RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Insight into the structural and functional features of myoglobin from *Hystrix cristata* L. and *Rangifer tarandus* L..

Antonella M. A. Di Giuseppe,[a] Jolanda V. Caso,[a] Valeria Severino,[b] Sara Ragucci,[a] Angela Chambery,[a] Rosita Russo,[a] Roberto Fattorusso,[a] Josè M. Ferreras,[c] Luigi Russo[a] and Antimo Di Maro [a,] *

[a]*Department of Environmental, Biological and Pharmaceutical Sciences and Technologies (DiSTABiF), Second University of Naples, Via Vivaldi 43, 81100-Caserta, Italy.*

[b]*Department of Internal Medicine Specialties, Geneva University, Rue Michel-Servet 1, 1211-Genève, Switzerland.*

[c]*Department of Biochemistry and Molecular Biology and Physiology, University of Valladolid, Campus Miguel Delibes, Paseo de Belén 17, 47011-Valladolid, Spain.*

*, E-mail: antimo.dimaro@unina2.it (ADM); *Tel*: +39 0823 274535, *Fax*: +39 0823 274571.

Abstract

The amino acid sequence, structural and functional features of two novel myoglobins (Mbs) isolated from crested porcupine (*Hystrix cristata* L.) and reindeer (*Rangifer tarandus* L.) were determined. The primary structure was achieved by using a combined approach based on *de novo* sequencing by ESI-Q-TOF MS/MS and peptide mapping by MALDI-TOF MS. This strategy allowed us to determine the primary structure of crested porcupine and reindeer Mbs. To go deeper, 3D modeling studies followed by structural characterization by NMR on both myoglobins demonstrate that reindeer Mb shows slightly different orientation of F, G and H α-helices. As a consequence, reindeer Mb may differently modulate the heme environment, facilitating oxygenation as well as ensuring that the heme iron remains in a ferrous state. Finally, reindeer Mb shows a less stable conformation with respect to crested porcupine Mb (T_m 353.7 K *vs* T_m 356.3 K, respectively).

Keywords: circular dichroism, *Hystrix cristata*, mass spectrometry, myoglobin, protein purification, *Rangifer tarandus*.

1. Introduction

The oxygen requirement by tissues demands the evolution of macromolecules necessary for its transport and/or storage [1]. This necessity is particularly enhanced in vertebrates muscle since that the high oxygen demand cannot be only satisfied by a continuous passive or active transport. A particularly specialized family of proteins, the globins [2], has been evolved to reversibly bind to the molecular oxygen, due to the presence of the heme group. During the evolution of the circulatory system, the globins family has given rise to tetrameric proteins, as hemoglobin, that carry oxygen [3], and monomeric proteins, as myoglobin, that are necessary in the storage and cellular diffusion of oxygen [4].

Myoglobin (Mb) is expressed in cardiac myocytes and oxidative skeletal muscle fibers [5] and consists of eight alpha helices connected by various loops [6]. Mb binds oxygen by its heme residue, a porphyrin ring:iron ion complex. The polypeptide chain (~150 amino acids) is folded and cradles the heme prosthetic group, positioning it between two histidinyl residues, His64 and His93. The iron ion interacts with six ligands, four of which are provided by the nitrogen atoms of the four pyroles. The imidazole side chain of His93 provides the fifth ligand, stabilizing the heme group and slightly displacing the iron ion away from the plane of the heme. The sixth ligand position, unoccupied in deoxymyoglobin, serves as binding site for O_2, as well as for other potential ligands such as CO or NO.

Mb has been studied for long time because of: i) its role in several human diseases [7, 8]; ii) understanding structure/function relationships in proteins [9]; and iii) its scavenging activity against reactive oxygen species and bioactive nitric oxide [10]. Furthermore, myoglobin is extensively studied in food and quality control fields considering its role in determining the meat colour and consequently its attractiveness for the consumers [11]. Indeed, Mbs of skeletal and cardiac muscle can assume three forms: oxymyoglobin (OxyMb), deoxymyoglobin (deoxyMb) and metmyoglobin (metMb), whose relative amounts determine the colour of fresh meat. *In vivo* the metMb is in low amounts due to the presence of a specific reducing system. The amount of metMb increases in the absence of metabolic energy when this system does not work [12]. Moreover, Mbs isolated from several species have been studied by our research group because of their potential use as molecular markers for the detection of fraudulent addition of undeclared species in raw meat [13, 14].

In this framework, the characterization of Mbs isolated from species whose meats is used in human diet is challenging, especially if integrated with studies on their structural features. In fact, these proteins result attractive in phylogenetic studies, since few residue substitutions can alter their structural and functional properties [15, 16].

In this work, primary structures of two novel myoglobins isolated from crested porcupine (*Hystrix cristata* L.) and reindeer (*Rangifer tarandus* L.) were determined by using a combined approach based on LC-ESI MS/MS and MALDI-TOF MS. At first, in order to fully describe the structural characteristics, 3D modeling analysis was performed for both Mbs. Next, the 3D models were validated using experimental data obtained by Nuclear Magnetic Resonance (NMR) spectroscopy. Our data showed that the two myoglobins adopt a similar compact structure with small but significant structural differences. In particular, due to a displacement of helices F, G and H, reindeer Mb is characterised by a slightly less compact fold with respect to crested porcupine. Moreover, to characterize the functional proprieties of both Mbs a series of autoxidation kinetics and thermal stability measurements were performed. The higher autoxidation rate of reindeer Mb with respect to crested porcupine one could be explained by its less stable conformation as confirmed by 3D models analysis. Our results suggest that the highlighted structural differences between reindeer and crested porcupine myoglobins might play an important role, facilitating oxygenation, in the modulation process of the heme environment.

2. Experimental

2.1. Meat species

Crested porcupine (thigh muscles) meat was collected from local hunters and kept at -20°C until use. Reindeer (thigh muscles) meat was purchased from a London (UK) food store and kept at -20°C until use.

2.2. Enzymes and chemicals

Cyanogen bromide and endoproteinases (trypsin, chymotrypsin, Asp-N and Glu-C) were purchased from Sigma-Aldrich (Milan, Italy). Solvents for RP-HPLC were supplied by Carlo Erba (Milan, Italy). Bicinchoninic acid (BCA) kit was purchased from Pierce (Rockford, IL, USA). Materials for chromatography were described elsewhere [17, 18]. The following solvents were used for RP-HPLC: solvent A, 0.1% TFA in water; solvent B, acetonitrile containing 0.1% TFA. All other reagents and chemicals were of analytical grade.

2.3. Extraction and purification procedure

Mbs were isolated from crested porcupine and reindeer muscles as previously described [19, 20]. Briefly, following partial removal of the fat and connective tissues, meat samples (about 50 g) were homogenized in 10 mM Tris•HCl buffer, pH 8.8 (1:2; w:v) using a Waring blender (Waring Products, Torrington, CT, USA), at 4°C. The homogenate was centrifuged, filtered through Miracloth paper (Calbiochem, San Diego, CA, USA) and dialysed in same buffer. Mbs were then gel-filtered on Sephacryl S-100 HR (GE Healthcare, Milan, Italy), and subjected to anion exchange chromatography on Source™ 15Q FPLC column (GE Healthcare), using the AKTA prime 100 FPLC (GE Healthcare). During purification procedure the absorbance was measured at 280 and 409 nm to monitoring heme-proteins, and the protein homogeneity checked by SDS-PAGE [21].

2.4. Preparation of apomyoglobin

The separation of apo-Mb from the heme group was performed by reversed-phase HPLC (RP-HPLC) on a C-4 column (4.6 x 150 mm, Alltech, Sedriano, Milan, Italy) as previously reported [20].

2.5. Analytical procedures and peptides separation

Chemical fragmentation with cyanogen bromide (CNBr) was performed in 70% formic acid [22]. Digestions with trypsin and chymotrypsin were performed as previously reported [23, 24]. Digestion with endoproteinase Glu-C was performed by two additions of the enzyme with a final enzyme-to-substrate ratio of 1:50 (w:w). Following incubation (37 °C for 24 h), digested samples were centrifuged at 15,800*g* for 10 min (GS-15R centrifuge; Beckman Coulter, Milan, Italy).

When needed, separation of endoproteinase and CNBr peptides by RP-HPLC was performed on a Breeze instrument (Waters S.p.A, Vimodrone, Milan, Italy), equipped with Symmetry C-18 column (0.46 x 150 mm; 5 µm particle size; Waters SpA) or C-4 column (0.46 x 250 mm; 5 µm particle size; Phenomenex, Castel Maggiore, BO, Italy), respectively, as previously reported [17, 25].

2.6. Mass spectrometry analysis

The relative molecular masses (M*r*) of whole myoglobins were determined by mass spectrometry using a quadrupole time of flight (Q-TOF) mass spectrometer (Q-TOF Micro, Waters, Manchester, UK) equipped with an electrospray ionisation (ESI) source. The capillary source voltage and the cone voltage were set at 3000 and 43 V, respectively. The source temperature was kept at 80 °C and nitrogen was used as drying gas (flow rate about 50 L/h). Samples from RP-HPLC were diluted to a concentration of 10 pmol/L with acetonitrile containing 0.1% formic acid in water (50:50, v/v) and infused into the system at a flow rate of 20 L/min. Peptides were separated by means of a modular CapLC system (Waters) directly connected with the ESI source. Samples were loaded onto a C-18 precolumn (5 mm length x 300 µm ID) at a flow rate of 20 µL/min and desalted for 5 min with solution of 0.1% formic acid. Peptides were then directed onto a symmetry-C18 analytical column (10 cm x 300 µm ID) using 5% CH_3CN, containing 0.1% formic acid at a flow rate of 5 µL/min. Elution was obtained by increasing the CH_3CN/0.1% formic acid concentration from 5% to 55% over 60 min. The precursor ion and the associated fragment ions present in the mass spectra of the tryptic peptides were measured with the mass spectrometer directly coupled to the chromatographic system. The time-of-flight analyser of the mass spectrometer was externally calibrated with a multi-point calibration using selected fragment ions of the collision induced dissociation (CID) of human [Glu1]-fibrinopeptide B [500 fmol/µL in CH_3CN:H_2O (50:50), 0.1% formic acid] at an infusion rate of 5 µL/min in the TOF MS/MS mode. Electrospray mass spectra and tandem MS/MS data were acquired on the Q-TOF mass spectrometer operating in the positive ion mode.

