Journal of Inorganic Biochemistry 246 (2023) 112289

Contents lists available at ScienceDirect

Journal of Inorganic Biochemistry

journal homepage: www.elsevier.com/locate/jinorgbio

Structural and dynamic characterization of the hexa-coordinated globin from *Spisula solidissima*

Alessandra Pesce [a,1], Katerina Barmpidi [b,1], Sylvia Dewilde [c], Carolina Estarellas [b], Luc Moens [c], Martino Bolognesi [d], Francisco Javier Luque [b], Marco Nardini [d,*]

[a] Department of Physics, University of Genova, Via Dodecaneso 33, I-16146 Genova, Italy
[b] Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació, Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQTC-UB), Universitat de Barcelona, Campus de l'Alimentació de Torribera, Santa Coloma de Gramenet, Spain
[c] Department of Biomedical Sciences, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk, Belgium
[d] Department of Biosciences, University of Milano, Via Celoria 26, I-20133 Milan, Italy

ARTICLE INFO

Keywords:
Nerve hemoglobins
Oxygen binding
Heme hexa-coordination
Dimeric hemoglobins
Crystal structure
Molecular dynamics

ABSTRACT

High energy consumption in the nervous system requires a continuous supply of O_2. This role is assisted by proteins from the globin super-family in the nerve cells of invertebrates, where 'nerve hemoglobins' (nHbs) are mainly present at mM concentrations and exhibit oxygen affinities comparable to those of vertebrate myoglobins. To gain insight into the structural bases of this function, we report the crystal structure of nHb from the Atlantic surf clam *Spisula solidissima* (*Ss*Hb), previously suggested to display a bis-histidyl hexa-coordinated heme in the deoxy state, high O_2 affinity, and ligand binding cooperativity when assayed *in situ*. The crystallized protein forms a dimer through packing of a 4-helix bundle involving helices E and F of each subunit. The *Ss*Hb 'classic' globin fold displays bis-histidyl (His71(E7) and His103(F8)) hexa-coordination of the heme-Fe atom, with structural and dynamics variations found in the inter-helix hinge regions. Molecular Dynamics simulations of both monomeric and dimeric species in the bis-histidyl hexa-coordinated, deoxy penta-coordinated, and O_2-bound hexa-coordinated states reveal distinct structural rearrangements at the interface between subunits in the dimer; these would affect the magnitude of the conformational fluctuations observed between monomer and dimer, and the topology of cavities within the protein matrix and at the interface. These results point to a distal site opening mechanism allowing access of the exogenous ligand to the heme and cast hypotheses on the dimer interface structural and dynamic properties that may support ligand binding cooperativity in dimeric *Ss*Hb.-

1. Introduction

Since the nervous system of most animals demands elevated energy consumption, a continuous supply with sufficient O_2 must be granted to avoid hypoxia and ensuing neuronal damages. The discovery of proteins from the globin super-family in the nerve cells of mammals and other vertebrates aroused much interest in their cellular functions, and thus in potential biomedical implications [1,2]. Historically, however, "nerve hemoglobins" (nHbs) were first observed in invertebrates, such as molluscs, arthropods, nemerteans, and nematode species [3,4].

Invertebrate nHbs are mainly present in glial cells and neurons [5–8]; usually, they are expressed at high concentrations (mM) and exhibit moderate oxygen affinities, with P_{50} values comparable to those of vertebrate myoglobins (Mbs) (P_{50} ~ 1–4 Torr) [9]. Indeed, during temporary hypoxia, invertebrate nHbs have been shown to support O_2 storage and/or transport to the highly metabolically active neurons [3,4,10]. Spectral analyses revealed two different types of nHbs based on the coordination at the heme-Fe atom. The nHbs of *Aplysia* spec.,

Abbreviations: 5c, penta-coordinated state; 6cHis, bis-histidyl hexa-coordinated state; $6cO_2$, oxygenated state; Hbs, hemoglobins; *h*Cygb, human cytoglobin; *Lp*HbII, *Lucina pectinata* HbII; Mbs, myoglobins; MD, Molecular Dynamics; MM/GBSA, Molecular Mechanics/Generalized Born Surface Area; MR, Molecular Replacement; nHbs, nerve hemoglobins; r.m.s.d., root mean square deviation; r.m.s.f., root mean square fluctuation; SAS, Solvent Accessible Surface; *Ss*Hb, *S. solidissima* nHb; *Sw*Mb, sperm whale Mb.

* Corresponding author.
E-mail address: marco.nardini@unimi.it (M. Nardini).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.jinorgbio.2023.112289
Received 24 April 2023; Received in revised form 31 May 2023; Accepted 7 June 2023
Available online 12 June 2023

A. aculeata and *C. lacteus* show a penta-coordinate heme geometry in the deoxygenated species, as for vertebrate hemoglobins (Hbs) and Mbs, whereas those of the bivalve *Tellina alternata* and *Spisula solidissima* are hexa-coordinate (in addition to the proximal His(F8), the heme-Fe atom is coordinated by distal His(E7) residue) and exhibit cytochrome *b*-type absorption spectra [10–12].

The nHb from the Atlantic surf clam *S. solidissima* (*Ss*Hb) has been extensively characterized biochemically and phylogenetically and can be considered as a prototype for invertebrate hexa-coordinated nHbs [10]. Phylogenetic analysis indicates that *Ss*Hb forms a common clade with the mollusc Mbs and Hbs, but not with the vertebrate neuroglobin, and consistently groups with the gill Hbs of the bivalve mollusc *L. pectinata* [10]. *Ss*Hb optical and Resonance Raman spectral properties are typical for ferrous low spin heme complexes with hexa-coordination of the heme-Fe atom; a minor high spin component was also identified, as for other bis-histidyl hexa-coordinated globins [13], in agreement with formation of a small photodissociated fraction. Furthermore, two *Ss*Hb CO coordination configurations were distinguished, with Fe-CO stretching mode frequencies comparable to those reported for Mb when (i) His71(E7) swings out of the heme pocket (open conformation) and CO has weak interactions with the surrounding amino acids, or (ii) the heme pocket is closed, due to the stabilization of CO by His71(E7) [14].

The ligand binding rates for CO and O_2 are both quite high (7.5×10^7 and 13×10^7 M^{-1} s^{-1}, respectively), with O_2 showing the higher rate, as generally observed for heme proteins [10]. The distal histidine association rate is an order of magnitude higher than that observed for other bis-histidyl hexa-coordinated globins; because of these elevated His binding rates, there is significant overlap with the CO binding phase. Indeed, after photodissociation, competitive binding of CO and the endogenous protein ligand (distal His) takes place. At high CO levels, one observes mainly a single phase of CO recombination. At lower CO levels, His binding becomes competitive, and a second, more prominent phase appears, the rate of the second phase corresponding to the rate of His(E7) replacement by CO [10].

Kinetic and equilibrium measurements showed that recombinant *Ss*Hb has a high O_2 affinity, with a P_{50} value of ~0.6 Torr, and that O_2 is bound in a non-cooperative manner as indicated by a slope (n) of ~1 in the Hill plot. However, it has been shown that the Hill coefficient for recombinant *Ss*Hb is dependent on the experimental conditions (temperature and pH) attaining a value of 1.5 [15]; moreover, *in situ* measurements on intact nerves, at higher protein concentrations, revealed a P_{50} of 2.3 Torr and an n value of 2.1 [16]. The differences in P_{50} and n values have been ascribed to *in vivo Ss*Hb association at the mM concentrations encountered *in vivo*, and/or to the existence of *in vivo* allosteric effectors that would lower oxygen affinity. Such behavior is in contrast with non-cooperative O_2 binding observed for the nHbs of other invertebrates, such as *A. aculeata*, despite the homodimeric structure observed by gel filtration [17]. While this suggests that the two identical subunits are functionally independent in the latter nHb, the cooperative behavior observed for *Ss*Hb *in vivo* may be consistent with heme-heme interactions promoted by an oligomeric structure [16,17].

To shed light on such functional aspects, we set out to characterize the *Ss*Hb three-dimensional structure, its bis-histidyl heme-Fe coordination and its oligomerization properties. The discovery of a dimeric *Ss*Hb in the crystal structure here presented allowed us to focus on the nature of interactions at the subunits interface, particularly regarding the geometrical arrangement of the hemes, and on the crosstalk between heme-propionates and protein residues to explore their potential roles on the cooperative behavior reported for this globin *in vivo*. Further, the structural and dynamical features of the monomeric and dimeric species of the protein, in different ligated states (bis-histidyl hexa-coordinate, deoxy penta-coordinate, O_2-bound hexa-coordinated), have been examined by means of extended molecular dynamics (MD) simulations. This analysis allows to disclose the structural features that may support the hexa- to penta-coordination transition and association-dissociation of the dimeric *Ss*Hb.

2. Materials and methods

2.1. Crystallization and structure determination

Ferric *Ss*Hb was expressed and purified as described previously [10]. In brief, its cDNA was cloned into the expression vector pET3a. The protein was expressed in *E. coli* BL21(DE3)pLysS cells, and purified by using: (i) ammonium sulfate precipitation (40–90% saturation), from which the 90% pellet was dissolved and dialyzed against 5 mM Tris-HCl, pH 8.5, (ii) a DEAE-Sepharose fast flow column (step elution in 5 mM Tris-HCl, pH 8.5, 200 mM NaCl), and (iii) a Sephacryl S200 gel filtration column in 5 mM Tris-HCl, pH 8.5.

Suitable *Ss*Hb crystals were grown in hanging drops, against a reservoir solution containing 2.5 M ammonium sulfate, and 50 mM sodium acetate (pH 5.5) at 277 K. X-ray diffraction data were collected to a maximum resolution of 1.7 Å at the ID-14-1 beamline of the ESRF synchrotron source, Grenoble (France), using one cryo-cooled crystal (at 100 K). The collected intensities were reduced and scaled using MOSFLM [18] and Scala [19]. *Ss*Hb crystals belong to the triclinic space group *P*1, with cell constants a = 48.2 Å, b = 50.7 Å, c = 74.2 Å, α = 75.7°, β = 88.1°, γ = 85.3°, and an estimated solvent content of 48.8% (*v*/v), accounting for four *Ss*Hb molecules per unit cell.

The 3D structure was determined by molecular replacement (MR) using the program Phaser [20] as implemented in the CCP4 program package [21]. The crystal structure of L. *pectinata* HbII (*Lp*HbII; PDB code 2OLP) [22] was used as a monomeric search model considering both the sequence coverage (91%) and sequence similarity (identity of 30%, and similarity of 56%) with the sequence of *Ss*Hb. In the MR model, side chains were truncated to Ala in cases of mismatch between the two amino acid sequences. Four MR solutions were found, with good crystal packing. Several cycles of manual rebuilding, using the program Coot [23], and refinement, using the program REFMAC (rigid body and restrained refinement) [24], were carried out to improve the electron density map. B-factors were refined isotropically. At the end of the refinement stage, the final model consists of 603 protein residues (8–157 in chain A, and 8–158 in chains B—D), 908 water molecules, and 4 glycerol molecules (Fig. 1A, Table 1). Although the recombinant protein was reported to be partially phosphorylated [10], no phosphorylated residues are visible in the electron density map. This might suggest that only the non-phosphorylated protein fraction was prone to crystallize or that phosphorylation occurs at Ser159, the only phosphorylatable residue in the regions not visible in the electron density map (1–7, 159–162) (Fig. 2). The quality of the refined model was checked with the program Molprobity [25]. Data processing and refinement statistics are reported in Table 1. The program PISA [26] was used to identify putative quaternary assemblies within the crystal unit cell. Atomic coordinates and structure factors have been deposited with the Protein Data Bank [27], with entry code 8OUP (www.rcsb.org/pdb).

