Research Articles

Computational analysis of vertebrate myoglobins reveals aggregation resistance in aquatic birds and higher surface hydrophobicity in fish

Comparative physiology of myoglobins

Syed M. Rizvi[1], Wei Zheng[1], Chengxin Zhang[1], and Yang Zhang[1,2]*

[1]Department of Computational Medicine and Bioinformatics

[2]Department of Biological Chemistry

University of Michigan, Ann Arbor, MI 48109 USA

*Author for Correspondence: Yang Zhang, Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, USA, phone: (734) 647-1549, fax: (734) 615-6553, email: zhng@umich.edu

Abstract

Myoglobin is the major oxygen carrying protein in vertebrate muscle. Previous studies identified in secondarily aquatic mammalian lineages high myoglobin net charge, which serves to prevent aggregation at the extremely high intracellular myoglobin concentrations found in these species. However, it is unknown how aquatic birds that dive for extended durations prevent myoglobin aggregation at their high intracellular myoglobin concentrations. It is also unknown whether secondarily aquatic lineages reduced the surface hydrophobicity of their myoglobins to prevent aggregation. Here, we used a deep learning-predicted distance-based protein folding algorithm to model the tertiary structures of 302 vertebrate myoglobin orthologs and performed a comparative analysis of their predicted net charge and surface hydrophobicities. The results suggest that aquatic avian divers, such as penguins and diving ducks, evolved highly charged myoglobins to reduce aggregation propensity and allow greater storage of oxygen for extended underwater foraging. High myoglobin net charge was also identified in golden eagles, a species that routinely suffers high-altitude hypoxia. Although no general association was found between myoglobin surface hydrophobicity and intracellular concentration, comparison of predicted net charge and surface hydrophobicities revealed significant differences between major vertebrate classes; bird myoglobins are the most positively charge, reptile myoglobins are the most negatively charged, and the myoglobins of ray-finned fish (*Actinopterygii*) have higher surface hydrophobicity than those of lobe-finned fish (*Sarcopterygii*). Our findings indicate the convergent evolution of high myoglobin net charge in aquatic birds and mammals, and offer novel insights into the diversification of myoglobin among vertebrate clades.

Introduction

Myoglobin, the metalloprotein responsible for the red color of muscle tissues, is a model protein for understanding intracellular diffusion dynamics and metalloprotein structure-function relationships.[1,2] Studies on orthologs of this protein from non-human vertebrates have provided important insights on its structure and function.[3–5] Furthermore, because protein aggregation is implicated in several degenerative diseases and also complicates the development of soluble biopharmaceuticals,[6–8] the ability of myoglobins from secondarily aquatic mammals to avoid aggregation at the dangerously high concentrations found in these species is of special interest.[9]

The muscle myoglobin concentrations of secondarily aquatic vertebrate divers (e.g. cetaceans and penguins) are often much higher than those of terrestrial vertebrates.[10–12] Although high concentrations of myoglobin in skeletal muscle allow greater storage of oxygen for use during prolonged dives, they also increase the risk of protein aggregation.[13] Aggregation is prevented in mammalian divers by the high net charge of their myoglobin, which increases electrostatic repulsion between interacting monomers and hence reduces the propensity to aggregate.[9,14] However, because previous studies of interspecific variation in myoglobin net charge examined only mammals, it is unknown how aquatic birds that dive for long durations (e.g. penguins and diving ducks) prevent the myoglobin aggregation expected at high intracellular concentrations.

In addition to net charge, hydrophobicity is another factor which affects the aggregation propensity of proteins.[15,16] Solvent-accessible hydrophobic residues are susceptible to associate with each other and can result in the formation of protein aggregates.[16,17] However, differences in myoglobin surface hydrophobicity between terrestrial and secondarily aquatic species have not been investigated. Furthermore, despite known differences in oxygen transport and globin

affinities between major classes of vertebrates (e.g. mammals versus reptiles),[18,19] any significant inter-class differences in myoglobin net charge and surface hydrophobicity remain to be identified.[20]

In this study, we modeled the tertiary structures of 302 vertebrate myoglobin orthologs and found significant differences in net charge and surface hydrophobicity between major classes of vertebrates. Furthermore, elevated net charge was found in those birds that routinely experience hypoxia, such as penguins, diving ducks, and golden eagles, indicating that mammalian and avian hypoxia-tolerant lineages convergently evolved high myoglobin net charge as a physicochemical defense against protein aggregation.

Materials and methods

Tertiary structure modeling by D-I-TASSER. To accurately predict myoglobin net charge using the three-dimensional structure-based net charge prediction algorithm PROPKA3,[21] D-I-TASSER[22] was used to first create tertiary structure models of the 311 vertebrate myoglobin orthologs on the NCBI Orthologs website.[23] D-I-TASSER[22] is an extended pipeline of I-TASSER[24] which integrates deep convolutional neural network-based distance and hydrogen-bonding network prediction to guide the assembly of template fragments into full-length model by Replica-Exchange Monte Carlo (REMC) simulations. The pipeline consists of four consecutive steps. First, D-I-TASSER uses DeepMSA[25] to iteratively search the query protein sequence against the whole-genome and metagenome sequence databases to obtain a multiple sequence alignment (MSA). Next, the selected MSA is used as the input for DeepPotential,[26] a newly developed deep residual neural-network-based predictor, in order to create multiple spatial restraints including (1)

distance-maps for both Cα and Cβ atoms; (2) C$\alpha\alpha$-based hydrogen-bonding networks; (3) Cα-Cβ torsion angles. Meanwhile, LOMETS3, a newly developed meta-server program containing both profile- and contact-based threading programs, is used to identify structural templates from a non-redundant PDB structural library. In the third step, continuous fragments extracted from the template structures by LOMETS3 are assembled into the full-length decoys by a REMC simulation under the guidance of a composite force field consisting of the deep learning-predicted spatial restraints, template-derived distance restraints, and knowledge-based energy terms calculated based on statistics of PDB structures. Finally, decoys from low-temperature replicas generated by the REMC simulations are clustered by SPICKER.[27] The SPICKER clusters are refined at the atomic level using fragment-guided molecular dynamic (FG-MD) simulations,[28] with the side-chain rotamer structures repacked by FASPR.[29]

Estimation of structure model quality. The global quality of structural models is usually appraised by the TM-score[30] between modeled and experimental structures:

$$TM = \frac{1}{L}\sum_{i=1}^{L_{ali}} \frac{1}{1+(d_i/d_0)^2} \tag{1}$$

where L is the number of residues, d_i is the distance between the i-th aligned residue pair, and $d_0 = 1.24 \cdot \sqrt[3]{L-15} - 1.8$ is a scaling factor. TM-score ranges between 0 and 1, and a TM-score greater than 0.5 indicates a structure model of correct global topology.[31]

Unfortunately, in the present study, we do not have the experimental structures for 98.7% of the myoglobin orthologs. Therefore, instead of the actual TM-score, we estimate the TM-score (*eTM-score*) using threading alignment quality, contact map fulfillment rate, and simulation convergence in D-I-TASSER:

$$eTM - score = a \cdot C^2 + b \cdot C + c \tag{2}$$

$$C = w_1 ln\left(\frac{M}{M_{total}} \cdot \frac{1}{<RMSD>}\right) + w_2 \sum_m ln\left(\frac{Z(m)}{Z_0(m)}\right) + w_3 ln\left(\frac{O(CM^{model}, CM^{pred})}{N(CM^{pred})}\right) \quad (3)$$

where a = 0.00098, b = 0.10770, and c= 0.79 are free parameters retrieved by regression, C is the c-score, M_{total} is the total number of decoy conformations used for clustering, M is the number of decoys in the top cluster, <*RMSD*> is the average RMSD of decoys in the same cluster, $Z(m)$ is the score of the top template by threading method m, $Z_0(m)$ is the cutoff for reliable templates, $N(CM^{pred})$ is the number of predicted contacts that guide the REMC simulation, and $O(CM^{model}, CM^{pred})$ is the number of overlapping contacts between final models and predicted contacts. $w_1 = 0.77$, $w_2 = 1.36$ and $w_3 = 0.67$ are free parameters. Estimated TM-score (*eTM-score*) highly correlates with real TM-score, with Pearson Correlation Coefficient (PCC) 0.81 on 797 training proteins.[32]

Data selection. To ensure accuracy of the final models, six models with estimated TM-score[30] <0.5 were removed from the analysis. Following strict statistics of structures in the PDB, TM-scores higher than 0.5 assume generally the same fold.[31] Unusually short orthologs (< 140 residues) were also discarded. Because some of the models included long N-terminal tails which resulted in unrealistically high net charge, STRIDE[33] was used to assign the N-terminal coil/loop regions for all models, and these regions were manually removed. After these curations, the final analysis included a total of 302 vertebrate myoglobins, of which 134 were from mammals, 73 from birds, 21 from reptiles, 72 from fish, and 2 from amphibians.