For MALDI-TOF analysis, 1 µL of digestion mixtures or each peptide solution was mixed with 1 µL of saturated α-cyano-4-hydroxycinnamic acid matrix solution [10 mg/mL in acetonitrile:0.1% TFA (1:1; v/v)] or sinapinic acid [10 mg/mL in acetonitrile/0.1% TFA (2:3; v/v)] [26]. Thus, a droplet of the resulting mixture (1 µL) was placed on the mass spectrometer's sample target and dried at room temperature. Once the liquid was completely evaporated, samples were loaded into the mass spectrometer and analysed. The instrument was externally calibrated using a tryptic alcohol dehydrogenase digest (Waters) in reflectron mode. For linear mode, a four-point external calibration was applied using an appropriate mixture (10 pmol/mL) of insulin, cytochrome C, horse Mb and trypsinogen as standard proteins (Sigma). A mass accuracy near to the nominal (50 and 300 ppm in reflectron and linear modes, respectively), was achieved for each standard.
All spectra were processed and analysed using MassLynx 4.0 software.

2.7. Autoxidation rate measurement

The autoxidation of OxyMb to metMb was monitored by recording the changes of the absorption spectrum in the 500- 700 nm range and estimating the absorbance decrease at 582 nm or 581 nm (the OxyMb α-peak) for crested porcupine and reindeer Mbs, respectively [20], using Synergy HT Multi-Mode Microplate Reader (BioTek, Bad Friedrichshall, Germany). All experiments were performed in triplicate with freshly prepared OxyMb. For the characterisation of the autoxidation process, spectra were collected every 10 min for 5.5 h. Ferrous and ferric Mb derivatives were prepared as previously described [27].

2.8. 3D structure modeling method

The 3D models for both Mbs were predicted by the I-TASSER software on the basis of their amino acid sequences. I-TASSER (Interactive Threading ASSEmbly Refinement) is a computational method that uses a combinatorial approach, employing all three conventional methods for structure modeling: comparative modeling, threading, and *ab initio* modeling [28]. The obtained models were evaluated and visualized using the softwares PROCHECK [29], MolProbity [30], PyMol [31], MOLMOL [32] and Chimera [33]. The estimation of the secondary structure using the predicted models was performed using the software DSSP [34]. The cavity volumes were estimated by CASTp [35] and Kfinder [36] software.

188

189 **2.9. NMR spectroscopy**

190

191 All NMR experiments were carried out at 500 MHz using a Varian Unity 500 spectrometer located
192 at the DiSTABiF in Caserta (Italy). NMR samples typically contained 0.5 mM of crested porcupine
193 or reindeer Mbs, 20 mM phosphate buffer (pH 6.8), 0.2 M NaCl and 90% H_2O/ 10% ^2H_2O. NMR
194 experiments for collecting structural information were performed at 298 K referenced to external
195 TMS (δ= 0 ppm). Deuterium oxide (^2H_2O) was purchased from Cambridge Isotope Laboratories
196 (Andover, MA, USA). Mono (1D) and two dimensional (2D) spectra were accumulated with a
197 spectral width of 7000 Hz. 2D experiments TOCSY [37] and NOESY [38] were recorded using the
198 States-Haberkorn method. Water suppression was achieved by DPFGSE sequence [39]. TOCSY and
199 NOESY were acquired with mixing times of 70 and 100 ms, respectively. Typically, 64 transients
200 of 1K data points were collected for each of the 256 increments; the data were zero filled to 2K in
201 ω1. Squared shifted sine-bell functions were applied in both dimensions prior to Fourier
202 transformation and baseline correction. Data were processed and analyzed using NMRPIPE [40] and
203 CARA software [41]. The hydrodynamic proprieties were estimated using the translational diffusion
204 coefficient (D_t) measured by Pulsed-field gradient spin-echo DOSY experiments [39]. The R_h was
205 estimated from the Stokes-Einstein equation: $(K_BT)/6\pi\eta D_t$, where K_B is the Boltzmann constant, T
206 is the temperature in Kelvin and η is the viscosity of the solution in Pa s The rotational correlation
207 time (τc) was estimated, considering a spherical globular protein, through the hydrodynamic radius
208 (R_h) from the Stokes-Einstein equation: $\tau c \sim (4\eta\pi R_h^3)/ 3 K_BT$. The hydrodynamic proprieties (D_t,
209 R_h) were also evaluated from the predicted 3D models using the software HYDROPRO [42-44].

210

211 **2.10. CD spectroscopy**

212

213 Crested porcupine and reindeer Mb samples were prepared in 4 mL of 20 mM phosphate buffer
214 containing 0.2 M NaCl at pH 6.8. The thermal denaturation of the two proteins was evaluated using
215 a JASCO-815 CD spectropolarimeter equipped with Peltier temperature control. CD spectra were
216 measured at 5 K intervals in the 278-368 K range (additional point at 371 K). After the final
217 measurement at 371 K, the samples were cooled to 298 K, and final spectra were acquired. The data
218 were collected using a quartz cuvette with a 1 cm path-length in the 200-260 nm wavelength range
219 with a data pitch of 1 nm. All data were recorded with a bandwidth of 1 nm with a scanning speed
220 of 50 nm/min and normalized against reference spectra to remove the background contribution of
221 buffer. The data obtained were fitted into two-state folding model. The fraction of unfolded protein

at each temperature was calculated from the observed ellipticity (θ_{obs}) and the ellipticity of the folded (θ_F) and the unfolded(θ_U) species using the following equation:

$$Keq= (\theta_F - \theta_{obs})/ (\theta_{obs} - \theta_U)$$

Next the standard Gibbs energy (ΔuG°) for unfolding of myoglobin at each temperature was calculated using: | $\Delta uG° = -RT\ ln\ Keq$ | where R is the ideal gas constant and T is the specific temperature. Then from the plot of lnKeq versus 1/T the van't Hoff equation was employed to obtain ΔuH°. The estimation of the secondary structure content was performed using the K2D3 server [45].

2.11. Bioinformatic tools and homology studies

All the used amino acid sequences of myoglobins were retrieved and analysed using the program BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and the NCBI taxonomy browser (http://www.ncbi.nlm.nih.gov/taxonomy/). Alignments were performed by Clustal*W* at EMBnet-CH (http://www.ch.embnet.org/software/ClustalW.html) and with MEGA[46] software. The similarity/identity matrix was obtained using the BOXSHADE program (http://mobyle.pasteur.fr/cgi-bin/MobylePortal/portal.py?form=boxshade). The standard one-letter code was used for the amino acid residues.

Homo sapiens L. (human, AC: P02144); *Sus scrofa* L. (pig, AC: P02189), *Physeter microcephalus* L. (Physeter catodon, AC: P02185), *Equus caballus* L. (horse, AC: P68082); *Caretta caretta* L. (Loggerhead sea turtle, AC: P56208), *Thunnus albacares* L.(yellowfin tuna, AC: P02205) myoglobin sequences have been used.

3. Results and discussion

3.1. Myoglobin isolation

Mbs were purified from *Hystrix cristata* L. and *Rangifer tarandus* L. as described in the Experimental section. Total proteins were extracted from meat homogenates in Tris•HCl buffer. Soluble proteins were fractionated by gel-filtration and anion exchange chromatography according to a previously published procedure [19, 20]. Homogeneity of both purified Mbs was confirmed by the presence of single peaks eluted from analytical FPLC and by SDS-PAGE analysis (**Fig. 1A and B**). Primary structural studies were carried out on the apo-Mbs, isolated by RP-HPLC as reported in the Experimental section.

3.2. Determination of the primary structure of crested porcupine Mb

The amino acid sequence of crested porcupine Mb was obtained by a general strategy based on the combined use of tandem mass spectrometry (ESI-MS/MS) and peptide mapping by MALDI-TOF MS[20]. In particular, the following experimental steps were carried out: i) determination of accurate relative molecular mass (M*r*) by ESI/Q-TOF MS of apo-Mbs; ii) enzymatic cleavage with trypsin followed by ESI-MS/MS analysis of the resulting tryptic peptides; iii) alignment of the sequenced peptides with the homologous reference protein and, iv) sequence completion by MALDI-TOF MS mapping of peptides from chymotrypsin, endoproteinase Glu-C hydrolysis or chemical fragmentation (CNBr).

The first step was to determine the M*r* of crested porcupine Mb by ESI/Q-TOF mass spectrometry (M*r* 16867.25±0.02; **Fig. 2A**). In the second step, apo-myoglobin was subjected to tryptic cleavage, and the resulting tryptic peptides were analyzed by tandem mass spectrometry. The tryptic mixture was analyzed by ESI/Q-TOF-coupled CapLC, recording automatically the MS/MS spectra on the three most intense mass peaks generated in each scan. The MS/MS data were first processed automatically by using the Biolynx application of MassLynx 4.0 software and then all MS/MS spectra leading to protein identification were manually double checked to verify sequence assignments. Amino acid sequences of crested porcupine Mb peptides obtained by tandem mass spectrometry are reported in **Table 1**. The *de novo* sequencing was supported by comparative homology analyses with the *Ctenodactylus gundi* Mb (gundi; AC: P20856), on the basis of the high sequence identity with crested porcupine Mb. Considering the gundi Mb sequence as reference protein, a coverage of about 61% was obtained from *de novo* sequencing analysis. Since the amino

acid residues at positions 32-62, 78-79, 97-102 and 134-153 were not determined, we decided to map the entire sequence analysing a new set of peptides obtained from chymotrypsin and endoproteinase Glu-C hydrolysis or chemical fragmentation with CNBr (**Table S1 and Fig. 3A**). The crested porcupine Mb sequence accounts for a calculated molecular mass of 16867.36 Da, which is in good agreement with the value obtained ESI/Q-TOF MS on the apo-myoglobin (16867.25±0.02 Da). Finally, the amino acid sequences of crested porcupine and gundi Mbs were compared each other (**Fig. 3A**). The primary structure of both Mbs have 83.7% identity (88.2% similarity). In particular, with respect to gundi Mb, we found eight amino acid substitutions (A13V, S51A, K56R, N74G, E83A, E116Q, G121A and A127T), whereas proximal (position 93, α-helix F) and distal histidinyl residues (position 64, α-helix E7) are conserved.

