The Fe–S proteome of Escherichia coli: prediction, function, and fate

Marine Lénon, Rodrigo Arias-Cartín, Frédéric Barras

To cite this version:

HAL Id: pasteur-04110552
https://pasteur.hal.science/pasteur-04110552
Submitted on 30 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
The Fe-S proteome of *Escherichia coli*: prediction, function and fate.

Marine Lénon#, Rodrigo Arias-Cartín#, and Frédéric Barras*

Institut Pasteur, Université Paris Cité, CNRS UMR6047, Stress Adaptation and Metabolism Unit, Department of Microbiology, F-75015 Paris, France.

# These authors contributed equally to this work

* Corresponding author

* To whom correspondence should be addressed to Frédéric Barras, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15. Email: fbarras@pasteur.fr
Abstract

Iron-sulfur (Fe-S) clusters are inorganic ubiquitous and ancient cofactors. Fe-S bound proteins contribute to most cellular processes, including DNA replication and integrity, genetic expression and regulation, metabolism, biosynthesis and most bioenergetics systems. Also, Fe-S proteins hold a great biotechnological potential in metabolite and chemical production, including antibiotics. From classic biophysics and spectroscopy methodologies to recent development in bioinformatics, including structural modeling and chemoproteomics, our capacity to predict and identify Fe-S proteins has spectacularly increased over the recent years. Here, these developments are presented and collectively used to update the composition of *Escherichia coli* Fe-S proteome, for which we predict 181 occurrences, *i.e.* 40 more candidates than in our last catalog (Py and Barras, 2010), and equivalent to 4% of its total proteome. Besides, Fe-S clusters can be targeted by redox active compounds or reactive oxygen and nitrosative species, and even be destabilized by contaminant metals. Accordingly, we discuss how cells handle damaged Fe-S proteins, *i.e.* degradation, recycling or repair.
1. **Introduction**

Fe-S clusters are assemblies of iron and sulfur atoms and stand among the most frequently used protein cofactors in biology [1]. Fe-S clusters arise in various forms, wherein 2, 3 or 4 iron atoms are linked to sulfide ions, yielding to the typical Fe₃S₂ (rhombic), Fe₃S₄ (cuboidal) or Fe₄S₄ (cubane) clusters and some atypical types such as Fe₄S₃, Fe₄S₅, Fe₈S₈ or Fe₈S₉ clusters [2–6]. Fe-S clusters were discovered as devices implicated in electron transfer in enzymes participating in photosynthesis and respiration, but we now know that they contribute in indispensable functions to nearly all cellular processes (see below). Fe-S proteins are present in both prokaryotes and eukaryotes where their pleiotropic role extends their influence onto general traits such as pathogenicity, CRISPR immunity systems and antibiotic resistance in bacteria, aging, cancer or ataxia in humans, plant growth and even replication of coronaviruses such as the SARS-CoV-2, the causal agent of the CoVID-19 pandemic [7–10].

Fe-S clusters undergo one-electron redox processes and can exhibit various redox states. This makes them ideal catalysts for intra- and inter-molecular electron transfer processes. Their redox potential value ranges from -600 mV (S.H.E.), depending upon the chemical nature of their coordination environment and electronic properties of their immediate surrounding within the polypeptide. Fe-S clusters with low redox potentials can be used as catalysts for thermodynamically unfavorable reactions. This is best illustrated by radical–S-adenosylmethionine (SAM) enzymes, which use a Fe₄S₄ cluster to inject one electron at a low potential, generating reactive free radicals exploited in a myriad of biosynthetic and metabolic reactions [11–15]. Fe-S clusters can also allow access of small compounds to ferric ions with strong Lewis acidity properties. Such clusters can be used in non-redox catalysis, as in dehydratases [16]. Last, reversible interconversions between cluster forms, with different redox states or different nuclearity, is used in Fe-S bound transcriptional regulators that function as sensors of oxygen, superoxide, nitric oxide, or even in proteins controlling DNA integrity [17–22].