Net charge calculation and visualization of surface electrostatics. Net charge of the folded state for each of the myoglobin models was calculated using the PROPKA3 algorithm,[21] under the physiologically relevant pH of 7. Visualization of surface electrostatics was performed using the APBS Electrostatics plugin.[34]

Phylogenetic history. Evolutionary relationships between the bird orders in this study (figure 3) were extracted from an established comprehensive phylogeny construction of birds based on 259 nuclear loci and >390,000 bases of genomic sequence data from each of 198 avian species, highly supported by bayesian and maximum likelihood analyses.[35]

Surface hydrophobicity calculation. We define surface hydrophobicity as the percentage of the solvent-accessible surface that is contributed by the side chains of hydrophobic amino acids (alanine, isoleucine, leucine, valine, and phenylalanine). Surface hydrophobicity was calculated for each of the 302 vertebrate myoglobins using PyMOL scripts (supplementary text S1).

[**Mb] values.** Maximal skeletal muscle myoglobin concentrations of 31 mammalian species were retrieved from the supplementary materials of Mirceta et al. for testing associations between concentration, net charge, and surface hydrophobicity.[9]

Results and discussion

D-I-TASSER model quality. The global quality of structural models can be assessed by the TM-score[30] between modeled and experimental structures of the target protein. TM-score ranges between 0 and 1, with TM-score > 0.5 meaning structure models of correct global topology.[31] At the time of this writing, experimentally determined structures for four of the vertebrate species in our dataset were available in the Protein Data Bank (PDB): *Homo sapiens* (PDB ID: 3RGK), *Equus caballus* (PDB ID: 1WLA), *Phoca vitulina* (PDB ID: 1MBS), and *Physeter catodon* (PDB ID: 1VXH). TM-scores between our predicted models and the experimentally determined experimental structures were greater than 0.8 for all four species (figure 1A). As the experimental structures for the 298 remaining orthologs are not available, we estimated the TM-score (*eTM-*

score) of the D-I-TASSER models using a combination of threading alignment quality, contact satisfaction rate, and convergence of the structure assembly simulations (see details in Materials and Methods).[36] In the current study, the estimated TM-score was greater than 0.8 for 94.4% of the 302 myoglobin models (figure 1B), and the estimated TM-score was close to the actual TM-score for the four species with experimentally determined structures (figure 1A).

Positive correlation between experimentally determined muscle myoglobin content and computationally modelled net charge. We examined the relationship between experimentally determined muscle myoglobin concentration and our predicted net charge for 31 of the mammalian species in our dataset for which measured myoglobin concentrations existed in previous literature. A positive Pearson correlation coefficient of 0.68 (p-value = 3.1×10^{-5}) was found, which suggests a strong positive linear correlation between experimentally measured myoglobin concentration and our computationally predicted myoglobin net charge (figure 2). This indicates the role of increased electrostatic repulsion in preventing aggregation at high intracellular myoglobin concentrations.

Myoglobin net charge differences between diving and non-diving mammals. The methods employed in this study successfully replicated the results of a previous study on mammalian myoglobins which found higher net charge in aquatic diving species compared to non-diving terrestrial species.[9] Wilcoxon rank sum test confirmed that the higher median net charge of diving mammals (+2.84*e*) was statistically significantly different (p-value = 2.1×10^{-11}) from the lower median net charge of terrestrial mammals (+0.78*e*). 76.9% of the species in the 90th percentile of net charge were aquatic diving species, such as cetaceans, pinnipeds, platypus (*Ornithorhynchus anatinus*), and sea otter (*Enhydra lutris*).

The median net charge of pinniped myoglobins (+3.66*e*) was found to be higher than that of cetacean myoglobins (+2.81*e*), and a Wilcoxon rank sum test confirmed that this difference is

statistically significant (p-value = 0.01). To the best of our knowledge, differences in net charge between pinniped and cetacean myoglobin have not been identified in previous literature. We hypothesized that the difference might be due to higher muscle concentrations of myoglobin in pinnipeds, which would require a greater need for electrostatic repulsion produced by myoglobin net charge. We compared experimentally measured muscle myoglobin concentrations of pinnipeds and cetaceans from previous literature, and found that the median muscle myoglobin concentration of pinnipeds (60.84 mg/g wet mass) was indeed higher than that of cetaceans (35.34 mg/g wet mass), and a Wilcoxon rank sum test confirmed that this difference was statistically significant (p-value = 0.003). These results further establish the physiological link between intracellular myoglobin concentration and myoglobin net charge.

Elevated myoglobin net charge in diving birds. Significant differences in myoglobin net charge were found between diving and non-diving birds. Specifically, two major diving bird lineages, the *Sphenisciformes* (penguins) and *Anseriformes* (waterfowl), had higher average myoglobin net charge than the other bird orders (figure 3). Wilcoxon rank sum test revealed a statistically significant difference (p-value = 0.04) between the higher median net charge of diving species (+3.55*e*) and the lower median net charge of non-diving species (+2.11*e*), as shown in figure 4. Diving birds in the 90th percentile of net charge included the Emperor penguin (*Aptenodytes forsteri*, +3.57*e*), Adélie penguin (*Pygoscelis adeliae*, +3.55*e*), and the tufted duck (*Aythya fuligula*, +6.56*e*), which is a species of diving duck. The 90th percentile of net charge also included aquatic species which do not typically forage underwater for long durations, in particular the Dalmatian pelican (*Pelecanus crispus*, +3.24*e*) and domestic goose (*Anser cygnoides domesticus*, +3.28*e*). Comparison of surface electrostatics between diving and non-diving bird

myoglobin models revealed obvious differences in charge intensity and localization, especially in the region dorsal to the heme binding site (figure 5).

The link between myoglobin net charge and diving capacity depends on the assumption that the muscles of diving species possess unusually high concentrations of myoglobin. Although information on bird muscle myoglobin content is scarce, literature searches confirmed that three of the species in the 90th percentile of myoglobin net charge in our study indeed possess unusually high muscle myoglobin content: tufted diving duck (*Aythya fuligula*), Emperor penguin (*Aptenodytes forsteri*), and Adélie penguin (*Pygoscelis adeliae*). Penguins are known to possess one of the highest myoglobin concentrations in the entire vertebrate clade,[37] and the pectoral muscle myoglobin concentration of non-diving muscovy ducks is 2.2 mg/g,[38] which is significantly lower than the 7.4 mg/g for dive-trained diving ducks.[39] These associations between aquatic bird diving capacity, muscle myoglobin concentration, and myoglobin net charge suggest that mammalian and avian divers convergently evolved high net charge as a defense against myoglobin aggregation.

High myoglobin net charge in the golden eagle. The third most highly charged avian myoglobin (+4.09*e*) in the dataset belonged to the golden eagle (*Aquila chrysaetos*), a high-altitude dwelling species. A previous study on golden eagle hemoglobin identified unique structural features that increase its oxygen affinity and hence allow the species to survive high-altitude hypoxia.[40] The high net charge of golden eagle myoglobin might be another molecular adaptation for hypoxic stresses at high altitudes. Given the established association between myoglobin net charge and muscle myoglobin concentration,[9] and the fact that high-altitude vertebrates tend to possess higher concentrations of muscle myoglobin compared to their lowland counterparts,[41–43] we predict that golden eagles possess unusually high concentrations of myoglobin in their muscles,

which requires the high net charge for protection against aggregation. Experimental measurement of muscle myoglobin content in this species is needed to confirm this prediction.

Inter-class differences in myoglobin net charge. Generally, we found that birds possess the most highly charged myoglobins while reptiles possess the most negatively charged myoglobins (figure 6). The median and average myoglobin net charge for all vertebrates is +1.35*e* and +1.37*e*, respectively. Wilcoxon rank sum tests confirmed that the median net charge of bird myoglobin (+2.11*e*) was statistically significantly higher than that of mammalian myoglobin (+1.14*e*), reptilian myoglobin (-0.34*e*), and fish myoglobin (+1.41*e*) at the 5% significance level (p-values = 5.7×10^{-8}, 2.4×10^{-9}, and 0.02 for comparison of birds with mammals, reptiles, and fish, respectively), and that the median net charge of reptilian myoglobin (-0.34*e*) was statistically significantly lower than that of the other classes (p-values = 1.2×10^{-5}, 2.4×10^{-9}, and 2.3×10^{-4} for comparison of reptiles with mammals, birds, and fish, respectively).