The crested porcupine Mb sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession number C0HJQ9.

3.3. Determination of the primary structure of reindeer Mb

The strategy employed for the determination of the primary structure of reindeer Mb was basically similar to that used for crested porcupine Mb. ESI/Q-TOF mass spectrometry analysis of reindeer apo-myoglobin showed that its M*r* was 16924.06±0.02. The transformed mass spectrum is reported in **Fig. 2B**. The structural characterization was initially performed by *de novo* peptide sequencing by tandem MS as reported in paragraph 3.2. Amino acid sequences of reindeer Mb peptides obtained by tandem mass spectrometry are reported in **Table 2**. The *de novo* sequencing was supported by comparative homology analyses with the *Cervus elaphus* L. Mb (red deer; AC: P02191), on the basis of the high sequence identity with reindeer Mb. All triptic peptides shown in **Table 2** have the same amino acid sequence of the corresponding peptides from red deer Mb with the exception of peptide T-5. In **Fig. 4** is reported the tandem mass spectrum of the doubly charged ion at m/z 759.80 (precursor ion: 1517.58 Da, expected molecular mass: 1517.66 Da; Δ = 0.08) from T-5 peptide (sequence position 119-133), containing the substitution N122D. The *de novo* sequencing data allowed us to obtain a coverage of about 57%. The complete overlapping of Mb peptides was achieved by MALDI-TOF MS analysis of peptides from chymotrypsin digestions or CNBr fragmentation (**Table S2**). The final sequence of reindeer Mb shows an experimental M*r* (16923.46±0.08) that is in very good agreement with theoretical one (M*r* 16923.38). As reported in **Fig. 3B**, one residue substitution (N122D) characterizes reindeer Mb with respect to red deer Mb. On the other hand, proximal (position 93, α-helix F) and distal histidinyl residues (position 64, α-helix E7) are conserved.

The reindeer Mb sequence data reported in this paper will appear in the UniProt Knowledgebase under the accession number C0HJR0.

3.4. Crested porcupine and reindeer Mbs autoxidation rate

Metmyoglobin formation from oxymyoglobins at pH 7.4 and 37 °C (physiological condition) is presented in **Fig. 5** for crested porcupine and reindeer. The metmyoglobin percentage increased over time in both Mb species with different autoxidation rate. A higher autoxidation rate was observed in reindeer oxyMb respect to crested porcupine one. In particular, the first order rate constant (k) was 0.0429 h^{-1} and 0.0336 h^{-1} for reindeer and crested porcupine, respectively.

These findings shown a higher autoxidation rate of reindeer Mb with respect to crested porcupine one, and that several amino acid residues (*i.e.*, Leu29, Lys45, Thr67, Val68) [47] involved in the autoxidation mechanism are conserved. Thus, the variations observed in the primary structure (twenty five amino acid substitutions: Ala/Asp, Glu/Ala, Lys/Ala and His/Pro at position 53, 83, 87 and 88, respectively; **Fig. S1)** of reindeer Mb compared to crested porcupine Mb may induce a different functional behaviour. In this framework, further investigations in the primary structure of both Mbs are needed to identify specific differences that could potentially influence their autoxidation time course.

3.5. The 3D structural model of reindeer and crested porcupine Mbs

The three-dimensional structure of a protein can be very informative and useful to understand functional characteristics of proteins. Therefore, in order to provide the molecular details of reindeer and crested porcupine Mbs, we computationally determined the 3D structure of both proteins using the I-TASSER algorithm which build 3D models on the base of multiple threading alignment Lometes and Illterative Tasser simulations [28]. In particular, the structure prediction by I-TASSER rely on template proteins with known structures obtained from database and the prediction procedures is based on matching the query sequence against a non-redundant sequence database. The computational modeling of reindeer Mb and crested porcupine Mb structures were performed and five models for both proteins were generated using the I-TASSER algorithm with C-scores ranging from -5 to 1.28 and from -5 to 1.29, respectively. The C-score is a confidence score for estimating the quality of predicted models and ranges from -5 to 2, with higher scores representing higher confidence in the model. The Model 1 for reindeer (C-score 1.28) and crested porcupine Mbs (C-score 1.29) was used as reference structure for all analysis described below (**Fig. 6A, B**). The

good quality of the predicted models for both Mbs were determined by evaluating the Ramachandran plot (Fig. SI2) using the software PROCHECK[28]. A comparison of reindeer ($RMSD_{bb}^{1-153}$= 0.708 Å) and crested porcupine ($RMSD_{bb}^{1-153}$ = 0.729 Å) Mbs predicted models with the structure of myoglobin (PDB code: 1MBN)[48] resolved by X-Ray crystallography indicates that both proteins show the classical globular fold of Mbs. Notably, the three myoglobins share a similar hydrophobic cleft, in terms of structural features, in which is inserted the heme prosthetic group (**Fig. 6C**).

Overall, reindeer and crested porcupine Mbs exhibit the typical topology of myoglobins with most of the hydrophobic amino acid residues buried in the interior and many of the polar residues on the surface. The tertiary structure is composed of eight α-helices joined by short non-helical regions (**Fig. 6A, B**) that provide a rigid structural framework for the heme pocket. As expected, the two myoglobins adopt a similar compact structure ($RMSD_{bb}^{1-153}$ = 0.306 Å) with small but significant structural differences. Therefore, in order to quantify the conformational dissimilarity between both Mbs, we evaluated the orientation of the secondary structure elements by measuring the inter-helical angles and inter-helical distances of reindeer and crested porcupine Mbs, respectively. Based on the data reported in the table (**Fig 6D**), due to a different orientation of α-helices that produce a slightly displacement, helices F, G and H display the most important structural variations. On the contrary, the hydrophobic pocket in which is located the heme prosthetic group, considering also the side-chain orientation of the distal and proximal histidinyl residues (His64, His93), does not show any significant structural difference. All together our data demonstrate that reindeer Mb, while retaining the typical structural organization of Mbs, adopts a fold that appears to be slightly less compact of crested porcupine Mb. As a consequence of these structural observations combined with the faster autoxidation rate of reindeer OxyMb with respect to porcupine, we can hypothesize that in the porcupine Mb, the Fe-O_2 group is more protected in the cavity than in the structure of reindeer Mb. This scenario is further supported by computational data, reported in a previous publication [49, 50], demonstrating that the residues of helix F, G and H (I99, I107, S108, F138 and Y146) have an significant impact on internal gas migration rates since they have a strong influence to the cavity network topology.

3.6. NMR spectroscopy

To further investigate the structural characteristics of reindeer and crested porcupine Mbs, we performed a solution structural characterization by Nuclear Magnetic Resonance spectroscopy. The ^{1}H monodimensional spectra (**Fig. 7A**) acquired at 298K indicate considering the good chemical

shifts dispersion in the amide, aromatic, or methyl regions, that both Mbs adopt a stable tertiary structure.

This finding was further confirmed by two dimensional NMR experiments as NOESY[38] (**Fig. 7B**) that represent a powerful method to obtain structural information regarding protein. In particular, in spite of the low resolution of spectra, 2D NOESY experiments of both Mbs in the amide region, show a considerable number of inter-residue HN-HN connectivities indicating that MBs adopt in solution a folded conformation with the presence of α-helix secondary structure. In order to characterize the size and shape of reindeer and crested porcupine Mbs we investigated the hydrodynamic proprieties by Nuclear Magnetic Resonance experiments. The translational diffusion coefficient (D_t) of both Mbs were measured by diffusion-order spectroscopy (DOSY) NMR experiments at different concentrations. The measured diffusion coefficients of reindeer and crested porcupine Mbs are concentration independent (**Fig. 7C**) , with a mean ± SD value of $1.16 \pm 0.04\ 10^{-10}\ m^2s^{-1}$ and $1.17 \pm 0.05\ 10^{-10}\ m^2s^{-1}$, respectively. These results clearly indicate that both myoglobins exist predominantly in a single state under the analysis conditions. By use of the Stokes-Einstein equation, the measured diffusion coefficients for reindeer and crested porcupine Mbs correspond to a hydrodynamic radius (R_h) of 2.06 nm and 2.04 nm, respectively (**Fig. 7C**). Moreover, we calculated, as reported in the Material and Methods section, the correlation time (τ_c) that represents the time for a protein to rotate one radian. The obtained correlation time by the NMR data, using Stokes-Einstein equation as reported in the experimental part, for reindeer ($\tau_c = 8.11 \pm 0.06$) and crested porcupine ($\tau_c = 7.89 \pm 0.06$) Mbs (**Fig. 7C**) are in a excellent agreement with that reported for different Mb from other species in a previous publication [51, 52]. Then, to estimate the molecular weight (MW) using the NMR experimental data we compared the obtained correlation time for both Mbs to a standard curve of τc versus protein molecular weight measured at the same temperature on a series of known monomeric proteins of varying size (Fig. SI3)[53]. The molecular weight of both Mbs ($MW_{\text{reindeer Mb}}$ ~16.9 kDa, $MW_{\text{crested porcupine Mb}}$ ~16.5 kDa) obtained from NMR data is in good agreement with that measured by gel-filtration (data not-shown) or by mass spectrometry analysis. In conclusion, hydrodynamic data clearly indicate that reindeer and crested porcupine Mbs are monomeric under the analysed conditions.