As a general trend, Fe-S cluster ligation sites are composed of cysteine residues whose thiol side chains provide bonding to Fe atoms or form sulfide bridges. Hence, a four Cys-containing motif, although admitting uncertainties on the space between individual Cys residues, is often seen as a predictor of a Fe-S cluster binding site. However, this can be misleading for two reasons. First, these sites may bind other metals, such as Zinc. Second, Fe-S cluster coordination with oxygen-based (aspartate, tyrosine or glutamate) and nitrogen-based (histidine or arginine) residues have also been described [23]. Hence, if the richness and diversity of cluster environment is an asset for biology as it allows versatility, it prevents simple
and straightforward Fe-S cluster signature to be applied from primary sequence analysis. In this article, we review the last advances in the search, identification, and assessment tools applied to Fe-S proteins with an emphasis on the *Escherichia coli* Fe-S proteome for which we present an updated version of the catalog of Fe-S proteins. We also address the question of the relationship between O$_2$ and Fe-S-based biology and provide original data that led us to envision different fates for damaged Fe-S proteins.

2. **Experimental assessment of the presence of an Fe-S cluster in a protein**

Experimental methods are required to assess whether a given protein hosts an Fe-S cluster or not (for review see [24, 25]). An arsenal of *in vitro* strategies is available, but they constitute a long and tedious path before a conclusion can be reached. In most cases, the biochemistry requires manipulation in the absence of O$_2$ inside glove boxes and high amounts of either the as-isolated holo-form protein or an apo-form in which an Fe-S cluster is chemically or enzymatically reconstituted. Once the purified holo-protein is obtained, methods for quantifying iron and sulfur are the first steps to suggest the presence of a cluster. Next, complementary biophysical analyses, such as spectroscopy (UV-visible, Electron Paramagnetic Resonance - EPR, Resonance Raman, Mössbauer, etc.), mass spectrometry, and crystallography must be used to establish the presence of a cluster and to characterize its properties and chemical environment, in particular the nature of its ligands [25]. Importantly, those spectroscopic techniques usually require biochemical and/or genetic manipulations and are complemented by functional studies. For in-depth study of Fe-S clusters, EPR and Mössbauer spectroscopy are widely used since they provide a solid characterization of the Fe-S cluster type, its environment, and redox state. Mössbauer spectroscopy is the more suited method since it allows characterization of iron complexes in any oxidation or spin state. However, this type of spectroscopy requires $^{57}$Fe isotope incorporation into the sample, demanding careful manipulation during cell culture or protein reconstitution. EPR studies can bypass this step using the natural Fe isotope, however it cannot detect silent Fe-S clusters (Spin = 0). For organisms amenable to genetic manipulation, use of strains lacking specific factors required for building and inserting clusters into proteins can be used to predict whether a given protein would hold a cluster and to assess its functional importance for activity *in vivo* [25].

3. **Bioinformatic prediction of Fe-S cluster containing proteins**

*In silico* identification of Fe-S binding sites from primary sequences or structures of proteins remains challenging. Yet, several bioinformatic analyses [26–28] have proven decisive in
allowing identification of new Fe-S binding proteins further demonstrated by experimental methods, giving significant insights into the detection of metabolic pathways on poorly characterized species [29–32].

