life

MDPI

Article

Quantifying Mineral-Ligand Structural Similarities: Bridging the Geological World of Minerals with the Biological World of Enzymes

Daniel Zhao [1,2] **and Stuart Bartlett** [1,3,]* **and Yuk L. Yung** [1,4]

[1]   Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA 91125, USA; dxczhao20@gmail.com (D.Z.); yly@gps.caltech.edu (Y.L.Y.)
[2]   Department of Mathematics, Harvard University, Massachusetts Hall, Cambridge, MA 02138, USA
[3]   Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan
[4]   NASA Jet Propulsion Laboratory, Oak Grove Dr, La Cañada Flintridge, CA 91011, USA
*   Correspondence: sjbart@caltech.edu

check for updates

Abstract: Metal compounds abundant on Early Earth are thought to play an important role in the origins of life. Certain iron-sulfur minerals for example, are proposed to have served as primitive metalloenzyme cofactors due to their ability to catalyze organic synthesis processes and facilitate electron transfer reactions. An inherent difficulty with studying the catalytic potential of many metal compounds is the wide range of data and parameters to consider when searching for individual minerals and ligands of interest. Detecting mineral-ligand pairs that are structurally analogous enables more relevant selections of data to study, since structural affinity is a key indicator of comparable catalytic function. However, current structure-oriented approaches tend to be subjective and localized, and do not quantify observations or compare them with other potential targets. Here, we present a mathematical approach that compares structural similarities between various minerals and ligands using molecular similarity metrics. We use an iterative substructure search in the crystal lattice, paired with benchmark structural similarity methods. This structural comparison may be considered as a first stage in a more advanced analysis tool that will include a range of chemical and physical factors when computing mineral-ligand similarity. This approach will seek relationships between the mineral and enzyme worlds, with applications to the origins of life, ecology, catalysis, and astrobiology.

1. Introduction

Life and its environment, the biosphere and geosphere, have always been inextricably linked. Despite the epistemological separation between organisms and their environs, an understanding of life is fundamentally incomplete without an understanding of the non-living aspects of its surroundings, with all the myriad interactions entailed therewith. The vast majority of known life resides within or upon Earth's geosphere, which provides physical structure, chemical free energy, and material substrates, without which life as we know it would be impossible.

There are myriad feedbacks between life and the geosphere. While we often view the physical environment as providing evolutionary challenges and constraints upon which organisms can persist, life's existence and chemical byproducts also shape which minerals, rocks and gases are most readily produced. Hence it is impossible to study life out of its environmental context, and any analysis of geospheric evolution is futile if life's influence is overlooked. Enzymes are arguably the most

sophisticated and versatile inventions of biological learning and evolution, which have enabled the colonization of a vast range of niches, and can facilitate nanoscale chemical and physical processes that are far beyond the reach of human engineering. Enzymes are primarily composed of amino acids, a relatively small but apparently universal set of 20 organic compounds. However, enzymes rely upon inorganic components for their function, particularly active site catalytic power. These inorganic molecules or clusters, shape and are shaped by the geosphere and abiotic environment. Their raw components must be sourced and sequestered from the environment, but their catalytic action is often aimed at chemically altering or metabolizing other compounds from the environment. The byproducts of these processes in turn feed back on the chemical and physical conditions of the environment.

A prime example is chemolithoautotrophic organisms, which exploit reduced mineral compounds as a source of free energy for metabolism. However, these minerals are not just consumed by such organisms, their own chemical evolution is also strongly affected. In the case of ferrous iron [Fe(II)]-oxidizing bacteria, chemical weathering has been shown to be enhanced due to the presence of such bacteria. In fact, it was found that "chemolithotrophic Fe(II)-oxidizing bacteria are likely contributors in the transformation of rock to regolith" [1,2].

In addition to the coevolution of life and environment over the last 3.5 billion years, the abiotic conditions at the time of the origins of life played a fundamental role in fostering and constraining those crucial events. Mineral compounds have been highlighted as candidate materials for catalysis [3–11], information storage [12–14], and compartmentalization, concentration and polymerization [5,11,15–17], among others [4,5,13,18]. Focusing on catalysis, many authors have highlighted the compelling structural similarities between key inorganic enzyme clusters, and minerals [7,8,19–26]. Iron-sulfur clusters are employed across a vast range of essential biochemical functions [27,28]. For this reason, significant efforts have explored the potential connections between minerals, and the enzyme families of ferredoxins, nitrogenases, NADH dehydrogenase, and CO-dehydrogenase, among others.