The inter-class differences in myoglobin net charge cannot be definitively attributed to differences in muscle myoglobin content, because studies of avian and reptilian muscle did not find myoglobin concentrations that are particularly different from those of mammals.[19,43,44] The low myoglobin net charge of reptiles and amphibians (figure 6) might be related to their ectothermy, but an exact mechanism by which a higher myoglobin net charge would be beneficial for homeothermic vertebrates remains to be identified. Perhaps the lower myoglobin net charge of reptiles is a byproduct, or even a cause for, the lower oxygen affinities of reptilian globins compared to those of mammals.[18,19]

Due to the abundance of mammalian myoglobin orthologs, differences in net charge between major mammalian orders were also investigated, excluding aquatic species. The results showed that myoglobin net charge for different mammalian orders is quite similar, although the myoglobin

net charge for even-toed ungulates (order *Artiodactyla*) is statistically significantly lower than that of primates, carnivorans, and chiropterans (Wilcoxon rank sum p-values = 4.3×10^{-4}, 0.009, and 0.049, respectively).

Surface hydrophobicity is not associated with high intracellular myoglobin concentrations. Although greater hydrophobicity is associated with increased aggregation propensity of proteins,[15,16] we found little correlation (a Pearson correlation coefficient of -0.08) between myoglobin surface hydrophobicity and mammalian muscle myoglobin content (figure 7). This suggests that species with dangerously high intracellular concentrations of myoglobin (e.g. cetaceans and pinnipeds) reduced their myoglobin aggregation propensity by modifying net charge alone instead of modifying both net charge and solvent-accessible surface area of hydrophobic side chains. Interestingly, for mammals and birds, the myoglobins with the highest surface hydrophobicity were also among the most intensely charged myoglobins in their classes. Within mammals, the ferret (*Mustela putorius furo*) myoglobin was the most negatively charged (-5.84*e*) and also had the highest surface hydrophobicity (22.40%). Among birds, the North Island brown kiwi (*Apteryx mantelli*) myoglobin had the highest surface hydrophobicity (20.64%) and was also the second most highly charged avian myoglobin (+5.31*e*). We speculate that the increased aggregation propensity due to high surface hydrophobicity is compensated by elevated net charge in these outlier species.

Elevated myoglobin surface hydrophobicity in fish. Differences in myoglobin surface hydrophobicity between the major classes of vertebrates were investigated. We found that the myoglobins of mammals, birds, and reptiles have similar levels of surface hydrophobicity, while fish myoglobin is characterized by much higher surface hydrophobicity (figures 8 and 9). Fish myoglobins had a median percentage of hydrophobic solvent-accessible surface area of 21.44%,

which was statistically significantly higher than that of mammals (13.82%), birds (13.34%), and reptiles (13.49%), as confirmed by Wilcoxon rank sum tests between fish and each of these three other vertebrate classes (p-values = 6.6×10^{-28}, 8.1×10^{-23}, and 3.2×10^{-11} for comparison between fish and mammals, birds, and reptiles, respectively). The two amphibians in our dataset had values similar to the other tetrapod classes (12.95% and 11.66% for *Microcaecilia unicolor* and *Rhinatrema bivittatum*, respectively), and so did the only non-tetrapod sarcopterygian fish (11.97% for the coelecanth *Latimeria chalumnae*). Interestingly, the coelecanth myoglobin was the lone outlier among the fish myoglobins shown in figure 8, which suggests that low myoglobin surface hydrophobicity might have evolved in the ancestors of lobe-finned fish (*Sarcopterygii*) even before the first tetrapods emerged from Devonian waters. Future studies may uncover the physiological or evolutionary reasons for the development of low myoglobin surface hydrophobicity in lobe-finned fish.

Conclusions

Our systematic study of 302 vertebrate myoglobins found that aquatic avian divers such as penguins and diving ducks evolved high myoglobin net charge as a defense against protein aggregation, similar to their mammalian counterparts. The high net charge in these species allows them to pack enough myoglobin-bound oxygen to prevent underwater hypoxia. We also predicted the same adaptations for golden eagles, a species which routinely experiences high-altitude hypoxia. Furthermore, significant differences in myoglobin net charge and surface hydrophobicity were found between mammals, birds, reptiles, and fish. Birds possess the most positively charged mammalian myoglobins while reptiles possess the most negatively charged ones. Lobe-finned fish

(*Sarcopterygii*) possess higher myoglobin surface hydrophobicity than ray-finned fish (*Actinopterygii*), and there seems to be interplay between net charge and surface hydrophobicity in the myoglobins of select mammalian and avian species.

Acknowledgements

This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the National Science Foundation grant ACI1548562.

Funding

This work was supported in part by National Institutes of Health (GM083107, GM136422, AI134678, OD026825 to Y.Z.), and the National Science Foundation (DBI1564756 and IIS1901191 to Y.Z. and MTM2025426 to Y.Z.).

References

1. Gros G, Wittenberg BA, Jue T. Myoglobin's old and new clothes: From molecular structure to function in living cells. *J Exp Biol*. 2010;213:2713-2725. doi:10.1242/jeb.043075

2. Lin Y, Wang J, Lu Y. Functional tuning and expanding of myoglobin by rational protein design. *Sci China Chem*. 2014;57(3):346-355. doi:10.1007/s11426-014-5063-5

3. Samuel PP, Smith LP, Phillips GN, Olson JS. Apoglobin stability is the major factor

governing both cell-free and in vivo expression of holomyoglobin. *J Biol Chem*. 2015;290:23479-23495. doi:10.1074/jbc.M115.672204

4. Van Der Meer DLM, Van Den Thillart GEEJM, Witte F, et al. Gene expression profiling of the long-term adaptive response to hypoxia in the gills of adult zebrafish. *Am J Physiol - Regul Integr Comp Physiol*. 2005;289(5):R1512-R1519. doi:10.1152/ajpregu.00089.2005

5. Kendrew JC, Dickerson RE, Strandberg BE, et al. Structure of myoglobin: A three-dimensional fourier synthesis at 2 . resolution. *Nature*. 1960;185(4711):422-427. doi:10.1038/185422a0

6. Aguzzi A, O'Connor T. Protein aggregation diseases: Pathogenicity and therapeutic perspectives. *Nat Rev Drug Discov*. 2010;9(3):237-248. doi:10.1038/nrd3050

7. Cleland JL, Powell MF, Shire SJ. The development of stable protein formulations: A close look at protein aggregation, deamidation, and oxidation. *Crit Rev Ther Drug Carrier Syst*. 1993;10(4):307-377.

8. Mitragotri S, Burke PA, Langer R. Overcoming the challenges in administering biopharmaceuticals: Formulation and delivery strategies. *Nat Rev Drug Discov*. 2014. doi:10.1038/nrd4363

9. Mirceta S, Signore A V., Burns JM, Cossins AR, Campbell KL, Berenbrink M. Evolution of mammalian diving capacity traced by myoglobin net surface charge. *Science (80-)*. 2013;340:1234192. doi:10.1126/science.1234192

10. Noren SR, Williams TM. Body size and skeletal muscle myoglobin of cetaceans:

Adaptations for maximizing dive duration. *Comp Biochem Physiol - A Mol Integr Physiol*. 2000;126(2):181-191. doi:10.1016/S1095-6433(00)00182-3

11. De Miranda MA, Schlater AE, Green TL, Kanatous SB. In the face of hypoxia: Myoglobin increases in response to hypoxic conditions and lipid supplementation in cultured Weddell seal skeletal muscle cells. *J Exp Biol*. 2012;215(5):806-813. doi:10.1242/jeb.060681

12. Weber RE, Hemmingsen EA, Johansen K. Functional and biochemical studies of penguin myoglobin. *Comp Biochem Physiol -- Part B Biochem*. 1974;49(2):197-204. doi:10.1016/0305-0491(74)90154-0

13. Siefker J, Biehl R, Kruteva M, Feoktystov A, Coppens MO. Confinement Facilitated Protein Stabilization As Investigated by Small-Angle Neutron Scattering. *J Am Chem Soc*. 2018;140(40):12720–12723. doi:10.1021/jacs.8b08454

14. Lawrence MS, Phillips KJ, Liu DR. Supercharging proteins can impart unusual resilience. *J Am Chem Soc*. 2007;129(33):10110–10112. doi:10.1021/ja071641y

15. Pedersen JS, Dikov D, Otzen DE. N- and C-terminal hydrophobic patches are involved in fibrillation of glucagon. *Biochemistry*. 2006;45(48):14503–14512. doi:10.1021/bi061228n

16. Iannuzzi C, Maritato R, Irace G, Sirangelo I. Misfolding and amyloid aggregation of apomyoglobin. *Int J Mol Sci*. 2013;4(7):14287–14300. doi:10.3390/ijms140714287

17. Mishra A, Ranganathan S, Jayaram B, Sattar A. Role of solvent accessibility for aggregation-prone patches in protein folding. *Sci Rep*. 2018;8(1):12896. doi:10.1038/s41598-018-31289-6

18. Pough FH. Blood oxygen transport and delivery in reptiles. *Integr Comp Biol*. 1980;20(1):173-185. doi:10.1093/icb/20.1.173

19. Weber RE, Johansen K, Abe AS. Myoglobin from the burrowing reptile Amphisbaena alba. Concentrations and functional characteristics. *Comp Biochem Physiol -- Part A Physiol*. 1981;68(2):159-165. doi:10.1016/0300-9629(81)90336-4

20. Ferreras JM, Ragucci S, Citores L, Iglesias R, Pedone P V., Di Maro A. Insight into the phylogenetic relationship and structural features of vertebrate myoglobin family. *Int J Biol Macromol*. 2016;93:1041-1050. doi:10.1016/j.ijbiomac.2016.09.065

21. Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical p<italic>K</italic>$_a$ Predictions BT - Journal of Chemical Theory and Computation. *J Chem Theory Comput*. 2011;7(2):525-537.

