Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

The energetics and evolution of oxidoreductases in deep time

Vikas Nanda[1], Kenneth N. McGuinness[2], Nolan Fehon[3], Ryan Feehan[4], Michelle Miller[3], Andrew Mutter[5], Justin Nam[1], Jenna E. AbuSalim[1], Joshua T. Atkinson[6], Hirbod Heidari[7], Natalie Losada[1], J. Dongun Kim[3], Ronald L. Koder[5], Yi Lu[7], Joff Silberg[6], Joanna Slusky[4], and Paul Falkowski[3]

[1]Center for Advanced Biotechnology and Medicine
[2]Caldwell University
[3]Rutgers University Institute of Marine and Coastal Sciences
[4]The University of Kansas
[5]The City College of New York Department of Physics
[6]Rice University
[7]The University of Texas at Austin Department of Chemistry

May 18, 2023

Abstract

The core metabolic reactions of life drive electrons through a class of redox protein enzymes, the oxidoreductases. The energetics of electron flow is determined by the redox potentials of organic and inorganic cofactors as tuned by the protein environment. Understanding how protein structure affects oxidation-reduction energetics is crucial for studying metabolism, creating bioelectronic systems, and tracing the history of biological energy utilization on Earth. We constructed ProtReDox (https://protein-redox-potential.web.app), a manually curated database of experimentally determined redox potentials. With over 500 measurements, we can begin to identify how proteins modulate oxidation-reduction energetics across the tree of life. By mapping redox potentials onto networks of oxidoreductase fold evolution, we can infer the evolution of electron transfer energetics over deep-time. ProtReDox is designed to include user-contributed submissions with the intention of making it a valuable resource for researchers in this field.

The energetics and evolution of oxidoreductases in deep time

Short Title: Protein Redox Potentials

Authors:

Kenneth N. McGuinness[1,2*], Nolan Fehon[3], Ryan Feehan[4], Michelle Miller[3], Andrew C. Mutter[5], Justin Nam[2], Jenna E. AbuSalim[2], Joshua T. Atkinson[6], Hirbod Heidari[7], Natalie Losada[2], J. Dongun Kim[3], Ronald L. Koder[5], Yi Lu[7], Jonathan J. Silberg[6], Joanna S.G. Slusky[4,8], Paul G. Falkowski[3,9*], Vikas Nanda[2,10*]

Affiliations:

[1] Department of Natural Sciences, Caldwell University, Caldwell NJ 07006

[2] Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway NJ 08854

[3] Environmental Biophysics and Molecular Ecology Program, Department of Marine and Coastal Sciences, Rutgers University, New Brunswick NJ 08901

[4] Computational Biology Program, The University of Kansas, Lawrence KS 66047

[5] Department of Physics, The City College of New York, New York NY 10016

[6] Department of Chemical and Biomolecular Engineering, Rice University, Houston TX 77005

[7] Department of Chemistry, University of Texas at Austin, Austin TX 78712

[8] Department of Molecular Biosciences, The University of Kansas, Lawrence KS 66045

[9] Department of Earth and Planetary Sciences, Rutgers University, New Brunswick NJ 08901

[10] Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, Rutgers University, Piscataway NJ 08854

*correspondence should be addressed to: KNM kmcguinness@caldwell.edu, PGF falko@marine.rutgers.edu or VN vik.nanda@rutgers.edu

Acknowledgements and Funding: The authors thank Jennifer Timm, and Jan Siess for useful discussions, and extend our sincere appreciation to our colleagues who meticulously measured and published the redox potentials surveyed in this work. This work was supported by a NASA Astrobiology Institute Grant 80NSSC18M0093 to J.J.S., P.G.F. and V.N. K.N.M. was supported by the NIH IRACDA Postdoctoral Fellowship Program Grant K12GM093854. J.N. was supported by the RISE program at Rutgers University.

Keywords: enzymes, metabolism, electron transfer, database, oxidation-reduction

Data Availability Statement: The data that support the findings of this study are openly available in https://protein-redox-potential.web.app

Conflict of Interest Statement: The authors have declared no conflict of interest.

ABSTRACT:

The core metabolic reactions of life drive electrons through a class of redox protein enzymes, the oxidoreductases. The energetics of electron flow is determined by the redox potentials of organic and inorganic cofactors as tuned by the protein environment. Understanding how protein structure affects oxidation-reduction energetics is crucial for studying metabolism, creating bioelectronic systems, and tracing the history of biological energy utilization on Earth. We constructed ProtReDox (https://protein-redox-potential.web.app), a manually curated database of experimentally determined redox potentials. With over 500 measurements, we can begin to identify how proteins modulate oxidation-reduction energetics across the tree of life. By mapping redox potentials onto networks of oxidoreductase fold evolution, we can infer the evolution of electron transfer energetics over deep-time. ProtReDox is designed to include user-contributed submissions with the intention of making it a valuable resource for researchers in this field.

Keywords: oxidation-reduction; proteins; energy metabolism; electrons; oxidoreductases, enzymes and coenzymes

INTRODUCTION

On a global scale from an electron perspective, all organisms are electronic half-cells, powered by circuits plugged into electron sources and sinks in the environment [1-3]. For example, in aerobic respiration, which is probably most familiar to us, as that is our source of energy, the oxidation of organic matter leads to a flux of electrons and protons through metabolic pathways to reduce oxygen to water and CO_2. This, like all metabolic pathways, is a half-cell terms of chemical oxidation-reduction pathways. In the case of aerobic respiration, the other half cell is oxygenic photosynthesis, where sunlight is used to oxidize water and the electrons and protons drive reduction of CO_2 to organic matter. The voltage potential between the anode (e.g., organic matter; in its simplest form, sugars) and the cathode (e.g., oxygen) provides over 1 volt of

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

energy. That is the most energy available for life on this planet - but life existed long before there was molecular oxygen.