2.2. Molecular dynamics simulations

The dynamical behavior of the *SsHb* monomer and dimer was studied by means of extended classical Molecular Dynamics (MD) simulations focusing on the ferrous bis-histidyl hexa-coordinated (6cHis), penta-coordinated (5c) and the oxygenated ($6cO_2$) heme states. The 6cHis systems were built using the X-ray structure reported here for the AB dimer of the *Ss*Hb structure (see above). The starting structures of the 5c and $6cO_2$ forms were built by homology modeling using the SwissModel webserver [28] and the crystallographic structure of oxygenated *Lp*HbII as template, similarly to that used for the MR model (see above). This strategy was adopted to account for the structural rearrangements triggered upon breakage of the heme-Fe—His71(E7) coordination bond, particularly regarding potential shifts in helix E and the CD loop. The classical MD simulations would thus provide structural information

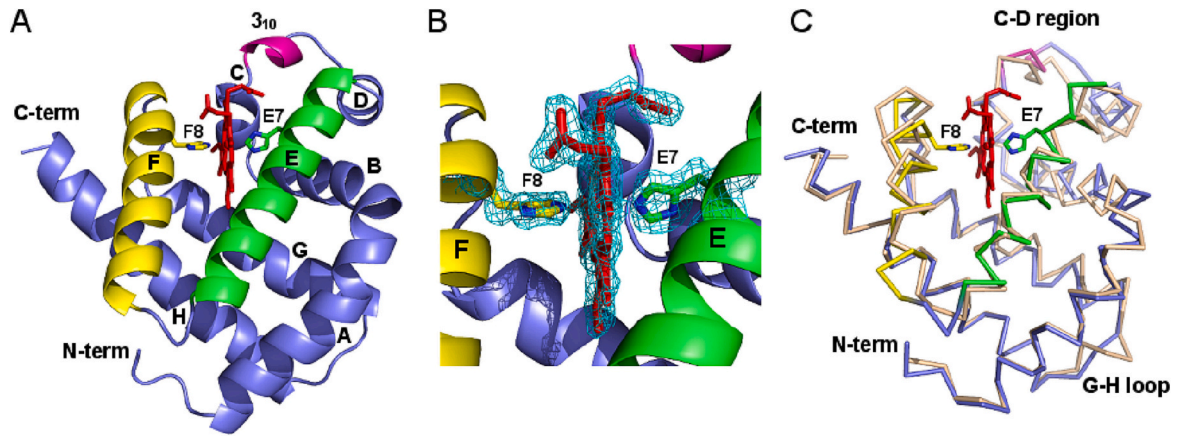

Fig. 1. A) Ribbon representation of *Ss*Hb tertiary structure. Helices and residues are labelled according to the conventional globin fold nomenclature [40]. The *E*- and F-helices are highlighted in green and yellow, respectively, the 3_{10} helix in the CD hinge in magenta. Distal His71(E7) and proximal His103(F8), coordinated to the heme-Fe, are shown in stick representation. B) 2Fo − Fc electron density map (cyan mesh), contoured at 1σ, for the heme region, including the distal His71(E7) and proximal His103(F8). C) Structural overlay of *Ss*Hb (colour code as in panel A) onto *Sw*Mb (wheat trace). Regions with low structural similarity are labelled. The *Sw*Mb heme is not shown for clarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Data collection and refinement statistics.

Data collection	
Spacegroup	*P*1
Wavelength	0.93400 Å
Resolution	49.0 Å – 1.7 Å (1.79 Å – 1.70 Å)[#]
Cell parameters	
a, b, c,	48.2 Å, 50.7 Å, 74.2 Å
α, β, γ	75.7°, 88.1°, 85.3°
N° of unique reflections	70,881 (10,388) [#]
Completeness	95.0% (94.8%) [#]
Multiplicity	3.9 (4.0) [#]
I/σ(I)	24.4 (11.9) [#]
R_{merge}[¶]	0.036 (0.120) [#]
Refinement	
R-factor[§] / R-free[$]	15.0% / 19.7%
Protein residues	603 (8–157 in chain A, 8–158 in chains B, C, D)
Average B factors	A: 18.8 Å^2; B: 15.6 Å^2; C: 18.0 Å^2; D: 19.2 Å^2
Heme groups	4
Average B factors	16.5 Å^2
Water molecules	908
Average B factors	32.1 Å^2
Glycerol molecules	4
Average B factors	35.0 Å^2
r.m.s.d. from ideality:	
Bond lengths	0.011 Å
Bond angles	1.72°
Ramachandran plot:	
Favored	100%
Disallowed regions	0%

[#] Values relative to the highest resolution shell are indicated in parentheses.
[¶] $R_{merge} = \Sigma_h \Sigma_i \mid I_{hi} - < I_h > \mid / \Sigma_h \Sigma_i I_{hi}$.
[§] R-factor $= \Sigma_h \mid\mid F_{obs}\mid - \mid F_{calc}\mid\mid / \Sigma_h \mid F_{obs}\mid$, with F_{obs} being the observed and F_{calc} the calculated structure factor amplitudes.
[$] R-free is calculated from 5% randomly selected data for cross-validation.

about the thermal fluctuations that can be sampled in the monomeric and dimeric states of 6cHis, 5c, and $6cO_2$ species. However, processes such as the 6cHis—5c transition, which imply the breakage/formation of the bond between the heme-Fe and the imidazole NE2 of His103(F8), would require the usage of computationally more demanding quantum mechanical computations and will not be explicitly considered in this study.

MD simulations were performed using AMBER20 package [29], the ff19SB force field for the protein [30], and the heme parameters developed previously [31,32]. The standard protonation state at physiological pH was assigned to ionizable residues, and capping groups were added due to the lack of precise structural information for the first residues at the N- and C-terminus of the protein in the X-ray structure. The simulated systems were neutralized and immersed in a pre-equilibrated octahedral box of OPC3 water molecules [33]. The final systems comprised the protein, around 10,000 water molecules, and 1 or 2 Na^+ counterions (for monomeric and dimeric species, respectively) to maintain charge neutrality, thus leading to simulated systems of *ca.* 30,000 and 40,000 atoms for the monomeric and dimeric chains, respectively.

Simulations were performed in the NPT ensemble for equilibration and the NVT ensemble for production runs [34,35]. The systems were simulated with periodic boundary conditions and Ewald sums (grid spacing 1 Å) for treating long-range electrostatic interactions. The initial system was minimized using a multistep protocol, which involves energy minimization of hydrogen atoms (2000 cycles of steepest descent +8000 cycles of conjugate gradient), then water molecules and counterions, and finally all atoms in the system (4000 cycles for steepest descent +1000 cycles of conjugate gradient in these two latter steps). The system was thermalized from 0 to 100 K in 20 ps at NVT ensemble, followed by four additional steps increasing the temperature from 100 to 300 K (50 ps/step) at NPT conditions. Finally, a last step of 5 ns was run to equilibrate the density of the system at a constant temperature (300K) and pressure (1 bar). The final structure of the equilibration process was used as a starting point for MD simulations using the NVT canonical ensemble. All the systems were simulated in duplicate, covering a production time of 1 μs per replica. To check the convergence of the dynamical behavior of the most flexible (5c and $6cO_2$) systems, MD simulations were extended up to 2 μs per replica.

The structural stability was evaluated from the positional root mean square deviation (r.m.s.d.) of selected atoms relative to the initial structure, and fluctuations were assessed using the root mean square fluctuation (r.m.s.f.) of protein residues. The analysis of inner cavities was performed using MDpocket [36,37], which uses a fast geometrical algorithm based on a Voronoi tessellation centered on the atoms and the associated alpha spheres to detect cavities and channels in the protein matrix. Analyses were performed using 10,000 snapshots taken equally spread over the full trajectories. The minimum and maximum alpha sphere radiuses were 2.8 Å and 5.0 Å, respectively. The identified pathways were superposed in time and space and a density map was generated from this superposition. Free cavities were then represented by using 3D isocontours.

The dimerization free energy was estimated by using the Molecular Mechanics Generalized Born Surface Area (MM/GBSA) method in

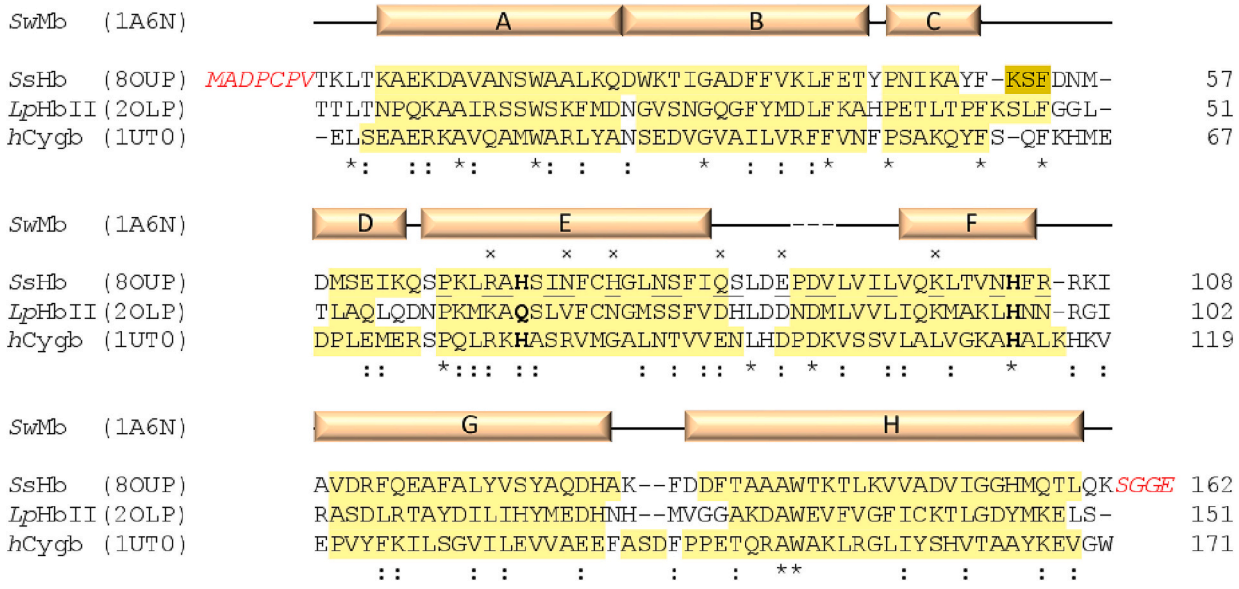

Fig. 2. Structure-based sequence alignments of *Ss*Hb *versus* hemoglobin II from *Lucina pectinata* (*Lp*HbII) [22], and human cytoglobin (*h*Cygb) [43]. Sperm whale myoglobin (*Sw*Mb) is taken as a reference for conventional secondary structure topological assignment (shown in the top line in wheat rectangles). Residues belonging to helices are shaded in light yellow, with those of the *Ss*Hb 3_{10} helix in dark yellow. Residues at topological position E7 and F8, relevant for heme-Fe coordination, are highlighted in bold. Identical and similar residues are indicated by asterisk and colon, respectively. Residues at the dimerization interface are underlined and those involved in polar interactions are indicated by a cross. PDB codes are in parentheses. In red *Italics*, *Ss*Hb residues not visible in the electron density map. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

conjunction with the end-point approach. Accordingly, an ensemble of 250 snapshots were taken from the last 250 ns of the MD simulations performed for the 6cHis, whereas a subset of 1,000 snapshots taken along the last microsecond was used for the 5c and $6cO_2$ states of the *Ss*Hb dimer. The binding free energy was determined by substracting the free energy of the separate subunits A and B (G_A,G_B) to the free energy of the dimeric species (G_{AB}; Eq. (1)). Each term was estimated by combining gas-phase and solvation contributions (Eq. (2)). The former includes the Coulombic (ΔE_{Coul}) and van der Waals (ΔE_{vdW}) interaction energy between the atoms of the protein, and the latter combines the electrostatic component of the solvation in water (ΔG_{GB}), which was determined using the Generalized Born method, and the non-electrostatic component (ΔG_{SAS}), which was estimated by using the relationship with the solvent-accessible surface (SAS) [38]. Since the main aim was to compute the changes in the interaction (free) energy between monomers for the distinct coordination states of the *Ss*Hb dimer, the conformational entropy term associated in the dimerization of the monomers was accounted for by means of Normal Mode Analysis calculations, which were performed for a subset of 11 snapshots for the 6cHis species, and 21 snapshots for the 5c and $6cO_2$ states, taken regularly along the trajectories of the two replicas run for each system. We note that the changes in the translational and rotational entropy components roughly cancel out in the comparison of the different states of dimeric *Ss*Hb. Solvent molecules and charged ions were deleted from each MD snapshot. Calculations were performed using the python script MMPBSA.py [39] using sander as implemented in AmberTools package [29].

$$\Delta G_{bind} = G_{AB} - G_A - G_B \tag{1}$$

$$\begin{aligned}\Delta G_{bind} &= \Delta E_{MM} + \Delta G_{sol} - T\Delta S_{conf} \\ &= \Delta E_{Coul} + \Delta E_{vdW} + \Delta G_{GB} + \Delta G_{SAS} - T\Delta S_{conf}\end{aligned} \tag{2}$$

3. Results and discussion

3.1. Quality of the model and overall structure

The structure of *Ss*Hb was solved on crystals belonging to the triclinic space group *P*1 (four molecules per unit cell, referred to as A, B, C, and D, respectively) by means of molecular replacement, based on the model of *Lp*HbII [22]. The *Ss*Hb structure was refined at 1.7 Å resolution, to final R-factor and R-free values of 15.0% and 19.7%, respectively, with ideal stereochemical parameters (Table 1). No electron density is observed for the first seven N-terminal residues, and for the last four C-terminal residues (five, in chain A). Superposition of 150 Cα atoms (residues 8–157) of the four independent *Ss*Hb chains yields r.m.s.d. values ranging between 0.14 Å and 0.42 Å; in particular, chains A and C show the closest match, with an r.m.s.d. of 0.14 Å. The only notable structural differences are present for residues at the N- and C-termini, and in the CD-E region (residues 52–69). However, all results discussed below apply to all chains unless otherwise stated.

