Non-heme iron through the three domains of life

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · MAY 2007

Impact Factor: 2.63 · DOI: 10.1002/prot.21324 · Source: PubMed

5 AUTHORS, INCLUDING:

Antonio Rosato
University of Florence
113 PUBLICATIONS **3,653** CITATIONS

SEE PROFILE

PROTEINS: Structure, Function, and Bioinformatics 67:317–324 (2007)

Non-Heme Iron Through the Three Domains of Life

Claudia Andreini, Lucia Banci, Ivano Bertini,* Sara Elmi, and Antonio Rosato
Magnetic Resonance Center (CERM) and Department of Chemistry, University of Florence, 50019 Sesto Fiorentino, Italy

ABSTRACT **Metalloproteins are proteins capable of binding one or more metal ions, which are often required for their biological function or for regulation of their activities or for structural purposes. In high-throughput genome-level protein investigation efforts, such as Structural Genomics, the systematic experimental characterization of metal-binding properties (i.e. the investigation of the metalloproteome) is not always pursued, and remains far from trivial. In the present work we have applied a bioinformatic approach to investigate the occurrence of (putative) non-heme iron-binding proteins in 57 different organisms spanning the entire tree of life. It is found that the non-heme iron-proteome constitutes between 1% and 10% of the entire proteome of an organism. However, the iron-proteome constitutes a higher fraction of the proteome in archaea (on average 7.1% ± 2.1%) than in bacteria (3.9% ± 1.6%) and in eukaryota (1.1% ± 0.4%). The analysis of the function of each putative iron-protein identified suggests that extant organisms have inherited the large majority of their iron-proteome from the last common ancestor. Proteins 2007;67:317–324.**

Key words: iron; metalloprotein; electron transfer; metalloproteome; evolution

INTRODUCTION

Metal ions play crucial roles in most of the biochemical processes at the basis of Life.[1] Indeed, many proteins need to bind one or more metal ions to be able to perform their function (hence called metalloproteins), either because the metal ion is involved in the catalytic mechanism or because it stabilizes/determines the protein tertiary or quaternary structure. Metal ions are also very important for the structure and function (in the case of RNA) of nucleic acids. The intracellular concentration of several metals (e.g. copper, zinc, iron, ...) as well as their distribution among the various cell compartments and their incorporation into metalloproteins is tightly controlled.[2–4] A proper balance of the equilibria involved in these control processes is necessary for a healthy phenotype.

Genome sequencing projects have provided us with the sequences of all the proteins that the various organisms can produce (proteome). To fully exploit these data, functional annotation is performed based on gene prediction algorithms and homology searches.[5–7] The result in terms of attributing a function to each protein potentially produced by an organism is thus determined by whether a previously characterized homologue is found, the level of homology detected (e.g. determined by the E-value, or by a sequence identity threshold), how many homologues are identified, etc. The same concepts can be applied to metalloproteins, where however the ability to bind a metal is not straightforwardly evident from sequence analysis with the standard methods. The presence of a motif suitable for the protein to bind the metal ion(s) can be exploited for a more reliable identification and classification (including functional inferences) of these proteins.[8–10] Genome-wide identification and classification of metalloproteins may be difficult even experimentally,[11] although some systematic efforts in this direction have been deployed.[12–14]

Iron is essential for Life and is the most abundant transition metal ion in living organisms. In cells, iron is normally found in the +2 and/or +3 oxidation states. Higher oxidation states may be generated transiently in the course of the catalytic cycle of enzymatic reactions. Besides single iron ions, proteins can bind a range of iron-containing cofactors, such as heme or iron–sulfur clusters, just to mention the most commonly found.

In this work, a bioinformatic approach based on the use of non-heme Iron-Binding Patterns (IBP's)[10] in combination with the analysis of the occurrence of protein domains known to bind non-heme iron is applied to improve the identification and classification (including functional inferences[8]) of non-heme iron-binding proteins for the human proteome as well as for a selected number of other eukaryotic and prokaryotic proteomes. Heme-binding proteins (cytochromes and globins) have not been taken into account, as binding of the cofactor is strongly influenced by the interaction of the polypeptide chain with the porphyrin moiety, and thus the coordination properties of the iron are comparatively somewhat less important.

METHODS

The protocol used to retrieve non-heme iron-binding proteins in the proteomes is depicted in Figure 1, and

Abbreviations: EC, enzyme classification; MBP, metal-binding pattern.

The Supplementary Material referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat/

Grant sponsor: MIUR (projects PRIN-COFIN 2005, FIRB).