In deep time, a set of enzymes evolved to facilitate electron transport – the oxidoreductases or EC 1 proteins. Biological electronic circuits require the movement of electrons over sub-nanometer distances through an electron transfer chain that powers life. The movement of electrons is governed by physical laws [4-6]. Oxidoreductases organize the positions and relative energetics of chains of redox-active cofactors, assuring the rapid, directional flow of electrons [7]. The energetic tendency of a redox-active group to gain electron or lose electrons can be experimentally measured as the redox potential, expressed in volts (V), relative to a reference such as the standard hydrogen electrode, at a standard pH. Redox-active groups that contribute to the redox potential can be cofactors such as iron-sulfur clusters, hemes, or flavins, or amino acid residues such as cysteine, methionine, or tryptophan. The relative stability of cofactor oxidation states are largely determined by the cofactor itself [8] but are further modulated by the protein matrix. Electrostatic interactions, such as proximity of positively charged basic amino acids, can stabilize a redox cofactor in the reduced state [9, 10]. The protein can modulate oxidation-reduction energetics through hydrogen bonding [11, 12], hydration [13] and dynamical features [14] of the protein-cofactor environment. Groups of oxidoreductases form metabolic pathways, powering cellular-scale circuits where the current depends on the rate of catalysis and diffusion of substrates [15]. It is critical to study how the protein environment modulates the energetics of oxidation-reduction reactions in order to understand how electron transfer is coupled to metabolism.

The connection between oxidoreductase structure and energetics is central to the deep-time evolution of metabolism. Oxidoreductases must have been among the first proteins at the origins of life over 3.5 billion years ago providing the spark for metabolism [2, 16-20]. Due to its fundamental electrical nature, the evolution of metabolism, and the associated oxidoreductases, was strongly coupled with changes in the redox state of the planet, which has become increasingly oxidized over time due to both geochemical and biological processes [2, 21, 22].

Modern oxidoreductases are massive nanomachines – far too complex to have arisen early in metabolism. Various structure-based bioinformatics approaches have been applied to identify universal sub-folds or domains within larger proteins that may have derived from early protein forms [16, 17, 23-32]. In previous work focused on the evolution of oxidoreductases, we found that modern, large enzymes were largely derived from just a few minimal protein-cofactor building blocks [16, 17, 33]. In addition to identifying core cofactor binding folds, we used a structure-derived criterion for electron transfer based on cofactor-cofactor distances [7] to map a network of electron transfer pathways between the different folds – which we refer to as the Spatial Adjacency Network, SpAN. A notable feature of the SpAN was the abundance of more reducing cofactor-binding folds in the network center and more oxidizing cofactor-folds at the periphery [17]. This suggests a time axis in the SpAN from the center to the periphery of the network reflecting the adaptation of protein redox energetics to emerging electron sources and sinks made available by an oxidizing planetary environment over geologic time. Mapping quantitative estimates of protein redox energetics onto the SpAN would allow us to potentially constrain the age of various protein folds based on redox information in the geologic record [2, 34, 35].

Computational approaches for prediction of redox energetics based on protein structures is an ongoing challenge. Current methods span many levels of theory from quantum-mechanical to empirical [36] and recent advances using machine learning [37]. Site-directed mutagenesis studies on natural oxidoreductases [38-40] and protein engineering [41-44] have been used to test molecular hypothesis of how the protein environment tunes redox energetics. Large datasets of protein structures, including oxidoreductases, are on the horizon with advances in functional annotation from genomic and metagenomic datasets [20, 45] combined with recent advances in structure prediction [46-48] including bound cofactors [49]. Effective models that can predict redox energetics based on structural information will become increasingly valuable for understanding bioenergetics, evolution of metabolism and engineering of bioelectronic pathways [42, 50].

Motivated by the need to design and train better models and the goal of mapping redox energetics onto the SpAN to study oxidoreductase evolution, we develop ProtReDox, a manually curated database of pro-

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

tein redox potentials. We examined literature reports of oxidoreductase energetics and identified the cofactor type, redox potential, UniProt and PDB (if available) identifiers, and experimental metadata such as potentiometric measurement technique, pH and buffer conditions. ProtReDox version one is available at*https://protein-redox-potential.web.app*. We apply this dataset to explore how redox energetics is modulated by cofactor-type, protein environment, experimental conditions and finally how energetics mapped onto the SpAN inform geochemical constraints on deep-time oxidoreductase evolution.

METHODS

Data collection and curation . The dataset includes 514 redox potentials from 239 unique enzyme/cofactor pairs consisting of metal ions (Cu, Fe, Mo), flavins, hemes, and multinuclear iron-sulfur clusters. Proteins are indexed by their UniProt ID, and approximately half are associated with high-resolution structures deposited in the Protein Data Bank [51]. Redox potentials are normalized to the standard hydrogen electrode and pH-corrected to pH 7.0. Redox potentials were only included if the midpoint potential could unequivocally be assigned to a particular cofactor.

ProtReDox database construction . Redox potentials and associated data are stored in a Google Firebase Cloud Firestore database [52]. The ProtReDox website is rendered using the Firebase Web v.9 modular JavaScript SDK in combination with React.js (v. 18.2) (react.dev). The website user interface comprises a navigation, logo, searchable redox dataset table, and a form to input new redox potentials and associated information. User-contributed additions to the dataset will be marked for review and evaluated manually.

Feature-redox correlation analysis **.** To better understand the key features controlling redox potential, 486 features were calculated as previously described[53] for a set of 42 protein structures with type 1 copper sites with experimentally identified reduction potentials. These features covered ten categories of physicochemical properties based on how they were calculated: solvation, electrostatics, hydrogen bonding, van der Waals, geometry, pocket void, secondary structure of the backbone region of the protein directly interacting to the redox site, amino acid angles, pocket lining, and surface area. The property values for sites on different chains of the same protein structure were averaged. Any features for which all structures had the same value were removed, leaving 446 features. Pearson correlation coefficients between features and reduction potential were then calculated using the python library SciPy [54]. For each structure, the reduction potential with an experimental pH closest to the crystallization pH was selected. When no crystallization pH was available, the reduction potential with the most neutral pH was selected. These reduction potentials were then normalized to pH 7.0 for further analysis (eq. 1).