Estellon et al. developed a sophisticated machine learning tool to design and assess a penalized linear model to predict Fe-S proteins using primary amino acid sequences [26]. Their approach was based on an ensemble of descriptors from a tailored non-redundant database of Fe-S-specific HMM (Hidden Markov Model) profiles and a curated selection of Fe-S coordinating domains and signatures, which were used to guide their machine learning algorithm on a large training dataset of protein sequences (PDB70, a data set of sequence alignments with identity <70% and using PDB structures as query excluding *E. coli* K12 sequences). This Fe-S predictive model of 67 descriptors reached 87.9% of precision and 80.1% of a recall (or sensitivity) on the *E. coli* K12 proteome, which are values of higher performance when compared to other motif-based analysis (Prosite, Pfam, InterPRO, etc.). Despite the inability of their model to detect 27 known Fe-S proteins at the time of the publication, the capacity of their software was demonstrated by the prediction and experimental validation of YhcC and YdiJ. Furthermore, the model was tested on 556 proteomes from bacterial and archaeal species and predicted an averaged content of Fe-S proteins at 2.37 ±1.31% of those prokaryotic total proteomes. A related method was created by Valasatava et al. [27] (available in the MetalPredator webserver). In this case, the search for metal-binding motifs by HMM-profiles combines domain-based predictions [33] and the local nature of Minimal Function Sites [34]. This tool has a similar precision (85.2%) and slightly higher recall (86.5%) when compared to the results of the IronSulfurProteHome [26] on the *E. coli* proteome. Validation of their work was based on 3D homology on seven of their predictions.

The search of Fe-S binding sites using 3D protein structures has been limited by the number of structures or models available. However, a very recent study from Wehrspan et al. [28] elaborated a new ligand-search algorithm to identify Fe-S cluster or Zn binding sites exploiting the novel AlphaFold2 structure database [35, 36]. Their method is based first in the compilation of a repertoire of all sidechain or backbone atoms that could potentially coordinate Fe-S cofactors or Zn ions in the AlphaFold2 database, which are later grouped and used as potential binding regions to be examined using a standard single-linkage clustering algorithm (within 8 Å) to superimpose each ligand type (Fe$_2$S$_2$, Fe$_3$S$_4$ or Fe$_4$S$_4$) and their coordinating atoms in all possible combinations. Next, those solutions are evaluated using a root-mean-squared deviation (RMSD) of the ligands and checked for steric clashes. Results with poor RMSD scores or steric clashes are removed, the remaining combinations filtered, the one with the
lowest RMSD retained and the ligand is placed into the structure. This study identified thousands of potential Fe-S clusters in the proteomes of 21 organisms listed in the AlphaFold2 structures. When this analysis was applied to the *E. coli* proteome, it obtained a recall rate of 74%, which is lower than the sensitivity of Valasatava *et al.* (86.5%) or Estellon *et al.* (80.1%), however it found a good overlap on the predicted Fe-S proteins or the false positives predicted by those two previous studies. Moreover, this tool was able to calculate if some proteins are more likely to accommodate Fe$_2$S$_2$, Fe$_3$S$_4$ or Fe$_4$S$_4$ clusters. Interestingly, the 3D scan made by this work predicted that YjiM, YcbX, CyuA (YhaM) and PreT (YeiT) are indeed Fe-S containing proteins, which were classified as false positives by Estellon *et al.* Likewise, it agreed with Valasatava *et al.* that seven of their false positives could coordinate Fe-S clusters. In addition, this work could place Fe$_4$S$_4$ clusters on the structural models of the newly characterized DppD or on the *E. coli* U32 protease homologs that have been experimentally confirmed (UbiU, UbiV and TrhP) or strongly suspected (RlhA) as Fe-S proteins [15, 37, 38]. Interestingly, none of these computational tools predicted MnmA as a Fe-S binding protein, an issue that remains controversial [35, 36]. Noteworthy, no bias or specific enrichment in a functional class of Fe-S protein or type of Fe-S binding motif, was found to be associated with one particular algorithm. It is clear that precision and sensitivity of any of those *in silico* approaches are still limited, due to the strength of their descriptors and conditioned to the curation and size of the datasets used for training and testing of the algorithms. Thus, those predictive methods will only improve as long as the factors involved in the Fe-S cluster binding and assembly -such as plasticity, abundance and nature- are elucidated by experimental and functional studies.