*Correspondence to: Prof. Ivano Bertini, Magnetic Resonance Center, University of Florence, Via L. Sacconi, 6, 50019 Sesto Fiorentino, Italy. E-mail: ivanobertini@cerm.unifi.it

Received 18 April 2006; Revised 16 October 2006; Accepted 7 November 2006

Published online 7 February 2007 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21324

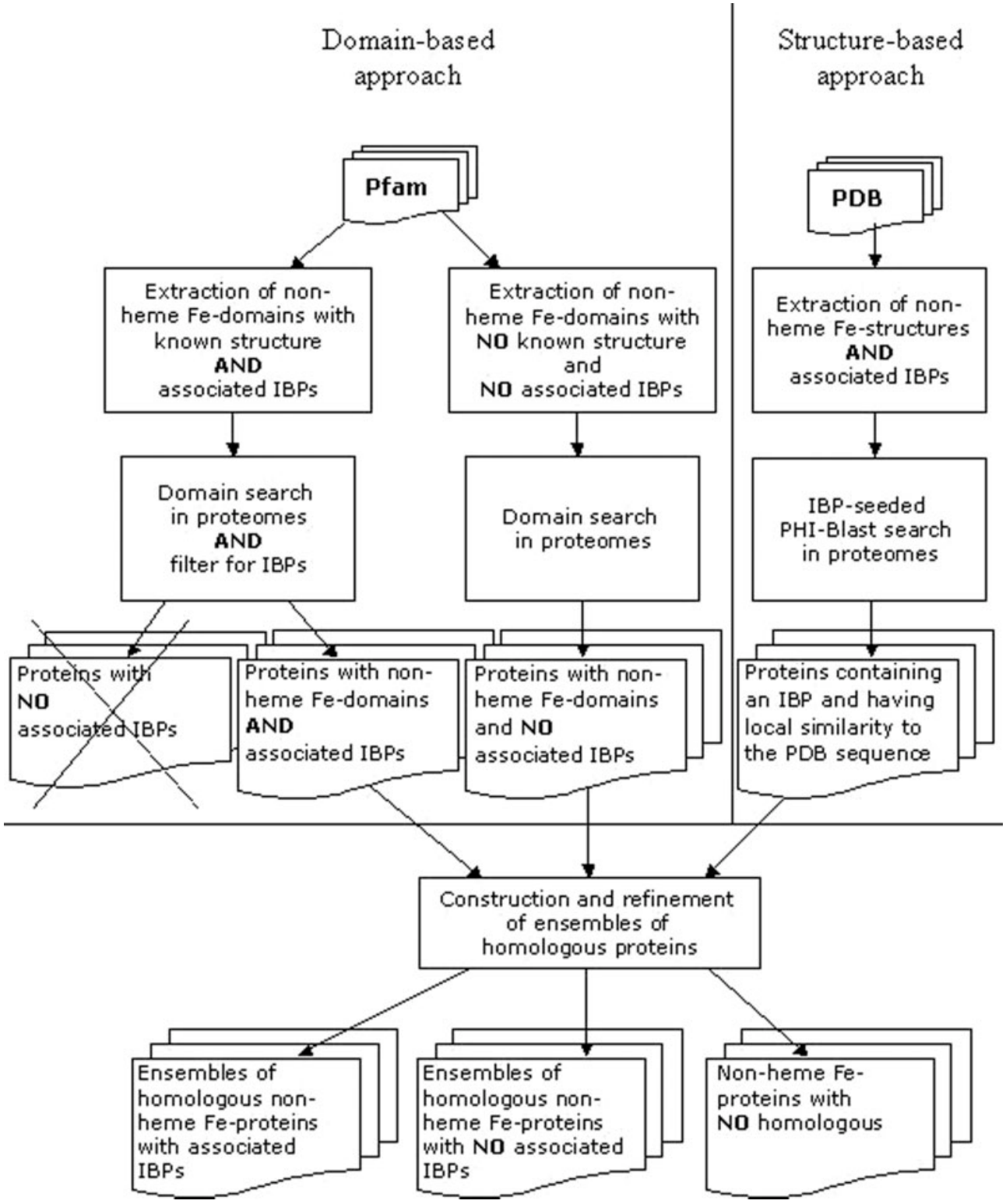

Fig. 1. Scheme of the protocol used for the identification of non-heme iron-binding proteins.

is described later. A detailed graphical scheme of the protocol is also given in Supplementary Figure S1. Figure S1A describes the preparation of input files, and Figure S1B describes the production of output (that is, the retrieval of putative iron-binding proteins).

Preparation of Input Files

All the structures deposited in the PDB[15] as of January 2006 and containing at least one iron ion in the coordinate file were extracted. Only one representative was kept for all proteins binding the metal ion or metallic cofactor (e.g. iron–sulfur clusters) with the same pattern(s) and having a sequence identity greater than 98% (nonredundant iron-protein structures). Heme-binding proteins and proteins binding iron in a nonphysiologically relevant manner (e.g. due to the presence of iron in the crystallization buffer) were manually removed. The resulting representative ensemble contained 330 iron-binding structures. The coordinate files of each structure were used to identify the residues coordinating the metal ion(s), which define the Iron Binding Pattern (IBP), with an approach already applied to other classes of metalloproteins.[10,16] Every residue having at least one heavy atom at a distance shorter than 2.8 Å from the metal was defined as a metal ligand. Polymetallic clusters, such as Fe_4S_4 in ferredoxins, were considered as a single center and thus all the ligands to the various iron ions formed a single IBP. Note that more than one IBP can be associated to a single protein, depending on the number of metal ions bound. As a result of this analysis, a library of 318 distinct IBP's was assembled (Table S1).