Mapping redox energetics on the SpAN **.** The SpAN is a network representation of protein electron transfer pathways with nodes corresponding to classes of protein microenvironments surrounding the redox cofactor (termed modules) and edges reflecting instances in the PDB where two modules are within electron transfer proximity (cofactor-cofactor distance < 14 Å). The generation of this network was described in our previous work [16, 17]. The 2020 version of the SpAN was used in this study.

RESULTS AND DISCUSSION

Cofactor-type is the primary determinant of redox energetics . Redox potentials included in ProtRedox span almost 2 V, ranging from the -675 mV 2[4Fe-4S] binding bacterial ferredoxin of *E. coli*[55] to the +1301 mV chlorophyll A in PS II within *T. elongatus* [56]. Within this broad range, the cofactor type is the primary determinant of redox potential (**Fig. 1**). Cofactor types are designated based primarily on the PDB-derived nomenclature. Cofactors from most reducing to most oxidizing were 4Fe-4S (SF4), 2Fe-2S (FES), flavins, mononuclear iron sites (Fe), iron-bound hemes (HEM) and copper sites (Cu). These ranges are consistent with previous analyses of protein electron carriers [8].

Molecular features that determine energetics . Protein redox potential is a complex property that is affected by features of the redox site first and second shell environment: solvation, hydrogen-bonding, ligand interactions, metal coordination, electrostatic interactions [57, 58] and corresponding enthalpic and entropic energy terms [59]. Redox potential can be directly calculated from first principle quantum mechanics calculations [60, 61],

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

however these calculations are expensive and are not practical for protein design. To better understand the protein features that determine redox potential, we calculated the correlation between 433 physicochemical features (including energy and geometry features) and reduction potentials (**Fig. 2**) for copper proteins with ReProDox.

The categories of features with best correlations tended to be those related to electrostatics and solvation. These include solvation features that describe Lazaridis-Karplus solvation energies both isotropic and anisotropic contributions for various distance cutoffs within 9 Å. The significant electrostatic features include calculations for Coulombic electrostatic potential as well as features describing the theoretical titration curve of surrounding residues. In contrast, other categories of features are more statistical. For example, eight of the nine significant "amino acid angle" features are Dunbrack rotamer energies of residues within 5 Å, indicating the use of some more common and some less common rotamers. In addition to further evaluating significant features that correlate with protein redox potentials found in ProtRedox, we expect these features can be used to train models [37, 53] for high throughput redox active protein design.

Coupling redox energetics to pH. Comparing protein redox potentials is challenging due to the numerous experimental conditions under which redox potentials are measured. Experimental pH is known to be a significant factor affecting redox processes accompanied by protonation/deprotonation events [62], which is commonly observed among Cu redox proteins [62-65]. To compare experimental redox potentials values are normalized to a reference pH (7.0) using the Nernst equation,

Eq. 1: $E_{\text{red}} \ = \ E + \ (59.16 \ mV * n * (pH \ - 7))$

where 59.16 mV is the Nernst constant relating pH to redox potential. E_{red} is the normalized reduction potential of each protein at pH 7 and E is the reduction potential measured at the literature pH. The variable n, assumed to be one, is the electron and proton *ratio involved* in the redox reaction, respectively.

For copper proteins with an azurin fold, we observed a correlation between pH and redox potential with a slope of -51 mV/pH unit (**Fig 3A**), near what is expected if the reactions followed Nernstian behavior (-59 mV/pH unit), assuming one electron transfer per reaction. Normalization removes the slope of this correlation (**Fig. 3B**). Experimental pH conditions showed the strongest positive Pearson coefficient with redox potential, above the computed factors from structure described earlier. However, there is a very large variance in observed potentials, clearly indicating that no one parameter can fully explain redox energetics.

Redox gradients and oxidoreductase evolution. Many of the ProtReDox entries are associated with an experimentally determined three-dimensional structure deposited in the PDB. This allowed us to map the redox energetics onto the SpAN – an existing network mapping electron transfer pathways in oxidoreductases of known structure [16, 17]. Nodes in the SpAN correspond to redox-cofactor binding protein microenvironments – termed modules. Edges reflect the existence spatial proximity of cofactor atoms in a pair of modules in one or many structures, providing a pathway for electron transfer. Cofactor edge-to-edge distances less than 14.0 Å were considered electron-transfer competent [7].

The full SpAN contains 133 modules [17]. We identified 18 modules with specified redox energetics (**Fig. 4, Table 1**). These modules formed a fully connected sub-graph within the SpAN with the exception of the heme-binding cytochrome-C fold module 140. Within this network, there is a clear downhill redox energetic gradient, starting from 4Fe-4S coordinating ferredoxin folds (module 85) with an average potential of -430 mV and ending with more oxidizing hemes (modules 1737 at +168 mV; 1746 at +70 mV), the molybdenum containing module 16 (+204 mV) and copper module 72 at +325 mV. One can envision electrons percolating from the center of this network to the periphery, driving redox-coupled reactions along a metabolic pathway.

Multiple features of the SpAN suggest its structure provides insight into the evolution of oxidoreductases in addition to their metabolic function. Network models of growing systems indicate that nodes with high centrality and connectivity are the first to arise [66-68]. In the ProtReDox annotated sub-graph of the SpAN, flavin module 7 and 4Fe-4S module 85 are reducing such that they are energetically matched with the early Earth redox environment. It is informative that the annotated modules form a connected sub-graph within

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

the SpAN. Most of these modules correspond both to isolated protein electron carriers [45] as well as being domains within larger oxidoreductases. Assignment of potentials is easier within an isolated domain versus a larger, multi-cofactor enzyme. Small, isolated modules would be useful building blocks of larger enzymes, forming multi-domain structures through duplication and diversification. Metal utilization for central versus peripheral modules is largely consistent with metal availability through geologic history [21, 69, 70], with early folds incorporating iron-containing cofactors and later folds binding molybdenum and copper.