### 4. The *E. coli* Fe-S proteome: an update

In the specific interest to establish the *E. coli* Fe-S proteome, we used two validated bioinformatic tools to predict a list of Fe-S bound candidate proteins: MetalPredator and IronSulfurProteHome websites [26, 27]. For *E. coli*, those tools predicted 141 iron-sulfur cluster containing proteins and we added 21 proteins, for which evidences of Fe-S binding were recently published. Furthermore, 19 additional proteins were predicted by 3D modelling using Alphafold database [28]. This yielded to 181 entries represented in Figures 1 and 2. For each protein, type of evidence, either computational (Figure 1A) or biophysical (Figure 1B), supporting the notion that it is a Fe-S bound protein is given in Supp Table 1. Roughly half of the 181 *E. coli* Fe-S proteins have been validated as such by biophysical methods (Figure 1B).
*E. coli* makes use of Fe-S proteins in most, if not all, processes such as DNA replication, repair and transcription, RNA modification and translation, amino acid, vitamin and metabolite biosynthesis, and bioenergetics (Figure 2). Respiration is the most represented cellular process as *E. coli* allocates 28% of its Fe-S proteome to respiratory pathways (Figure 2). Biosynthesis of metabolites and cofactors, including Fe-S themselves, constitute the second most populated group. As expected, Fe-S bound proteins allow stress sensing and adaptation. Interestingly, a few Fe-S proteins are predicted to be involved in transport. Interestingly, there are 15 Fe-S proteins for which no functional prediction, neither from homology search or genetic screen was found in the literature. Attribution of function for these proteins might unearth activities of interest.

5. **Fe-S & Oxygen**

Besides their intrinsic redox chemical properties, the reason why Fe-S clusters were retained in such a wide range of proteins might be related to the geological conditions prevailing at the onset of life, namely abundance of Fe and S and anoxic atmosphere. The Earth’s Great Oxygenation Event, probably due to the activity of Fe-S containing photosynthesis apparatus, caused shortage in bioavailable iron as it precipitated from the soluble ferrous form into insoluble ferric form. Besides, Fe$^{2+}$ can act as a catalyst for production of reactive oxygen species (ROS), via Fenton reaction, which destabilize Fe-S cluster. Therefore, the long-admitted view is that sophisticated Fe-S cluster biogenesis machineries emerged to build and insert Fe-S clusters into proteins [39–41]. Indeed, Fe-S cluster-based biology and aerobic life might be seen as mutually exclusive, unless a great investment was put to control Fe-S biogenesis and its use in the cell. Support for this view was put forward by Andreini and collaborators in a bioinformatic analysis of Fe-S proteins predicted content of prokaryotic genomes [42]. A set of 434 prokaryotes genomes, including 18 genomes from obligate aerobes, 29 from obligate anaerobes, 214 from aerobes, 130 from aerotolerant anaerobes and 43 from facultative anaerobes was analyzed for occurrence of Fe-S proteins as predicted by bioinformatic protocols. Remarkably, there are more predicted Fe-S proteins per genome from anaerobes than aerobes, even when corrected with genome size. In aerobes the number of predicted Fe-S proteins correspond to less than 3% of the total predicted proteome whereas in genomes of anaerobes (obligate, facultative, aerotolerant) predicted Fe-S proteins amount to more than 3% of the total predicted proteins. Accordingly, we observe that *E. coli* Fe-S proteome is 4% and some recent *in silico* studies in archaea suggest that in methanogens and alkanotrophs 5-10% of the proteome corresponds to Fe-S proteins [29]. Both aerobes and
anaerobes exhibit shared family of Fe-S proteins involved in energy production and conversion, amino acid metabolism, nucleotide and coenzyme metabolism, and Fe-S biogenesis. Interestingly, the number of paralogs involved in energy production and conversion is much higher in anaerobes than in aerobes, presumably reflecting the multiplicity of potential electron acceptor chemical recruited for anaerobic respiratory chains instead of the unique O$_2$ in aerobic ones. Another important observation was that the number of proteins using Fe$_2$S$_2$ cluster, which are more resistant to O$_2$ and less demanding in iron, appeared to increase in aerobes vs anaerobes. Hence, this study stands as a validation of the expected negative interaction between O$_2$ and Fe-S-based biology. It also provides a rationale of why organisms developed strategy to mitigate deleterious consequences of enhanced O$_2$ level in the atmosphere after emergence of photosynthesis, either by reducing the use of Fe-S cluster containing proteins, or by evolving dedicated assisting biogenesis machineries.