While these structural analogies are highly compelling, there is a relative dearth of systematic, quantitative comparisons between the structures of minerals and metalloenzyme clusters. Understanding the origins of life requires a simultaneous application of constraints from the ancient geological world and the known biological world. As such, it is essential to systematically assess how biological machinery and geological compounds are related. Structurally, we must quantify the similarities between inorganic enzyme active sites and minerals, to ascertain whether biochemical catalysts are truly derived from mineral analogues. Functionally, we must establish whether minerals can function as catalysts, and for those that are weakly catalytic, predict the modifications (by amino acid addition or otherwise) required to achieve catalytic efficacy.

The present work was motivated by these questions. We aimed to lay an analytical foundation for the large scale comparison of minerals and metalloenzyme ligands. In this context, we refer to "ligands" as the active centers of metalloenzymes often involved in catalytic activity. Given the unprecedented recent advances in chemoinformatics, rational ligand design and big data [29,30], methods for mineral-enzyme comparison applied to large datasets have vast potential to extract hitherto unknown connections between the geological and biological worlds, predict ancient catalytic compounds relevant to life's origins, and could even help constrain life-environment relations on other planets. It could also reveal mineral compounds with functional potential that had previously not been considered.

In this paper we report on a first set of mineral-enzyme structural comparisons using a mature chemoinformatic toolbox known as RDKit [31,32]. We sourced mineral structures from the Crystallography Open Database [33], and enzyme data from the Protein Data Bank [34,35]. In this initial phase, the focus was Iron-Sulfur and Nickel-Iron-Sulfur ligands, since these play such a crucial role in biochemistry and possibly its origins. Comparisons with a range of iron-sulfur mineral structures confirmed many previous structural analogies but also highlighted some negative results. We propose the presented methodology as a powerful technique that can now be developed to include chemical

characteristics, and applied to large mineral and enzyme databases to build networks of relationships between the geological and biological worlds.

Section 2 presents the main aspects of our methodology, including mineral structure extraction and similarity index computation. Section 3 presents our similarity index results and compares them with previous speculations on mineral-enzyme structural analogies. We conclude in Section 4.

2. Methodology

This work aims to establish the groundwork of a method to objectively compare mineral and ligand structures. A unique challenge with making mathematical comparisons between the structures of mineral-ligand pairs, as opposed to traditional molecular similarity comparisons, is the arrangement of atoms in the mineral crystal lattice, which does not define a single molecule for comparison. Unit cells are often used as a basis for visual comparison between minerals and ligands that share alike structural motifs. However, establishing a mathematical index of similarity requires a more methodical approach, as the scope and size of the unit cell is variable and oftentimes incongruous. The aim of our method is to overcome this problem by extracting a molecular substructure of similar size to the target ligand from the crystal lattice, in order to make an analogous comparison with the ligand.

Using crystallographic data available on the Open Crystallography Database, we generated the structural coordinates of supercells of various minerals. To estimate the size of the ligands, we constructed an oriented bounding box around the target ligand using its structural coordinates found on the Protein Data Bank. The dimensions of the bounding box were normalized to the average bond length of the ligand and resized to the coordinate space of the mineral lattice. We iterated the bounding box through several increments of axis rotations and translations across the lattice, selecting all atoms and bonds within the bounding box as the molecular substructure for that iteration. Once a substructure was selected, we calculated its molecular similarity index with the ligand.

Over several transformation iterations, the index with the largest value and its respective substructure was retained as the final similarity index, as depicted in Figure 1.
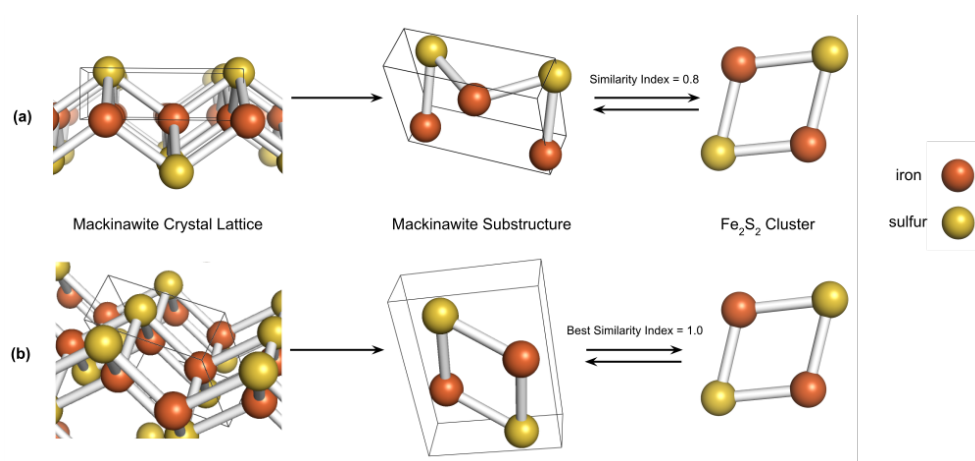

