (1995) 1–41.
- [33] J.W. Ponder, D.A. Case, *Adv. Prot. Chem.* 66 (2003) 27–85.
- [34] H.J.C. Berendsen, J.P.M. Postma, W.F. Van Gunsteren, A. Di Nola, J.R. Haak, *J. Chem. Phys.* 81 (1984) 3684–3690.
- [35] J. Wang, P. Cieplak, P.A. Kollman, *J. Comput. Chem.* 21 (2000) 1049–1074.
- [36] A. Crespo, M.A. Marti, S.G. Kalko, A. Morreale, M. Orozco, J.L. Gelpi, F.J. Luque, D.A. Estrin, *J. Am. Chem. Soc.* 127 (2005) 4433–4444.
- [37] C. Jarzynski, *Phys. Rev. Lett.* 78 (1997) 2690–2693.
- [38] M.O. Jensen, S. Park, E. Tajkhorshid, K. Schulten, *Proc. Natl. Acad. Sci. USA* 99 (2002) 6731–6736.
- [39] A. Crespo, M.A. Marti, D.A. Estrin, A.E. Roitberg, *J. Am. Chem. Soc.* 127 (2005) 6940–6941.
- [40] J.M. Soler, E. Artacho, J. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, *J. Phys. Cond. Mat.* 14 (2002) 2745–2779.
- [41] M.A. Martí, D.A. Scherlis, F.A. Doctorovich, P. Ordejón P, D.A. Estrin, *J. Biol. Inorg. Chem.* 8 (2003) 595–600.
- [42] M.A. Martí, L. Capece, A. Crespo, F. Doctorovich, D.A. Estrin, *J. Am. Chem. Soc.* 127 (2005) 7721–7728.
- [43] N. Troullier, J.L. Martins, *Phys. Rev. B* 43 (1991) 1993–2006.
- [44] S.G. Louie, S. Froyen, M.L. Cohen, *Phys. Rev. B* 26 (1982) 1738–1742.
- [45] J.P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 77 (1996) 3865–3868.
- [46] M. Eichinger, P. Tavan, *J. Chem. Phys.* 110 (1999) 10452–10467.
- [47] A. Crespo, D.A. Scherlis, M.A. Martí, P. Ordejón, A.E. Roitberg, D.A. Estrin, *J. Phys. Chem. B* 107 (2003) 13728–13736.
- [48] J.N. Harvey, *Struct. Bond.* 112 (2004) 151–183.
- [49] D.M.A. Smith, M. Dupuis, T.P. Straatsma, *Mol. Phys.* 103 (2005) 273–278.

- [50] R.J. Deeth, N.J. Fey, *Comput. Chem.* 25 (2004) 1840–1848.
- [51] D.A. Scherlis, D.A. Estrin, *Int. J. Quant. Chem.* 87 (2002) 158–166.
- [52] C. Rovira, K. Kunc, J. Hutter, P. Ballone, M. Parrinello, *J. Phys. Chem. A* 101 (1997) 8914–8925.
- [53] M. Couture, S. Yeh, B.A. Wittenberg, J.B. Wittenberg, Y. Ouellet, D.L. Rousseau, M. Guertin, *Proc. Natl. Acad. Sci. USA* 96 (1999) 11223–11228.
- [54] D.A. Scherlis, D.A. Estrin, *J. Am. Chem. Soc.* 123 (2001) 8436–8437.
- [55] E.S. Peterson, S. Huang, J. Wang, M.L. Miller, G. Vidujiris, A.P. Kloek, D.E. Goldberg, M.R. Chance, J.B. Wittenberg, J.M. Friedman, *Biochemistry* 36 (1997) 13110–13121.
- [56] T.E. Meyer, M.D. Kamen, *Adv. Protein Chem.* 35 (1982) 105–212.
- [57] H. Iwasaki, T. Yoshimura, S. Suzuki, Shidara. *Biochim. Biophys. Acta* 1058 (1991) 79–82.
- [58] T.M. Schnidt, A.A. DiSpirito, *Arch. Microbiol.* 154 (1990) 453–458.
- [59] J.W.B. Moir, *Biochim. Biophys. Acta* 1430 (1999) 65–72.
- [60] S. Kundu, B. Snyder, K. Das, P. Chowdhury, J. Park, J.W. Petrich, M.S. Hargrove, *Proteins* 46 (2002) 268–277.
- [61] M.A. Martí, D.E. Bikiel, A. Crespo, M. Nardini, M. Bolognesi, D.A. Estrin. *Proteins: Structures, Function and Bioinformatics*, in press.
- [62] F. Schotte, M. Lim, T.A. Jackson, A.V. Smirnov, J. Soman, J.S. Olson, G.N. Phillips, M. Wulff, P.A. Anfinrud, *Science* 300 (2003) 1944–1947.
- [63] M. Milani, A. Pesce, Y. Ouellet, S. Dewilde, J. Friedmann, P. Ascenzi, M. Guertin, M. Bolognesi, *J. Biol. Chem.* 279 (2004) 21520–21525.