The *Ss*Hb protein structure matches well the three-over-three α-helical sandwich fold of classical Hbs and Mbs, where the eight helices building up the globin fold are conventionally labelled A through H, according to their sequential order (Figs. 1A and 2) [40]. Structural overlay of the *Ss*Hb structure on sperm whale Mb (*Sw*Mb; PDB code 1A6N) [41], taken as the prototypical globin structure, yields an r.m.s.d. of 1.3 Å (calculated over 145 Cα pairs). Compared to the classical globin fold, however, *Ss*Hb displays a 3_{10} helix turn in the CD loop region, the *E*- and F-helices region (in particular, the F helix) is more extended, and the GH hinge loop is shorter (Figs. 1C and 2). The structure of the E-F region is crucial because the E and F helices host the distal and proximal His residues (at topological positions E7 and F8, respectively) coordinating the heme-Fe atom in the bis-histidyl hexa-coordinated form of the protein (Fig. 1A,B), and because they are responsible for dimerization of the protein (Fig. 3A) (see below). The PISA program [26] in fact identified the dimer formed by chains A and B within the crystal unit cell (or by chains C and D) as a stable species ($\Delta G^{int} = -43.7$ kcal mol^{-1}), with a SAS buried after complex formation of 4294 Å^2. The AB and CD dimers superimpose well, with a r.m.s.d. of 0.3 Å, calculated over 300 Cα pairs.

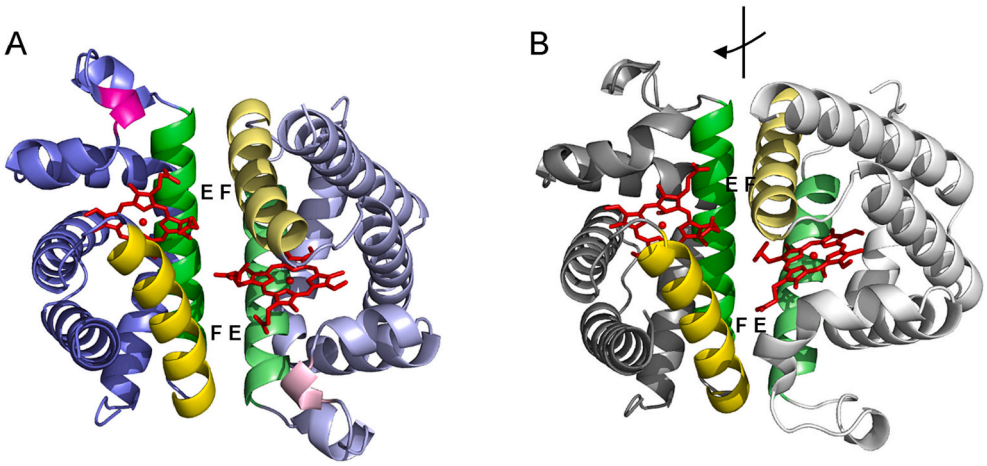

Fig. 3. Quaternary structure: A) *Ss*Hb dimer (coloured as in Fig. 1) and B) *Lp*HbII dimer (gray: PDB: 2OLP, oxy form). For both globins, the EF-type dimer is shown with the interface *E*- and F-helices labelled and coloured in green and yellow, respectively. The colour of one of the two subunits within the dimer is paler. The heme groups are shown in red. The different rotation of the two subunits in the *Ss*Hb and *Lp*HbII dimers orientation (~15°) is indicated in panel B. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

This dimer is defined as of the "EF-type" because it stems from the formation of a sort of 4-helix bundle at the protein interface involving helices E and F of the two subunits (Fig. 3A). The average distance between the heme-Fe atoms within the dimeric subunits is 17.6 Å, and because of the quaternary structure achieved their porphyrin ring planes are almost orthogonal. Such relatively short inter-heme distance is a feature characteristic of the EF-type globin dimers found in other molluscan Hbs [22,42]. The dimer interface includes 17 residues from each subunit (Fig. 2), six of which showing side chain hydrogen bonds/salt bridges. In particular, in the AB dimer these side chain interactions involve residues Arg69(E5)A–Gln84B, Arg69(E5)A–Glu88B, Asn74(E10)A–Lys98(F3)B, His77(E13)A–His77(E13)B, His77(E13)A–Gln84B, Gln84A–Arg69(E5)B, Gln84A–His77(E13)B, Glu88A–Arg69(E5)B, Lys98(F3)A–Asn74(E10)B. In the CD dimer such interactions are essentially conserved; little changes due to different orientations of Lys30(B3)C and Arg69(E5)C side chains are observed, Lys30(B3) replacing Arg69(E5) in the interaction with Gln84.

3.2. Heme distal and proximal sites in 6cHis SsHb

One of the most evident features of the *Ss*Hb overall structure is the bis-histidyl hexa-coordination of the heme-Fe atom, which is observed in the crystal structure in agreement with expectations based on the protein spectral properties [10]. His103(F8) and His71(E7) are the proximal and distal heme-Fe ligands, respectively (Fig. 1A,B).

The proximal His103(F8) provides a coordination bond (2.0–2.1 Å, depending on the chain) through its NE2 atom to the heme-Fe atom; the side chain imidazole ring is staggered relative to the heme pyrrole nitrogen atoms, being oriented toward the methinic bridge CHB and CHD atoms of the porphyrin ring. The orientation of the His103(F8) side chain is further stabilized, as often found in globins [44], by a hydrogen bond between the histidyl ND1 atom and Leu99(F4) carbonyl oxygen atom (2.8–3.0 Å, depending on the chain). The proximal coordination bond is orthogonal to the porphyrin plane, in accordance with the conventional binding pattern of globins [44]. The *Ss*Hb distal site displays a coordination bond connecting the heme-Fe atom to the distal His71(E7) NE2 atom, with a bond length of 2.0–2.1 Å, depending on the chain. The His71(E7) imidazole ring is almost perfectly staggered relative to the heme pyrrole nitrogen atoms, being oriented toward the methinic bridge CHA and CHC atoms of the porphyrin ring; thus, the distal and the proximal His imidazole rings display close to orthogonal (~75°) relative azimuthal orientations.

The average heme (pyrrole)N–Fe distance is 2.0 Å; the iron atom appears well within the heme (pyrrole)N atoms plane, the Fe–His71(E7) coordination bond being almost orthogonal to the heme plane. Besides van der Waals contacts (about 60 < 4.0 Å) with residues in the heme crevice, heme binding to *Ss*Hb is stabilized by ionic interactions involving the porphyrin propionates. Indeed, the analysis of the heme propionates in the four molecules of the asymmetric unit reveals a certain degree of conformation variability, although some interactions are fully conserved. On the proximal side, propionate A is engaged in a bidentate salt bridge with Arg(106). Furthermore, in chains A and C only, the propionate O1A atom is further stabilized by a H-bond with the ND2 atom of Asn102(F7) of the facing partner in the dimer (chains B and D, respectively). A further structural crosstalk between subunits of the dimer is evident at the heme distal side, where propionate D (in particular the OD2 atom) is H-bonded to the NZ atom of Lys98(F3) of the partner chain in the dimer. In chain B, the OD2 atom is H-bonded to a water molecule, instead. Of note, Lys98(F3) is always stabilized in its orientation by a H-bond interaction between its NZ atom and Asn74(E10) OD1 atom of the facing subunit within the dimer (Fig. 4). Interestingly, a similar H-bond network is found in the heme cavity of *Lp*HbII. In particular, Arg100(F11) and Lys92(F3), corresponding to *Ss*Hb Arg106(F11) and Lys98(F3), establish H-bonds with the propionates of the heme group, and Lys92(F3) has been proposed as contributor to *Lp*HbII dimer stability [22].

3.3. Structural homologs

A DALI analysis [45] reveals that *Ss*Hb 3D structure closely matches the tertiary structures of other globins, in particular *Lp*HbII (DALI *Z*-score of 23.1, residue identity of 29%) [22], and human cytoglobin (*h*Cygb, DALI Z-score of 22.5, residue identity of 55%) [43]. The structural matches of *Ss*Hb with *Lp*HbII and *h*Cygb are interesting since *Lp*HbII is a protein phylogenetically related to *Ss*Hb, sharing with it high sequence similarity and a similar dimerization mode but is penta-coordinated, while *h*Cygb is the bis-histidyl hexa-coordinated globin with the highest structural homology with *Ss*Hb. A backbone comparison among these Hbs highlights an excellent conservation of secondary structure elements with an r.m.s.d. of 1.3 Å calculated over 146 Cα pairs for *Lp*HbII (PDB code 2OLP), and 1.4 Å calculated over 149 Cα pairs for *h*Cygb (PDB code 1UT0) (Figs. 2 and 5).

The major backbone differences between *Ss*Hb with *Lp*HbII are localized in the CD loop, at the first N-terminal half of the E helix (residues 52–69), and in the GH hinge (residues 128–133). The backbone difference in the C-D-E region accounts for the different heme-Fe coordination states of the two proteins (bis-histidyl hexa-coordinated in *Ss*Hb *vs* O_2-bound in *Lp*HbII). In particular, distal coordination of *Ss*Hb His71(E7) to the heme-Fe atom results in a shift of the N-terminal part of helix E toward the heme of about 1.7 Å compared to O_2-bound *Lp*HbII, and is achieved through a tilt of about 7° of the helix pivoting on its C-terminal end (Fig. 5A). Tilting the *E*-helix from the hexa- to the penta-coordinated location, to grant access of the exogenous ligand to the distal site, would require some flexibility in the CD loop region

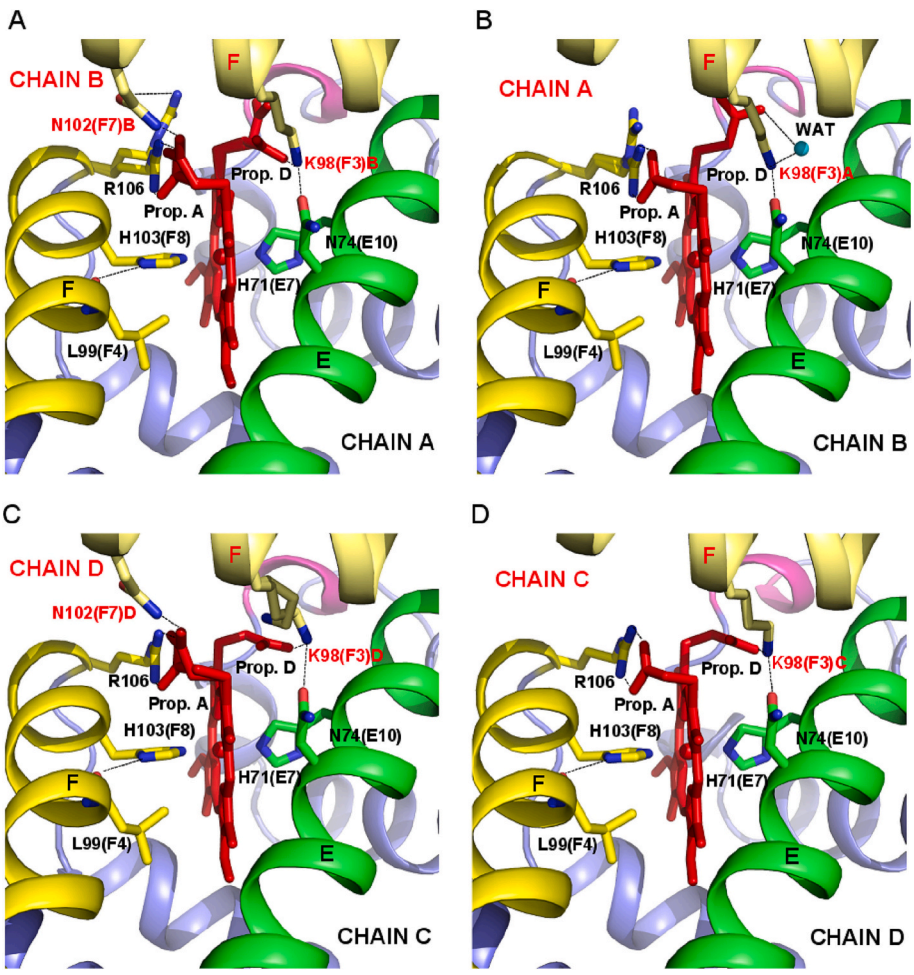

Fig. 4. Heme pocket of 6cHis *Ss*Hb. The four chains building the AB and CD dimers are shown in the four panels. Key residues in the heme cavity are shown as sticks, coloured as in Fig. 1, and labelled according to their topological positions. Residues belonging to the same dimeric partner are labelled in red with the corresponding chain letter. Hydrogen bonds are indicated with dashed lines. The heme group is shown in red and a water molecule (chain B) in cyan. Heme propionates A and D are labelled. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