Figure 1. Example of substructure search in the crystal lattice of mackinawite, using the Fe_2S_2 cluster as the target ligand. Shown in (**a**), an incongruent substructure which leads to a sub-optimal similarity index, while the substructure encompassed in (**b**) is chemically identical to the ligand.

The similarity indices were calculated using RDKit-generated molecular fingerprints [36] and the Sorensen-Dice similarity coefficient. RDKit fingerprints are topological representations of molecular structure generated by hashing all subgraphs within a specified range of path sizes into bit IDs. Many fingerprinting algorithms exist, but we chose RDKit's implementation for its versatility in dealing with non-aromatic molecules and its relatively low computation time, since many nested iterations of substructure search require significantly longer running times for most 3D fingerprinting methods.

For the same reason, Dice similarity [37,38] was chosen as an effective and quick metric to quantify similarity. The usefulness of Dice similarity and comparable mathematical metrics have been documented in several studies from a wide range of fields, including chemoinformatics [39,40] and medicine [41,42]. Deferring to standard chemoinformatic practices [43,44], we used a similarity index of 0.8 as a rough threshold for inferring two molecules were structurally correlated, while we considered an index of 0.65–0.8 as suggestive of a potential correlation. However, as explored in Baldi and Nasr [45], the usefulness of such a threshold depends largely on the dataset. Especially in this context, the "significant" range of similarity is defined essentially for qualitative use since mineral compositions and relative ligand sizes can vary greatly.

3. Results

Most studies on catalysis at the origins of life have focused on a select group of ligands involved with key metabolic pathways, and their potential mineral predecessors. For example, Russell and Hall [25] proposed that nickel-substituted greigite served as a catalytic precursor to the active C-cluster of acetyl-CoA synthase (ACS/CODH) and potentially other metalloenzyme cofactors, noting its structural similarity as suggestive of similar functionality. In Wächtershäuser's "iron-sulfur world" hypothesis, ferrous surfaces near hydrothermal vents play a pivotal role in primitive metabolic chemistry, specifically in electron transport and catalysis [10]. Notably, it was also proposed that the first iron-sulfur clusters arose from pyrite-like surfaces and were partially retained as ligands in ancient ferredoxins, suggesting a structural relatedness between the two. We now know that pyrite has very limited catalytic ability, but in the last two decades interest has shifted toward iron sulfide minerals [7,19] and iron hydroxides [3,20,46].

Given the aforementioned importance of iron sulfide clusters, significant work has aimed at synthetically bridging the gap between iron sulfide minerals and simple enzymes, both in the origins field [21,26,47–49] and chemical engineering [50,51]. Ferredoxins are considered one of the most ancient protein structures in the biosphere [52–54], and make use of the $[Fe_2S_2]$, $[Fe_3S_4]$, and $[Fe_4S_4]$ clusters. In the following, all three of these structures were compared with key minerals that have been propounded as prebiotic catalysts. Using our described method, we generated a similarity index matrix of relevant ligands and metal-sulfide minerals, presented in Figure 2.

As expected, mackinawite and greigite share many structural similarities with the iron-sulfur clusters. This structural affinity has been documented in multiple studies [23,25], and Russell and Hall [24] proposed that a protoferredoxin cluster could have arisen in an intermediate oxidation state between mackinawite and greigite. The alikeness of both minerals towards the $[Fe_2S_2]$ and $[Fe_4S_4]$ clusters supports this hypothesis. The similarity indices of unity (perfect match) imply structural indistinguishability, but notably not functional equivalence. While some structural resemblance is expected in mineral-ligand pairs of similar function, the converse is not true, as molecular structure alone may not predict chemical bonding behavior.

The less well-studied minerals marcasite and troilite demonstrate an unexpectedly high structural affinity with iron-sulfur clusters as well, providing a new avenue to study the poorly-understood bridge between minerals and metalloenzyme clusters. Though less abundant than pyrite, these minerals still compose a significant portion of metal sulfides in locality counts [7]. This example highlights the utility of our method to reveal less studied minerals that could be more important to early metabolic catalysis than previously thought.