3.7. Validation of porcupine and reindeer structural models

One strategy for assessing the accuracy of calculated ensemble conformers is the cross-validation. We performed a cross-validation analysis for crested porcupine and reindeer Mbs predicted models using the hydrodynamic proprieties that were not considered in the computational

modeling. In particular, we back-calculated for both myoglobins the hydrodynamic radius using the HYDROPRO software[42-44]. A comparison for both Mbs of the calculated hydrodynamic radius (**Fig. 7C**; R_h = 2.11 nm for reindeer or R_h =2.10 nm for crested porcupine Mb) with the experimental value (R_h = 2.06 ± 0.09 nm for reindeer, R_h = 2.04 ± 0.09 nm for crested porcupine Mb) indicates that the predicted models can properly describe the experimental data observed in solution. Moreover, to further validate the predicted models, we estimated from CD data (Fig. 8) the protein secondary structure for both Mbs. The data indicate that the α-helix secondary structure content for reindeer (72.7 %) and crested porcupine (72.6%) is in a good agreement with the α-helix amount (Mb reindeer = 72.5 %, Mb crested porcupine= 71.2%) obtained from the predicted Mb models using the software DSSP [54]. Overall, our analysis demonstrated that the predicted model for reindeer and crested porcupine Mbs represent a realistic picture of the tertiary structure that the protein adopt in solution.

3.8. Thermal stability monitored by CD spectroscopy

Thermal unfolding of reindeer (**Fig. 8A, B**) and crested porcupine (**Fig. 8C, D**) Mbs has been investigated by CD spectroscopy. In both cases, the reversible thermal unfolding encompassing the temperature range between 283 and 371 K can be fitted to a classical two-state model with melting temperature of T_m= 353.7 K for reindeer Mb and T_m= 356.3 K for crested porcupine Mb (**Table 3**). Assuming a two-state mechanism for the thermal denaturation of both Mbs, the Gibbs free energies of protein unfolding (ΔuG) were calculated as reported in Materials and Methods. As a criterion of the thermal stability of the two Mbs we estimated the Gibbs free energy at 298 K. Our data, illustrated in the **Table 3**, indicate that the two Mbs under investigation, while having a similar globular structure, show a different thermal stability. Overall, the thermal unfolding data indicate that reindeer Mb adopts a less stable conformation than crested porcupine Mb. This finding may be due to the small but significant structural differences highlighted by the predicted and validated 3D models showing that reindeer Mb presents a slightly less compact fold with respect to crested porcupine Mb.

3.9. Comparison of Myoglobins across species

The amino acid sequences of crested porcupine Mb and reindeer Mb were compared each other and with Mbs isolated from *H. sapiens*, *S. scrofa*, *P. microcephalus*, *E. caballus*, *C. caretta* and *T. albacares* (**Fig. 9A**). The analysis shows that proximal (position 93) and distal (position 64)

histidinyl residues are present in both sequences, whereas several amino acid residues characterising the binding site for heme (i.e., Thr39, Lys42, Phe43, Ser92, His97, Ile99, Tyr103, Leu104) [55] are conserved. Furthermore, the primary structure of crested porcupine Mb and reindeer Mb have 83.7% identity (88.2% similarity). When compared with other myoglobin (**Fig. 9B**), they showed a range of identity/similarity of among 43.8 and 92.2% identity (54.9 and 96.1% similarity). In order to further understand the relationship between structure-function for reindeer and crested porcupine Mbs we analyzed the tertiary structure and investigated the network of O_2 pathways for a set of myoglobins from different species. The main structural characteristic of myoglobin is the presence of four internal cavities in its native state, as depicted in **Fig. 9C** [56]. The X-ray structure (PDB code 1J52) of the Sperm-whale myoglobin obtained in the presence of 7 atm of Xenon demonstrated the occupation by Xe of the internal cavities (named Xe1, Xe2, Xe3, Xe4). These cavities with a radius larger than 1.2 Å, are lined by hydrophobic residues and are recognized to play an important role for the uptake of ligands [57]. Moreover, an additional cavity is located in proximity of the distal histidine (DP). Several studies demonstrated the ability of myoglobin to reversibly combine with small ligands such as O_2, CO and NO [49, 58-61]. For the case of O_2 pathways Cohen and co-workers [49], using a computational approach for studying gas migration, indentified the residues that affect gas ligand transport (**Fig. S4**). As illustrated in the **Fig. 9B**, the sequence identity of the myoglobins under investigation is between 50% and 95%. Additionally, the superposition of 3D structures indicates that the globular fold is well conserved among the analyzed myoglobins and their secondary and tertiary structure is near identical with only small structural differences. Moreover, in according with the results reported by Cohen and Schulten [60] we analyzed for reindeer and crested porcupine Mbs the content of less solvent exposed residues having a high propensity to create O_2 favorable regions. Interestingly, our investigation indicates that reindeer Mb with respect to the crested porcupine Mb present a larger number of residues promoting the formation of cavities. Therefore, since the tertiary structure is mainly conserved across the species we explored for the studied myoglobins the dimension of the internal cavities close to the residues having an important role in the O_2 migration pathways. Interestingly, our analysis indicates that the various myoglobins exhibit cavity locations and dimensions which are completely different from one protein to another.

At first, in order to better explore the structural differences between myoglobins we analyzed the 3D structures evaluating located the dimension of the internal cavities locatedclose to the networks of O_2 pathways. As reported in the table (**Fig. S5**) , despite to a similar tertiary structure all myoglobins show high variability considering the volume of the cavities. Then, we compared the cavities detected for reindeer and crested porcupine Mbs. Notably, our analysis shows (**Fig. 10**) that

reindeer Mb presents a larger number of cavities than crested porcupine. This finding appear to be correlated to the different amino acid composition, as mentioned above, of the two myoglobins and may partially explain the different functional behaviour in terms of autoxidation rate. Of course, to fully understand the functional proprieties of both myoglobins a detailed dynamical description is required. In fact, due to the thermal fluctuations of the residues there is a possibility to have additional random cavities where oxygen molecules can fit in.

Conclusions and perspectives

Myoglobins isolated from crested porcupine (*Hystrix cristata* L.) and reindeer (*Rangifer tarandus* L.) with MW of 16867 Da and 16923 Da, respectively, show different functional proprieties. In particular, reindeer Mb has higher autoxidation rate with respect to crested porcupine Mb. This finding, may be related to the differences observed in the primary structure of the two proteins. Moreover, the 3D models predicted and successively validated using experimental NMR data indicate that reindeer Mb presents a slightly less compact fold with respect to crested porcupine Mb. Additionally, thermal unfolding measurements demonstrated that reindeer Mb adopts a less stable conformation than crested porcupine Mb. Overall, our study suggests that, considering the small but significant structural differences combined with the conformational motions, reindeer Mb with respect to crested porcupine Mb may differently modulate the heme environment, facilitating oxygenation. In fact, our results may be useful to deeply understand the very complex gas diffusion process for both Mbs. Finally, our study may represent a suitable model to describe how proteins modulate the response activity to different external environmental conditions.

Acknowledgements

This research was supported by funds from the Second University of Naples.

References

1. D. L. Gilbert, in *Oxygen in the Animal Organism*, eds. F. Dickens and E. Neil, Pergamon Press, New York, 1964, vol. 31, pp. 641-654.
2. S. N. Vinogradov, D. Hoogewijs, X. Bailly, R. Arredondo-Peter, J. Gough, S. Dewilde, L. Moens and J. R. Vanfleteren, *BMC Evol Biol*, 2006, 6, 31.
3. A. Maton, H. J., M. C. W., J. S., W. M. Q., D. LaHart and J. D. Wright, *Human Biology and Health*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1993.
4. D. L. Nelson and M. M. Cox, *Lehninger Principles of Biochemistry (6rd Ed.)*, Freeman, W. H. & Company, New York, USA, 2013.
5. G. A. Ordway and D. J. Garry, *J Exp Biol*, 2004, 207, 3441-3446.
6. J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips, *Nature*, 1958, 181, 662-666.
7. M. Weber, M. Rau, K. Madlener, A. Elsaesser, D. Bankovic, V. Mitrovic and C. Hamm, *Clin Biochem*, 2005, 38, 1027-1030.
8. D. W. Kehl, N. Iqbal, A. Fard, B. A. Kipper, A. De La Parra Landa and A. S. Maisel, *Transl Res*, 2012, 159, 252-264.
9. M. Brunori, *Protein Sci*, 2010, 19, 195-201.
10. U. Flogel, M. W. Merx, A. Godecke, U. K. Decking and J. Schrader, *Proc Natl Acad Sci U S A*, 2001, 98, 735-740.
11. S. P. Suman and P. Joseph, *Annu Rev Food Sci Technol*, 2013, 4, 79-99.
12. M. N. Antoniewski and S. A. Barringer, *Crit Rev Food Sci Nutr*, 2010, 50, 644-653.
13. N. Giarretta, A. M. Di Giuseppe, M. Lippert, A. Parente and A. Di Maro, *Food Chem*, 2013, 141, 1814-1820.
14. A. M. Di Giuseppe, N. Giarretta, M. Lippert, V. Severino and A. Di Maro, *Food Chem*, 2015, 169, 241-245.
15. J. T. Lecomte, D. A. Vuletich and A. M. Lesk, *Curr Opin Struct Biol*, 2005, 15, 290-301.
16. L. Ma, X. Shao, Y. Wang, Y. Yang, Z. Bai, Y. Zhao, G. Jin, Q. Ga, Q. Yang and R. L. Ge, *Gene*, 2014, 533, 532-537.
17. A. Di Maro, A. Chambery, A. Daniele, P. Casoria and A. Parente, *Phytochemistry*, 2007, 68, 767-776.
18. A. Di Maro, I. Terracciano, L. Sticco, L. Fiandra, M. Ruocco, G. Corrado, A. Parente and R. Rao, *J Biotechnol*, 2010, 147, 1-6.
19. R. Dosi, A. Di Maro, A. Chambery, G. Colonna, S. Costantini, G. Geraci and A. Parente, *Comp Biochem Physiol B Biochem Mol Biol*, 2006, 145, 230-238.
20. R. Dosi, A. Carusone, A. Chambery, V. Severino, A. Parente and A. Di Maro, *Food Chemistry*, 2012, 133, 1646-1652.
21. U. K. Laemmli, *Nature*, 1970, 227, 680-685.
22. E. Gross, *Methods in Enzymology*, 1967, 11, 238-255.
23. A. Chambery, A. Di Maro, M. M. Monti, F. Stirpe and A. Parente, *Eur J Biochem*, 2004, 271, 108-117.
24. A. Di Maro, A. Chambery, V. Carafa, S. Costantini, G. Colonna and A. Parente, *Biochimie*, 2009, 91, 352-363.
25. A. Di Maro, P. Ferranti, M. Mastronicola, L. Polito, A. Bolognesi, F. Stirpe, A. Malorni and A. Parente, *J Mass Spectrom*, 2001, 36, 38-46.
26. V. Severino, A. Chambery, M. Vitiello, M. Cantisani, S. Galdiero, M. Galdiero, L. Malorni, A. Di Maro and A. Parente, *J Proteome Res*, 2010, 9, 1050-1062.
27. E. E. Di Iorio, *Methods Enzymol*, 1981, 76, 57-72.
28. A. Roy, A. Kucukural and Y. Zhang, *Nat Protoc*, 2010, 5, 725-738.

