Highlight Review

A Genomic Frontier in Bioinorganic Chemistry

Ivano Bertini* and Antonio Rosato

(Received May 13, 2004; CL-048005)

Abstract

Genome sequencing projects are providing researchers with an unprecedented wealth of information. This information has the potential to make a deep impact on how experiments are planned and the physiology of living organisms is investigated. One field where the availability of genome sequence has not been thoroughly exploited yet is bioinorganic chemistry. The latter discipline deals with the interaction of biological molecules with inorganic compounds. In this review we give an overview of some aspects of the interplay of genome sequencing projects and bioinorganic chemistry.

♦ Introduction

Nearly ten years have passed since the first complete genome sequence of a bacterium became available.[1] The technology of genome sequencing has progressed tremendously since then, allowing researchers to successfully undertake formidable tasks such as the investigation of the human genome,[2] or the simultaneous determination of the genome sequences of all organisms present in environmental samples (e.g. sea water).[3] These technological advances have resulted in the accumulation of an enormous wealth of data in gene banks (see www.ncbi.nlm.nih.gov/Genomes). The analysis of the above genetic information provides for each organism the complete list (and the primary sequence) of all the proteins that, in principle, an organism can produce.

The problems then arise from attributing a function to each of the proteins potentially produced in the organism, understanding the reciprocal interactions among the various proteins, quantifying the quantity of each protein produced, evaluating the environmental conditions that trigger or inhibit protein production, etc. In higher organism, such as human, the additional issues of tissue-specific expression, and intracellular localization arise, as well as alternative splicing or post-translational modification. All of the above aspects are important for the understanding of the functioning of the entire cells, i.e. of the molecular mechanisms of Life.

In this contribution focus is made on the topic of the interaction between proteins and metals. It is known that a number of proteins in vivo bind metal ions or metal-containing cofactors (such as heme) for their function. Incorporation of the correct metallic cofactor in the correct oxidation state with the correct geometry and stoichiometry is often crucial to the physiological function. Proteins deprived of their metallic cofactor(s) may acquire a number of undesirable features, besides being simply inactive: for example, either they are partly or completely unfolded or may show altered substrate specificity. Alterations in the various intra- and extra-cellular equilibria involving metal ions (such as metal uptake, metal release, metal storage, intra-cellular metal distribution) have been shown in a number of instances to be associated with pathological states.

Researchers can exploit the results of genome sequencing projects to obtain a complete view of the multiple pathways of metal ions inside and outside cells, and of their relevance to good or bad phenotypes. Comparative analysis of the genomes of different organisms can shed light on the molecular aspects of physiological processes, and on how they are influenced by metal ions. In the following, we will discuss the huge potentialities already available in these directions, as well as some of the limitations we are still facing.

♦ The Problem of Genome Annotation

As mentioned in the Introduction, a genome sequencing project provides with the list and the primary sequence of all the proteins that the organism investigated can produce. The question then arises from which function is performed by each protein on the list. If, as it most often happens, there are no experimental data available for a given protein, the only possibility is to try and look for similar proteins that have been experimentally characterized to infer functional information. Here, similar necessarily means "similar in primary sequence," as the primary sequence is all that is known from genome sequencing. In this respect, the most used approach is probably that of detecting homology through sequence alignments; in this approach, the sequence of each protein lacking experimental characterization in the list provided by a newly finished genome sequencing project is individually aligned against a database of protein sequences already characterized (somehow). The quality of alignments is evaluated statistically, and alignments scoring better than a given threshold are then assumed to be indicative of functional homology. Database searches can efficiently be carried out through the program BLAST,[4] or some variant of it.[5] The functional homology detected is used to "annotate" the protein in the genome-associated protein list. For example, if the alignment of the sequence of an uncharacterized protein in a new genome and of a known peroxidase features 40% identical amino acids in aligned positions, then the uncharacterised protein will be dubbed as a (putative) peroxidase. From that moment on, the annotation will always be made available together with the protein sequence.

The approach outlined above is only one of those that can be used to annotate (i.e. attach some information to) uncharacterised proteins whose sequence is predicted as the result of a successful genome sequencing project. However, nearly all of the methods focus on functional rather than chemical features. This may cause some prominent aspects such as metal-binding properties to be completely overlooked, because annotators have

Magnetic Resonance Center, Via L Sacconi 6, 50019 Sesto Fiorentino, Italy

looked e.g. for proteases rather than for the capability of binding zinc (despite the well-documented existence of zinc-binding proteases). While the capability of binding metal ion(s) is clearly not per se of functional relevance, it does provide information on the properties of a protein, as well as indicates the involvement of a protein in the physiological pathways of metal trafficking.[6]