Another noticeable result is that most ligands have relatively low similarity indices with pyrite and chalcopyrite compared to the other iron-sulfur minerals, supporting the notion that pyrite surfaces are functionally disparate from iron-containing clusters, and had suboptimal catalytic properties for early metabolic chemistry. One explanation for this is provided by Russell and Hall [24], who argued that pyrite's tetragonal crystal structure inhibited nucleation and reduction–common features of iron-sulfide clusters in metalloenzymes. The preferred hypothesis of Krupp [55] involves the oxidation of mackinawite to greigite facilitated through electron conduction, motivated in part by their closely

related crystal structure. Unlike the catalytically inert pyrite, the oxidized greigite can participate further in protometabolism.
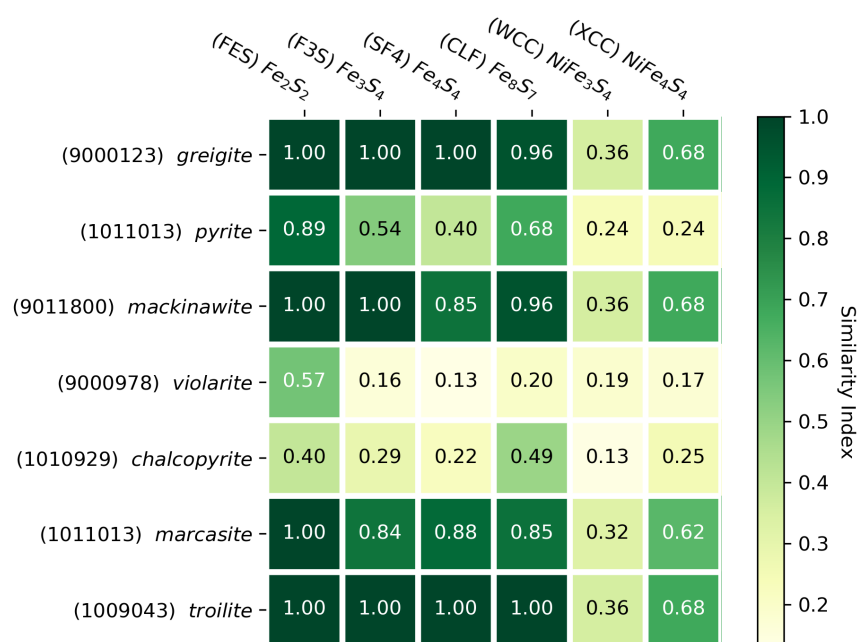

	(FES) Fe_2S_2	(F3S) Fe_3S_4	(SF4) Fe_4S_4	(CLF) Fe_8S_7	(WCC) $NiFe_3S_4$	(XCC) $NiFe_4S_4$
(9000123) *greigite*	1.00	1.00	1.00	0.96	0.36	0.68
(1011013) *pyrite*	0.89	0.54	0.40	0.68	0.24	0.24
(9011800) *mackinawite*	1.00	1.00	0.85	0.96	0.36	0.68
(9000978) *violarite*	0.57	0.16	0.13	0.20	0.19	0.17
(1010929) *chalcopyrite*	0.40	0.29	0.22	0.49	0.13	0.25
(1011013) *marcasite*	1.00	0.84	0.88	0.85	0.32	0.62
(1009043) *troilite*	1.00	1.00	1.00	1.00	0.36	0.68

Figure 2. Heatmap of similarity coefficients between selected minerals and ligands with associated COD and PDB ligand identifiers.

It is noted that nickel-containing ligands and minerals show significantly lower similarities relative to their pure iron-sulfur counterparts. This is likely partially an artefact of the underlying similarity method. Unlike many network-based 3D fingerprinting methods, the RDKit fingerprint is based on atom and bond arrangements and does not consider the spatial geometry of the molecule. It is therefore especially sensitive to differences in atom types, which explains the large drop in similarity index when iron is substituted with nickel. With respect to its catalytic properties, as shown by Cody et al. [19] and Huber et al. [56], these nickel-containing minerals exhibit significantly different behaviors in organic chemistry (ie. enhancement of the Koch reaction) compared to iron-sulfur minerals, and are perhaps involved in different metabolic pathways. The substitution of a non-iron transition metal atom could drastically alter its catalytic potential, so the observed drop in similarity due to nickel-iron mismatches is not necessarily detrimental from a functional perspective. Regardless, we acknowledge this potential bias of the similarity algorithm and stress the importance of examining chemical and electronic properties of the minerals in conjunction with molecular structure.