22. Li Y, Zheng W, Zhang C, et al. Protein 3D Structure Prediction by D-I-TASSER in CASP14. In: *CASP 14 Abstract Book*. ; 2020:339-341.

23. NCBI. NCBI Orthologs - Myoglobin. https://www.ncbi.nlm.nih.gov/gene/4151/ortholog/?scope=7776.

24. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER suite: Protein structure and function prediction. *Nat Methods*. 2015;12(1):7-8. doi:10.1038/nmeth.3213

25. Zhang C, Zheng W, Mortuza SM, Li Y, Zhang Y. DeepMSA: Constructing deep multiple sequence alignment to improve contact prediction and fold-recognition for distant-homology proteins. *Bioinformatics*. 2020;36:2105-2112.

doi:10.1093/bioinformatics/btz863

26. Li Y, Zhang C, Zheng W, et al. Learning deep statistical potentials for protein folding. In: *CASP 14 Abstract Book*. ; 2020:72-73.

27. Zhang Y, Skolnick J. SPICKER: A clustering approach to identify near-native protein folds. *J Comput Chem*. 2004;25(6):865-871. doi:10.1002/jcc.20011

28. Zhang J, Liang Y, Zhang Y. Atomic-level protein structure refinement using fragment-guided molecular dynamics conformation sampling. *Structure*. 2011;19(12):1784-1795. doi:10.1016/j.str.2011.09.022

29. Huang X, Pearce R, Zhang Y. FASPR: An open-source tool for fast and accurate protein side-chain packing. *Bioinformatics*. 2020;36(12):3758–3765. doi:10.1093/bioinformatics/btaa234

30. Zhang Y, Skolnick J. Scoring function for automated assessment of protein structure template quality. *Proteins Struct Funct Genet*. 2004;57:702-710. doi:10.1002/prot.20264

31. Xu J, Zhang Y. How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics*. 2010;26(7):889-895. doi:10.1093/bioinformatics/btq066

32. Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9(1):40. doi:10.1186/1471-2105-9-40

33. Frishman D, Argos P. Knowledge-based protein secondary structure assignment. *Proteins Struct Funct Bioinforma*. 1995;23:566-579. doi:10.1002/prot.340230412

34. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*.

2001;98(18):10037-10041. doi:10.1073/pnas.181342398

35. Prum RO, Berv JS, Dornburg A, et al. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature*. 2015;526:569–573. doi:10.1038/nature15697

36. Zhang C, Zheng W, Cheng M, Omenn GS, Freddolino PL, Zhang Y. Functions of Essential Genes and a Scale-Free Protein Interaction Network Revealed by Structure-Based Function and Interaction Prediction for a Minimal Genome. *J Proteome Res*. 2021. doi:10.1021/acs.jproteome.0c00359

37. Ponganis PJ, Costello ML, Starke LN, Mathieu-Costello O, Kooyman GL. Structural and biochemical characteristics of locomotory muscles of emperor penguins, Aptenodytes forsteri. *Respir Physiol*. 1997;109(1):73-80. doi:10.1016/S0034-5687(97)84031-5

38. Kagen LJ, Linder S. Immunological Studies of Duck Myoglobin. *Proc Soc Exp Biol Med*. 1968;128(2):438-441. doi:10.3181/00379727-128-33032

39. Stephenson R, Turner DL, Butler PJ. The Relationship Between Diving Activity and Oxygen Storage Capacity in the Tufted Duck (Aythya Fuligula). *J Exp Biol*. 1989;141:265-275.

40. Ali A, Rehman T, Roshan A. Analysis of Oxygen Affinity of the Major Hemoglobin Component HbA from Aquila chrysaetos (Golden Eagle). *J Nat Sci Res*. 2014;4(9):55-61.

41. Dawson NJ, Ivy CM, Alza L, et al. Mitochondrial physiology in the skeletal and cardiac muscles is altered in torrent ducks, Merganetta armata, from high altitudes in the Andes. *J Exp Biol*. 2016;219(23):3719-3728. doi:10.1242/jeb.142711

42. VAUGHAN BE, PACE N. Changes in myoglobin content of the high altitude acclimatized rat. *Am J Physiol*. 1956;185(3):549-556. doi:10.1152/ajplegacy.1956.185.3.549

43. Xin Y, Tang X, Wang H, et al. Functional characterization and expression analysis of myoglobin in high-altitude lizard Phrynocephalus erythrurus. *Comp Biochem Physiol Part - B Biochem Mol Biol*. 2015;188:31-36. doi:10.1016/j.cbpb.2015.06.004

44. Pages T, Planas J. Muscle myoglobin and flying habits in birds. *Comp Biochem Physiol -- Part A Physiol*. 1983;74(2):289-294. doi:10.1016/0300-9629(83)90602-3

Figure legends

Figure 1. **(A)** TM-align superposition of models and experimental structures from human, horse, harbor seal, and sperm whale. **(B)** Histogram depicting the distribution of estimated TM-scores for the 302 D-I-TASSER models used in this study.

Figure 2. Predicted myoglobin net charge and experimentally determined muscle myoglobin concentrations in mammals.

Figure 3. Cladogram and average predicted myoglobin net charge of bird orders. Branches between internal nodes were colored orange because the net charge range for 56.7% of bird orders falls between 1.5 and 2.5*e*.

Figure 4. Differences in predicted myoglobin net charge between non-diving and diving birds.

Figure 5. Visualization of surface electrostatics for the myoglobin models from **(A)** chicken, **(B)** Emperor penguin, **(C)** mallard, and **(D)** tufted duck. The side shown is dorsal to the heme binding site.

Figure 6. Predicted median myoglobin net charge in mammals, birds, reptiles, and fish.

Figure 7. Predicted myoglobin surface hydrophobicity and experimentally determined muscle myoglobin concentrations in mammals.

Figure 8. Predicted myoglobin surface hydrophobicity in mammals, birds, reptiles, and fish.

Figure 9. Hydrophobic surface areas (red) on myoglobin models from human, chimpanzee, cow, ferret, pigeon, kiwi, coelecanth, and Atlantic cod.

Figure 1.