In parallel, we extracted protein domains defined as nonheme iron-binding from the Pfam library of domains.[17] All domains whose description contained the words "iron" or "Fe" have been selected and manually analyzed to collect a list of iron-binding domains constituting the input for searches in each proteome. Heme-binding domains were manually removed. To enhance our coverage, iron-binding protein structures retrieved from the PDB have been analyzed to identify domains always or commonly containing a iron-binding site not already annotated as non-heme iron-binding in Pfam (e.g., TauD[18]). This procedure also allowed us to associate IBPs to Pfam domains having at least one structurally characterized representative. In this way we created an ensemble of iron-binding domains, most of which with an associated IBP (Fig. 1).

Proteomes Selection and Search

Proteome sequences for the selected organisms (Table S2 in the Supplementary Material) have been downloaded from the Entrez Genome (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome) web site at the National Center for Biotechnology Information. The organisms have been selected to map evenly the tree of life (Figure S2 in the Supplementary Material). We selected archaeal and bacterial organisms by taking one organism per order; when there was more than one organism sequenced in a order, we selected the organism sequenced first. 40 bacterial, 12 archaeal, and 5 eukaryotic proteomes were analyzed.

Each of the 57 selected proteomes was then searched for the protein sequences containing a non-heme iron binding domain, with the program HMMER[19] using as input the HMMs available from Pfam. For iron-binding domains with an associated IBP (or IBP's), we filtered the results by discarding the sequences containing the domain but lacking the IBP (Fig. 1). Some variability was allowed for the IBP, by letting the spacing between metal ligands vary within ±20% (or by ±1 residue when the spacing was less than 5 residues). As an independent analysis, we run PHI-BLAST[20] searches on the same proteomes using each IBP as a seed and the sequence of the protein in the PDB containing it as query (Fig. 1). Hits with a fraction of identical amino acids aligned by PHI-BLAST with respect to the length of the query protein (I_d^{global}) greater than 0.20 when the IBP had at least three ligands, or with $I_d^{global} > 0.30$ when the IBP had only two ligands, were kept as putative non-heme iron-binding proteins. This approach has been applied in previous studies.[10,16]

All putative iron-binding proteins retrieved were pooled together and grouped based on homology. Homologous proteins were identified by performing BLAST[21] searches for each putative iron-binding protein against all the other identified iron-binding proteins. Proteins were grouped in a way such that each ensemble contained all the hits within a given threshold (E-value calculated by BLAST $<10^{-3}$) for each of its members. An additional restraint imposed was that the length of the protein sequences had to be within a twofold. All ensembles of sequences were manually inspected to check whether two (or more) functionally different protein families had been grouped together. For sequences retrieved by Pfam domains having at least one representative of known structure, we also verified that in each ensemble all the proteins had a similar IBP (within the allowed flexibility). Grouping of different protein families occurred in less than 1% of the ensembles calculated; for these instances, the ensembles were rebuilt with a tighter BLAST threshold (10^{-5}). A profile was built for each ensemble of sequences and used to further search the proteomes in order to extend our coverage. Each profile thus describes a group of proteins similar in sequence and, when available, with a similar IBP. Newly identified proteins were then filtered on the basis of the presence of the expected IBP, and subsequently added to the ensemble corresponding to the profile. The latter was recalculated and profile searches repeated until no new sequences could be identified. At the end of the procedure described in this section, a database of putative iron-binding protein sequences was obtained. The procedure implemented produced ensembles of sequences sharing sequence similarity. The majority of ensembles also had an associated IBP, which was conserved with little variation in spacing in all proteins belonging to each ensemble (Fig. 1).

In a previous work of ours on human zinc-binding proteins,[16] the analysis of the gene annotations deposited together with the genome sequence was also useful. In the present study, gene annotations instead did not prove informative, because: *i*) the coverage and quality of annotations across the various organisms analyzed is quite dissimilar, preventing meaningful comparisons; *ii*) even in the best annotated organisms, such as human, the degree of accuracy was somewhat lower than what

observed for zinc, implying that results could not be handled automatically with an acceptable reliability.

Evaluation of Methods Performance

We used standard performance measures to evaluate the quality of our results. These measures included precision and recall. For a given target class C (e.g., non-heme iron binding proteins), let true positives (TP) be the number of examples correctly predicted as belonging to C, false positives (FP) the number of examples incorrectly predicted as belonging to C, and false negatives (FN) the number of examples of class C incorrectly assigned to another class. Precision is the number of true positives over the total number of examples assigned to the class, that is, TP/(TP + FP). Recall (or coverage) is the number of true positives over the total number of examples belonging to the class, that is, TP/(TP + FN). In the case of binary classification (i.e., yes/no as in the present case), recall for the positive class is also known as sensitivity.

The ensemble of positives was built taking from the PDB one representative sequence for each group of non-heme iron-binding proteins having a sequence identity greater than 90% (heme-binding and artefacts were manually removed). The ensemble of negatives was built similarly taking one representative for all other proteins in the PDB. These two datasets were searched using the input files prepared as detailed in the previous sections. The performance of our approach was evaluated for different HMMER E-values and different I_d^{Global} thresholds.