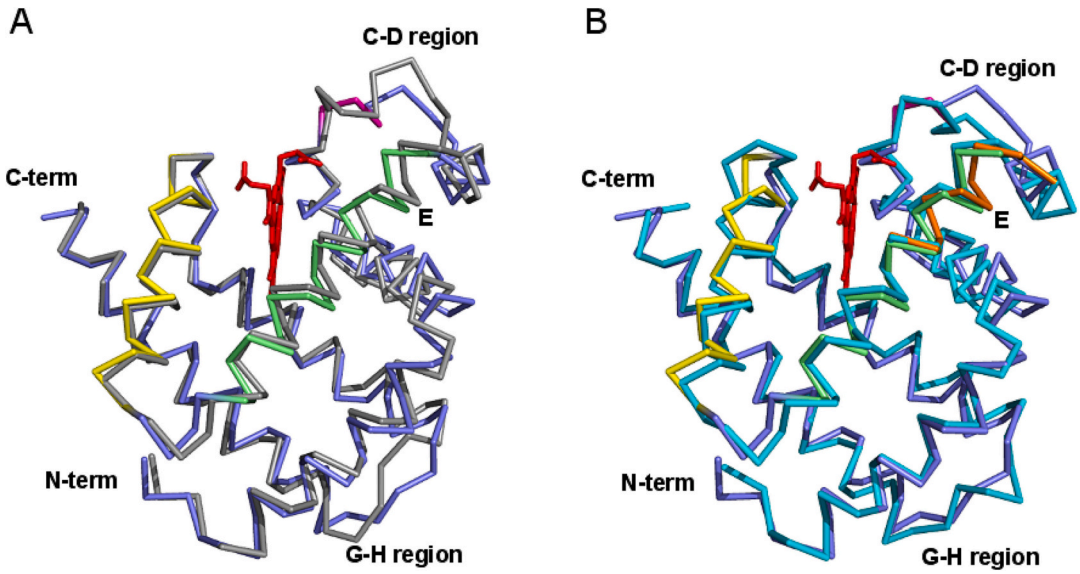

Fig. 5. Structural overlay of *Ss*Hb (colour code as in Fig. 1A) onto A) *Lp*HbII (gray trace), and B) *h*Cygb (cyan and orange traces for the closed and open form, respectively). Regions with low structural similarity are labelled, together with N- and C-term. The *Ss*Hb heme is shown in red. The *Lp*HbII and *h*Cygb hemes are not shown for clarity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

preceding helix E. Accordingly, when the four *Ss*Hb chains are compared with each other, the C-D-E region (in particular, residues 52–69) shows a r.m.s.d. (2.0 Å on average) which is about 2-fold higher than the values found in the protein core; also the B-factors (28.4 $Å^2$ on average) are ~1.6 times higher than the mean of the protein, both parameters being indicative of increased flexibility in that area.

The structure of the GH hinge (residues 128–133) is well conserved in the four *Ss*Hb chains (average r.m.s.d. 0.23 Å) and quite rigid (B-factors 16.0 $Å^2$ on average). The minor different backbone conformations of the GH hinge observed in *Ss*Hb and *Lp*HbII are likely due to different structural environments and sequences. In fact, residues from this loop are involved in a crystal contact in *Ss*Hb, but not in *Lp*HbII, and contain a flexible Gly-Gly motif in *Lp*HbII but not in *Ss*Hb (Fig. 2).

As anticipated above, *Ss*Hb and *Lp*HbII share a similar dimerization mode, based on the EF-interface (Fig. 3). However, the relative orientations of the two subunits in dimeric *Ss*Hb and *Lp*HbII are significantly different. Indeed, when the *Ss*Hb A chain is superimposed on *Lp*HbII chain A, the corresponding B chains diverge by a rotation of ~15° around an axis located at the dimeric interface (Fig. 3B). This is not surprising considering that helix E, at the dimeric interface, is differently oriented in the two globins that display different heme-Fe coordination states.

In consideration of the *Ss*Hb heme bis-histidyl hexa-coordination, the *Ss*Hb fold is then compared to the tertiary structure of hexa-coordinated *h*Cygb. The average r.m.s.d. of the four *Ss*Hb chains, taken individually, with the bis-histidyl hexa-coordinated *h*Cygb is 1.35 Å. As in the case of the comparison with *Lp*IIHb, major backbone differences are localized in the CD hinge region, where one extra residue is present in the *h*Cygb sequence (Fig. 2), at the first N-terminal half of the *E*-helix (residues 52–69), and in the GH hinge (due to a two-residue insertion). A shift of about 2.3 Å is present at the N-terminus of the B-helix. Of note, the *h*Cygb structure (PDB code 1UT0) has been reported also in an "open" conformation (coexisting in the crystal with the "closed" hexa-coordinated conformation), where the N-terminal half of the E-helix moves away from the distal site pocket. As a result of shifting the E1–E10 segment, including the distal HisE7(81) residue, in the *h*Cygb "open" conformation, the heme group is penta-coordinated [43]. Indeed, the hexa-coordinated structure of *Ss*Hb places the N-terminal half of the E-helix in a conformation that is intermediate between those reported for the "closed" (hexa-coordinated) and "open" (penta-coordinated) states of *h*Cygb. Regarding the quaternary structure, *h*Cygb is also a dimer, but its dimerization interface in unrelated to that of *Ss*Hb [43,46,47].

3.4. MD simulations of monomeric and dimeric SsHb species

To gain insight into the structure-function relationships in *Ss*Hb, extended MD simulations (up to 2 μs) were performed in duplicate for the monomeric and dimeric species in the 6cHis, 5c, and $6cO_2$ states. The structure of the simulated systems was generally stable along the sampled trajectories, as indicated by the r.m.s.d. profiles determined for the protein backbones (Fig. S1 and S2 in Supplementary Material). This is particularly evident for the 6cHis species, where the r.m.s.d. (averaged along the whole trajectory of the two independent replicas) is 1.3 Å and 1.6 Å for the monomeric and dimeric protein, respectively (Table 2).

Table 2
Average root mean square deviation (r.m.s.d., Å; standard deviation in parenthesis) determined from the two independent replicas run for monomeric and dimeric species in the 6cHis, 5c and $6cO_2$ states using the equilibrated structure as reference. For the 6cHis species the r.m.s.d. was also determined considering the X-ray crystal structure (PDB code 8OUP) as reference (values in italics).

Systems	1 μs			2 μs	
	6cHis	5c	$6cO_2$	5c	$6cO_2$
Monomer	1.3 (0.2) *1.1 (0.2)*	2.1 (0.5)	1.4 (0.2)	–	–
Dimer	1.6 (0.3) *1.4 (0.3)*	2.0 (0.2)	2.4 (0.4)	2.2 (0.3)	2.6 (0.4)
Dimer – Chain A	1.1 (0.2) *0.9 (0.1)*	2.2 (0.3)	1.3 (0.2)	2.2 (0.3)	1.5 (0.3)
Dimer – Chain B	1.1 (0.2) *0.9 (0.1)*	1.3 (0.2)	1.8 (0.2)	1.4 (0.3)	1.9 (0.2)

3.4.1. Monomeric and dimeric 6cHis SsHb species

In the 6cHis state, the *Ss*Hb fold is very stable, including the CD loop, with residue fluctuations along the MD simulations for both monomeric and dimeric species generally comprised between 0.7 and 1.4 Å (Fig. 6). Additionally, the side chains of the residues that shape the hydrophobic core of the protein occupy well-defined positions, this trait being also observed in the structural analysis of the subunits A and B within the dimer. Overall, the residues are tightly packed in the protein matrix, which limits the conformational fluctuations of the side chain around the position seen in the X-ray structure.

The relative position of the heme groups in the dimer (Fig. 3A) is maintained along the trajectories. The Fe—Fe distance is, on average, 18.1 ± 0.7 Å, which is close to the value found in the X-ray structure (17.6 Å). The heme propionate-A atoms O1A and O2A preserve their interaction with Arg106(F11) in the same chain (Fig. 4 and Fig. S3), as noted in average distances of 3.7 ± 1.3 Å (chain A) and 4.3 ± 1.8 Å (chain B), although transient fluctuations in the propionate-guanidinium distance occur along the trajectories (Table S1). The oxygen atoms of heme propionate-D maintain the interaction with Lys98 (F3) of the facing subunit within the dimer, as reflected in average distances of 4.4 ± 1.7 and 4.7 ± 2.2 Å (Fig. S4), but the interaction between Lys98(F3) and Asn74(E10) observed in the X-ray structure (Fig. 4) is lost during the simulations (distance >7 Å; Table S1). The H-bond between the proximal His103(F8) and Leu99(F4) carbonyl O is stably maintained along the simulations (average distance of 2.9 ± 0.2 Å; Fig. S5).

To complement the structural analysis of the interactions formed by the residues located at the interface, the MM/GBSA method was used to estimate the energetics of the interaction between monomers. The results obtained for the two replicas of the 6cHis system are highly similar and amounts, on average, to −17 kcal mol^{-1} (Table 3). Indeed, there is a close resemblance in the values of the distinct free energy components determined for the two replicas, which agrees with the structural rigidity of the 6cHis dimeric species along the MD simulations (see above). The charged nature of most residues at the interface is reflected in the large contribution of the Coulombic term (note also the large standard deviation), which is about 4-fold larger than the van der Waals component. This also explains the large contribution of the solvation term to the dimerization free energy. The interactions between subunits (A and B chains) at the dimeric interface are in general preserved along the two simulations run for the 6cHis dimer. The most relevant changes concerns residues Arg69(E5), Gln84 and Glu88, which are located close to the hinge between helices E and F, as noted in the weakening of the electrostatic interaction between Arg69(E5)A with Gln84B and Glu88B, and the strengthening of the interaction between Arg69(E5)B and Glu88A (Fig. 7).

3.4.2. Monomeric and dimeric SsHb 5c species

The starting structure of the 5c *Ss*Hb was built by homology modeling using the crystallographic structure of 5c *Lp*HbII as template. This was motivated by the need to account for the structural rearrangements triggered upon breakage of the Fe-His71(E7) bond, particularly regarding the potential shifts and rotation of helix E that may affect the arrangement of the AB and CD loops, and helices G and H (Fig. 8A). At the residue level, the most noticeable change observed in the 5c simulations concerns the orientation of Trp29(B2). In the 6cHis state this residue points to the interior of the protein (closed conformation), where the indole ring is confined against the side chains of Trp22(A11), Ile32(B5), Phe75(E11) and Tyr120(G12). In contrast, in the 5c state Trp29(B2) protrudes toward the solvent (open conformation), lying above helix E, close to Cys76(E12) (Fig. 8B). This conformational rearrangement avoids the steric clash that would occur between Ile32 (B5) and Trp29(B2) if the closed conformation of this latter residue were maintained in the 5c fold. Let us note that this feature cannot be attributed to the penta-coordinated *Lp*HbII used as template, since the same results are obtained when the cyanide-bound form of *h*Cygb (PDB