29. R. A. Laskowski, J. A. Rullmannn, M. W. MacArthur, R. Kaptein and J. M. Thornton, *J Biomol NMR*, 1996, 8, 477-486.
30. I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson and D. C. Richardson, *Nucleic Acids Res*, 2007, 35, W375-383.
31. W. L. DeLano, 2002.
32. R. Koradi, M. Billeter and K. Wüthrich, *J Mol Graph*, 1996, 14, 51-55, 29-32.
33. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, 25, 1605-1612.
34. R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend, *Nucleic acids research*, 2011, 39, D411-419.
35. J. Dundas, Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz and J. Liang, *Nucleic acids research*, 2006, 34, W116-118.
36. S. H. Oliveira, F. A. Ferraz, R. V. Honorato, J. Xavier-Neto, T. J. Sobreira and P. S. de Oliveira, *BMC Bioinformatics*, 2014, 15, 197.
37. R. Brüschweiler and R. R. Ernst, *J Magn Reson*, 1983, 53, 521-528.
38. A. Kumar, R. R. Ernst and K. Wüthrich, *Biochem Biophys Res Commun*, 1980, 95, 1-6.
39. Y. Cohen, L. Avram and L. Frish, *Angew Chem Int Ed Engl*, 2005, 44, 520-554.
40. F. Delaglio, S. Grzesiek, G. W. Vuister, G. Zhu, J. Pfeifer and A. Bax, *J Biomol NMR*, 1995, 6, 277-293.
41. R. L. J. Keller, 2004.
42. J. García De la Torre, M. L. Huertas and B. Carrasco, *J Magn Reson*, 2000, 147, 138-146.
43. J. García De la Torre, M. L. Huertas and B. Carrasco, *Biophys J*, 2000, 78, 719-730.
44. A. Ortega, D. Amorós and J. García de la Torre, *Biophys J*, 2011, 101, 892-898.
45. S. M. Kelly, T. J. Jess and N. C. Price, *Biochim Biophys Acta*, 2005, 1751, 119-139.
46. E. Pizzo, P. Buonanno, A. Di Maro, S. Ponticelli, S. De Falco, N. Quarto, M. V. Cubellis and G. D'Alessio, *J Biol Chem*, 2006, 281, 27454-27460.
47. R. E. Brantley, Jr., S. J. Smerdon, A. J. Wilkinson, E. W. Singleton and J. S. Olson, *J Biol Chem*, 1993, 268, 6995-7010.
48. H. C. Watson, *Prog Stereochem*, 1969, 4, 299.
49. J. Cohen, A. Arkhipov, R. Braun and K. Schulten, *Biophys J*, 2006, 91, 1844-1857.
50. X. Huang and S. G. Boxer, *Nat Struct Biol*, 1994, 1, 226-229.
51. D. Wang, U. Kreutzer, Y. Chung and T. Jue, *Biophys J*, 1997, 73, 2764-2770.
52. G. Gros, B. A. Wittenberg and T. Jue, *J Exp Biol*, 2010, 213, 2713-2725.
53. P. Rossi, G. V. Swapna, Y. J. Huang, J. M. Aramini, C. Anklin, K. Conover, K. Hamilton, R. Xiao, T. B. Acton, A. Ertekin, J. K. Everett and G. T. Montelione, *J Biomol NMR*, 2010, 46, 11-22.
54. R. P. Joosten, T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend, *Nucleic Acids Res*, 2011, 39, D411-419.
55. S. V. Evans and G. D. Brayer, *J Mol Biol*, 1990, 213, 885-897.
56. R. F. Tilton, Jr., I. D. Kuntz, Jr. and G. A. Petsko, *Biochemistry*, 1984, 23, 2849-2857.
57. M. Brunori, D. Bourgeois and B. Vallone, *J Struct Biol*, 2004, 147, 223-234.
58. C. Bossa, A. Amadei, I. Daidone, M. Anselmi, B. Vallone, M. Brunori and A. Di Nola, *Biophys J*, 2005, 89, 465-474.
59. E. E. Scott, Q. H. Gibson and J. S. Olson, *J Biol Chem*, 2001, 276, 5177-5188.
60. J. Cohen and K. Schulten, *Biophys J*, 2007, 93, 3591-3600.
61. M. Ceccarelli, R. Anedda, M. Casu and P. Ruggerone, *Proteins*, 2008, 71, 1231-1236.

Figure legends

Fig. 1. FPLC elution profiles of crested porcupine (A) and reindeer Mbs (B) on an AKTA Purifier System from anion exchange chromatography using a Source 15Q PE 4.6/100 column. Experimental conditions are described in the text. In the insets are reported SDS-PAGE analyses of the same purified Mbs (lanes 1 and 2, 1.5 and 3 µg, respectively; M, protein markers).

Fig. 2. Deconvoluted ESI/Q-TOF mass spectra of HPLC-purified apo- Mbs from crested porcupine (A) and reindeer (B).

Fig. 3. Amino acid sequences of crested porcupine Mb compared with *C. gundi* one (A) and reindeer Md compared with *C. elaphus* one (B). The overlapping peptides used for assembling protein sequences are reported. Residues differing among Mbs are in bold. Proximal (position 93, α-helix F) and distal histidinyl residues (position 64, α-helix E7) are reported in red. Abbreviations: CB, cyanogen bromide; C, chymotrypic peptide; E, endoproteinase Glu-C; T, tryptic peptides.

Fig. 4. Fragmentation spectrum of the doubly charged precursor ion at m/z 759.80 (precursor ion: 1517.58 Da) mapped on sequence position 119-133. (A) MS/MS spectrum annotated with the y and b ion series. (B) Fragmentation table showing the ion series matching the spectrum. The matched and unmatched a, b, y and z ions are shown, along with the mass differences between the theoretical and experimental values. The matching probability is also reported below the three-letter amino acid code for the sequenced peptide.

Fig. 5: Autoxidation rates of reindeer and crested porcupine Mbs.

Fig. 6. The 3D models of reindeer (blue) and crested porcupine (red) Mb superimposed to the X-ray structure (PDB code: 1MBN) of the sperm whale (*Physeter macrocephalus* L.) Mb (light gray) in two orientation (A, B) rotated of 180° around z-axis. The heme prosthetic group is shown in magenta. (C) Close-up view of the heme hydrophobic pocket. The distal and proximal histidinyl residues are shown in stick style. (D) Helix-Helix angles and distances for the three myoglobins under investigation.

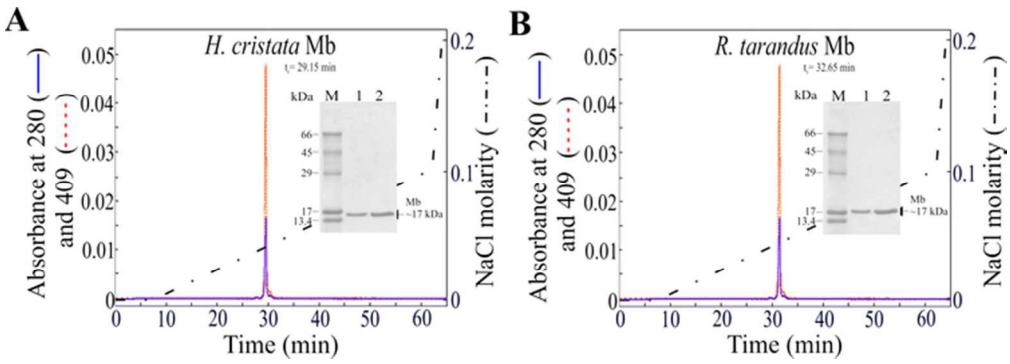
Fig. 7. NMR structural investigation. (A) 1D ^{1}H NMR spectrum acquired at pH 6.8, 298 K on 500 MHz spectrometer of reindeer (upper) and crested porcupine (lower) Mbs. (B) 2D ^{1}H-^{1}H NOESY

spectrum of reindeer (left) and crested porcupine (right) Mbs. (C) Hydrodynamic parameters of crested porcupine and reindeer Mbs. The values are obtained by DOSY NMR measurements and by HYDRO software using the predicted 3D models.