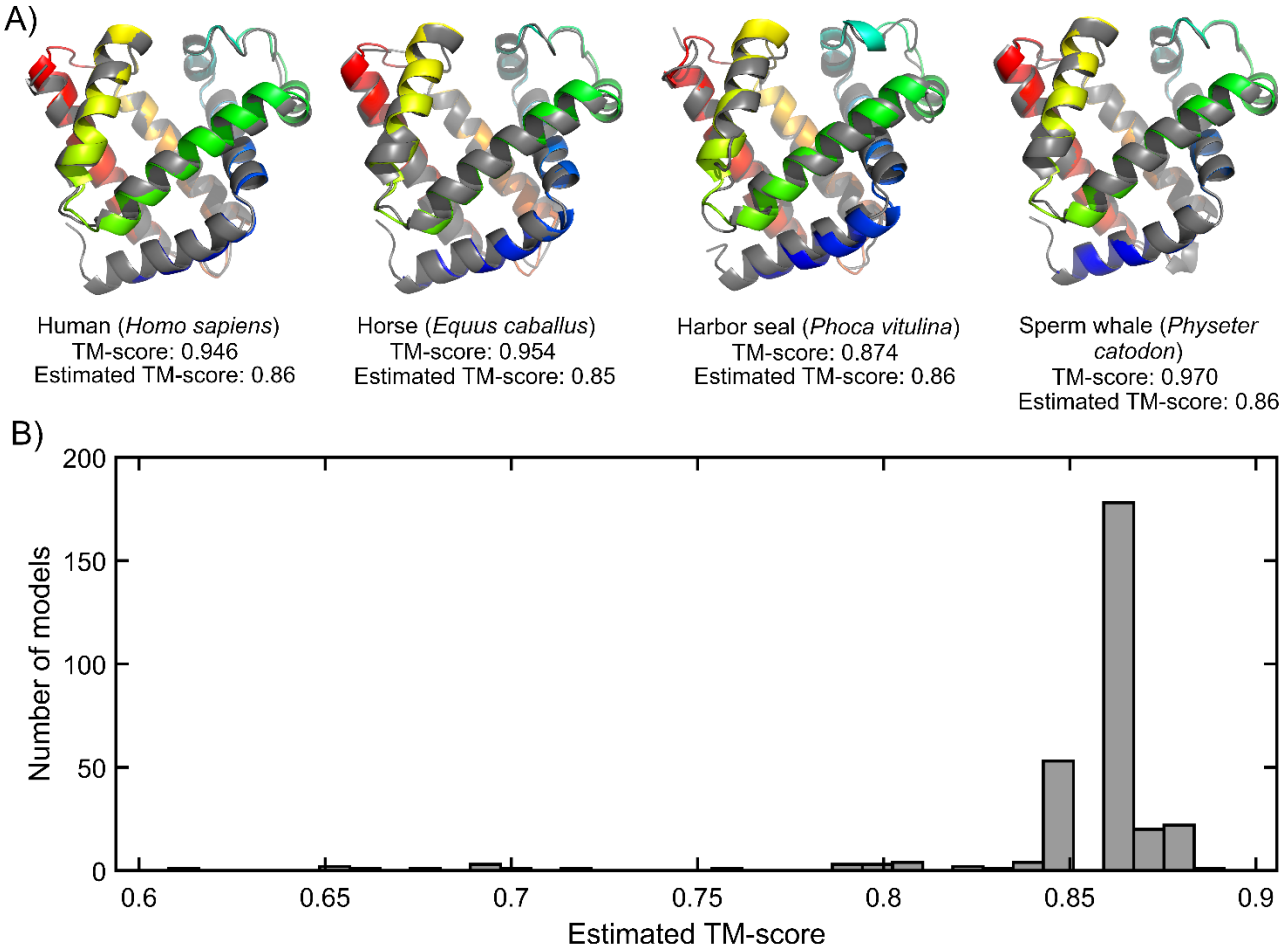

A)
Human (*Homo sapiens*)
TM-score: 0.946
Estimated TM-score: 0.86
Horse (*Equus caballus*)
TM-score: 0.954
Estimated TM-score: 0.85
Harbor seal (*Phoca vitulina*)
TM-score: 0.874
Estimated TM-score: 0.86
Sperm whale (*Physeter catodon*)
TM-score: 0.970
Estimated TM-score: 0.86
B)
Number of models
200
150
100
50
0
0.6
0.65
0.7
0.75
0.8
0.85
0.9
Estimated TM-score

Figure 2.

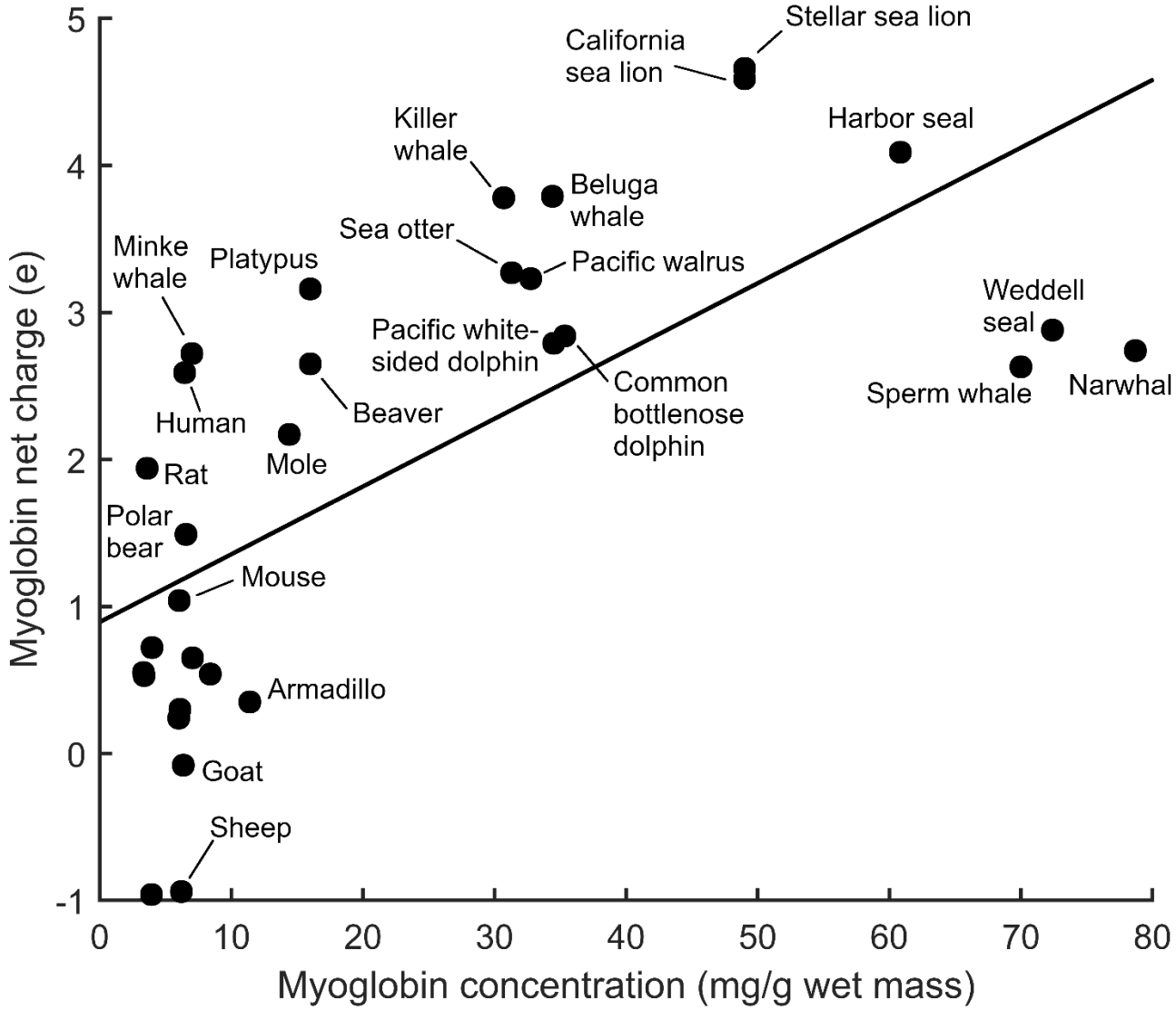

Stellar sea lion
California sea lion
Harbor seal
Killer whale
Beluga whale
Sea otter
Pacific walrus
Minke whale
Platypus
Weddell seal
Pacific white-sided dolphin
Common bottlenose dolphin
Sperm whale
Narwhal
Human
Beaver
Mole
Rat
Polar bear
Mouse
Armadillo
Goat
Sheep
Myoglobin net charge (e)
Myoglobin concentration (mg/g wet mass)
5
4
3
2
1
0
-1
0
10
20
30
40
50
60
70
80

Figure 3.