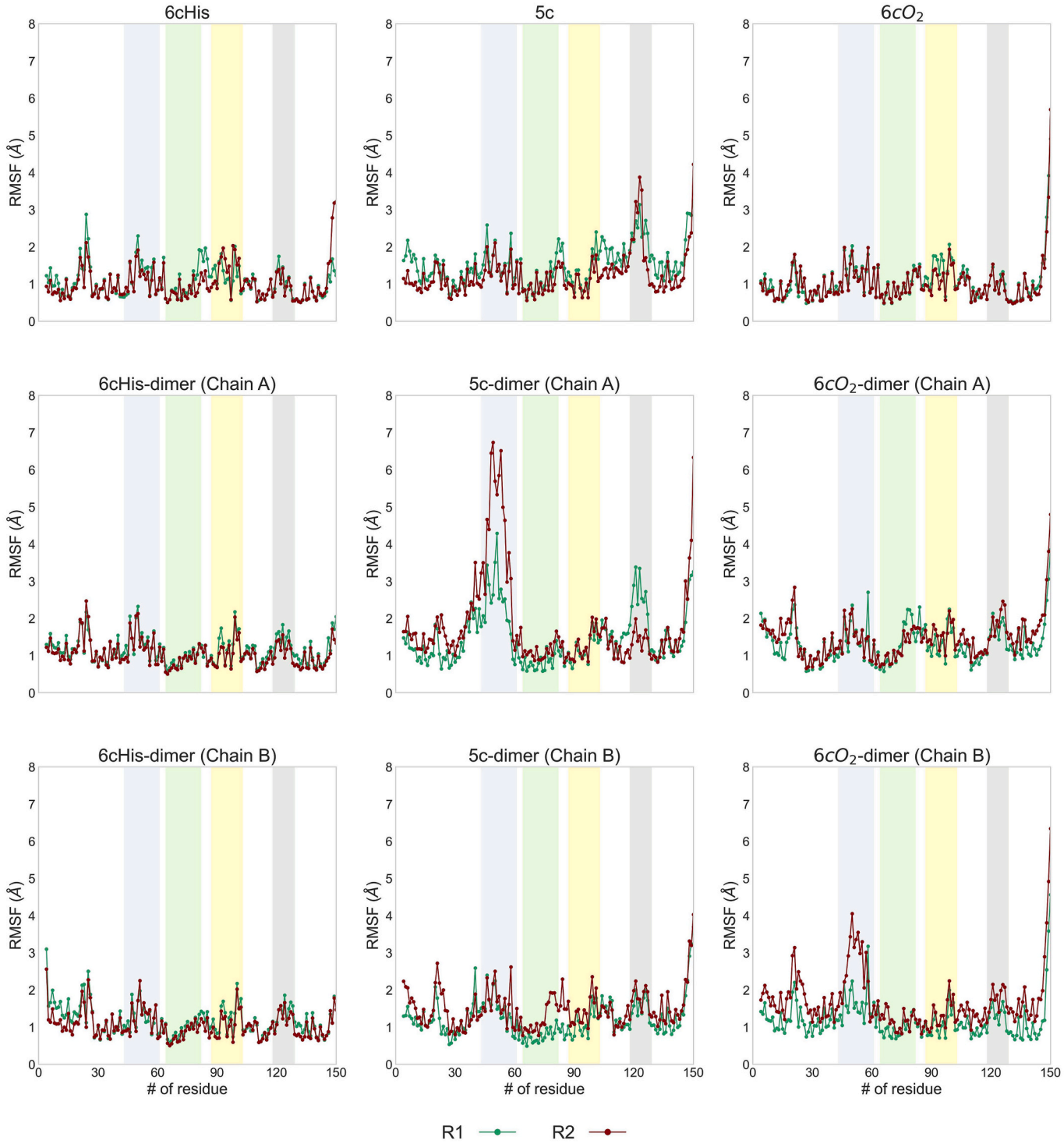

Fig. 6. Per-residue root mean square fluctuations (r.m.s.f., Å) determined from the two independent replicas (coloured in green and red) of (top) monomeric and (middle and bottom) dimeric *Ss*Hb in 6cHis, 5c and 6cO_2 states. The coloured bars correspond to the CD loop (purple), helix E (green) and F (yellow), and the GH loop (gray). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

code 4B3W) [48] is used as reference to build the homology model of 5c *Ss*Hb (data not shown).

In the 5c *Ss*Hb, the average r.m.s.d. is ~2 Å for both the monomeric and dimeric species (Table 2), thus suggesting increased structural changes (relative to the 6cHis species) due to the release of the restraint imposed by the Fe–His71(E7) coordination bond, especially in subunit A of the dimer (r.m.s.d. of 2.2 Å *versus* 1.3 Å for chain B). The largest fluctuations for both monomeric and dimeric 5c *Ss*Hb are localized in the inter-helical loops, especially the CD loop, the GH hinge, and at the C-terminus (Fig. 6). Although the r.m.s.f. is generally comprised between 1.0 and 2.0 Å, larger fluctuations are observed in the GH loop of both monomeric (r.m.s.f. of 3.9 Å) and dimeric (subunit A; r.m.s.f. of 3.4 Å) species, and in the CD loop of chain A of the dimer (r.m.s.f. of 6.8 Å). This highlights a distinct structural response of chains A and B along the trajectories sampled for the dimeric species.

Regarding the position of His71(E7) side chain during MD, the *out*

Table 3
Free energy changes (kcal mol^{-1}; standard deviation in parenthesis) for the formation of the dimeric species in the 6cHis, 5c, and 6cO_2 states. Values determined for an ensemble of 250 (6cHis) or 1000 (5c and 6cO_2) snapshots taken from the last 250 (6cHis) or 1000 (5c, 6cO_2) ns of the two independent trajectories (R1 and R2) using the MM/GBSA method (see Eq. (2) in Materials and methods).

Free energy	6cHis	5c	6cO_2
Replica 1			
ΔE_{MM}	−448 (62)	−632 (46)	−522 (58)
Coulombic	−364 (63)	−532 (45)	−438 (59)
van der Waals	−84 (6)	−100 (6)	−85 (7)
ΔG_{sol}	413 (55)	580 (42)	486 (53)
Electrostatic	425 (55)	595 (41)	498 (53)
Non-electrostatic	−12 (1)	−16 (1)	−12 (1)
$-T\Delta S_{conf}$	20 (5)	22 (5)	25 (4)
ΔG_{bind}	−15	−30	−11
Replica 2			
ΔE_{MM}	−431 (74)	−548 (70)	−603 (74)
Coulombic	−346 (75)	−457 (73)	−515 (75)
van der Waals	−85 (7)	−91 (9)	−87 (7)
ΔG_{sol}	393 (65)	514 (65)	562 (70)
Electrostatic	405 (65)	527 (65)	575 (70)
Non-electrostatic	−13 (1)	−13 (1)	−13 (1)
$-T\Delta S_{conf}$	18 (6)	17 (5)	30 (7)
ΔG_{bind}	−20	−17	−11
Combined			
ΔE_{MM}	−440	−590	−562
Coulombic	−355	−495	−477
van der Waals	−85	−95	−86
ΔG_{sol}	403	547	524
Electrostatic	415	561	537
Non-electrostatic	−12	−14	−13
$-T\Delta S_{conf}$	19	20	27
ΔG_{bind}	−17	−24	−11
$\Delta\Delta G_{bind}$	0	−7	+6

conformation (pointing toward the mouth of the heme cavity) found at the beginning of the simulation quickly changed to the *in* conformation (oriented to the protein interior) in the first 100 ns of the two independent simulations run for both monomeric and dimeric species (Fig. 9 and Fig. S6). The *in* conformation remained primarily stable in the distal pocket along the MD simulations, as the imidazole ring becomes enclosed in a hydrophobic cage shaped by Phe37(B10), Phe51, Phe75 (E11), and Leu68(E4), thus sterically hindering access to the heme-Fe atom. Nevertheless, thermal fluctuations in the distal cavity enabled conformational transitions of the His71(E7) side chain, which transiently could fluctuate between *in* and *out* conformations (Fig. S6), the distance from the heme-Fe to the center of the imidazole ring being increased from about 5–6 Å to about 9 Å (Fig. S7). As illustrated in Fig. 9C, these fluctuations arise from concerted shifts in the B helix and the CD loop, which in turn affect the packing of His71(E7) against the aromatic rings of Phe37(B10), Phe41(B14) and Phe51(CD1). Noteworthy, these structural rearrangements enhance accessibility to the heme-Fe atom, thus possibly facilitating the access of O_2 to the distal site from the solvent regions, as exemplified for selected snapshots in Fig. S8. Finally, it is worth noting that the frequency of these transient conformational changes is larger in the dimer compared to monomer, as noted in the analysis of the fluctuations of both the dihedral ($NC_\alpha C_\beta C_\gamma$) angle in His71(E7) side chain (Fig. S6) and of the distance from the heme-Fe atom to the center of the imidazole ring (Fig. S7), which can be expected from the enhanced structural flexibility of the dimeric species.

The average value of the dimerization free energy of the 5c dimeric *Ss*Hb is −24 kcal mol^{-1} (Table 3), which supports the stabilization of the interaction between subunits in the 5c species compared to the 6cHis form (−17 kcal mol^{-1}). However, the strengthening of the interaction is particularly relevant only in replica 1 (−30 kcal mol^{-1}), this trait being mainly attributed to the van der Waals component, which stabilizes the interaction between subunits by −9 kcal mol^{-1} relative to replica 2. In contrast, the dimerization free energy in replica 2 (−17 kcal mol^{-1}) coincides with the value obtained for the 6cHis dimer, thus reflecting the similar van der Waals interaction. This reflects the occurrence of diverse structural rearrangements at the interface between chains A and B upon transition from the 6cHis species to the 5c form, involving both rotation of the E and F helices and separation between subunits A and B in the dimer. In the 6cHis dimer, these helices form an angle of ~29.4°, which is in agreement with the angle found for the monomer, and the relative orientation of the pair of E-F helices in chains A and B is 124.4° (Table 4). However, in the 5c dimer, the angle formed by the E and F helices varies from 31.5° to 37.3°, reflecting a larger structural fluctuation relative to the 5c monomer, where the angle is ~31°. Furthermore, the angle between the *E*-F helices in subunits A and B is reduced to ~115°. On the other hand, the averaged Fe—Fe inter-heme distance increases from 18.1 ± 0.7 Å in the 6cHis state to 19.8 ± 0.7 Å in the 5c dimer (Fig. S9 and Table S1). The Fe—Fe distance increase is more evident in replica 2, where the Fe—Fe distance steadily increases from about 20 Å to 22.5 Å in the last 200 ns of the simulation (Fig. S9). On the other hand, a distinct behavior is observed for the interactions of residues located close to the hinge of helices E and F, such as the strengthening of the Arg69(E5)A–Glu88B contact in replica 1, and the weakening of the Glu88A–Arg69(E5)B interaction in replica 2 (Fig. 7).

These results reflect the increased conformational flexibility triggered by the 6cHis → 5c transition. The magnitude of the enhanced structural fluctuations is not identical in the two chains present in the 5c dimer, at least within the time scale (2 μs) considered for each MD simulation. Nevertheless, this is in contrast with the structural stability observed in the simulations run for the monomers.

3.4.3. *Monomeric and dimeric 6cO_2 SsHb species*

The average r.m.s.d. during MD simulations for the 6cO_2 monomeric species is 1.4 Å, which suggests a higher structural rigidity of the 6c-bound state relative to the 5c species (Table 2). The 6cO_2 dimeric form exhibits larger changes, with an average r.m.s.d. of 2.6 Å, slightly larger than the r.m.s.d. of the 5c dimer (2.2 Å), but also a notable asymmetry between the two subunits within the dimer, as noted in r.m.s.d. values of 1.5 and 1.9 Å for chains A and B, respectively. The largest fluctuations in the MD simulations for both monomeric and dimeric *Ss*Hb 6cO_2 species are localized in the inter-helical loops and at the N- and C-termini (Fig. 6). The r.m.s.f. are generally comprised between 0.5 and 1.5 Å in the monomer, and between 0.8 and 2.1 Å, in the dimer, but larger fluctuations (up to 4 Å) are observed in the CD loop of subunit B (replica 2). The heme-bound O_2 is hosted in a distal cage defined by Phe37(B10), Phe51(CD1) and Phe75(E11), besides the H-bond interaction with the distal His71(E7) (average distance of 3.1 Å; Table S1 and Fig. S10). Nevertheless, transient fluctuations in the His71(E7)-O_2 distance are observed, especially for chain B in replica 2 (Fig. S9). As noted above for the 5c dimeric species, this suggests a distinct structural response of the two protein subunits along the trajectories sampled for the 6cO_2 dimer.