Data Analysis

To obtain functional hints, all sequences retrieved were analyzed against the entire Pfam domain library as well as in terms of their Gene Ontology (GO) functional annotation.[22]

RESULTS AND DISCUSSION

Input Preparation and Evaluation of Performance

1001 protein structures containing at least one non-heme iron ion were retrieved from the PDB, corresponding to 330 nonredundant iron-binding protein sequences and 318 distinct IBPs (Table S1). 126 Pfam domains were identified as iron-binding (Table S3). The data were used as the starting point to search in the selected proteomes. The number of iron ligands contained in a single IBP was found to vary between two and six, with four being the most common coordination number. Note however that, because polymetallic clusters are treated as a single block this number does not always reflect the coordination geometry around each individual iron ion, but is actually correctly defined as the number of amino acids involved in binding the cofactor. In the case of particularly complex cofactors, even seven protein residues may be involved in binding.

To evaluate the performance of our protocol, we applied it to the entire PDB (thus, including both non-heme iron proteins and all other proteins) and computed performance measures. For PHI-BLAST searches, we removed from the results all hits to the query itself (which is a protein of known structure) and to its homologs in order not to introduce a favourable bias in the performance measures.[10] The structure-based part of the protocol alone had a recall (fraction of non-heme iron-proteins identified with respect to all non-heme iron-proteins in the PDB) of 0.62 with a precision of 0.95 (fraction of proteins correctly predicted to be non-heme iron binding). At a HMMER E-value threshold of 0.05 the domain-based searches filtered for the IBP's had instead a recall of 0.96 and a precision of 0.84. Both parameters were little affected when the E-value threshold was varied in the range 10^{-3} to 10^{1} (0.92–0.98 and 0.85–0.80, respectively). If no filtering was applied after Pfam searches, the recall remained unchanged, as expected, while the precision dropped to 0.54 and was also largely insensitive to the E-value threshold. This corresponds to roughly 30% of the hits retrieved by Pfam domains with known structure lacking the expected IBP and thus being discarded by the filtering procedure. Still, Pfam searches are able to identify non-heme iron proteins with a greater than 50% success chance in an ensemble, the entire PDB, containing a number of negative examples overwhelmingly larger than the number of positives. Pfam domain searches retrieved more true non-heme iron proteins than PHI-BLAST, hence the higher recall, whereas the number of false positives retrieved was, after filtering, slightly larger. More than 95% of PHI-BLAST hits were also identified by domain searches. False positives retrieved by Pfam domain searches were most often metalloproteins binding another metal instead of iron.

Searching for Iron-Binding Proteins in the Proteomes

In the 57 selected proteomes, 5398, 1733, and 1318 distinct putative iron-binding proteins have been identified respectively in bacterial, archaeal, and eukaryotic organisms (Tables S4–S6), including 1113 sequences from domains without any structure available (13% of all sequences retrieved). The most common IBP's are found to contain only cysteine ligands. Forty percent of all retrieved proteins had a pattern of this kind. This is not surprising, given the wide diffusion of iron–sulfur proteins. IBP's without cysteines, formed by a combination of histidine and aspartate/glutamate residues, are found in 52% of the putative iron-proteins identified here. IBP's containing cysteines together with other residues account for only 8% of the proteins.

As mentioned, not all the proteins initially retrieved by the structure-based method are identified also by Pfam domain searches coupled with IBP filtering. This can happen when there is sufficient local sequence similarity to a iron-protein of known structure around the IBP in a (family of) protein(s) lacking any structural characterization. The latter sequences may have been retrieved also by one of the Pfam domains lacking a structural representative. This