To re-examine past structural comparisons through a computational lens, we applied our method to a selection of mineral-ligand pairs that were compared visually in previous literature. These are shown in Figure 3.

The $[Fe_4S_4]$ cluster exhibits both high visual and molecular similarity with the cuboid structure of greigite. Likewise, as shown in (c), the visual comparison made in Nitschke et al. [8] between the CODH/ACS C-cluster and nickel-substituted greigite is also supported by its high similarity index. Both these factors raise the intriguing possibility that there could exist an intimate relationship between the two that is functional and evolutionary in nature.

Adversely, the structural disparity between the $[Fe_7MoS_9]$ cluster and the greigite substructure extracted in (e) seems to indicate a much weaker or indirect potential relationship, though the cuboid motif appears to be shared [25,57]. Other comparisons, such as those shown in (b) and (d), are visually compelling but have relatively low indices, suggesting that the actual structural affinity of the pair may be less than the apparent aesthetic match. This demonstrates the advantage of a quantitative metric like "Similarity Index" over intuition.
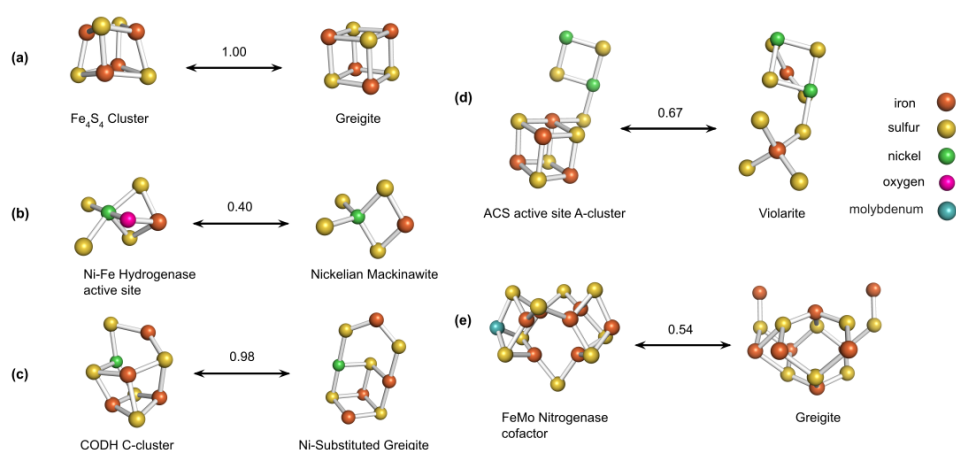
Figure 3. Structural comparisons using similarity index (number above arrows) of (**a**) $[Fe_4S_4]$ and greigite, (**b**–**d**) nickel-containing ligands and minerals from Nitschke et al. [8], (**e**) $[Fe_7MoS_9]$ nitrogenase cofactor and greigite [25].

4. Conclusions

Our study presents a foundational, mathematical approach to comparing mineral-ligand pairs relevant to the origins of life and to the large scale structural comparison of contemporary ligands and minerals. Visual comparisons are useful but limited in scope, and are not always reliable. Screening of large geological data for structural similarities kindles opportunities to identify new, potentially important minerals, and diversifies the existing pool of metal sulfides to investigate in a guided manner. A significant majority of minerals have yet to be studied–not because of their irrelevance, but rather due to the difficulty of choosing the right minerals and ligands to investigate. By harnessing cutting-edge chemoinformatics and drug discovery research in origins of life studies, we provide a powerful angle for exploring the myriad connections between minerals and enzymes.

However, the present study is a starting point with the aim of building a more comprehensive comparison tool that will include chemical factors such as redox potential, charge, magnetic state, kinetic properties and catalytic efficacy. Structural similarity is an important but clearly insufficient metric given the complex array of genotype-structure mappings found in the biological repertoire (structurally similar ligands that are phylogenetically distant and vice versa). Ultimately, the presented methodology may serve as a coarse-grained filter phase, feeding into further steps of chemical and physical comparison. This more comprehensive tool will be able to identify mineral-ligand pairs that are structurally similar but in fact share little to no chemical or catalytic similarity (which may be the case for some of the structurally similar pairs highlighted in this work).