The dimerization free energy is, on average, −11 kcal mol^{-1} (Table 3), which implies a reduction compared to the 5c dimer (−24 kcal mol^{-1}), and slightly less favorable relative to the 6cHis dimeric form (−17 kcal mol^{-1}). The same value is obtained for the two replicas (−11 kcal mol^{-1}; Table 3), which reflects the similar van der Waals energy components. The arrangement of the E and F helices at the interface in the 6cO_2 dimer reflects the trends discussed above for the 5c dimeric species. Thus, one may notice a higher diversity in the angle formed by E and F helices in chains A and B (ranging from 30.1° to 37.3°), which is in contrast with the similar value found for the two replicas of the 6cO_2 monomer (about 32.6°). Furthermore, the pair of E-F helices exhibits an intermediate arrangement between the angles found for the 5c (114.7° and 115.0°) and 6cHis (124.3° and 124.5°). The simulation run for dimeric 6cO_2 reveals an increased separation

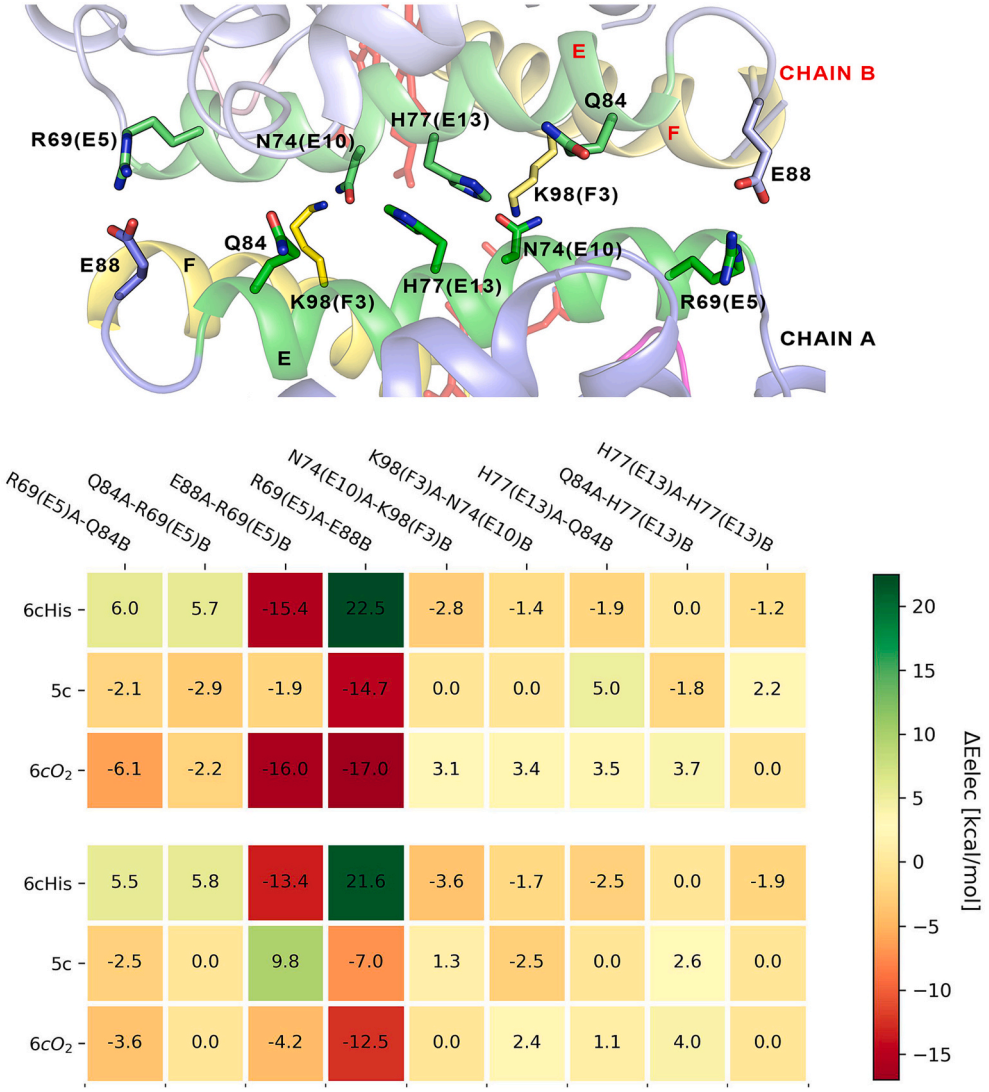

Fig. 7. Change (relative to the equilibrated structure) in the electrostatic interaction energy (kcal/mol) between residues located at the interface of subunits A and B (see upper panel). Results provided for the dimeric species of the 6cHis, 5c and $6cO_2$ states determined for the first (rows 1–3) and second (rows 4–6) independent MD simulations. Positive/negative values indicate weakening/strengthening of the interaction between residues.

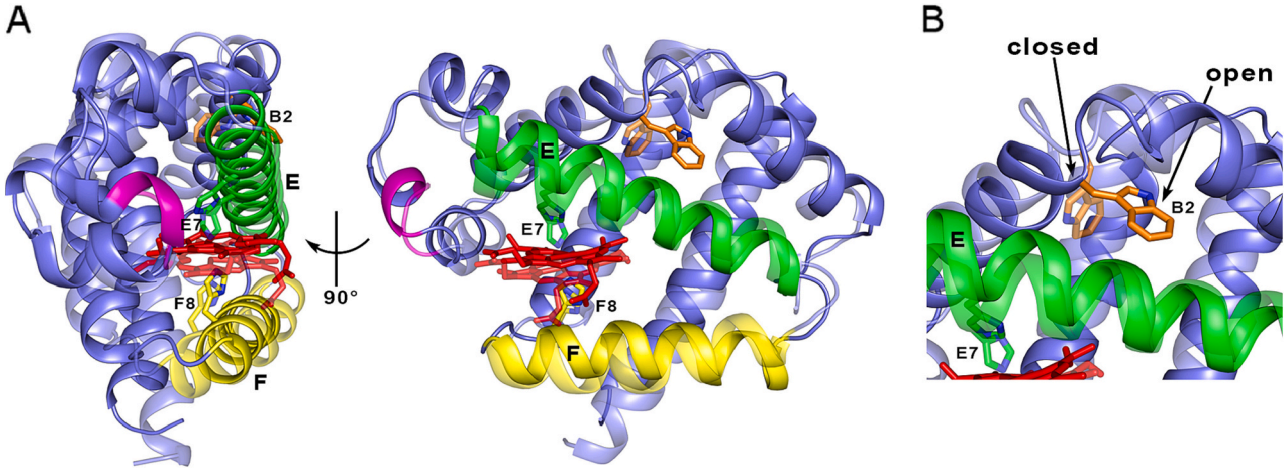

Fig. 8. (A) Two views of the rearrangement of helices E (green) and F (yellow) between 6cHis (transparent) and 5c (opaque) states. The *Ss*Hb heme is shown in red. (B) Arrangement of Trp29(B2) (shown as sticks with C atoms in orange) in the closed conformation embedded in the protein matrix found in the 6cHis state, and in the open conformation protruding toward the interface between subunits A and B found in the 5c state. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

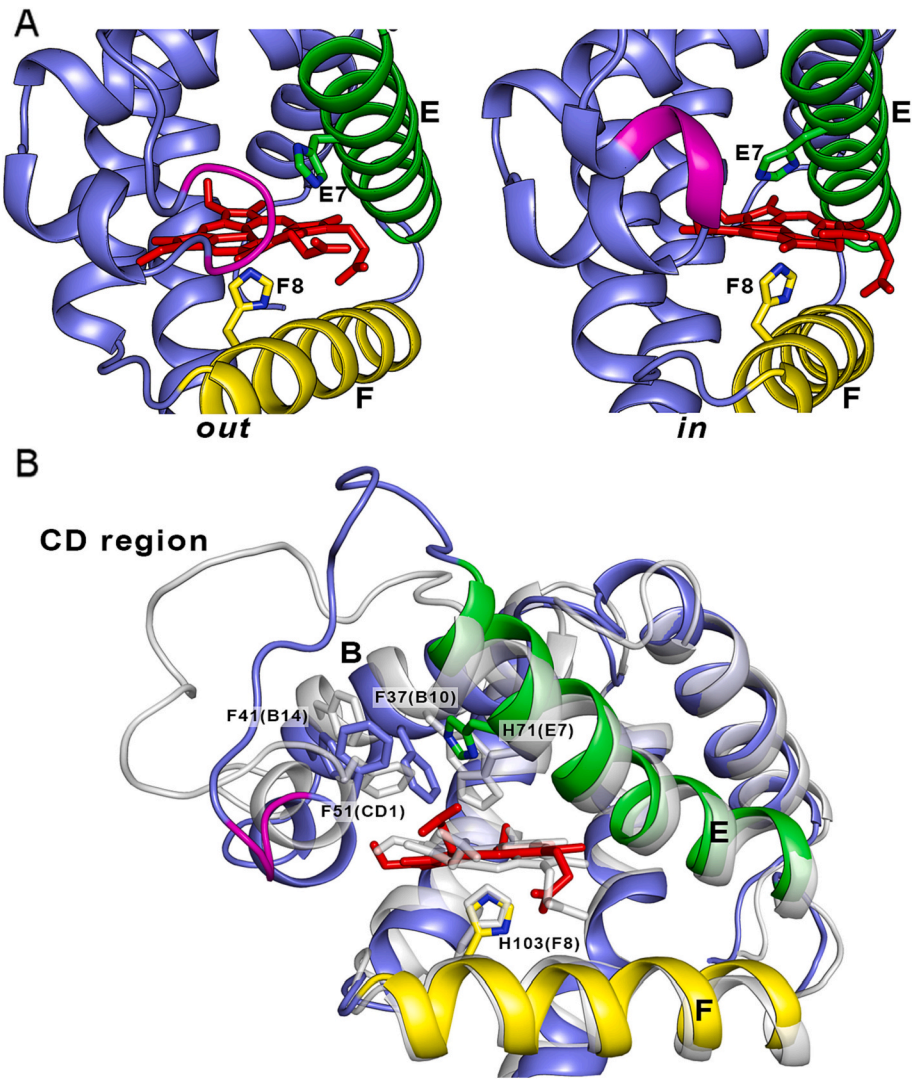

Fig. 9. (A) *Out* and *in* conformations of His71(E7) side chain into the distal pocket during MD simulations. (B) Comparison of selected snapshots (coloured and gray) that show the conformational change in helix B and the CD loop that enables the rearrangement of His71(E7), Phe37(B10), Phe41(B14) and F51(CD1) in the distal pocket.

between the hemes, as shown by the Fe—Fe distance of 20.7 ± 0.7 Å, although values close to 23 Å are transiently reached in replica 1 (Table S1 and Fig. S9). Furthermore, the interaction between propionate A and Arg106(F11) is lost at the beginning of the simulation and is not recovered along the trajectory for none of the two replicas (Table S1). Likewise, the interaction of Lys98(F3) with the propionate or either Asn74(E10) between monomers (Fig. 4) are also weakened or even lost along the simulations (Table S1). While these results point at looser crosstalk interactions between the residues that surround the heme groups in the dimeric $6cO_2$ state, the interaction between residues Arg69 (E5) and Glu88 are reinforced (Fig. 7).

3.5. Protein cavity system

The different coordination state of the heme not only has an impact on the dynamic behavior of the protein backbone but also on the distribution of cavities within the protein matrix. In the 6cHis state, a set of inner cavities are clearly delineated in the inner protein matrix, independently of the monomeric or dimeric species as well as between the two replicas run for the *Ss*Hb dimer, which can be explained by the low flexibility observed for the 6cHis state (Fig. 10 and Fig. S11). These include i) a front cavity located above the heme and defined by the Phe37(B10) and Phe41(B14) aromatic rings, together with Phe51(CD1) and Phe54 in the CD loop, ii) a central cavity located at the inner edge of the heme shaped by Phe36(B9) and Phe37(B10), Phe75(E11), Phe113 (G1), Phe117(G5) and Tyr120(G8), and iii) a back cavity shaped by residues in helices A, E, F and H (Fig. 10). The passage from the front cavity to the central one is primarily limited by Phe37(B10), which is packed against the distal hydrophobic cage formed by Phe36(B9), Phe41 (B14), His71(E7), and Phe75(E11); the transition between the central and back cavities is limited by the side chains of Leu86, Leu142(H12) and ValH145(H15). Analysis of the 6cHis *Ss*Hb dimer reveals that the front cavity protrudes from the distal pocket of each chain and reaches the dimer interface passing close to helix F toward the heme of the opposing subunit (Fig. 10A); nevertheless, continuity between the two cavities is interrupted by the heme propionates.

In contrast with the topology found for the 6cHis state, the increased conformational flexibility found for the 5c dimer is reflected by a volume reduction in the inner protein matrix cavities (Fig. 10B). Strikingly, this effect is opposite at the interface of the dimer, where one may notice the presence of a series of tunnels that would facilitate a connection between the front cavities of the two subunits. Thus, an S-shaped passage emerges from the heme cavity of a chain, slides through the edge defined by helices E and F of the opposing subunit, crosses the void space

Table 4
Angle (°) formed by the E and F helices in the monomeric and dimeric (chains A and B) species of *Ss*Hb in the 6cHis, 5c and 6cO_2 states, and relative orientation (°) of the vectors defined by the pair of E-F helices in the dimer. Values averaged for the last 250 ns of the MD simulation (standard deviation in parenthesis).

Coordination state	Replica	Monomer [#]	Dimer		
			Chain A [#]	Chain B [#]	Interface [¶]
6cHis	X-ray	–	28.9	28.9	122.8
	R1	29.4 (0.8)	29.4 (0.6)	29.6 (0.5)	124.3 (2.1)
	R2	28.7 (0.7)	29.2 (0.6)	29.2 (0.6)	124.5 (2.1)
5c	R1	30.8 (1.1)	33.5 (0.7)	32.9 (0.9)	114.7 (1.6)
	R2	31.8 (1.1)	31.5 (1.1)	37.3 (1.2)	115.0 (2.3)
6cO_2	R1	32.8 (0.9)	37.3 (0.8)	34.4 (0.8)	116.3 (2.2)
	R2	32.4 (0.9)	35.0 (0.9)	30.1 (0.8)	119.0 (1.9)

[#] The angle between E and F helices was estimated by considering the vectors defined by the position of the C_α atoms of His71(E7) and His103(F8) to the C_α atom of Glu88 at the EF hinge.