CONCLUSIONS

Knowledge of redox energetics of oxidoreductases is critical to understanding metabolic function and evolution. ProtReDox is intended to be a valuable tool in this regard as we and others contribute to its growth. Currently, the size of ProtReDox limits the extent to which structure-based predictive models can be trained on redox energetics. However, with further experimental investigations and as high-quality models of protein structures become readily accessible, these models should improve. This would allow more complete mapping of data structures such as the SpAN, providing a greater understanding of the evolution of redox energetics in metabolism through time.

REFERENCES

1. Jelen, B.I., D. Giovannelli, and P.G. Falkowski, *The role of microbial electron transfer in the coevolution of the biosphere and geosphere.* Annual review of microbiology, 2016. **70** : p. 45-62.

2. Moore, E.K., B.I. Jelen, D. Giovannelli, H. Raanan, and P.G. Falkowski, *Metal availability and the expanding network of microbial metabolisms in the Archaean eon.* Nature Geoscience, 2017.**10** (9): p. 629-636.

3. Falkowski, P.G., T. Fenchel, and E.F. Delong, *The microbial engines that drive Earth's biogeochemical cycles.* science, 2008.**320** (5879): p. 1034-1039.

4. Marcus, R.A. and N. Sutin, *Electron transfers in chemistry and biology.* Biochimica et Biophysica Acta (BBA)-Reviews on Bioenergetics, 1985. **811** (3): p. 265-322.

5. Gray, H.B. and J.R. Winkler, *Electron tunneling through proteins.* Quarterly reviews of biophysics, 2003. **36** (3): p. 341-372.

6. Moser, C.C., J.M. Keske, K. Warncke, R.S. Farid, and P.L. Dutton,*Nature of biological electron transfer.* Nature, 1992.**355** (6363): p. 796-802.

7. Page, C.C., C.C. Moser, X. Chen, and P.L. Dutton, *Natural engineering principles of electron tunnelling in biological oxidation–reduction.* Nature, 1999. **402** (6757): p. 47-52.

8. Atkinson, J.T., I. Campbell, G.N. Bennett, and J.J. Silberg,*Cellular assays for ferredoxins: a strategy for understanding electron flow through protein carriers that link metabolic pathways.*Biochemistry, 2016. **55** (51): p. 7047-7064.

9. Rogers, N.K., G.R. Moore, and M.J. Sternberg, *Electrostatic interactions in globular proteins: calculation of the pH dependence of the redox potential of cytochrome C551.* Journal of molecular biology, 1985. **182** (4): p. 613-616.

10. Sternberg, M.J., F.R. Hayes, A.J. Russell, P.G. Thomas, and A.R. Fersht, *Prediction of electrostatic effects of engineering of protein charges.* Nature, 1987. **330** (6143): p. 86-88.

11. Karlin, S., Z.-Y. Zhu, and K.D. Karlin, *The extended environment of mononuclear metal centers in protein structures.*Proceedings of the National Academy of Sciences, 1997. **94** (26): p. 14225-14230.

12. Adman, E., K.D. Watenpaugh, and L.H. Jensen, *NH−S hydrogen bonds in Peptococcus aerogenes ferredoxin, Clostridium pasteurianum rubredoxin, and Chromatium high potential iron protein.*Proceedings of the National Academy of Sciences, 1975. **72** (12): p. 4854-4858.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

13. Dey, A., F.E. Jenney Jr, M.W. Adams, E. Babini, Y. Takahashi, K. Fukuyama, K.O. Hodgson, B. Hedman, and E.I. Solomon, *Solvent tuning of electrochemical potentials in the active sites of HiPIP versus ferredoxin.* Science, 2007. **318** (5855): p. 1464-1468.

14. Beratan, D.N., C. Liu, A. Migliore, N.F. Polizzi, S.S. Skourtis, P. Zhang, and Y. Zhang, *Charge transfer in dynamical biosystems, or the treachery of (static) images.* Accounts of chemical research, 2015.**48** (2): p. 474-481.

15. Teo, J.J. and R. Sarpeshkar, *The merging of biological and electronic circuits.* Iscience, 2020. **23** (11): p. 101688.

16. Raanan, H., D.H. Pike, E.K. Moore, P.G. Falkowski, and V. Nanda,*Modular origins of biological electron transfer chains.* Proc Natl Acad Sci U S A, 2018. **115** (6): p. 1280-1285.

17. Raanan, H., S. Poudel, D.H. Pike, V. Nanda, and P.G. Falkowski,*Small protein folds at the root of an ancient metabolic network.*Proc Natl Acad Sci U S A, 2020. **117** (13): p. 7193-7199.

18. Timm, J., D.H. Pike, J.A. Mancini, A.M. Tyryshkin, S. Poudel, J.A. Siess, P.M. Molinaro, J.J. McCann, K.M. Waldie, R.L. Koder, P.G. Falkowski, and V. Nanda, *Design of a minimal di-nickel hydrogenase peptide.* Sci Adv, 2023. **9** (10): p. eabq1990.

19. Baymann, F., E. Lebrun, M. Brugna, B. Schoepp-Cothenet, M.T. Giudici-Orticoni, and W. Nitschke, *The redox protein construction kit: pre-last universal common ancestor evolution of energy-conserving enzymes.* Philos Trans R Soc Lond B Biol Sci, 2003. **358** (1429): p. 267-74.

20. Hoarfrost, A., A. Aptekmann, G. Farfañuk, and Y. Bromberg,*Deep learning of a bacterial and archaeal universal language of life enables transfer learning and illuminates microbial dark matter.*Nature communications, 2022. **13** (1): p. 2606.

21. Falkowski, P.G., T. Fenchel, and E.F. Delong, *The microbial engines that drive Earth's biogeochemical cycles.* Science, 2008.**320** (5879): p. 1034-9.