Astrobiology as a discipline has made remarkable advancements in many diverse branches of thought, including bottom-up laboratory experiments on mineral catalysis, top-down phylogenetic analyses, and topological studies of enzyme structure. We believe a data-driven, informatic approach to the field may help unite these many fragmented ideas and, in the mineral context, compartmentalize the geological world into a new system of groups as a basis for study. Classifying minerals based on their inherent catalytic abilities, functional similarity with biomolecules, or association with peptides would enable a systematic means of expanding the field and closing the gap between bottom-up geological insights, and top-down biological and phylogenetic studies. The approach presented herein could readily be expanded to fulfill this grand objective.

References

1. Liu, J.; Chakraborty, S.; Hosseinzadeh, P.; Yu, Y.; Tian, S.; Petrik, I.; Bhagi, A.; Lu, Y. Metalloproteins containing cytochrome, iron–sulfur, or copper redox centers. *Chem. Rev.* **2014**, *114*, 4366–4469. [CrossRef] [PubMed]

2. Napieralski, S.A.; Buss, H.L.; Brantley, S.L.; Lee, S.; Xu, H.; Roden, E.E. Microbial chemolithotrophy mediates oxidative weathering of granitic bedrock. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 26394–26401. [CrossRef] [PubMed]

3. Barge, L.M.; Flores, E.; Baum, M.M.; VanderVelde, D.G.; Russell, M.J. Redox and pH gradients drive amino acid synthesis in iron oxyhydroxide mineral systems. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 4828–4833. [CrossRef]

4. Cleaves, H.J., II; Scott, A.M.; Hill, F.C.; Leszczynski, J.; Sahai, N.; Hazen, R. Mineral–organic interfacial processes: Potential roles in the origins of life. *Chem. Soc. Rev.* **2012**, *41*, 5502–5525. [CrossRef]

5. Hazen, R.M.; Sverjensky, D.A. Mineral surfaces, geochemical complexities, and the origins of life. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a002162. [CrossRef] [PubMed]

6. Huber, C.; Eisenreich, W.; Hecht, S.; Wächtershäuser, G. A possible primordial peptide cycle. *Science* **2003**, *301*, 938–940. [CrossRef]

7. Li, Y.; Kitadai, N.; Nakamura, R. Chemical diversity of metal sulfide minerals and its implications for the origin of life. *Life* **2018**, *8*, 46. [CrossRef]

8. Nitschke, W.; McGlynn, S.E.; Milner-White, E.J.; Russell, M.J. On the antiquity of metalloenzymes and their substrates in bioenergetics. *Biochim. Biophys. Acta BBA Bioenerg.* **2013**, *1827*, 871–881. [CrossRef]

9. Vincent, L.; Berg, M.; Krismer, M.; Saghafi, S.T.; Cosby, J.; Sankari, T.; Vetsigian, K.; Cleaves, H.J.; Baum, D.A. Chemical ecosystem selection on mineral surfaces reveals long-term dynamics consistent with the spontaneous emergence of mutual catalysis. *Life* **2019**, *9*, 80. [CrossRef] [PubMed]

10. Wächtershäuser, G. Before enzymes and templates: Theory of surface metabolism. *Microbiol. Rev.* **1988**, *52*, 452. [CrossRef] [PubMed]

11. Wang, Q.; Barge, L.M.; Steinbock, O. Microfluidic production of pyrophosphate catalyzed by mineral membranes with steep pH gradients. *Chem. Eur. J.* **2019**, *25*, 4732–4739. [CrossRef] [PubMed]

12. Arrhenius, G.O. Crystals and life. *Helv. Chim. Acta* **2003**, *86*, 1569–1586. [CrossRef]

13. Cairns-Smith, A.G. Chemistry and the missing era of evolution. *Chem. Eur. J.* **2008**, *14*, 3830–3839. [CrossRef] [PubMed]

14. Greenwell, H.C.; Coveney, P.V. Layered double hydroxide minerals as possible prebiotic information storage and transfer compounds. *Orig. Life Evol. Biosph.* **2006**, *36*, 13–37. [CrossRef]

15. Erastova, V.; Degiacomi, M.T.; Fraser, D.G.; Greenwell, H.C. Mineral surface chemistry control for origin of prebiotic peptides. *Nat. Commun.* **2017**, *8*, 1–9. [CrossRef]

16. Ferris, J.P.; Hill, A.R.; Liu, R.; Orgel, L.E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **1996**, *381*, 59–61. [CrossRef]

17. Hanczyc, M.M.; Mansy, S.S.; Szostak, J.W. Mineral surface directed membrane assembly. *Orig. Life Evol. Biosph.* **2007**, *37*, 67–82. [CrossRef]