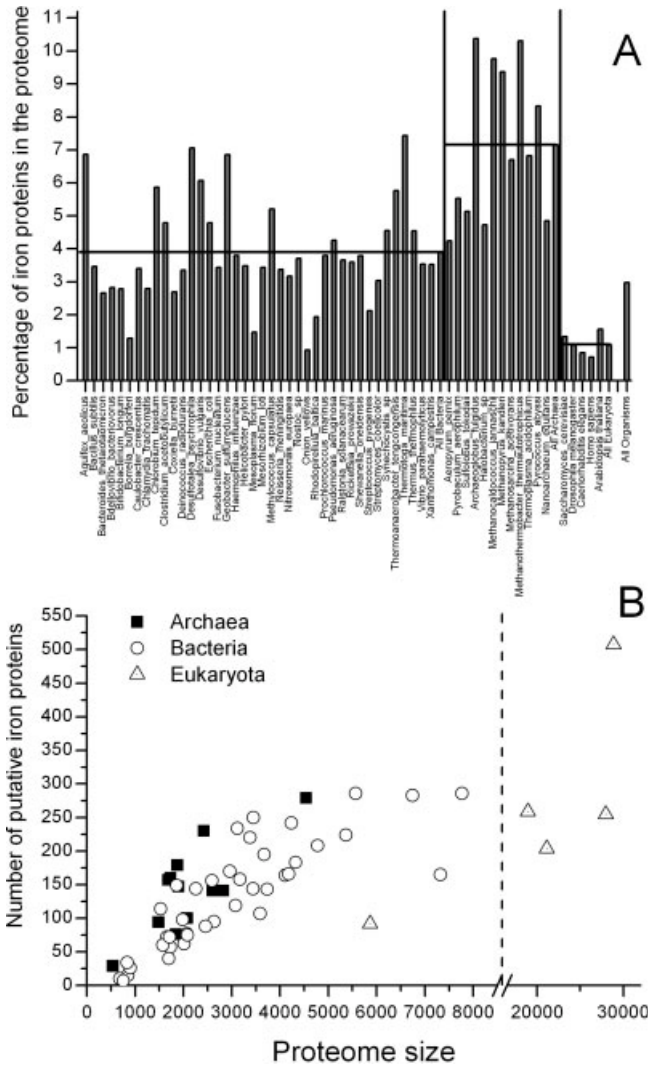

Fig. 2. (**A**) Percentage of iron binding proteins identified in each proteome analyzed. The average values for all bacterial, archaeal, and eukaryotic organisms are also shown. Vertical lines separate the three domains. Horizontal lines show the average values within each domain; (**B**) number of putative iron-binding proteins as a function of the proteome size (filled squares: bacteria, open circles: archaea, dotted triangles: eukaryota).

indeed happened in three instances. In all three cases, inspection of the resulting ensemble of proteins showed that the IBP was present in all proteins. In addition, we also verified that the residues in the IBP were conserved in the starting Pfam HMM, thus allowing us to associate an IBP directly to the Pfam domain.

Distribution of Non-Heme Iron-Proteins in the Three Domains of Life

Figure 2(A) reports the percentage of iron-binding proteins in each of the proteomes analyzed. It can be seen that the share of iron-proteins in each organism is quite variable. Bacterial proteomes contain between 1% and 7% iron-proteins, with an average of 3.9%. Archaeal proteomes have 4% to slightly more than 10% iron-proteins, with an average of 7.1%. Finally, eukaryotic proteomes contain between 0.7% and 1.6%, with an average of 1.1%. In humans, the share of putative iron-proteins identified is only 0.7%, which compares with about 10% predicted human zinc-binding proteins,[16] and an average of 8.8% zinc-proteins predicted for all eukaryota analyzed here.[23] There is an apparent variability in the fraction of the proteome that is non-heme iron-binding among the three domains of Life but also within each domain. This variability is particularly evident for bacteria, which could depend on the extreme variability of their life styles, and, probably to a lesser extent, on the larger number of organisms analyzed. The observed predominance of archaeal proteins did not result from a bias from the original selection of PDB structures, as only 4% of the non-heme iron-proteins of known structure are from archaea, whereas as many as 75% are bacterial.

The number of proteins identified as a function of the proteome size is plotted in Figure 2(B). It can be seen that the number of putative iron-binding proteins correlates well with the size of the proteome in the case of Archaea. Some proportionality is also observed for bacteria, albeit with significant deviations, while there is little, if any correlation, in the case of eukaryota. The bacterium with the highest share of putative non-heme iron-proteins is *Thermotoga maritima*, a non-spore forming, rod-shaped bacterium which was originally isolated from geothermal-heated marine sediment. *T. maritima* is one of the deepest lineages in Eubacteria, and this may suggest that use of iron was most widespread in the early stages of evolution. Conversely, the organisms with the lowest share of putative iron-proteins are parasites such as *Mycoplasma florum* and *Onion yellows phytoplasma*. Among the eukaryota investigated here, the largest fraction of the iron-proteome with respect to the entire proteome is observed in *A. thaliana*.

The putative iron-binding proteins identified have been grouped in 342 ensembles (Table S7), as described in the Methods section. 339 proteins could not be assigned to any ensemble (Table S8). Each ensemble may contain proteins from all three domains of life, or from two domains, or from one domain only. Figure 3(A) shows the fraction of ensembles containing proteins from three, two or just one domain: 17% contain proteins from all domains of life, while 29% contain proteins from both archaea and bacteria, and 24% contain only bacterial proteins. However, the data of Figure 3(A) do not take into account the size of each group. Figure 3(B) therefore shows the relative size of the various kinds of ensembles in terms of their number of iron-binding proteins. The resulting picture is quite different, showing that more than half of the proteins belong to ensembles spanning all three domains of life, while 23% and 13% belong to ensembles spanning respectively archaea and bacteria or eukaryota and bacteria.

Notably, when looking at the "average" non-heme-iron proteome of an organism in each domains of life, it is found that the share of proteins belonging to ensembles spanning all three domains of life is similarly important for all