22. Margulis, L. and J.E. Lovelock, *Biological modulation of the Earth's atmosphere.* Icarus, 1974. **21** (4): p. 471-489.

23. Eck, R.V. and M.O. Dayhoff, *Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences.*Science, 1966. **152** (3720): p. 363-366.

24. Romero Romero, M.L., A. Rabin, and D.S. Tawfik, *Functional Proteins from Short Peptides: Dayhoff's Hypothesis Turns 50.* Angew. Chem. Int. Ed., 2016. **55** : p. 15966-15971.

25. Todd, A.E., C.A. Orengo, and J.M. Thornton, *Evolution of function in protein superfamilies, from a structural perspective.*Journal of molecular biology, 2001. **307** (4): p. 1113-1143.

26. Kolodny, R., S. Nepomnyachiy, D.S. Tawfik, and N. Ben-Tal,*Bridging themes: short protein segments found in different architectures.* Molecular biology and evolution, 2021. **38** (6): p. 2191-2208.

27. Ferruz, N., F. Lobos, D. Lemm, S. Toledo-Patino, J.A. Farías-Rico, S. Schmidt, and B. Höcker, *Identification and analysis of natural building blocks for evolution-guided fragment-based protein design.*Journal of molecular biology, 2020. **432** (13): p. 3898-3914.

28. Abeln, S. and C.M. Deane, *Fold usage on genomes and protein fold evolution.* PROTEINS: Structure, Function, and Bioinformatics, 2005. **60** (4): p. 690-700.

29. Alva, V., M. Remmert, A. Biegert, A.N. Lupas, and J. Söding, *A galaxy of folds.* Protein Science, 2010. **19** (1): p. 124-130.

30. Alva, V., J. Soding, and A.N. Lupas, *A vocabulary of ancient peptides at the origin of folded proteins.* Elife, 2015. **4** : p. e09410.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

31. Dupont, C.L., A. Butcher, R.E. Valas, P.E. Bourne, and G. Caetano-Anolles, *History of biological metal utilization inferred through phylogenomic analysis of protein structures.* Proc Natl Acad Sci U S A, 2010. **107** (23): p. 10567-72.

32. Cheng, H., R.D. Schaeffer, Y. Liao, L.N. Kinch, J. Pei, S. Shi, B.-H. Kim, and N.V. Grishin, *ECOD: an evolutionary classification of protein domains.* PLoS computational biology, 2014. **10** (12): p. e1003926.

33. Bromberg, Y., A.A. Aptekmann, Y. Mahlich, L. Cook, S. Senn, M. Miller, V. Nanda, D.U. Ferreiro, and P.G. Falkowski, *Quantifying structural relationships of metal-binding sites suggests origins of biological electron transfer.* Sci Adv, 2022. **8** (2): p. eabj3984.

34. Lyons, T.W., C.T. Reinhard, and N.J. Planavsky, *The rise of oxygen in Earth's early ocean and atmosphere.* Nature, 2014.**506** (7488): p. 307-15.

35. Hummer, D.R., J.J. Golden, G. Hystad, R.T. Downs, A. Eleish, C. Liu, J. Ralph, S.M. Morrison, M.B. Meyer, and R.M. Hazen, *Evidence for the oxidation of Earth's crust from the evolution of manganese minerals.* Nature Communications, 2022. **13** (1): p. 960.

36. Chen, C.G., A.N. Nardi, A. Amadei, and M. D'Abramo,*Theoretical Modeling of Redox Potentials of Biomolecules.*Molecules, 2022. **27** (3): p. 1077.

37. Galuzzi, B.G., A. Mirarchi, E.L. Vigano, L. De Gioia, C. Damiani, and F. Arrigoni, *Machine Learning for Efficient Prediction of Protein Redox Potential: The Flavoproteins Case.* J Chem Inf Model, 2022. **62** (19): p. 4748-4759.

38. Mauk, A.G. and G.R. Moore, *Control of metalloprotein redox potentials: what does site-directed mutagenesis of hemoproteins tell us?* JBIC Journal of Biological Inorganic Chemistry, 1997. **2** : p. 119-125.

39. Iismaa, S.E., A. Vazquez, G. Jensen, P. Stephens, J. Butt, F. Armstrong, and B. Burgess, *Site-directed mutagenesis of Azotobacter vinelandii ferredoxin I. Changes in [4Fe-4S] cluster reduction potential and reactivity.* Journal of Biological Chemistry, 1991. **266** (32): p. 21563-21571.

40. Clark, K.M., Y. Yu, N.M. Marshall, N.A. Sieracki, M.J. Nilges, N.J. Blackburn, W.A. Van Der Donk, and Y. Lu, *Transforming a blue copper into a red copper protein: engineering cysteine and homocysteine into the axial position of azurin using site-directed mutagenesis and expressed protein ligation.* Journal of the American Chemical Society, 2010. **132** (29): p. 10093-10101.

41. Campbell, I.J., D. Kahanda, J.T. Atkinson, O.N. Sparks, J. Kim, C.-P. Tseng, R. Verduzco, G.N. Bennett, and J.J. Silberg,*Recombination of 2Fe-2S ferredoxins reveals differences in the inheritance of thermostability and midpoint potential.* ACS Synthetic Biology, 2020. **9** (12): p. 3245-3253.

42. Mutter, A.C., A.M. Tyryshkin, I.J. Campbell, S. Poudel, G.N. Bennett, J.J. Silberg, V. Nanda, and P.G. Falkowski, *De novo design of symmetric ferredoxins that shuttle electrons in vivo.*Proceedings of the National Academy of Sciences, 2019. **116** (29): p. 14557-14562.

43. Hosseinzadeh, P., N.M. Marshall, K.N. Chacón, Y. Yu, M.J. Nilges, S.Y. New, S.A. Tashkov, N.J. Blackburn, and Y. Lu, *Design of a single protein that spans the entire 2-V range of physiological redox potentials.* Proceedings of the National Academy of Sciences, 2016.**113** (2): p. 262-267.