[¶] The orientation between the pair of E-F helices was measured from the angle between the unit vectors obtained by the addition of the vectors in the E and F helices within each subunit.

between the heme groups, touches the edge between helices E and F of the original chain, and ends up in the front cavity of the opposing subunit. The formation of this passage may be attributed to the separation of the opposing helices E and F, as well as the separation between the Fe atoms in the two chains, since the Fe—Fe distance increases from 18.1 Å in the 6cHis dimer to 19.8 Å in the 5c dimeric state. In the monomer, the front cavity is limited to the interior of the distal cage, without protrusion toward the surface of the monomer. The central and back cavities are smaller compared to those found in the dimer. So, the front cavity would be the most affected by the dimerization process.

In the 6cO_2 state the increased structural rigidity of the backbone relative to the 5c state allows to delineate a larger number of bigger cavities in the inner protein matrix, which permits to envisage a series of cavities that link the heme distal cavity to the bulk solvent passing through helices F and H (Fig. 10C). On the other hand, the passage *via* secondary channels through the loop AB and the terminus of helix G are prevented by the steric hindrance exerted by Phe75(E11) and Tyr120 (G8). Compared to the dimer, the back cavity is larger in the monomer, whereas the front cavity is smaller.

4. Conclusions

Among nHbs, *Ss*Hb has been extensively characterized biochemically and phylogenetically and proposed to perform a Mb-like function by enhancing oxygen supply to neurons [10,17]. The *Ss*Hb three-dimensional structure reveals a substantially conserved classical myoglobin fold where the heme-Fe atom is hexa-coordinated, having the proximal His103(F8) and the distal His71(E7) as axial ligand residues, confirming previous spectroscopic observations and predictions based on sequence analysis [10]. *In vitro* (at low concentrations) the recombinant protein was reported to have a P_{50} value for O_2 of ~0.6 Torr, the ligand binding occurring in a non-cooperative manner (slope of ~1 in the Hill plot) [10]. However, other studies on recombinant *Ss*Hb indicated cooperativity on O_2 binding, reporting a Hill coefficient increasing with pH and decreasing with temperature, attaining a value of 1.5 [15] and a dimeric oligomerization state [49]. The latter results fit better the *in situ* measurements on intact nerves at high protein concentrations (mM), that revealed a P_{50} of 2.3 Torr, and an n value of 2.1, suggesting *in vivo* aggregation, probably higher than dimeric, triggered by the high cellular concentration and/or by the presence of allosteric effectors moderately lowering O_2 affinity [6].

Interestingly, the *Ss*Hb crystal structure here reported shows the protein as a stable dimeric species, in keeping with previously reported size exclusion chromatography data [49]. Considering that the vapor diffusion crystallization method is based on approaching protein saturation by slowly increasing its concentration, and that the *Ss*Hb dimerization mode found is of the EF-type, well known within the globin superfamily, it is conceivable that the *Ss*Hb quaternary assembly found in the crystalline state is representative of the basic oligomerization state found *in vivo*, at physiological *Ss*Hb concentration.

Of note, to the best of our knowledge, *Ss*Hb is the first EF-type dimeric structure displaying bis-histidyl heme coordination. This observation appears functionally relevant since bis-histidyl hexa-coordinated globins require a structural rearrangement (shift or tilt) of the E helix to allow binding of an exogenous ligand to the heme. In this scenario, an EF dimeric interface would support ligand binding cooperativity between the globin subunits provided that the dimeric packing allows sufficient dynamics. Indeed, the *Ss*Hb dimeric interaction is mediated mostly by polar resides (12 out of 17 interface residues), and a large number of water molecules (~70) are bound at the dimer association interface, mediating inter-subunit interactions. Such polar surface would explain, on one side, why the protein can stay monomeric in solution at low concentration; on the other, it supports a stable dimer that is able to breathe, thus enabling enough flexibility to allow the structural rearrangements needed in the transition from the bis-histidyl (6cHis state) to penta-coordination (5c state).

Our MD simulations confirmed the trend described above, showing that the penta-coordinate species is more flexible than the 6cHis and the 6cO_2 species. Overall, the 6cHis →5c transition is granted by the flexibility of the CD loop region, which allows the *E*-helix to tilt away from the heme-distal side thus breaking the bis-histidyl coordination. The E-helix move is coupled to a closed-open transition of Trp29(B2) side chain. This residue is embedded in the protein matrix in the 6cHis state and contributes, together with the heme-Fe—His71(E7) coordination, to restrain the magnitude of the structural fluctuations in both the protein backbone and side chains. In contrast, the steric hindrance with Ile32 (B5) caused by the E helix shift after breakage of the Fe—His71(E7) bond promotes the rearrangement of Trp29(B2) toward the subunits interface in the dimeric species, thus facilitating the accommodation of the E helix in the 5c state and providing a basis for enhanced residue fluctuations within the protein matrix.

The structural changes triggered by the 6cHis →5c transition have an impact on the balance between electrostatic and van der Waals interactions of the inter-subunit contact residues, and eventually are reflected in distinct shapes of the dimeric interface tunnels, a quaternary structure feature that would facilitate connections between the two heme groups. Related to such observation is the 'cross-talk' between heme propionates from one subunit and Asn102(F7)/Lys98(F3) of the facing chain, an interaction that is strong in the 6cHis state and tends to become looser in the 5c and 6cO_2 states, while the inter-heme distance increases.

Concerning the gating movement of His71(E7) at the distal pocket, we note that such structural effect would be in keeping with the reported *Ss*Hb reverse Bohr effect, ascribed in hexa-coordinated nerve globins to HisE7 protonation at low pH, that would increase O_2 affinity as the (protonated) residue swings out of the pocket [15]. At the end of the binding process, the heme-coordinated O_2 is hosted in a 'aromatic cage' defined by Phe37(B10), Phe51(CD1) and Phe75(E11) and, as expected, stabilized by a H-bond to the distal His71(E7) in the 6cO_2 state.

Both inter-subunit 'cross-talk' and hemes interconnection through tunnels at the dimeric interface are features that would be in keeping with the reported *in vivo* functional cooperativity reported for *Ss*Hb [6,17], while the progressive distancing of the subunits throughout the 6cHis → 5c transition would render the heme pocket more accessible, thus supporting high ligand association rates and the high O_2 affinity

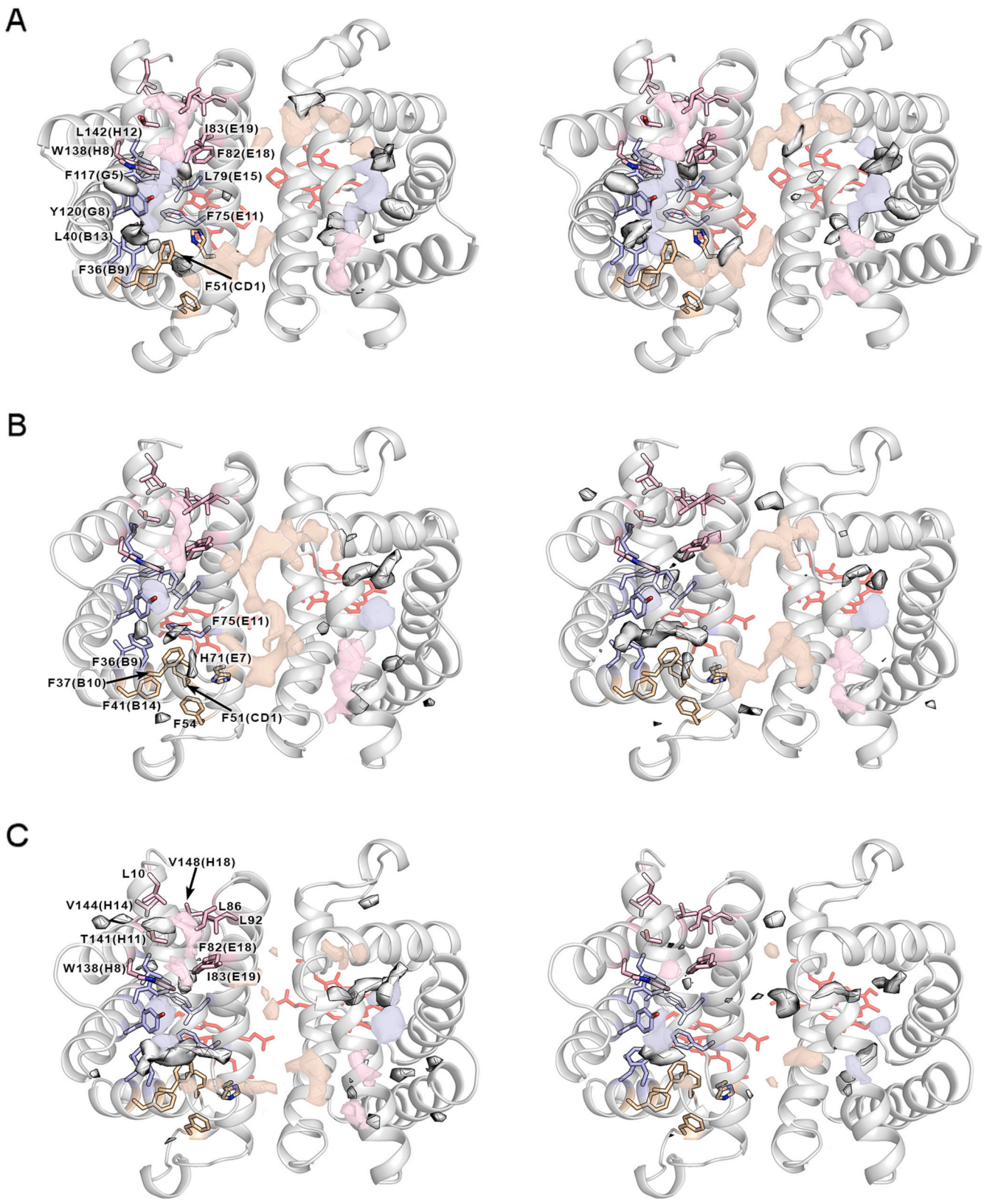

Fig. 10. Representation of the cavities located in the inner protein matrix and at the interface between subunits for the *Ss*Hb dimer in the (A) 6cHis state, (B) 5c state, and (C) 6cO_2 state. The front, central and back cavities are shown as isocontours coloured in yellow, violet and pink, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

observed for *Ss*Hb [15]. Finally, the unusually high His (intramolecular) association rate (100 μs) reported for *Ss*Hb [10] might be explained by an additionally increased inter-subunit distance, indicated in the $6cO_2$ state by MD simulations, where the E helix and His71(E7) are less constrained in their heme-rebinding shifts. Such scenario would also suggest that oxygenation and *in vitro* dilution may cause dissociation of the dimer into high-affinity monomers, as previously proposed for *Ss*Hb and reported for other invertebrate nHbs, such as *Cerebratulus lacteus* nHb [10,15].

CRediT authorship contribution statement

Alessandra Pesce: Methodology, Investigation, Writing – original draft. **Katerina Barmpidi:** Methodology, Investigation, Writing – original draft. **Sylvia Dewilde:** Conceptualization, Methodology. **Carolina Estarellas:** Methodology, Investigation, Writing – original draft. **Luc Moens:** Conceptualization, Supervision. **Martino Bolognesi:** Writing – review & editing, Supervision. **Francisco Javier Luque:** Conceptualization, Methodology, Investigation, Writing – review & editing, Funding acquisition. **Marco Nardini:** Conceptualization, Investigation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Coordinates and structure factors have been deposited in the Protein Data Bank under accession number 8OUP. MD data will be made available on request.

Acknowledgements

This work is dedicated to Sylvia Dewilde who was strongly involved in this project and sadly passed away on October 3, 2020, at the age of 47.

Funding: K.B., C.E. and F.J.L. acknowledge financial support from the State Research Agency/Spanish Ministry of Science and Innovation (AEI/10.13039/501100011033; grants PID2020-117646RB-I00 and CEX2021-001202-M), the Generalitat de Catalunya (grant 2021SGR00671). The Consorci de Serveis Universitaris de Catalunya (CSUC) is acknowledged for providing computational resources (Molecular Recognition project). MB and MN are grateful to the University of Milano NOLIMITS Center for continuous support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jinorgbio.2023.112289.