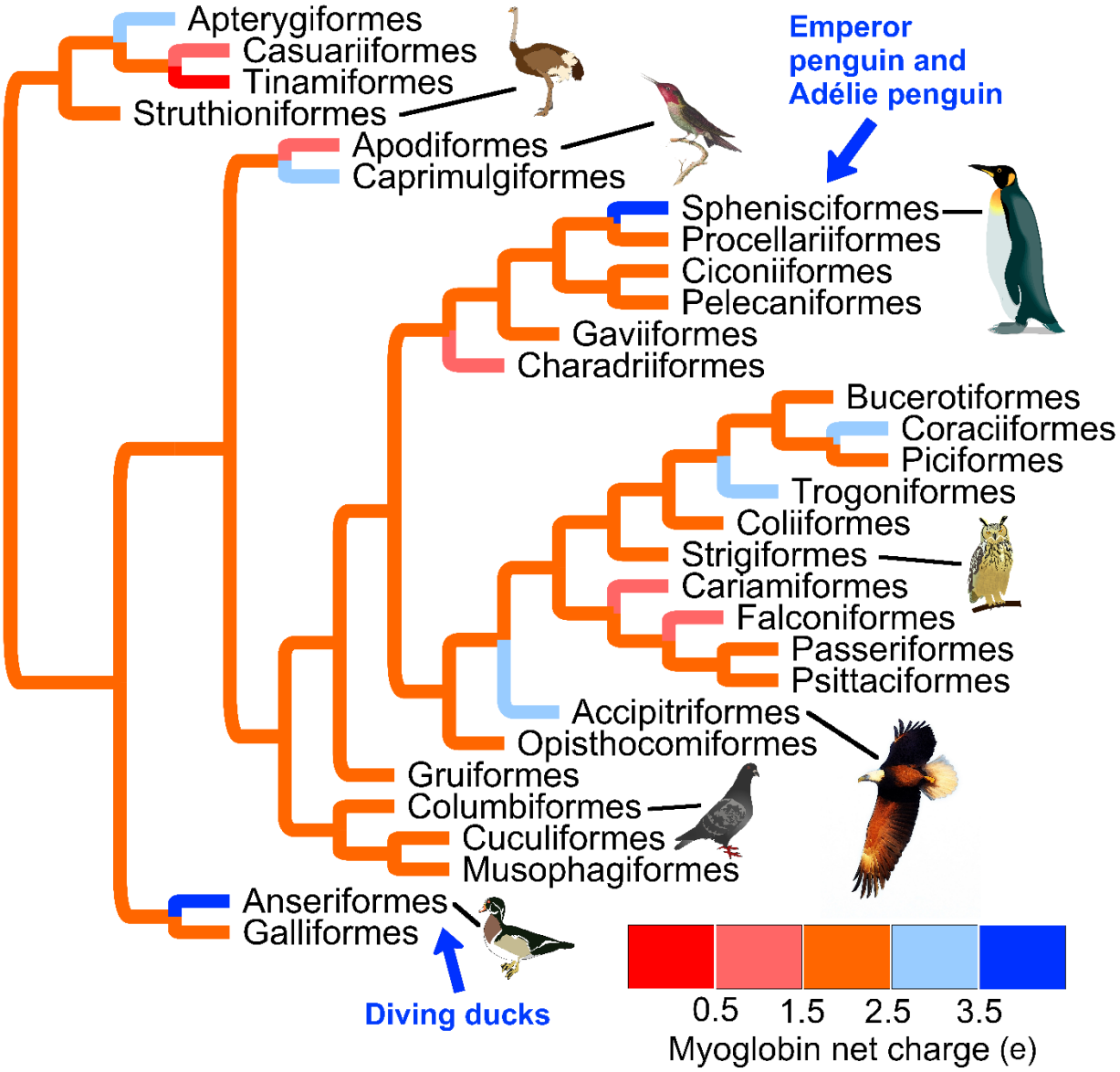

Apterygiformes
Casuariiformes
Tinamiformes
Struthioniformes
Apodiformes
Caprimulgiformes
Emperor penguin and Adélie penguin
Sphenisciformes
Procellariiformes
Ciconiiformes
Pelecaniformes
Gaviiformes
Charadriiformes
Bucerotiformes
Coraciiformes
Piciformes
Trogoniformes
Coliiformes
Strigiformes
Cariamiformes
Falconiformes
Passeriformes
Psittaciformes
Accipitriformes
Opisthocomiformes
Gruiformes
Columbiformes
Cuculiformes
Musophagiformes
Anseriformes
Galliformes
Diving ducks
0.5
1.5
2.5
3.5
Myoglobin net charge (e)

Figure 4.

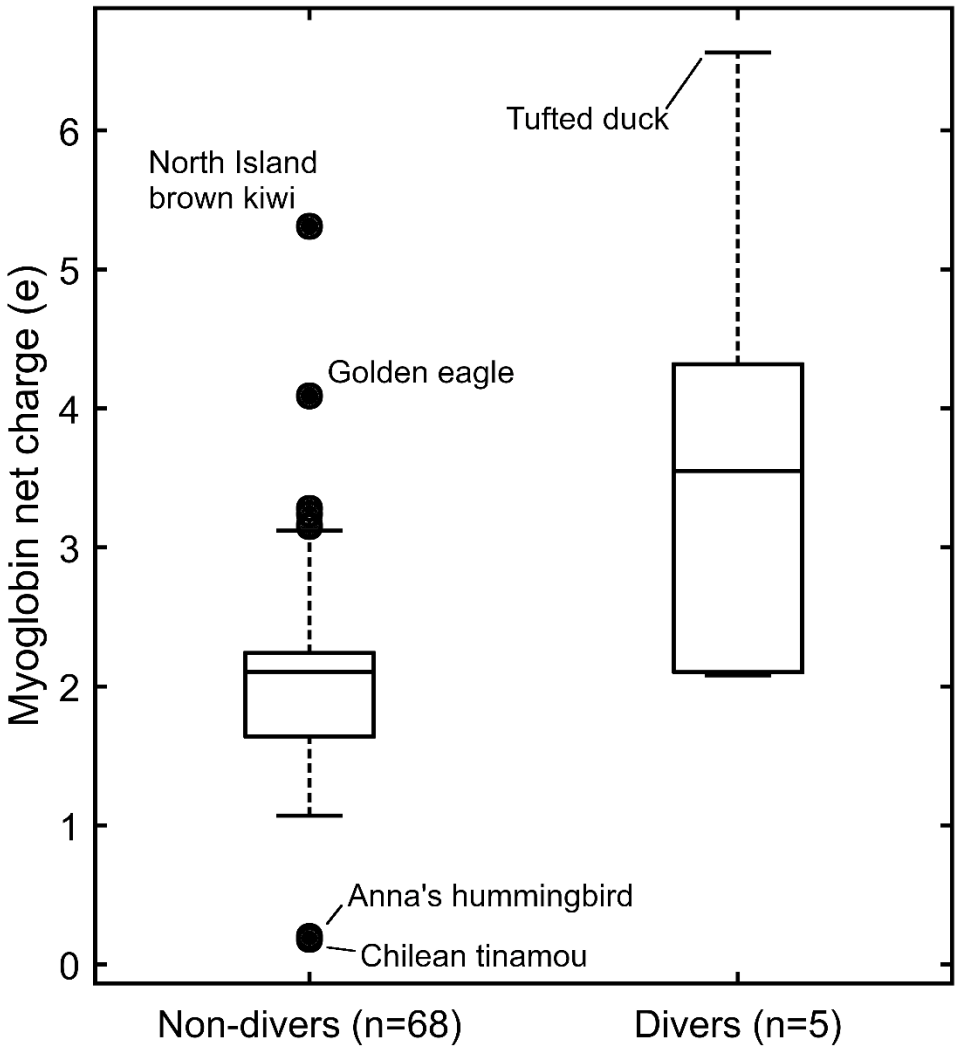

Tufted duck
North Island
brown kiwi
Golden eagle
Anna's hummingbird
Chilean tinamou
Myoglobin net charge (e)
0
1
2
3
4
5
6
Non-divers (n=68)
Divers (n=5)

Figure 5.