To tackle this latter aspect of genome annotation, we have developed a methodology for the identification of metalloproteins in sequence databanks.[7] The methodology relies on the exploitation of known metal binding-patterns (MBPs), experimentally available from the three-dimensional structures deposited in the Protein Data Bank (PDB).[8] MBPs are strings of the type $AX_nBX_mC\ldots$, where A, B, C, ... are the metal-binding amino acids, and n, m, ... the number of aminoacidic residues in between two subsequent ligands. One can in fact obtain an exhaustive list of binding patterns for a given metal ion (e.g. copper(I)/(II)) by retrieving from the PDB all structures where that metal has been experimentally observed. Then, by measuring all distances between the metal atom(s) contained in the structure and the non-hydrogen atoms of the protein, the donor atoms are readily identified and thereby the aminoacids binding to the metal (i.e. the MBP). This approach is quite similar to that implemented in the Metalloprotein Database, a collection of MBP's automatically extracted from the PDB.[9] To each metalloprotein in the PDB the methodology,[7] therefore, attaches a MBP. The ensemble of the primary sequence of the metalloprotein (the query) from the PDB and of the corresponding MBP is used as input for a variant of BLAST, PHI-BLAST,[10] which uses the MBP as a seed for alignment, to scan gene banks (or a complete genome sequence). In practice, PHI-BLAST extracts from the gene bank all proteins containing the MBP (the hits). Then, for each of these hits individually, the pattern is aligned with that contained in the query sequence, and the alignment extended around the MBP, stopping as soon as there is significant divergence between the two primary sequences. An example of the results obtained form the application of the method to search for copper-binding proteins in four different genomes is shown in Figure 1 (data from).[7] A statistical evaluation of the results obtained with this method when applied to the prediction of metal-binding properties for the proteins in the PDB has yielded quality parameters and ranges of confidence.[7] These however are (possibly) metal-dependent. For copper, it has been found that copper-binding capability can be predicted with good confidence when the amino acid identity around the MBP is at least 20%.[7] However, it is important to stress that there is significant overlap between the coordination properties of various metal ions of the 3*d* series, as extensively proved by metal substitution techniques.[11,12] Therefore, the methodology described above is more correctly described as being capable to predict metal-binding properties in vitro rather than in vivo. Indeed, the results indicate that the search for copper-binding protein in fact can retrieve a number of sequences that are more homologous to zinc-binding proteins.[7] Possible improvements of the above methodology by coupling it with secondary structure prediction methods or the evaluation of amino acid chemical similarity around the MBP are being explored in our lab.

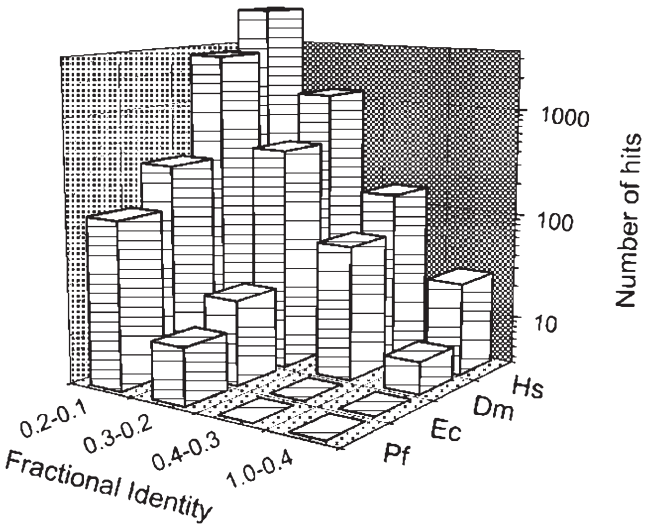

Figure 1. Number of hits retrieved by searching for copper-binding patterns in the complete genome sequences of *Pyrococcus furiosus (Pf)*, *Escherichia coli (Ec)*, *Drosophila melanogaster (Dm)*, and *Homo sapiens (Hs)* as a function of the value of sequence identity to the query PDB proteins around the MBP. The value for the human genome and identity between 0.1 and 0.2 (7627) is left out of scale for clarity. Note the logarithmic scale on the z axis.

♦ The Biological Role of Copper

Even if one had acquired knowledge of all the metalloproteins encoded in a genome, the question of their functional properties and reciprocal interactions would still be completely open. In this respect, an important biological problem that has received much attention recently is that of the control of intra-cellular metal concentration and localization, which is also intimately related to the control of the uptake, release and delivery to metalloproteins of metal ions. Copper is the metal for which most is known on the above topics. In fact, a quite detailed picture is available for what concerns copper homeostasis in yeast, from a number of biological, genetic, biochemical and structural data.[13–16] Bioinformatic analyses have shown that the knowledge available for yeast is in good part transferable also to other systems, such as bacteria or human.[17–19] However, even when obvious similarities and analogies are detected, there are also a number of variations that must be taken into account to properly understand the physiology of the living cell. A typical example is that of the comparison of the yeast and human copper uptake and transport system, which will be discussed in some detail in the next paragraph. It is also important to keep in mind that bacterial organisms that have evolved in very different environmental conditions may have evolved quite different systems to cope with, for instance, very high or very low metal bioavailability. Thus, there must be unrelated biological pathways in different organisms that can only be individually studied.