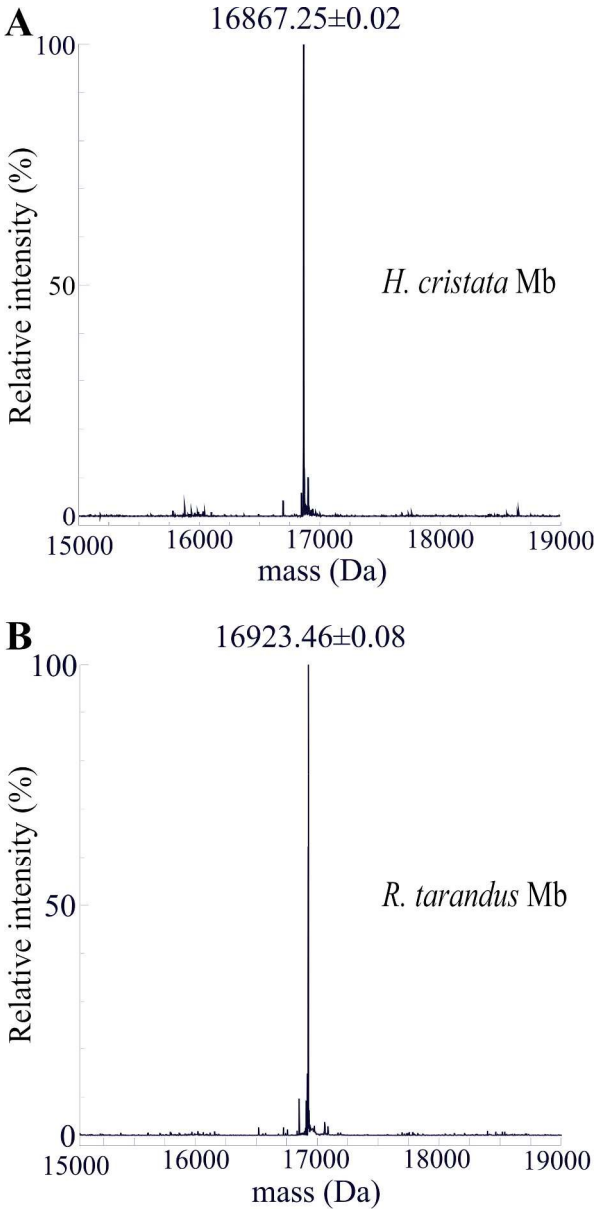
Fig. 8. Thermal unfolding of reindeer and crested porcupine Mbs followed by circular dichroism. (A, C) Thermal unfolding of crested reindeer and porcupine Mbs carried out in the range of 278-371 K. (B, D) Melting curve of reindeer and crested porcupine Mbs monitored by CD 222 nm. The data were fitted using a two-state model.

Fig. 9. (A), multiple alignment of myoglobin sequences from *H. cristata*, *R. tarandus*, *P. microcephalus*, *S. scrofa*, *E. caballus*, *H. sapiens*, *C. caretta* and *T. albacares*. Asterisk *, identical, double dots :, conserved and single dot ., semiconserved amino acid residues. Proximal (position 93) and distal histidinyl residues (position 64) are reported in red. (B), identity-similarity matrix of myoglobin sequences reported above. (C), structure of the studied myoglobins aligned and superimposed, demonstrating the very strong conservation of their secondary structure. The structure are *H. sapiens* (light gray) PDB ID: 3RGK, *Physeter macrocephalus* (orange) PDB ID: 1J52, *S. scrofa* (light pink) PDB ID: 1PMB, *E. caballus* (light green) PDB ID: 1WLA, *T. albacares* (yellow) PDB ID: 1MYT, *C. caretta* (violet) PDB ID: 1LHS, *H. cristata* (red), *R. tarandus* (blue). The Xe binding sites of *P. microcephalus* Mb are shown as dark gray spheres. The heme group is depicted in magenta.

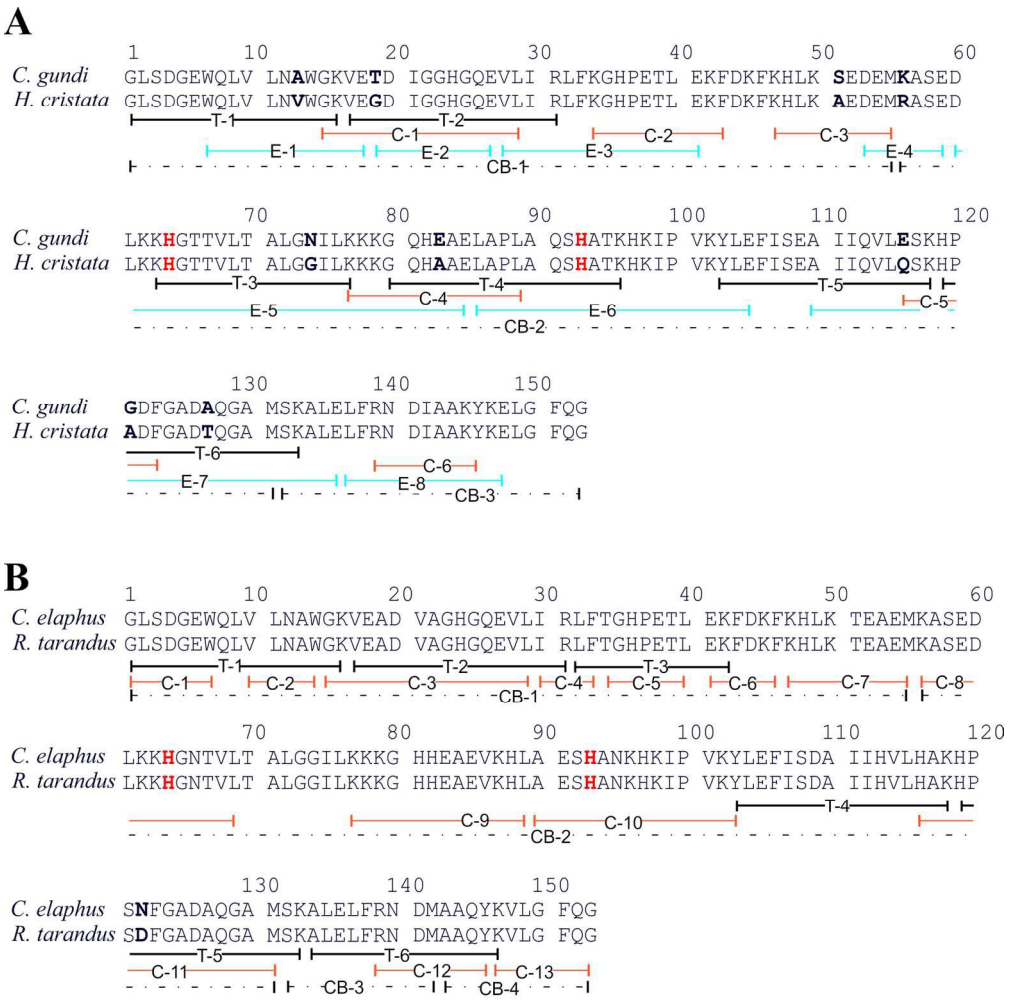
Fig. 10. The 3D models of reindeer (A) and crested porcupine (B) Mb in two orientation rotated of 180° around z-axis. The detected cavities are reported in dark gray. The volume of each cavity is also reported in parenthesis.



63x22mm (300 x 300 DPI)

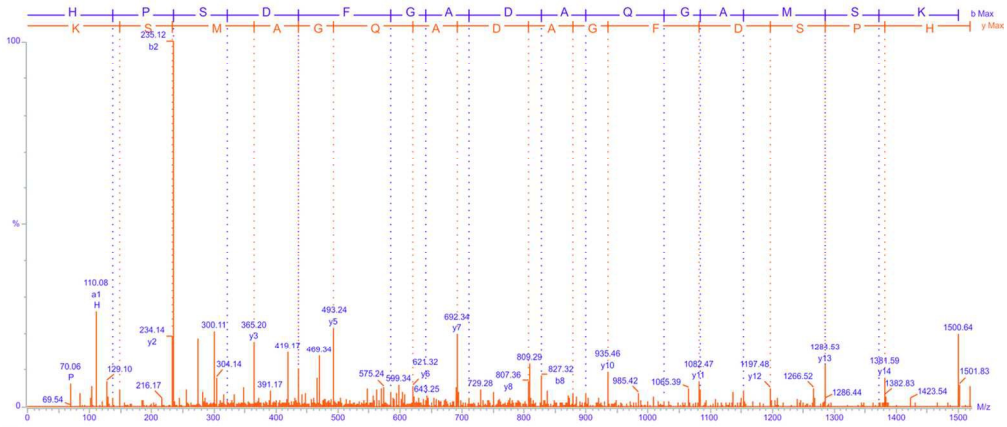


186x381mm (300 x 300 DPI)



157x155mm (300 x 300 DPI)