18. Mizuuchi, R.; Blokhuis, A.; Vincent, L.; Nghe, P.; Lehman, N.; Baum, D. Mineral surfaces select for longer RNA molecules. *Chem. Commun.* **2019**, *55*, 2090–2093. [CrossRef]

19. Cody, G.; Boctor, N.; Brandes, J.E.A.; Filley, T.; Hazen, R.; Yoder, H., Jr. Assaying the catalytic potential of transition metal sulfides for abiotic carbon fixation. *Geochim. Cosmochim. Acta* **2004**, *68*, 2185–2196. [CrossRef]

20. Duval, S.; Baymann, F.; Schoepp-Cothenet, B.; Trolard, F.; Bourrie, G.; Grauby, O.; Branscomb, E.; Russell, M.J.; Nitschke, W. Fougerite: The not so simple progenitor of the first cells. *Interface Focus* **2019**, *9*, 20190063. [CrossRef]

21. Kim, J.D.; Rodriguez-Granillo, A.; Case, D.A.; Nanda, V.; Falkowski, P.G. Energetic selection of topology in ferredoxins. *PLoS Comput. Biol.* **2012**, *8*, e1002463. [CrossRef] [PubMed]

22. Milner-White, E.J. Protein three-dimensional structures at the origin of life. *Interface Focus* **2019**, *9*, 20190057. [CrossRef] [PubMed]

23. Roldan, A.; Hollingsworth, N.; Roffey, A.; Islam, H.U.; Goodall, J.; Catlow, C.; Darr, J.; Bras, W.; Sankar, G.; Holt, K.; et al. Bio-inspired CO_2 conversion by iron sulfide catalysts under sustainable conditions. *Chem. Commun.* **2015**, *51*, 7501–7504. [CrossRef] [PubMed]

24. Russell, M.J.; Hall, A. The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc.* **1997**, *154*, 377–402. [CrossRef] [PubMed]

25. Russell, M.J.; Hall, A.J. The onset and early evolution of life. *Mem. Geol. Soc. Am.* **2006**, *198*, 1.

26. Scintilla, S.; Bonfio, C.; Belmonte, L.; Forlin, M.; Rossetto, D.; Li, J.; Cowan, J.A.; Galliani, A.; Arnesano, F.; Assfalg, M.; et al. Duplications of an iron–Sulphur tripeptide leads to the formation of a protoferredoxin. *Chem. Commun.* **2016**, *52*, 13456–13459. [CrossRef] [PubMed]

27. Johnson, M.K.; Smith, A.D. Iron–Sulfur proteins. In *Encyclopedia of Inorganic and Bioinorganic Chemistry*; Wiley Online Library: Hoboken, NJ, USA, 2011.

28. Rees, D.C.; Howard, J.B. The interface Between the Biological and Inorganic Worlds: Iron-sulfer Metalloclusters. *Science* **2003**, *300*, 929–931. [CrossRef]

29. Hazen, R. Deep Carbon through Deep Time. In *Deep Carbon*; Cambridge University Press: Cambridge, UK. 2019; pp. 620–652.

30. Hazen, R.M.; Downs, R.T.; Eleish, A.; Fox, P.; Gagné, O.C.; Golden, J.J.; Grew, E.S.; Hummer, D.R.; Hystad, G.; Krivovichev, S.V.; et al. Data-driven discovery in mineralogy: Recent advances in data resources, analysis, and visualization. *Engineering* **2019**, *5*, 397–405. [CrossRef]

31. Landrum, G. Rdkit documentation. *Release* **2013**, *1*, 1–79.

32. Lovrić, M.; Molero, J.M.; Kern, R. PySpark and RDKit: Moving towards Big Data in Cheminformatics. *Mol. Inform.* **2019**, *38*, 1800082. [CrossRef]

33. Gražulis, S.; Chateigner, D.; Downs, R.T.; Yokochi, A.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database–an open-access collection of crystal structures. *J. Appl. Crystallogr.* **2009**, *42*, 726–729. [CrossRef] [PubMed]

34. Bernstein, F.C.; Koetzle, T.F.; Williams, G.J.; Meyer, E.F., Jr.; Brice, M.D.; Rodgers, J.R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *Eur. J. Biochem.* **1977**, *80*, 319–324. [CrossRef] [PubMed]

35. Rose, P.W.; Prlić, A.; Altunkaya, A.; Bi, C.; Bradley, A.R.; Christie, C.H.; Costanzo, L.D.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* **2016**, *D1*, D271–D281.