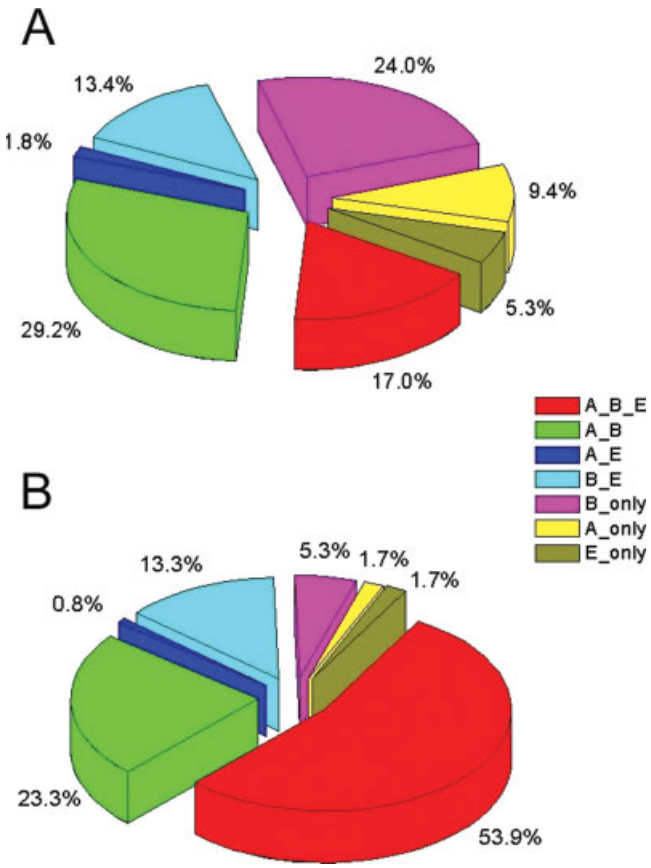

Fig. 3. (**A**) Composition of putative iron-binding protein ensembles with respect to the three domains of life (A: Archaea; B: Bacteria; E: Eukaryota). (**B**) Total fraction of putative iron-binding proteins belonging to ensembles whose members span all three, two or only one domain of life.

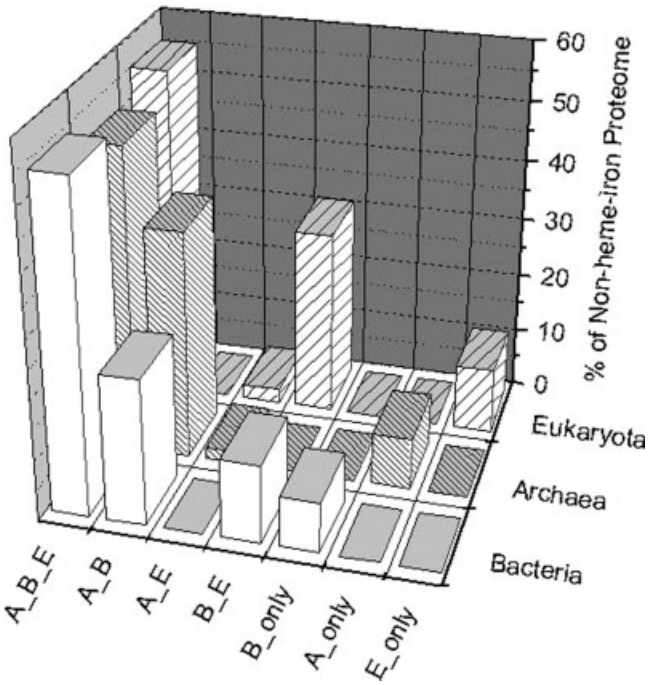

Fig. 4. Fraction (with respect to the entire non-heme-iron proteome) of putative iron-binding proteins in each domain of life belonging to ensembles whose members span all three, two or only one domain (A: Archaea; B: Bacteria; E: Eukaryota).

(Fig. 4). This indicates that the majority of iron-proteins has homologues in all domains of life. By comparing Figure 3(A,B), it appears that the ensembles containing only bacterial proteins (24% of all ensembles) contain only 5.3% of all sequences identified. Iron-binding proteins specific to bacteria constitute on average 10% of the iron-proteome of a bacterium, similarly to what observed for archaea-specific proteins in archaea and for eukaryota-specific proteins in eukaryota (Fig. 4). Ensembles containing domain-specific proteins (bacterial in particular) thus must be more scarcely populated than ensembles spanning all the three domains of life. Altogether, the data of Figures 3 and 4 thus suggest that the majority of putative iron-binding proteins are actually common to the three domains of life, and belong to a relatively restricted number of different families. Indeed, the non-heme-iron-proteome of an organism contains between 50% and 55% of proteins having homologs in the all domains of life (Fig. 4). This observation may be justified by assuming that (most) cellular machineries exploiting iron are evolutionarily ancient, and thus have been vertically inherited from the universal common ancestor by all organisms, whereas there has been little, if any, recruitment of non-heme iron proteins in cellular processes evolved more recently, such as intracellular transport or cell differentiation. This contention would be in agreement also with the fact that the total number of non-heme iron-proteins is relatively similar, between 150 and 300, for the majority of the organisms analyzed in spite of a 30-fold range of proteome sizes [Fig. 2(B)]. Finally, it is to be noted that the number of proteins belonging to ensembles spanning either archaea and bacteria or eukaryota and bacteria is significantly larger than that in ensembles spanning archaea and eukaryota. This is presumably indicative that some differentiation in terms of iron use has occurred between organisms belonging to the latter two domains of life. This differentiation could be a consequence of adaptation to extremely different life styles, which has led to differential gene losses.