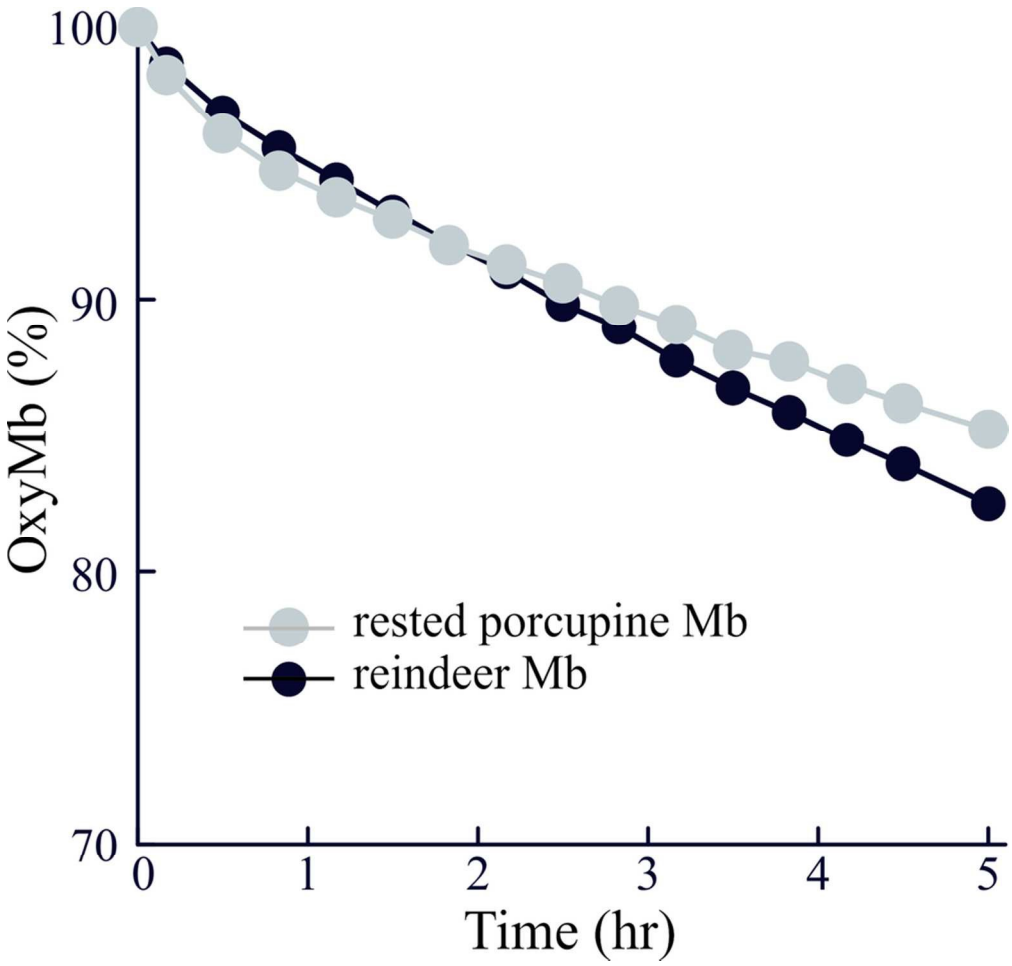
A



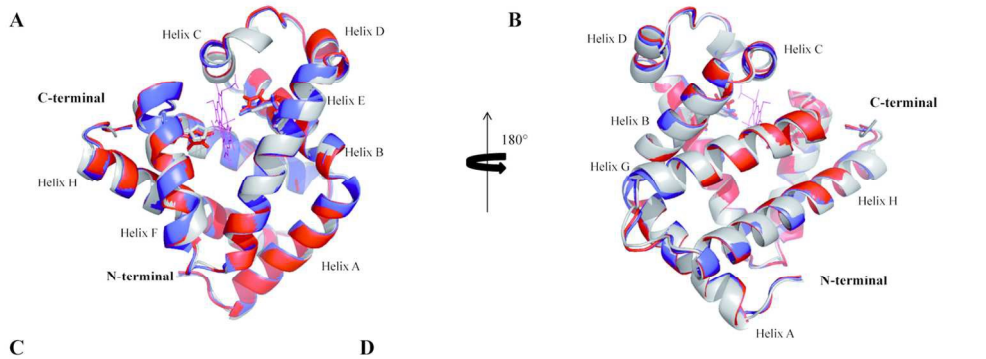
B

a	110.07	207.12	294.16	409.18	556.25	613.27	684.31	799.34	870.37	998.43	1055.45	1126.49	1257.53	1344.56	1472.66
	-0.00	---	---	---	0.05	---	0.05	---	0.05	---	0.02	0.03	-0.02	---	---
b	138.07	235.12	322.15	437.18	584.25	641.27	712.31	827.33	898.37	1026.43	1083.45	1154.49	1285.53	1372.56	1500.65
	-0.00	0.00	-0.06	0.03	0.01	-0.03	-0.08	0.01	-0.04	-0.01	-0.09	0.04	-0.13	0.05	---
	His	Pro	Ser	Asp	Phe	Gly	Ala	Asp	Ala	Gln	Gly	Ala	Met	Ser	Lys
	100	100	100	100	100	100	100	100	100	94	100	100	100	100	100
y	1518.66	1381.61	1284.55	1197.52	1082.49	935.43	878.40	807.37	692.34	621.30	493.24	436.22	365.19	234.15	147.11
	---	0.02	0.02	0.04	0.03	-0.03	0.07	0.00	0.00	-0.02	-0.00	0.01	-0.01	0.01	-0.01
z	1501.63	1364.58	1267.52	1180.49	1065.46	918.40	861.37	790.34	675.31	604.27	476.21	419.19	348.16	217.12	130.08
	---	---	-0.12	---	0.07	-0.03	0.00	0.05	0.03	0.00	-0.00	0.02	0.05	0.01	-0.01

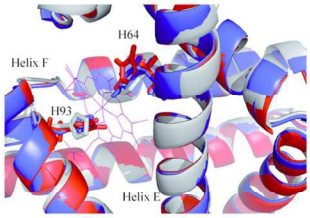
115x76mm (300 x 300 DPI)



84x80mm (300 x 300 DPI)



C

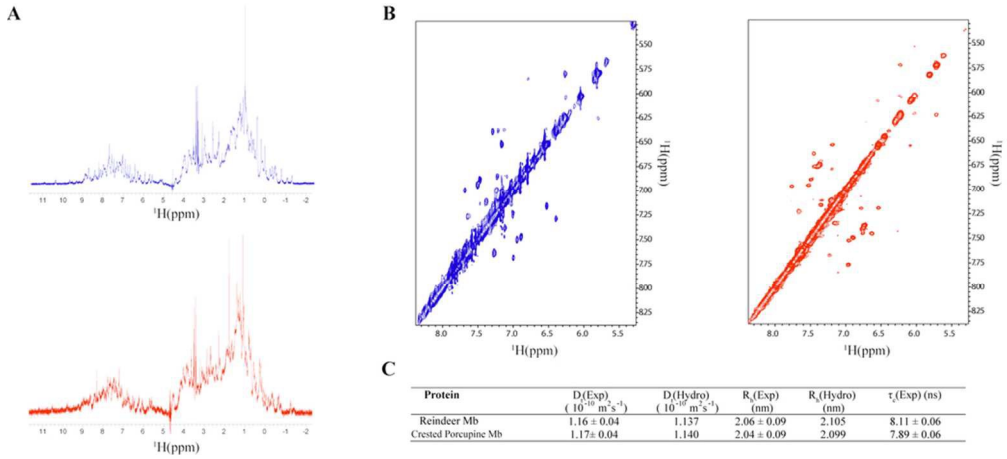


D

Protein	Interhelical angles (°)[1]							
Helix	A/B	B/C	C/D	D/E	E/F	F/G	G/H	A/H
1MBN[2]	100	-122	39	-67	-138	120	161	-94
Reindeer Mb	99	-120	41	-64	-140	120	159	-93
Crested Porcupine Mb	98	-113	54	-64	-140	122	160	-93
	Interhelical distances (Å)[1]							
Helix	A/B	B/C	C/D	D/E	E/F	F/G	G/H	A/H
1MBN[2]	20.5	12.9	15.4	17.6	15.3	19.2	11.7	15.9
Reindeer Mb	20.2	12.4	15.4	18.1	14.8	19.2	11.3	14.9
Crested Porcupine Mb	19.9	13.2	15.1	18.2	14.8	16.8	9.9	14.8

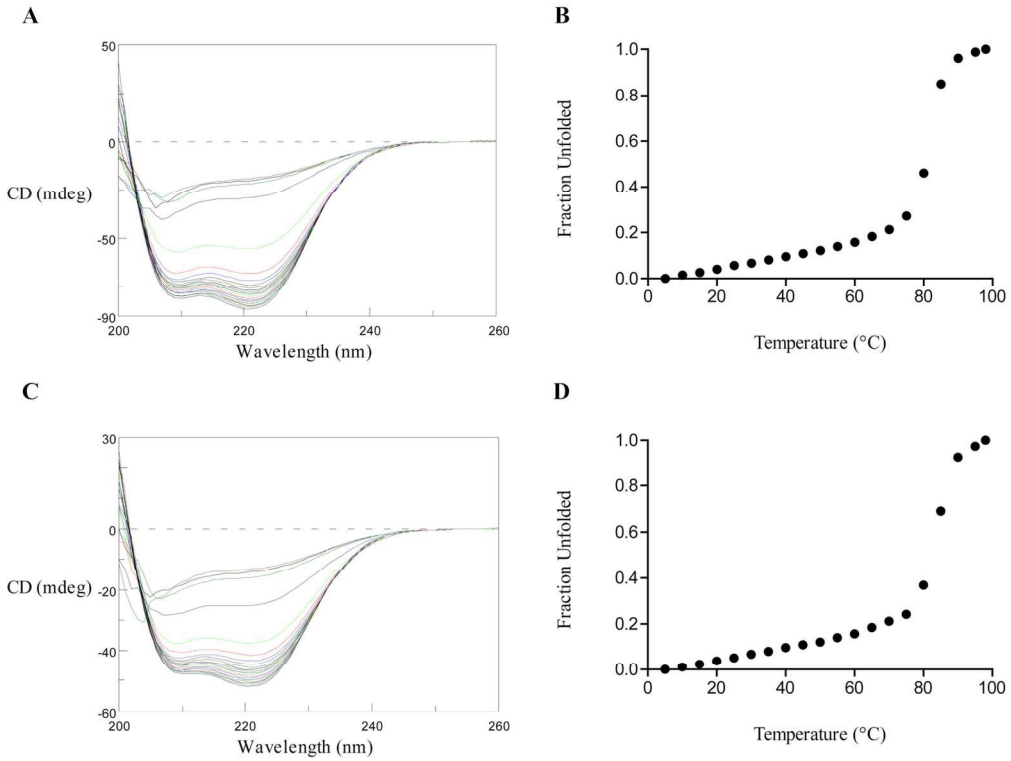
1Calculated using the software interhlx written by Kyoko Yap and Mitsuhiko Ikura (University of Toronto). Helices were defined using DSSP software.
2 PDB accession code of the oxymyoglobin structure at 1.6 angstroms resolution.