44. Prabhulkar, S., H. Tian, X. Wang, J.J. Zhu, and C.Z. Li,*Engineered proteins: redox properties and their applications.*Antioxid Redox Signal, 2012. **17** (12): p. 1796-822.

45. Campbell, I.J., G.N. Bennett, and J.J. Silberg, *Evolutionary relationships between low potential ferredoxin and flavodoxin electron carriers.* Frontiers in energy research, 2019. **7** : p. 79.

46. Jumper, J., R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, and A. Potapenko,*Highly accurate protein structure prediction with AlphaFold.*Nature, 2021. **596** (7873): p. 583-589.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

47. Baek, M., F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G.R. Lee, J. Wang, Q. Cong, L.N. Kinch, and R.D. Schaeffer,*Accurate prediction of protein structures and interactions using a three-track neural network.* Science, 2021. **373** (6557): p. 871-876.

48. Lin, Z., H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, N. Smetanin, R. Verkuil, O. Kabeli, and Y. Shmueli, *Evolutionary-scale prediction of atomic-level protein structure with a language model.* Science, 2023.**379** (6637): p. 1123-1130.

49. Hekkelman, M.L., I. de Vries, R.P. Joosten, and A. Perrakis,*AlphaFill: enriching AlphaFold models with ligands and cofactors.*Nature Methods, 2023. **20** (2): p. 205-213.

50. Atkinson, J.T., I.J. Campbell, E.E. Thomas, S.C. Bonitatibus, S.J. Elliott, G.N. Bennett, and J.J. Silberg, *Metalloprotein switches that display chemical-dependent electron transfer in cells.* Nature chemical biology, 2019. **15** (2): p. 189-195.

51. Berman, H.M., J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne, *The Protein Data Bank.*Nucleic Acids Res, 2000. **28** (1): p. 235-42.

52. 03/14/2023]; Available from: https://firebase.google.com/docs/firestore.

53. Feehan, R., M.W. Franklin, and J.S.G. Slusky, *Machine learning differentiates enzymatic and non-enzymatic metals in proteins.* Nat Commun, 2021. **12** (1): p. 3712.

54. Virtanen, P., R. Gommers, T.E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S.J. van der Walt, M. Brett, J. Wilson, K.J. Millman, N. Mayorov, A.R.J. Nelson, E. Jones, R. Kern, E. Larson, C.J. Carey, I. Polat, Y. Feng, E.W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E.A. Quintero, C.R. Harris, A.M. Archibald, A.H. Ribeiro, F. Pedregosa, P. van Mulbregt, and C. SciPy, *SciPy 1.0: fundamental algorithms for scientific computing in Python.* Nat Methods, 2020. **17** (3): p. 261-272.

55. Saridakis, E., P. Giastas, G. Efthymiou, V. Thoma, J.M. Moulis, P. Kyritsis, and I.M. Mavridis, *Insight into the protein and solvent contributions to the reduction potentials of [4Fe-4S]2+/+ clusters: crystal structures of the Allochromatium vinosum ferredoxin variants C57A and V13G and the homologous Escherichia coli ferredoxin.* J Biol Inorg Chem, 2009. **14** (5): p. 783-99.

56. Ishikita, H., B. Loll, J. Biesiadka, W. Saenger, and E.W. Knapp,*Redox potentials of chlorophylls in the photosystem II reaction center.* Biochemistry, 2005. **44** (10): p. 4118-24.

57. Moore, G.R., G.W. Pettigrew, and N.K. Rogers, *Factors influencing redox potentials of electron transfer proteins.* Proc Natl Acad Sci U S A, 1986. **83** (14): p. 4998-9.

58. Essex, D.W., M. Li, R.D. Feinman, and A. Miller, *Platelet surface glutathione reductase-like activity.* Blood, 2004.**104** (5): p. 1383-5.

59. Di Rocco, G., G. Battistuzzi, M. Borsari, C.A. Bortolotti, A. Ranieri, and M. Sola, *The enthalpic and entropic terms of the reduction potential of metalloproteins: Determinants and interplay.*Coordination Chemistry Reviews, 2021. **445** .

60. Hughes, T.F. and R.A. Friesner, *Development of Accurate DFT Methods for Computing Redox Potentials of Transition Metal Complexes: Results for Model Complexes and Application to Cytochrome P450.* J Chem Theory Comput, 2012. **8** (2): p. 442-59.

61. Jafari, S., Y.A. Tavares Santos, J. Bergmann, M. Irani, and U. Ryde,*Benchmark Study of Redox Potential Calculations for Iron-Sulfur Clusters in Proteins.* Inorg Chem, 2022. **61** (16): p. 5991-6007.

62. Krishtalik, L.I., *pH-dependent redox potential: how to use it correctly in the activation energy analysis.* Biochim Biophys Acta, 2003. **1604** (1): p. 13-21.

63. Jeuken, L.C., R. Camba, F.A. Armstrong, and G.W. Canters, *The pH-dependent redox inactivation of amicyanin from Paracoccus versutus as studied by rapid protein-film voltammetry.* J Biol Inorg Chem, 2002.**7** (1-2): p. 94-100.

64. Battistuzzi, G., M. Borsari, G.W. Canters, E. de Waal, A. Leonardi, A. Ranieri, and M. Sola, *Thermodynamics of the acid transition in blue copper proteins.* Biochemistry, 2002. **41** (48): p. 14293-8.

65. Hulsker, R., A. Mery, E.A. Thomassen, A. Ranieri, M. Sola, M.P. Verbeet, T. Kohzuma, and M. Ubbink, *Protonation of a histidine copper ligand in fern plastocyanin.* J Am Chem Soc, 2007.**129** (14): p. 4423-9.

66. Fell, D.A. and A. Wagner, *The small world of metabolism.*Nature biotechnology, 2000. **18** (11): p. 1121-1122.

67. Wagner, A., *The yeast protein interaction network evolves rapidly and contains few redundant duplicate genes.* Molecular biology and evolution, 2001. **18** (7): p. 1283-1292.