Functional Implications

It is intriguing to analyze, at least at a relatively general level, which is the repertoire of functions carried out by the putative iron-proteins identified. Not unexpectedly, the most common functional role for iron-binding proteins is in the catalysis of redox reactions and in electron transfer. The second most common role is in the catalysis of hydrolytic reactions. These observations are true for all three domains of life. Further details can be obtained by the analysis of the content of the putative iron-proteins identified in terms of known domains. Figure 5 shows an overview of this analysis. The iron-proteomes of both bacteria and archaea are enriched in proteins containing Fe_4S_4-ferredoxin domains as well as, to a similar extent, of

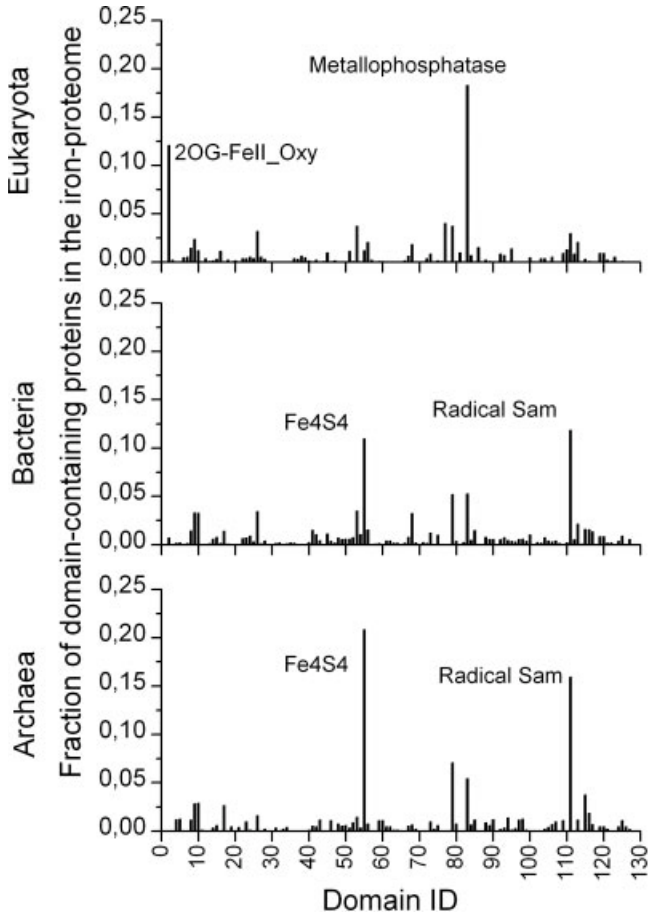

Fig. 5. Fraction of putative iron-binding proteins containing a given Pfam domain in the total iron-proteome of (from top to bottom) Eukaryotic, Bacterial and Archaeal organisms. Pfam domains have been listed alphabetically from 1 to 126 (see Table S3 in the Supplementary Material). Highly frequent domains have been labelled with their Pfam ID.

proteins containing a catalytic domain that, among other reactions, can generate a radical species by reductive cleavage of S-adenosylmethionine. The active site contains a Fe_4S_4 cluster with atypical iron coordination.[24,25] The picture is somewhat different in eukaryota where the most common domains are the metallophosphatase domain and, to a lesser extent, a domain associated with iron-dependent oxygenase activity. The exact nature of the metal ions in the active site of metallophosphatases is still debated. There is agreement that the site contains a pair of divalent metal ions, which could be zinc/iron or manganese/manganese.[26] At present the hypothesis that the relevant metallation state of these proteins is zinc/iron is receiving most support.[26,27]

The analysis of the Gene Ontology terms associated to the iron-proteins retrieved indicated that the range of cellular functions related to the non-heme iron-proteome does not change substantially in the three domains of Life (Figure S3). Between 40% and 45% of the iron-proteins are involved in electron transfer or in catalysing redox reactions. Another 17% are involved in the catalysis of hydrolytic reactions in archaea and bacteria, whereas this role belongs to 27% of eukaryotic proteins. In absolute terms, the latter difference corresponds to the per-organism average number of iron-dependent hydrolases being nearly tripled. Also these data point to the bulk of the non-heme iron-proteome being somewhat conserved in all domains of Life.

CONCLUSIONS

The present work shows that the fraction of the proteome predicted to require non-heme iron to carry out cellular function (the iron-proteome) varies as follows: archaea > bacteria > eukaryota (Fig. 2). For archaea and bacteria there is a fair correlation between the size of the entire proteome and the size of the iron-proteome, in agreement with available studies suggesting that the repertoire of different functions within a given proteome is also proportional to its size.[28]

The majority of the putative iron-binding proteins identified belongs to ensembles that span all three domains of life. On the other hand, the number of iron-proteins that pertain to only one domain is small (Fig. 4). These data may suggest that at least the bulk of proteins requiring iron for their function has appeared quite early during evolution. Thus, the differentiation of the content and function of iron-binding proteins among archaea, bacteria, and eukaryota is probably mainly due to gene loss events associated with changes in environmental conditions and life styles. Indeed, iron proteins are mainly involved in respiratory pathways or in the metabolism of substrates (Fig. 5). Changes in the environment leading to changed ways of respiration and/or to changed availability of nutrients would thus significantly shift evolutionary pressure for conservation of entire metabolic pathways. New functions developed by eukaryotic organisms to better sustain their life style in a mutated environment would likely use metals with better availability in the modern world than iron, leading to a poor expansion or even an erosion (e.g., as for yeast) of their iron-proteome despite the expansion of the whole proteome.