130x88mm (300 x 300 DPI)



Protein	D_t(Exp) (10^{-10} m^2s^{-1})	D_t(Hydro) (10^{-10} m^2s^{-1})	R_h(Exp) (nm)	R_h(Hydro) (nm)	τ_c(Exp) (ns)
Reindeer Mb	1.16 ± 0.04	1.137	2.06 ± 0.09	2.105	8.11 ± 0.06
Crested Porcupine Mb	1.17± 0.04	1.140	2.04 ± 0.09	2.099	7.89 ± 0.06

88x41mm (300 x 300 DPI)



143x108mm (300 x 300 DPI)

A

```
                1               20                40                60                80
                .               .                 .                 .                 .
H. cristata     GLSDGEWQLVLNVWGKVEGDIGGHGQEVLIRLFKGHPETLEKFDKFKHLKAEDEMRASEDLKKHGTTVLTALGGILKKKG
R. tarandus     GLSDGEWQLVLNAWGKVEADVAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGNTVLTALGGILKKKG
S. scrofa       GLSDGEWQLVLNVWGKVEADVAGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGNTVLTALGGILKKKG
E. caballus     GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEAEMKASEDLKKHGTVVLTALGGILKKKG
H. sapiens      GLSDGEWQLVLNVWGKVEADIPGHGQEVLIRLFKGHPETLEKFDKFKHLKSEDEMKASEDLKKHGATVLTALGGILKKKG
P. macrocephalus VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTEAEMKASEDLKKHGVTVLTALGAILKKKG
C. caretta      GLSDDEWNHVLGIWAKVEPDLSAHGQEVIIRLFQLHPETQERFAKFKNLTTIDALKSSEEVKKHGTTVLTALGRILKQKN
T. albacares    ----ADFDAVLKCWGPVEADYTTMGGLVLTRLFKEHPETQKLFPKFAGIAQADIAG-NAAISAHGATVLKKLGELLKAKG
                     ::: **  *. ** *    *  :: ***  **** : * :*  :         .  :. ** .**. ** :** *.

                81              100               120               140         153
                .               .                 .                 .           .
H. cristata     QHAAELAPLAQSHATKHKIPVKYLEFISEAIIQVLQSKHPADFGADTQGAMSKALELFRNDIAAKYKELGFQG
R. tarandus     HHEAEVKHLAESHANKHKIPVKYLEFISDAIIHVLHAKHPSDFGADAQGAMSKALELFRNDMAAQYKVLGFQG
S. scrofa       HHEAELTPLAQSHATKHKIPVKYLEFISEAIIQVLQSKHPGDFGADAQGAMSKALELFRNDMAAKYKELGFQG
E. caballus     HHEAELKPLAQSHATKHKIPIKYLEFISDAIIHVLHSKHPGDFGADAQGAMTKALELFRNDIAAKYKELGFQG
H. sapiens      HHEAEIKPLAQSHATKHKIPVKYLEFISECIIQVLQSKHPGDFGADAQGAMNKALELFRKDMASNYKELGFQG
P. macrocephalus HHEAELKPLAQSHATKHKIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
C. caretta      NHEQELKPLAESHATKHKIPVKYLEFICEIIVKVIAEKHPSDFGADSQAAMKKALELFRNDMASKYKEFGFQG
T. albacares    SHAAILKPLANSHATKHKIPINNFKLISEVLVKVMHEKAGLDAGG--QTALRNVMGIIIADLEANYKELGFSG
                 *   :  **:***.*****:: :::*.: :::*:  :   * *.  * *: :.: ::  *: ::** :*:.*
```

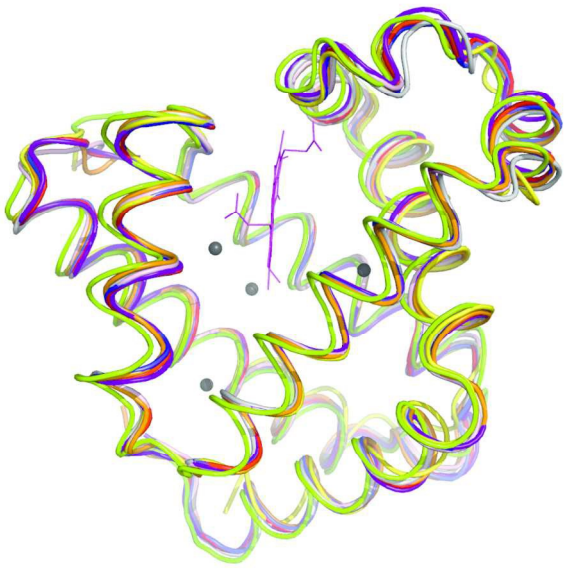
B

Similarity

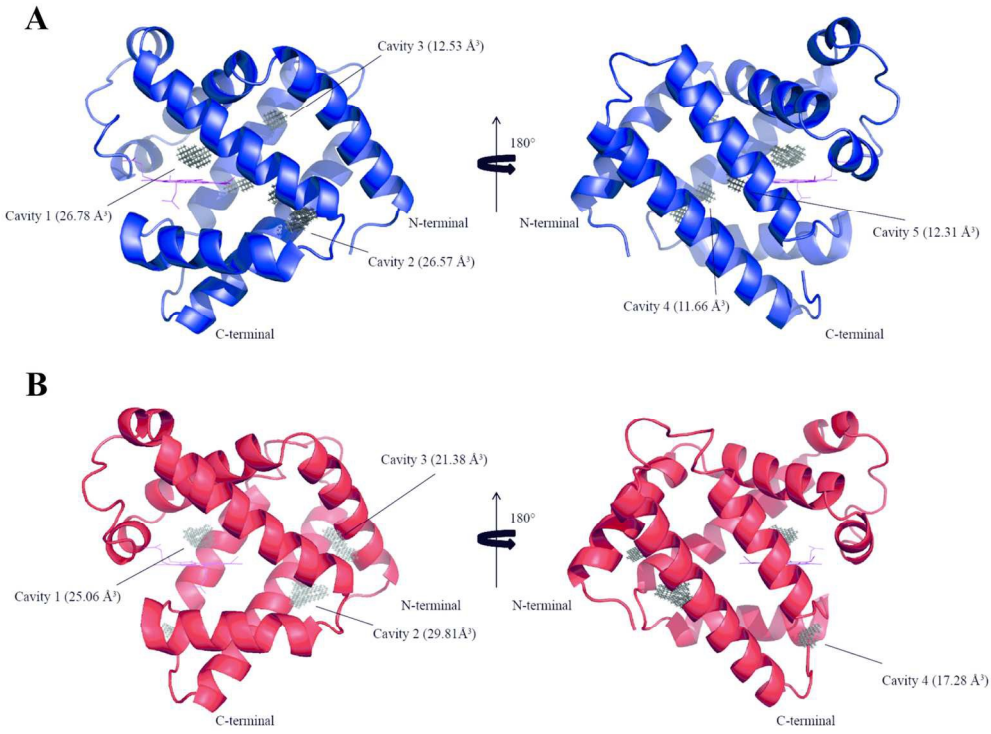
H. cristata	----	88.2	96.1	92.8	92.2	90.8	79.7	54.9
R. tarandus	83.7	----	91.5	92.8	88.9	90.2	77.1	52.3
S. scrofa	92.2	89.5	----	94.8	94.8	93.5	79.1	54.2
E. caballus	88.2	89.5	90.8	----	91.5	94.1	79.1	53.6
H. sapiens	88.9	85.6	93.5	88.2	----	92.8	79.1	55.6
P. macrocephalus	81.7	82.4	86.3	87.6	84.3	----	77.8	54.9
C. caretta	68.6	68.0	69.9	68.6	69.3	64.7	----	52.9
T. albacares	43.8	40.5	43.8	43.1	44.4	41.8	41.2	----

Identity

C



185x273mm (300 x 300 DPI)



136x99mm (300 x 300 DPI)

Table 1: Amino acid sequences of tryptic peptides from crested porcupine myoglobin, obtained by tandem mass spectrometry. Sequence position, experimental masses of precursor ions, charge state and molecular weight of tandem MS/MS sequence deduced from y series, together with mass accuracy, are reported.

	Tryptic peptide	Sequence position	Precursor ion (Da)	Charge state	MW from *de novo* sequence	Δ(Da)
T-1	GLSDGEWQLVLNVWGK	1-16	1799.92	2	1799.86	-0.06
T-2	VEGDLGGHGQEVLIR	17-31	1577.82	2	1577.79	-0.03
T-3	KHGTTVLTALGGILK	63-77	1507.91	2	1507.88	-0.03
T-4	GQHAAELAPLAQSHATK	80-96	1728.89	2	1728.85	-0.04
T-5	YLEFISEAIIQVLQSK	103-118	1880.03	2	1879.91	-0.12
T-6	HPADFGADTQGAMSK	119-133	1531.67	2	1531.59	-0.08

Table 2: Amino acid sequences of tryptic peptides from reindeer myoglobin, obtained by tandem mass spectrometry. Sequence position, experimental masses of precursor ions, charge state and molecular weight of tandem MS/MS sequence deduced from y series, together with mass accuracy, are reported.

	Tryptic peptide	Sequence position	Precursor ion (Da)	Charge state	MW from *de novo* sequence	Δ(Da)
T-1	GLSDGEWQLVLNAWGK	1-16	1771.89	2	1772.82	-0.07
T-2	VEADVAGHGQEVLIR	17-31	1591.83	2	1591.82	-0.01
T-3	LFTGHPETLEK	32-42	1270.65	2	1270.57	-0.08
T-4	YLEFISDAIIHVLHAK	103-118	1868.03	3	1868.07	0.04
T-5	HPSDFGADAQGAMSK	119-133	1517.66	2	1517.58	-0.08
T-6	ALELFRNDMAAQYK*	134-147	1668.84	3	1668.77	-0.07

*, missed cleavage

Table 3. Summary of thermodynamic data for the unfolding of crested porcupine and reindeer myoglobins

Protein	Tm (K)	ΔuH° (kJ/mol)	ΔuG° (298 K) (kJ/mol)
Reindeer Mb	353.7	257	38.8
Crested porcupine Mb	356.3	267	42.2