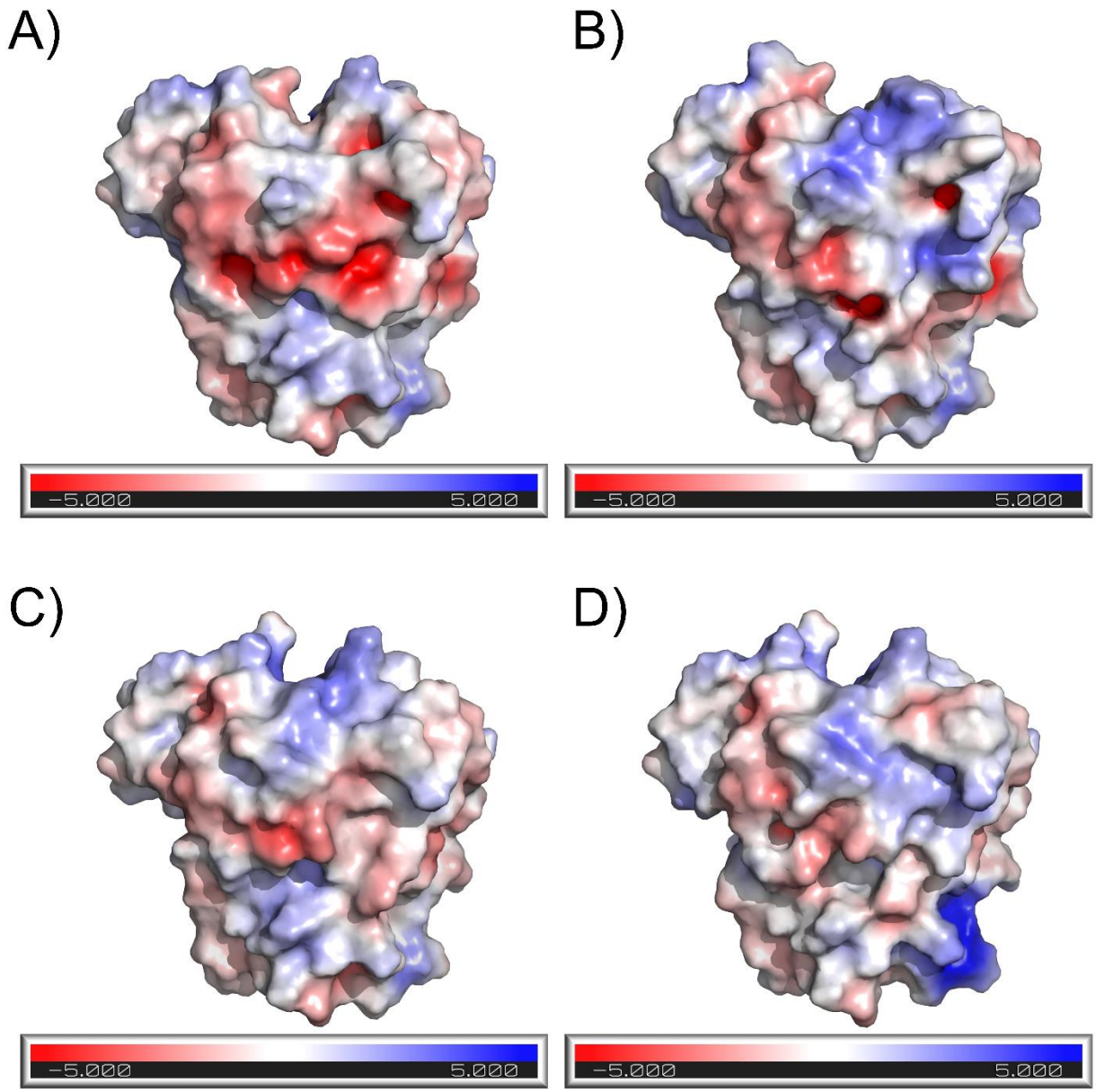


Figure 6.

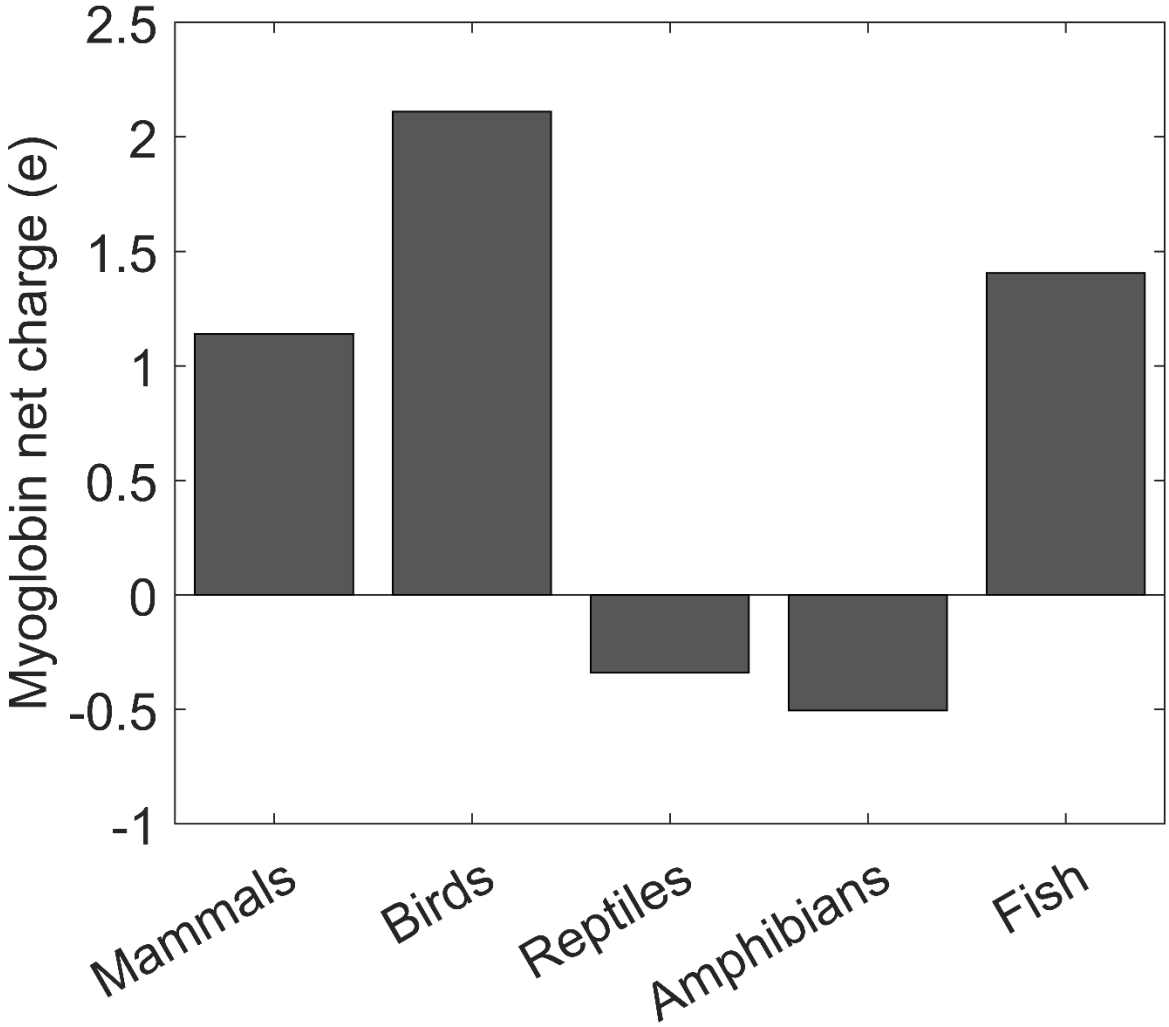

Myoglobin net charge (e)
2.5
2
1.5
1
0.5
0
-0.5
-1
Mammals
Birds
Reptiles
Amphibians
Fish

Figure 7.