68. Barabási, A.-L., *The new science of networks.* Cambridge: Perseus, 2002.

69. Williams, R.J.P., *The bakerian lecture, 1981 natural selection of the chemical elements.* Proceedings of the Royal Society of London. Series B. Biological Sciences, 1981. **213** (1193): p. 361-397.

70. Anbar, A.D., *Elements and evolution.* Science, 2008.**322** (5907): p. 1481-1483.

71. Senn, S., V. Nanda, P. Falkowski, and Y. Bromberg,*Function-based assessment of structural similarity measurements using metal co-factor orientation.* Proteins, 2014. **82** (4): p. 648-56.

FIGURES AND TABLES:

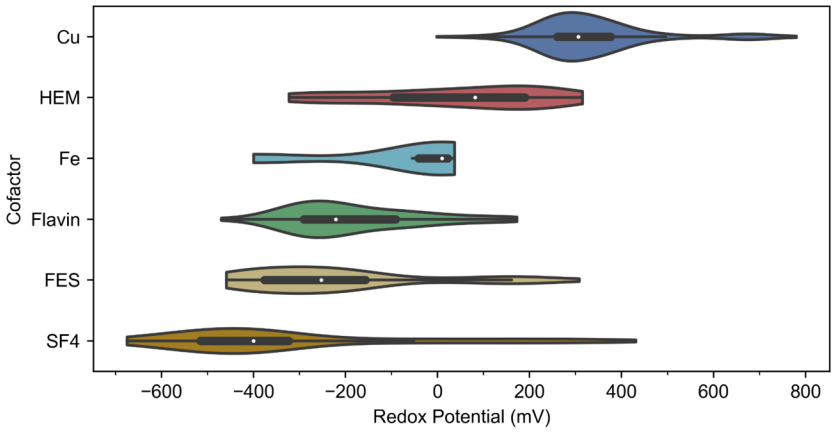

Figure 1. Distribution of redox potentials for the most abundant cofactor types found in ProtReDox. Sorted according to mean redox potential cofactors and displayed vertically from most reducing to most oxidizing along with corresponding atomic structures. SF4 (4Fe4S; 82) -329 ± 268 mV; FES (2Fe2S; 64) -220 ± 199 mV; Flavin (114) -183 +/- 141 mV; Fe (10) -72 ± 173 mV; HEM (42) 43 ± 190 mV; Cu (147) 333 ± 129 mV. Count of each cofactor type is within parenthesis.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

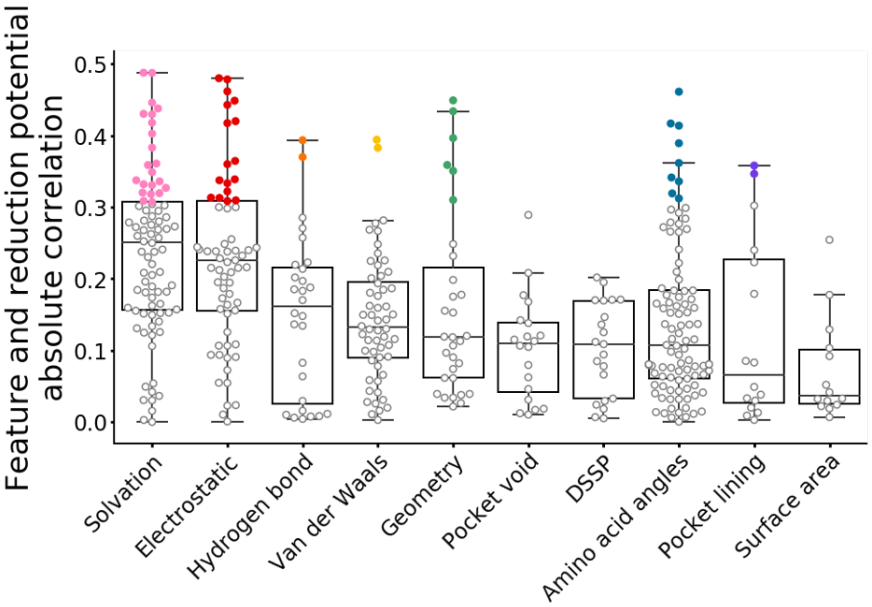

Figure 2. Correlation of reduction potential and physicochemical properties: Groups of physicochemical properties of metalloenzymes that can be measured from protein structure are shown along the x-axis. Each circle is the absolute value of the Pearson correlation coefficient for reduction potentials. Colored circles represent p-value [?] 0.05 for the correlation and empty circles represent p-value [?] 0.05. Horizontal box lines, from top to bottom, represent the upper quartile, median, and lower quartile correlation values for the respective property category. The whiskers display the range of correlation values for the respective property category, except for outliers, which are greater than 150% of the interquartile range from the box.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