6. What to do with damaged Fe-S proteins: to repair, recycle or degrade them?

The capacity of Fe-S clusters to detect O$_2$, ROS or reactive nitrogen species (RNS) has its drawbacks, as Fe-S clusters can be altered or degraded by such compounds, which might lead to inactivation and destabilization of the hosting polypeptide. One question is to know whether all Fe-S proteins are similarly destabilized in the face of ROS or RNS. The answer is evidently negative since alteration of Fe-S cluster will depend upon its location into the structure of the hosting polypeptide. Well buried, solvent inaccessible clusters are likely to be more stable. On the contrary, well exposed cluster might be more prone to targeting by toxics and ensuing alteration. This was well studied with members of the dehydratase family, such as aconitase or fumarase.

The nature of the ligands holding the cluster can also be a determining factor for its stability. The case of fumarase had been investigated in detail [43]. Fumarases contain Fe$_4$S$_4$ bound via a C$_X$nCXXC motif. Like all dehydratases, cluster in the \textit{E. coli} fumarase exhibits a labile exposed catalytically active Fe atom, which is freed upon oxidation of the cluster, yielding to an inactive Fe$_3$S$_4$-bound enzyme$^{44}$. In contrast, the \textit{Bacteroides thetaiotaomicron} fumarase harbors a C$_X$xC motif, which cannot be reactivated when oxidized by H$_2$O$_2$. This is due to the generation of radical species that carbonylate the peptide chain $^{44}$. Thus, composition of the ligating motif itself could be essential for Fe-S proteins in the resistance to oxidative environments. Besides, the number of ligand cysteines could also be important for tolerance to oxygen attack, as reported in the case of the six cysteines liganded Fe$_4$S$_3$ cluster of hydrogenase from \textit{Ralstonia eutropha} $^6$. 
Another example of Fe-S cluster targeted by environmental redox conditions is given by the transcriptional regulator Fnr that controls anaerobiosis/aerobiosis switch [45–47]. The Fe₄S₄ cluster integrity of Fnr controls the monomer-dimer equilibrium and eventually the expression of hundreds of genes. Briefly, its Fe₄S₄-bound homodimeric form binds to operator regions of genes, repressing those involved in aerobic metabolism and activating those required for anaerobic metabolism. Exposure of the cluster to O₂ leads to the Fe₄S₄ and subsequent conversion to Fe₃S₄ and Fe₂S₂ causing the dissociation of the dimer in inactive apo-monomers.

Another question concerns the fate of the “damaged” (or oxidized) Fe-S proteins. Anew, studies with Fnr as a model provided important insights. Indeed, it was found that in aerobic condition, apo-Fnr was degraded by the chaperone/protease ClpXP [45, 48, 49] but it was also shown that reconversion of apo-Fnr to Fe₄S₄-Fnr is possible upon switch to anaerobic growth [49]. This perfectly illustrates how cells handle “damaged” Fe-S protein, either degrading them or reactivating them. This last possibility suggests that apo-forms of some proteins are stable long enough to engage in a new cycle of maturation via their recognition by ISC or SUF machineries. This view is consistent with an early study aiming at investigating whether SUF was a repairing system [50]. Interestingly, besides Fnr, other Fe-S proteins (AcnB, IscU, IscR, LipA and MoaA) have been identified as substrates of ClpXP in E. coli [51, 52], which may suggest that this protease could be involved in a broader spectrum in the homeostasis of the Fe-S proteome. Whether ClpXP acts on these Fe-S proteins specifically when they are damaged remain to be established.

A distinct possibility put forward for long, is the existence of dedicated factors that would “repair” damaged clusters. Interestingly, Clp might play a role in the repairing of damaged Fe-S proteins by helping iron release from Dps sequestration [53]. YtfE might also be such a repair factor acting by re-metalating clusters having lost Fe atoms [54] like those invoked in the progressive dismantlement of Fnr or in the case of dehydratases. For instance the oxidatively-damaged cluster Fe₃S₄ of the E. coli fumarase can be reactivated if an iron source is provided [55].