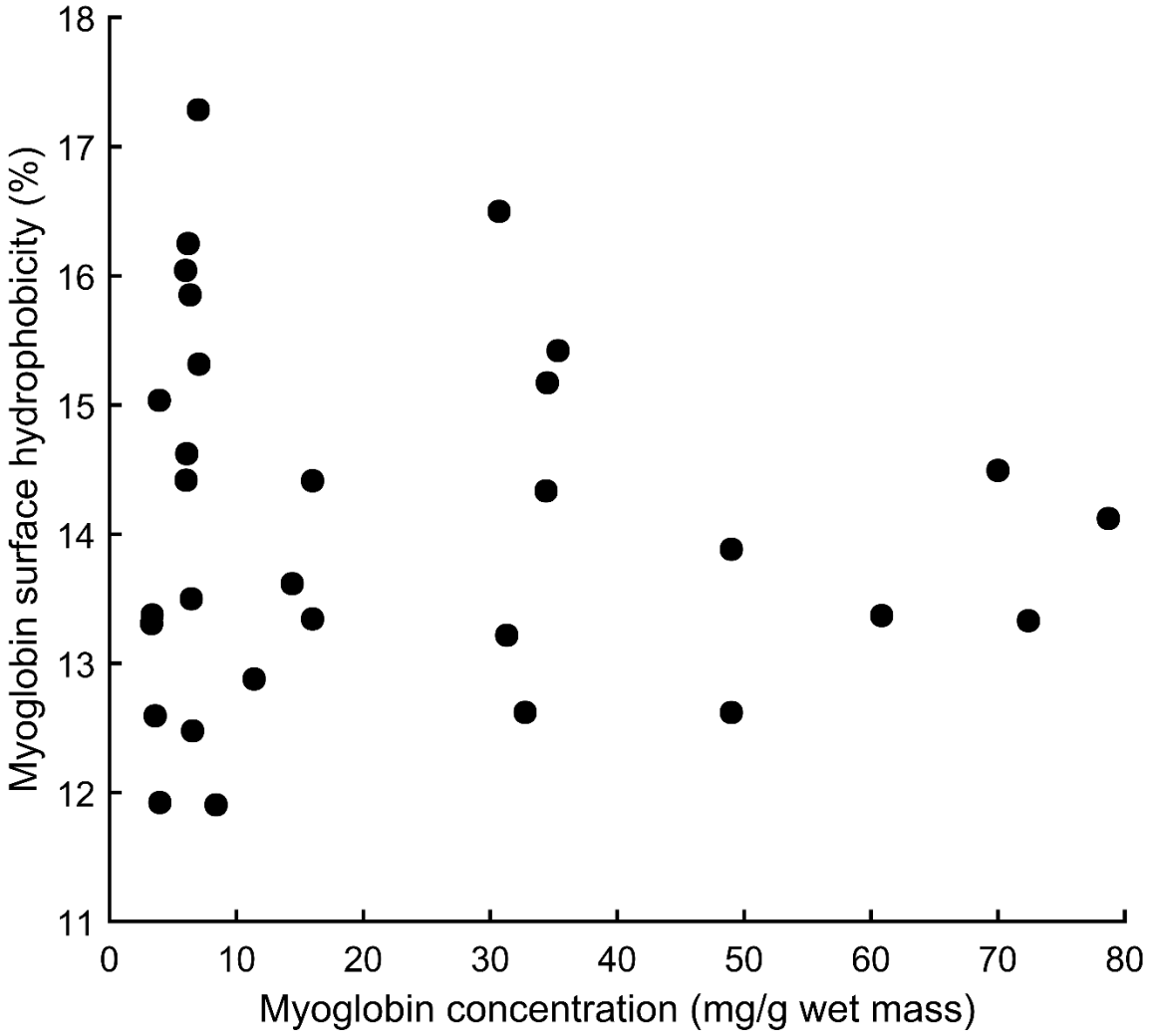

Myoglobin surface hydrophobicity (%)
18
17
16
15
14
13
12
11
0
10
20
30
40
50
60
70
80
Myoglobin concentration (mg/g wet mass)

Figure 8.

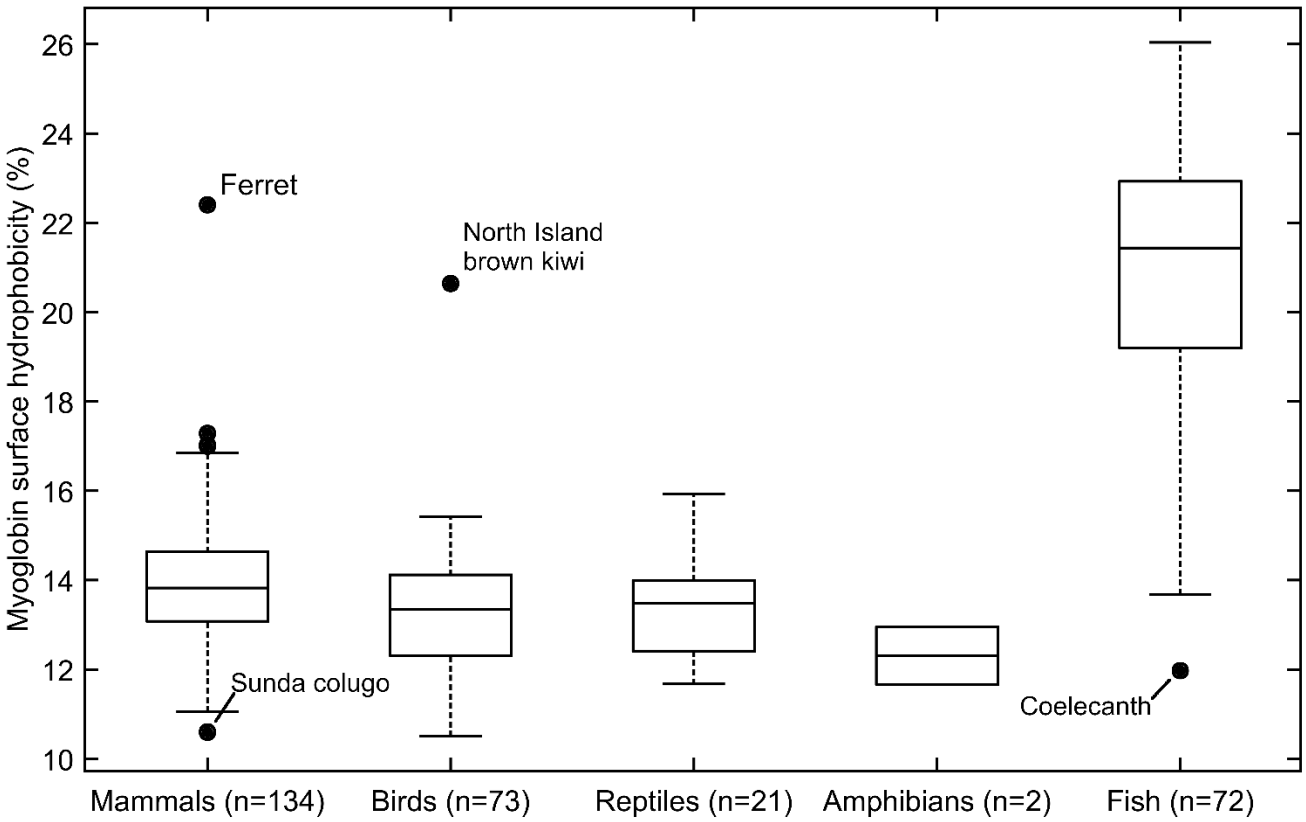
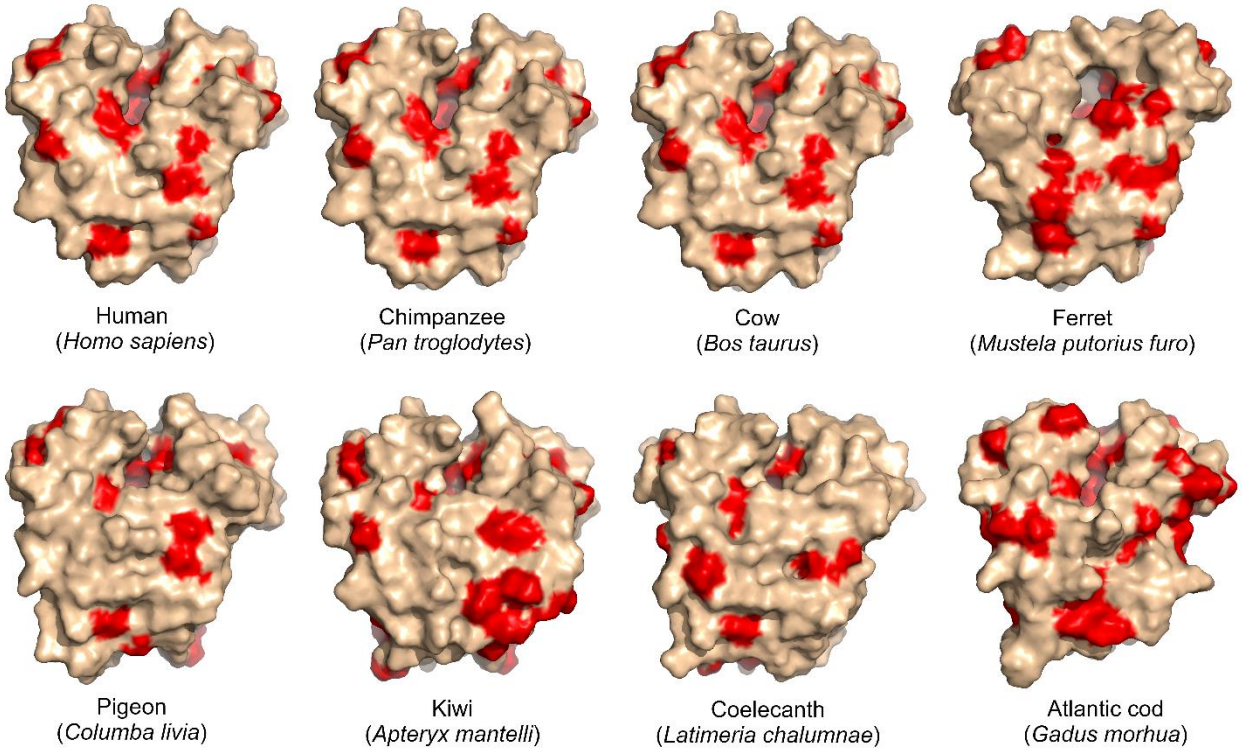

Myoglobin surface hydrophobicity (%)
26
24
22
20
18
16
14
12
10
Ferret
North Island
brown kiwi
Sunda colugo
Coelecanth
Mammals (n=134)
Birds (n=73)
Reptiles (n=21)
Amphibians (n=2)
Fish (n=72)

Figure 9.


Human
(*Homo sapiens*)
Chimpanzee
(*Pan troglodytes*)
Cow
(*Bos taurus*)
Ferret
(*Mustela putorius furo*)
Pigeon
(*Columba livia*)
Kiwi
(*Apteryx mantelli*)
Coelecanth
(*Latimeria chalumnae*)
Atlantic cod
(*Gadus morhua*)