References

[1] T. Burmester, B. Weich, S. Reinhardt, T. Hankeln, A vertebrate globin expressed in the brain, Nature 407 (2000) 520–523, https://doi.org/10.1038/35035093.
[2] T. Burmester, T. Hankeln, Neuroglobin and other nerve haemoglobins, in: M. Bolognesi, G. di Prisco, C. Verde (Eds.), Dioxygen Binding and Sensing Proteins. Protein Reviews 9, Springer, Milano, 2008, pp. 211–222, https://doi.org/10.1007/978-88-470-0807-6_18.
[3] J.B. Wittenberg, Functions of cytoplasmic hemoglobins and myohemerythrin, in: C. P. Mangum (Ed.), Blood and Tissue Oxygen Carriers. Advances in Comparative and Environmental Physiology 13, Springer-Verlag, Berlin, Heidelberg, 1992, pp. 59–85, https://doi.org/10.1007/978-3-642-76418-9_3.
[4] R.E. Weber, S.N. Vinogradov, Nonvertebrate hemoglobins: functions and molecular adaptations, Physiol. Rev. 81 (2001) 569–628, https://doi.org/10.1152/physrev.2001.81.2.569.
[5] A. Arvanitaki, N. Chalazonitis, Photo potentials d'excitation et d'inhibition de differents somata identifiables (*Aplysia*). Activation monocromatiques, Bull. Inst. Océanogr. Monaco 1164 (1960) 1–83.
[6] D.W. Kraus, J.E. Doeller, P.R. Smith, A physiological comparison of bivalve mollusc cerebro-visceral connectives with and without neurohemoglobin. I. Ultrastructural and electrophysiological characteristics, Biol. Bull. 174 (1988) 54–66, https://doi.org/10.2307/1541759.
[7] I. Schindelmeiser, D. Kuhlmann, A. Nolte, Localization and characterization of hemoproteins in the central nervous tissue of some gastropods, Comp. Biochem. Physiol. B 64B (1979) 149–154, https://doi.org/10.1016/0305-0491(79)90153-6.
[8] T.L. Vandergon, C.K. Riggs, T.A. Gorr, J.M. Colacino, A.F. Riggs, The mini-hemoglobins in neural and body wall tissue of the nemertean worm *Cerebratulus lacteus*, J. Biol. Chem. 273 (1998) 16998–17011, https://doi.org/10.1074/jbc.273.27.16998.
[9] B.A. Wittenberg, R.W. Briehl, J.B. Wittenberg, Haemoglobins of invertebrate tissues. Nerve haemoglobins of *Aphrodite, Aplysia* and *Halosydna*, Biochem. J. 96 (1965) 363–371, https://doi.org/10.1042/bj0960363.
[10] S. Dewilde, B. Ebner, E. Vinck, K. Gilany, T. Hankeln, T. Burmester, J. Kreiling, C. Reinisch, J.R. Vanfleteren, L. Kiger, M.C. Marden, C. Hundahl, A. Fago, S. Van Doorslaer, L. Moens, The nerve hemoglobin of the bivalve mollusc *Spisula solidissima*: molecular cloning, ligand binding studies, and phylogenetic analysis, J. Biol. Chem. 281 (2006) 5364–5372, https://doi.org/10.1074/jbc.M509486200.
[11] D.W. Kraus, J.M. Colacino, Extended oxygen delivery from the nerve hemoglobin of *Tellina alternata* (bivalvia), Science 232 (1986) 90–92, https://doi.org/10.1126/science.232.4746.90.
[12] P. Strittmatter, H.B. Burch, The heme protein in ganglia of *Spisula solidissima*, Biochim. Biophys. Acta 78 (1963) 562–563, https://doi.org/10.1016/0006-3002(63)90928-4.
[13] M. Couture, T. Burmester, T. Hankeln, D.L. Rousseau, The heme environment of mouse neuroglobin: evidence for the presence of two conformations of the heme pocket, J. Biol. Chem. 276 (2001) 36377–36382, https://doi.org/10.1074/jbc.M103907200.
[14] J.T. Sage, D. Morikis, P.M. Champion, Spectroscopic studies of myoglobin at low pH: heme structure and ligation, Biochemistry 30 (1991) 1227–1237, https://doi.org/10.1021/bi00219a010.
[15] C. Hundahl, A. Fago, S. Dewilde, L. Moens, T. Hankeln, T. Burmester, R.E. Weber, Oxygen binding properties of non-mammalian nerve globins, FEBS J. 273 (2006) 1323–1329, https://doi.org/10.1111/j.1742-4658.2006.05158.x.
[16] D.W. Kraus, J.E. Doeller, A physiological comparison of bivalve mollusc cerebro-visceral connectives with and without neurohemoglobin. III. Oxygen demand, Biol. Bull. 174 (1988) 346–354, https://doi.org/10.2307/1541960.
[17] S. Dewilde, M. Blaxter, M.L. van Hauwaert, J. Vanfleteren, E.L. Esmans, M. Marden, N. Griffon, L. Moens, Globin and globin gene structure of the nerve myoglobin of *Aphrodite aculeata*, J. Biol. Chem. 271 (1996) 19865–19870, https://doi.org/10.1074/jbc.271.33.19865.
[18] T.G.G. Battye, L. Kontogiannis, O. Johnson, H.R. Powell, A.G.W. Leslie, iMosflm: a new graphical interface for diffraction-image processing with MOSFLM, Acta Crystallogr. D Biol. Crystallogr. 67 (2011) 271–281, https://doi.org/10.1107/S0907444910048675.
[19] P.R. Evans, Proceedings of the CCP4 Study Weekend on Data Collection and Processing, 114–122, CLRC Daresbury Laboratory, UK, 1993.
[20] L.C. Storoni, A.J. McCoy, R.J. Read, Likelihood-enhanced fast rotation functions, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 432–438, https://doi.org/10.1107/S0907444903028956.
[21] The CCP4 suite: programs for protein crystallography. Collaborative Computational Project, Number 4, Acta Crystallogr. D Biol. Crystallogr. 50 (1994) 760–763, https://doi.org/10.1107/S0907444994003112.
[22] J.A. Gavira, A. Camara-Artigas, W. De Jesús-Bonilla, J. López-Garriga, A. Lewis, R. Pietri, S.R. Yeh, C.L. Cadilla, J.M. García-Ruiz, Structure and ligand selection of hemoglobin II from *Lucina pectinata*, J. Biol. Chem. 283 (2008) 9414–9423, https://doi.org/10.1074/jbc.M705026200.
[23] P. Emsley, K. Cowtan, Coot: model-building tools for molecular graphics, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 2126–2132, https://doi.org/10.1107/S0907444904019158.
[24] G.N. Murshudov, A.A. Vagin, E.J. Dodson, Refinement of macromolecular structures by the maximum-likelihood method, Acta Crystallogr. D Biol. Crystallogr. 53 (1997) 240–255, https://doi.org/10.1107/S0907444996012255.
[25] V.B. Chen, W.B. Arendall, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L. W. Murray, J.S. Richardson, D.C. Richardson, MolProbity: all-atom structure validation for macromolecular crystallography, Acta Crystallogr. D Biol. Crystallogr. 66 (2010) 12–21, https://doi.org/10.1107/S0907444909042073.
[26] E. Krissinel, K. Henrick, Inference of macromolecular assemblies from crystalline state, J. Mol. Biol. 372 (2007) 774–797, https://doi.org/10.1016/j.jmb.2007.05.022.
[27] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I. N. Shindyalov, P.E. Bourne, The protein data bank, Nucleic Acids Res. 28 (2000) 235–242, https://doi.org/10.1093/nar/28.1.235.
[28] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T.A.P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, Nucleic Acids Res. 46 (W1) (2018) W296–W303, https://doi.org/10.1093/nar/gky427.
[29] D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A. W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, V. Man, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, A. Onufriev,

F. Pan, S. Pantano, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, L. Wilson, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, P.A. Kollman, Amber 2020, University of California, San Francisco, 2020.

[30] C. Tian, K. Kasavajhala, K. Belfon, L. Raguette, H. Huang, A. Migues, J. Bickel, Y. Wang, J. Pincay, Q. Wu, C. Simmerling, ff19SB: amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution, J. Chem. Theory Comput. 16 (2020) 528–552, https://doi.org/10.1021/acs.jctc.9b00591.

[31] A. Crespo, M.A. Marti, S.G. Kalko, A. Morreale, M. Orozco, J.L. Gelpi, F.J. Luque, D. A. Estrin, Theoretical study of the truncated hemoglobin HbN: exploring the molecular basis of the NO detoxification mechanism, J. Am. Chem. Soc. 127 (2005) 4433–4444, https://doi.org/10.1021/ja0450004.

[32] A. Bidon-Chanal, M.A. Marti, A. Crespo, M. Milani, M. Orozco, M. Bolognesi, F. J. Luque, D.A. Estrin, Ligand-induced dynamical regulation of NO conversion in *mycobacterium tuberculosis* truncated hemoglobin-N, Proteins 64 (2006) 457–464, https://doi.org/10.1002/prot.21004.

[33] Y. Xiong, P.S. Shabane, A.V. Onufriev, Melting points of OPC and OPC3 water models, ACS Omega 5 (2020) 25087–25094, https://doi.org/10.1021/acsomega.0c02638.

[34] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. Di Nola, J.R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690, https://doi.org/10.1063/1.448118.

[35] A.R. Leach, Molecular Modelling: Principles and Applications, 2nd ed., Pearson Education Limited, 2001.

[36] V. Le Guilloux, P. Schmidtke, P. Tuffery, Fpocket: an open source platform for ligand pocket detection, BMC Bioinformatics. 10 (2009) 168, https://doi.org/10.1186/1471-2105-10-168.

[37] P. Schmidtke, A. Bidon-Chanal, F.J. Luque, X. Barril, MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories, Bioinformatics 27 (2011) 3276–3285, https://doi.org/10.1093/bioinformatics/btr550.

[38] A. Onufriev, D. Bashford, D.A. Case, Exploring protein native states and large-scale conformational changes with a modified generalized born model, Proteins 55 (2004) 383–394, https://doi.org/10.1002/prot.20033.

[39] B.R. Miller, T.D. McGee, J.M. Swails, N. Homeyer, H. Gohlke, A.E. Roitberg, MMPBSA.py: an efficient program for end-state free energy calculations, J. Chem. Theory Comput. 8 (2012) 3314–3321, https://doi.org/10.1021/ct300418h.

[40] M.F. Perutz, Regulation of oxygen affinity of hemoglobin: influence of structure of the globin on the heme iron, Annu. Rev. Biochem. 48 (1979) 327–386, https://doi.org/10.1146/annurev.bi.48.070179.001551.

[41] J. Vojtechovský, K. Chu, J. Berendzen, R.M. Sweet, I. Schlichting, Crystal structures of myoglobin-ligand complexes at near-atomic resolution, Biophys. J. 77 (1999) 2153–2174, https://doi.org/10.1016/S0006-3495(99)77056-6.

[42] W.E. Royer, J.E. Knapp, K. Strand, H.A. Heaslet, Cooperative hemoglobins: conserved fold, diverse quaternary assemblies and allosteric mechanisms, Trends Biochem. Sci. 26 (2001) 297–304, https://doi.org/10.1016/s0968-0004(01)01811-4.

[43] D. de Sanctis, S. Dewilde, A. Pesce, L. Moens, P. Ascenzi, T. Hankeln, T. Burmester, M. Bolognesi, Crystal structure of cytoglobin: the fourth globin type discovered in man displays heme hexa-coordination, J. Mol. Biol. 336 (2004) 917–927, https://doi.org/10.1016/j.jmb.2003.12.063.

[44] M. Bolognesi, D. Bordo, M. Rizzi, C. Tarricone, P. Ascenzi, Nonvertebrate hemoglobins: structural bases for reactivity, Prog. Biophys. Mol. Biol. 68 (1997) 29–68, https://doi.org/10.1016/S0079-6107(97)00017-5.

[45] L. Holm, Dali server: structural unification of protein families, Nucleic Acids Res. 50 (W1) (2022) W210–W215, https://doi.org/10.1093/nar/gkac387.

[46] M. Makino, H. Sugimoto, H. Sawai, N. Kawada, K. Yoshizato, Y. Shiro, High-resolution structure of human cytoglobin: identification of extra N- and C-termini and a new dimerization mode, Acta Crystallogr. D Biol. Crystallogr. 62 (2006) 671–677, https://doi.org/10.1107/S0907444906013813.

[47] H. Sugimoto, M. Makino, H. Sawai, N. Kawada, K. Yoshizato, Y. Shiro, Structural basis of human cytoglobin for ligand binding, J. Mol. Biol. 339 (2004) 873–885, https://doi.org/10.1016/j.jmb.2004.04.024.

[48] M. Gabba, S. Abbruzzetti, F. Spyrakis, F. Forti, S. Bruno, A. Mozzarelli, F.J. Luque, C. Viappiani, P. Cozzini, M. Nardini, F. Germani, M. Bolognesi, L. Moens, S. Dewilde, CO rebinding kinetics and molecular dynamics simulations highlight dynamic regulation of internal cavities in human cytoglobin, PLoS One 8 (2013) 49770, https://doi.org/10.1371/journal.pone.0049770.

[49] J.E. Doeller, D.W. Kraus, A physiological comparison of bivalve mollusc cerebro-visceral connectives with and without neurohemoglobin. II. Neurohemoglobin characteristics, Biol. Bull. 174 (1988) 67–76, https://doi.org/10.2307/1541760.