Furthermore, studies in members of the radical-SAM family have enlarged our view of protein factors needed for the repair of their cannibalized Fe-S clusters. The NfuA carrier protein has been endowed with the capacity to regenerate the Fe₄S₄ cluster in the lipoyl synthase (LipA), which sulfur atom has been sacrificed during catalysis [56].

7. Perspectives
Since their discovery [1], Fe-S proteins have been studied by relying on a wide array of approaches, from chemistry to genetics via biochemistry, structural biology and spectroscopy. This has provided us with a deep insight in (i) the mechanism underlying reactivity of Fe-S proteins, (ii) the role played by the cluster, (iii) the contribution of Fe-S proteins to cellular homeostasis. The next step will be to apply high-throughput techniques aiming at getting an integrated description of the Fe-S proteome dynamics in vivo. Thermo proteome profiling methodology already contributed to a whole cell view by reporting the effect of genetic alteration of the ISC system [57]. Another methodology is chemoproteomic [37]. This technique is based in tracking Fe-S cysteine ligands by LC-MS/MS coupled to the labeling using iodoacetamide-alkyne derivative (termed isoTOP-ABPP and ReDiME). Its recent use to follow changes in Fe-S proteome as a response to iron-depletion or mutations in the ISC system opens highly promising possibility to an in vivo integrated description of cell response and adaptation to environmental challenges and genetic disorders. We recently showed that heterologous expression of some active Fe-S proteins requires the heterologous co-expression of the native Fe-S cluster assembly machinery [58]. Deciphering the recognition mode between machineries and targets might help in optimizing and broadening heterologous expression of Fe-S containing proteins, a goal of considerable interest in biotechnology.

Bioinformatic approaches have provided new vision regarding distribution of Fe-S proteins and their evolution. Yet, predicting whether a given polypeptide will host a cluster is still limited by the wide diversity of motifs allowing Fe-S binding. Another structural feature that might be used is the intrinsic information of Fe-S polypeptide to be recognized by Fe-S maturation factors. Indeed, studies in eukaryotes have unearthed the so-called LYR motif, which endows a subset of Fe-S proteins with the capacity to interact with Hsc20 ISC machinery component. Future studies should aim at investigating whether similar type of “recognition sequence” occurs and permits interaction between apo-targets and machineries in prokaryotes.

**Acknowledgements**

We thank the SAMe Unit members for discussions and suggestions. This work was supported by the French State Program ‘Investissements d’Avenir’ (Grant “IBEID” ANR-10-LABX-62), by CNRS and by Institut Pasteur.

**Data Availability Statement**

The data underlying this article are available in the article and in its online supplementary material.
References


REVIEW OF THE FE-S PROTEOME OF _ESCHERICHIA COLI_


Figure 1. Characterization of *E. coli* Fe-S proteome. A. Left panel: List of Fe-S proteins predicted by bioinformatic tools only, *i.e.* not experimentally assessed yet. Motif/Pattern stands for primary sequence motif and pattern search [26, 27] or dedicated sequence alignment (see Supp Table 1), AlphaFold [28] and 3D model obtained from structural analysis of homologous holo-proteins (see Supp Table 1). B. Right panel: List of Fe-S proteins characterized by biophysical approaches such as UV-visible (UV-vis), Electron Paramagnetic Resonance (EPR), Mössbauer and Resonance Raman spectroscopy and X-Ray crystallography (Xray) (see Supp Table 1).
Figure 2. The Fe-S proteome of *E. coli*. Distribution of the *E. coli* Fe-S proteins according to their association with a specific biological process. Fe-S proteins characterized by biophysical approaches are in bold. Fe, iron; S, sulfur; TCA, Tricarboxylic Acid Cycle. Details about all listed Fe-S proteins are in Supp Table 1.