REFERENCES

1. Bertini I, Sigel A, Sigel H. Handbook on metalloproteins, 1st ed. New York: Marcel Dekker; 2001. p 1.
2. Changela A, Chen K, Xue Y, Holshen J, Outten CE, O'Halloran TV, Mondragon A. Molecular basis of metal-ion selectivity and zeptomolar sensitivity by CueR. Science 2003;301:1383–1387.
3. Finney LA, O'Halloran TV. Transition metal speciation in the cell: insights from the chemistry of metal ion receptors. Science 2003;300:931–936.
4. Bertinato J, L'Abbe MR. Copper modulates the degradation of copper chaperone for Cu, Zn superoxide dismutase by the 26 S proteosome. J Biol Chem 2003;278:35071–35078.
5. Mathe C, Sagot MF, Schiex T, Rouze P. Current methods of gene prediction, their strengths and weaknesses. Nucleic Acids Res 2002;30:4103–4117.
6. Milanesi L, D'Angelo D, Rogozin IB. GeneBuilder: interactive in silico prediction of gene structure. Bioinformatics 1999;15:612–621.
7. Wiehe T, Gebauer-Jung S, Mitchell-Olds T, Guigo R. SGP-1: prediction and validation of homologous genes based on sequence alignments. Genome Res 2001;11:1574–1583.
8. Bertini I, Luchinat C, Provenzani A, Rosato A, Vasos PR. Browsing gene banks for Fe_2S_2 ferredoxins and structural modeling of 87 plant-type sequences: an analysis of fold and function. Proteins Struct Funct Genet 2002;46:110–127.

9. Degtyarenko K, Contrino S. COMe: the ontology of bioinorganic proteins. BMC Struct Biol 2004;4:3.
10. Andreini C, Bertini I, Rosato A. A hint to search for metalloproteins in gene banks. Bioinformatics 2004;20:1373–1380.
11. Bertini I, Rosato A. Bioinorganic chemistry in the post-genomic era. Proc Natl Acad Sci USA 2003;100:3601–3604.
12. Scott RA, Shokes JE, Cosper NJ, Jenney FE, Adams MWW. Bottlenecks and roadblocks in high-throughput XAS for structural genomics. J Synchrotron Rad 2005;12:19–22.
13. Hogbom M, Ericsson UB, Lam R, Bakali HM, Kuznetsova E, Nordlund P, Zamble DB. A high throughput method for the detection of metalloproteins on a microgram scale. Mol Cell Proteomics 2005;4:827–834.
14. Shi W, Zhan C, Ignatov A, Manjasetty BA, Marinkovic N, Sullivan M, Huang R, Chance MR. Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics. Structure (Camb) 2005;13:1473–1486.
15. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucl Acids Res 2000;28:235–242.
16. Andreini C, Banci L, Bertini I, Rosato A. Counting the zinc proteins encoded in the human genome. J Proteome Res 2006;5:196–201.
17. Bateman A, Coin L, Durbin R, Finn RD, Hollich V, Griffiths-Jones S, Khanna A, Marshall M, Moxon S, Sonnhammer EL, Studholme DJ, Yeats C, Eddy SR. The Pfam protein families database. Nucleic Acids Res 2004;32 (Database issue):D138–D141.
18. Muller I, Kahnert A, Pape T, Sheldrick GM, Meyer-Klaucke W, Dierks T, Kertesz M, Uson I. Crystal structure of the alkylsulfatase AtsK: insights into the catalytic mechanism of the Fe(II) α-ketoglutarate-dependent dioxygenase superfamily. Biochemistry 2004;43:3075–3088.
19. Eddy SR. Profile hidden Markov models. Bioinformatics 1998;14: 755–763.
20. Zhang Z, Schaffer AA, Miller W, Madden TL, Lipman DJ, Koonin EV, Altschul SF. Protein sequence similarity searches using patterns as seeds. Nucleic Acids Res 1998;26:3986–3990.
21. McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res 2004; 32:W20–W25.
22. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25:25–29.
23. Andreini C, Banci L, Bertini I, Rosato A. Zinc through the three domains of life. J Proteome Res 2006;5:3173–3179.
24. Hanzelmann P, Schindelin H. Crystal structure of the S-adenosylmethionine-dependent enzyme MoaA and its implications for molybdenum cofactor deficiency in humans. Proc Natl Acad Sci USA 2004;101:12870–12875.
25. Sofia HJ, Chen G, Hetzler BG, Reyes-Spindola JF, Miller NE. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. Nucleic Acids Res 2001;29:1097–1106.
26. Swingle MR, Honkanen RE, Ciszak EM. Structural basis for the catalytic activity of human serine/threonine protein phosphatase-5. J Biol Chem 2004;279:33992–33999.
27. Namgaladze D, Hofer HW, Ullrich V. Redox control of calcineurin by targeting the binuclear Fe^{2+}-Zn^{2+} center at the enzyme active site. J Biol Chem 2002;277:5962–5969.
28. Ranea JA, Buchan DW, Thornton JM, Orengo CA. Evolution of protein superfamilies and bacterial genome size. J Mol Biol 2004; 336:871–887.