36. RDKit: Open-Source Cheminformatics. Available online: http://www.rdkit.org (accessed on 11 April 2013).

37. Dice, L.V. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [CrossRef]

38. Søorenson, T.J. A method of stabilizing groups of equivalent amplitude in plant sociology based on the similarity of species content and its application to analyses of the vegetation on Danish commons. *R. Dan. Acad. Sci. Lett.* **1948**, *5*, 134.

39. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Chemoinform.* **2015**, *7*. [CrossRef]

40. Willett, P.; Barnard, J.M.; Downs, G.M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 983–996, doi:10.1021/ci9800211. [CrossRef]

41. Hull, R.D.; Fluder, E.M.; Singh, S.B.; Nachbar, R.B.; Kearsley, S.K.; Sheridan, R.P. Chemical Similarity Searches Using Latent Semantic Structural Indexing (LaSSI) and Comparison to TOPOSIM. *J. Med. Chem.* **2001**, *44*, 1185–1191. [CrossRef]

42. Maggiora, G.; Vogt, M.; Stumpfe, D.; Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **2013**, *57*, 3186–3204. [CrossRef]

43. Willett, P. Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* **2003**, *31 Pt 3*, 603–606. [CrossRef]

44. Zahoránszky-Kõhalmi, G.; Bologa, C.G.; Oprea, T.I. Impact of similarity threshold on the topology of molecular similarity networks and clustering outcomes. *J. Chemoinform.* **2016**, *8*. [CrossRef]

45. Baldi, P.; Nasr, R. When is Chemical Similarity Significant? The Statistical Distribution of Chemical Similarity Scores and Its Extreme Values. *J. Chem. Inf. Model.* **2010**, *50*, 1205–1222. [CrossRef] [PubMed]

46. Russell, M.J. Green rust: The simple organizing seed of all life? *Life* **2018**, *8*, 35. [CrossRef] [PubMed]

47. Kim, J.D.; Pike, D.H.; Tyryshkin, A.M.; Swapna, G.V.T.; Raanan, H.; Montelione, G.T.; Nanda, V.; Falkowski, P.G. Minimal Heterochiral de Novo Designed 4Fe4S Binding Peptide Capable of Robust Electron Transfer. *J. Am. Chem. Soc.* **2018**, *140*, 11210–11213. [CrossRef]

48. Moore, E.K.; Jelen, B.I.; Giovannelli, D.; Raanan, H.; Falkowski, P.G. Metal availability and the expanding network of microbial metabolisms in the Archaean eon. *Nat. Geosci.* **2017**, *10*, 629–636. [CrossRef]

49. Szilagyi, R.K.; Hanscam, R.; Shepard, E.M.; McGlynn, S.E. Natural selection based on coordination chemistry: Computational assessment of [4Fe-4S]-maquettes with non-coded amino acids. *Interface Focus* **2019**, *9*. [CrossRef]

50. Holm, R.; Lo, W. Structural conversions of synthetic and protein-bound iron–sulfur clusters. *Chem. Rev.* **2016**, *116*, 13685–13713. [CrossRef]

51. Venkateswara Rao, P.; Holm, R.H. Synthetic Analogues of the Active Sites of Iron-sulfur Proteins. *Chem. Rev.* **2004**, *104*, 527–560. [CrossRef]

52. Bruschi, M.; Guerlesquin, F. Structure, function and evolution of bacterial ferredoxins. *FEMS Microbiol. Lett.* **1988**, *54*, 155–175. [CrossRef]

53. Eck, R.V.; Dayhoff, M.O. Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* **1966**, *152*, 363–366. [CrossRef]

54. Raanan, H.; Poudel, S.; Pike, D.H.; Nanda, V.; Falkowski, P.G. Small protein folds at the root of an ancient metabolic network. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 7193–7199. [CrossRef] [PubMed]

55. Krupp, R.E. Phase relations and phase transformations between the low-temperature iron sulfides mackinawite, greigite, and smythite. *Eur. J. Mineral.* **1994**, *6*, 265–278. [CrossRef]

56. Huber, C.; Wächtershäuser, G. Activated Acetic Acid by Carbon Fixation on (Fe,Ni)S Under Primordial Conditions. *Science* **1997**, *276*, 45–247. [CrossRef] [PubMed]

57. Spatzal, T.; Aksoyoglu, M.; Zhang, L.; Andrade, S.L.; Schleicher, E.; Weber, S.; Rees, D.C.; Einsle, O. Evidence for interstitial carbon in nitrogenase FeMo cofactor. *Science* **2011**, *334*, 940. [CrossRef]