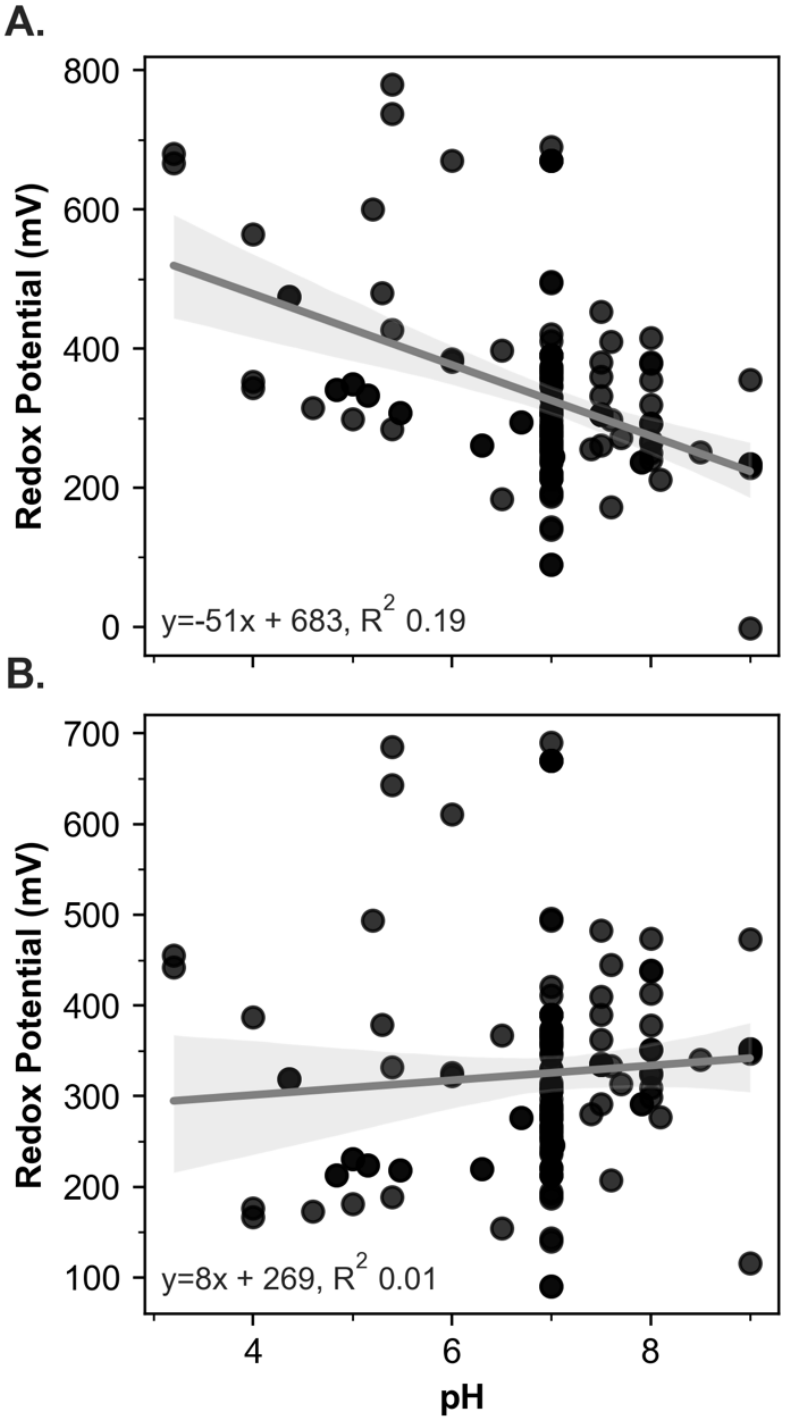

Figure 3. Cu cofactor type correlation of pH and redox potential A.) experimental results and B.) normalized using the Nernst equation (eq. 1) Figures include linear regression with a 95% confidence interval and the equation of best fit.

Posted on 18 May 2023 | The copyright holder is the author/funder. All rights reserved. No reuse without permission. | https://doi.org/10.22541/au.168439504.41026088/v1 | This a preprint and has not been peer reviewed. Data may be preliminary.

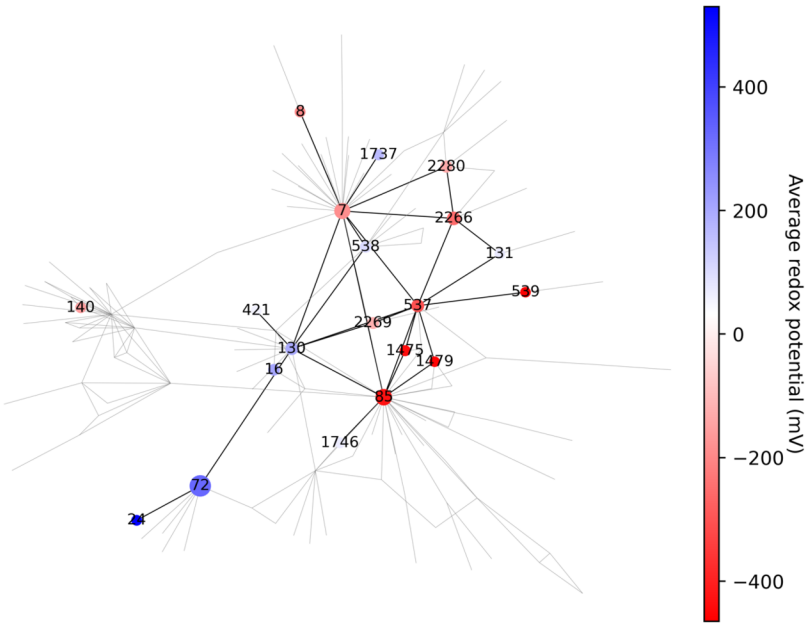

Figure 4. Network of redox active modules where the nodes are modules that share common and structural similarity [16, 17, 71], and the edges are connections between modules that are within the same protein and are capable of electron transfer between each module. Node colors represent a gradient of average redox potentials for each module. Gray edges are connections between modules without associated redox potentials within ProtReDox, and black edges are between modules with reported potentials. Node sizes are porportional to the number of redox potentials for each module. Nodes with greater than one measurement are shown. Size of the node is proportional to the number of experimental redox potentials for each node. Labeled nodes represent the redox-active cluster type: SF4 (85, 539, 1475, 1479), FES (537, 538, 539), Flavin (7, 8, 2266, 2269, 2280), heme (130, 131, 140, 421, 1737, 1746), Mo (16), and Cu (24, 72). See Table 1 for values.

Table 1: Redox energetics associated with SpAN modules. Module numbers defined in [17]. Potentials are the arithmetic mean of values in the database.

Hosted file

`image5.emf` available at https://authorea.com/users/619752/articles/644090-the-energetics-and-evolution-of-oxidoreductases-in-deep-time

Data Availability Statement : The data that support the findings of this study are openly available on the ProtReDox site: https://protein-redox-potential.web.app

Conflict of Interest: Authors do not declare any conflict of interests