Metalloproteomes: a bioinformatic approach

Claudia Andreini, Ivano Bertini, Antonio Rosato*

Magnetic Resonance Center (CERM) and Department of Chemistry – University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy Phone: + 39 055 4574270; Fax: + 39 055 4574271.

Claudia Andreini: andreini@cerm.unifi.it

Ivano Bertini: ivanobertini@cerm.unifi.it

Antonio Rosato: rosato@cerm.unifi.it

- Supplementary Figure Captions

- Supplementary Table Captions

- Supplementary methods

SUPPLEMENTARY FIGURE CAPTIONS

Figure S1. Number of putative zinc-binding proteins as a function of the proteome size [7].

Figure S2. Percentages of transition metal-binding (A) proteins and (B) SCOP [27] families in the PDB, as of June 2008. The sum of percentages referring to each metal-ion is larger than the global percentages of transition metalloproteins because of the presence of structures which bind more than one metal ion. The ensemble of transition metal-binding proteins was built as described in the Methods section. In the SCOP classification system, families include proteins that are clearly evolutionarily related, so that the pairwise sequence identity between any pair of proteins within the same family is usually at least 30% [27]. The parseable file in http://scop.mrc-lmb.cam.ac.uk/scop/parse/dir.cla.scop.txt_1.73 was used to associate PDB structures with families.

SUPPLEMENTARY TABLE CAPTIONS

Table S1. List of Pfam metal-binding domains (zinc, copper, non-heme iron)

Table S2. List of organisms analyzed

Table S3. List of putative bacterial metallo-proteins, grouped by organism. The fields of the table contain: (i) NCBI code, (ii) protein length, (iii) brief description as reported in the proteome release, (iv) potential zinc-binding pattern(s) in the sequence (v) domain composition (when a domain is followed by a pattern within brackets, the pattern is localized within the domain)

Table S4. List of putative archeal metallo-proteins, grouped by organism. See Caption to Table S3 for details.

Table S5. List of putative eukaryotic metallo-proteins, grouped by organism. See Caption to Table S3 for details.

SUPPLEMENTARY METHODS

All the protein structures deposited in the PDB as of June 2008 and containing at least one transition metal atom among V, Mo-W, Mn, Fe, Co, Ni, Cu and Zn were downloaded. However, not all the metal-binding structures in the PDB are necessarily metalloproteins (e.g., the metal may be an artefact of the protein crystallization procedure), and also the structures of metalloproteins in PDB may bind metals different from the physiological ones. To distinguish between physiological and non-physiological metal-binding structures we manually checked the primary literature, taking as physiological metals those used by the protein in its physiological context or, in the case of an enzyme, those able to activate it, at least *in vitro*. To reduce the extent of literature mining, metal-binding structures were grouped based on their classification in the CATH [26] and SCOP [27] databases, and literature analysis was conducted on a per-group basis rather than for individual structures. The procedure results in grouping together metal-binding sites of homologous proteins with similar structures and functions, and makes the detection of non-physiological sites easier in that such sites are typically spotted out as the only members of their respective groups. Approximately 15% of the initial ensemble of metalloprotein structures were discarded after examination of the relevant literature.