As mentioned above, there is a clear homology but also significant differences between the pathways for copper uptake and delivery to the trans-Golgi network in human and yeast. Several of these differences belong to the trans-membrane ATPase that pumps copper(I) from the cytoplasm into the trans-Golgi network. First of all, there is only one such enzyme in yeast (Ccc2) but two in humans (ATP7A, aka Menkes protein, and ATP7B, aka Wilson protein). There is significant sequence similarity among these three proteins. ATP7A is expressed in essen-

Published on the web (Advance View) July 5, 2004; DOI 10.1246/cl.2004.946

tially all tissues, while ATP7B is predominantly expressed in the liver. Second, human ATP7A and ATP7B both contain six cytoplasmic soluble domains (each potentially able of binding one copper(I) ion) against two in yeast Ccc2. The solution and solid state structure of several of these domains are available, both in the apo- and metal-loaded forms,[20–22] and indicate a substantial structure similarity also regardless of the organism. Third, ATP7A and ATP7B but not Ccc2 can relocate from the trans-Golgi membrane to the plasma membrane, where their function becomes that of exporting copper from the cytoplasm into the extracellular media.[23,24] Finally, more differences can be identified among the various soluble copper(I)-binding domains. Indeed, the two yeast domains are both strongly negatively charged (predicted pI 4.2–4.4). Instead, the six domains in ATP7A and ATP7B are much more different from one another: only domain IV in ATP7A and domains IV and V in ATP7B are as much negatively charged as the yeast domains. The other domains have pIs ranging from 5.0 to 7.0, with the notable exception of domain II in both proteins which is positively charged (pI 8.7 in ATP7A, 8.1 in ATP7B). The two physiological partners of the three ATPases are HAH1 in human (which can transfer copper(I) to both ATP7A and ATP7B) and Atx1 in yeast. HAH1 and Atx1 have a sequence identity of 38%, a very similar three-dimensional structure (Figure 2)[25] (Unpublished data from our laboratory), and pIs of, respectively, 6.7 and 8.6. Notwithstanding the difference in pI, the electrostatic potential of HAH1 and Atx1 at the protein surface in the region involved in the intermolecular interaction with the partner ATPase (determined by NMR for the yeast proteins)[26] is fairly similar (Figure 3, Top). The electrostatic potential at the corresponding surface of the various human ATPase domains is instead quite variable, and can be somewhat different from that observed in the two domains of Ccc2 (Figure 3, Bottom).

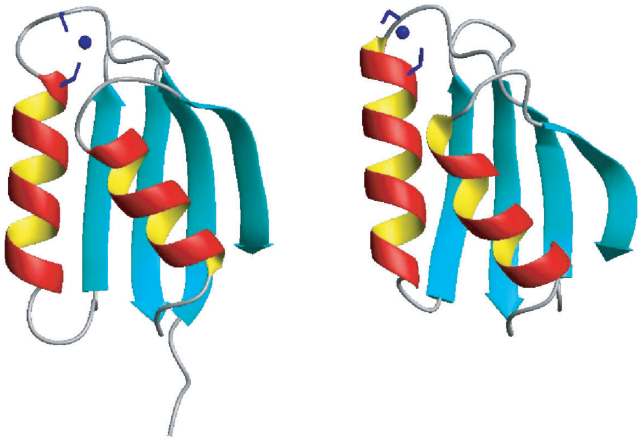

Figure 2. Comparison of the average solution structures of yeast copper(I)-Atx1 (left) (25) and human copper(I)-HAH1 (right) (Unpublished data from our laboratory). The copper site is shown.

All of the above differences concur to make the physiological role of ATP7A and ATP7B different, and more difficult to fully understand than for Ccc2. In particular, the role of the additional domains is quite unclear. Indeed, there are several open questions such as, "Is there a preferential point for the initial copper uptake from HAH1 or are all six domains equally competent (in ATP7B it appears that domain II may be the entry point)[27]? What is the extent of inter-domain interactions, and how do they depend upon the metallation state of ATP7A/ATP7B (in this respect, it has been proposed that inter-domain interactions may regulate the translocation of the protein from the trans-Golgi network to the plasma membrane)?"[28,29] Understanding in detail the answers to these and other questions is important to the understanding of the molecular bases of the physiology of the cell as well as of pathological states arising from malfunctioning of these pathways. It is known that a number of mutations in either ATP7A or ATP7B can give rise to serious illnesses. Mutations in ATP7A cause so-called Menkes disease, which is associated to defective intestinal absorption of copper.[30] Mutations in ATP7B cause so-called Wilson disease, which is associated to abnormally high hepatic copper content. In both cases most of the disease-causing mutations are located in the transmembrane regions of the protein; however some mutations also occur in the cytoplasmic domains or in the linker regions connecting them. The molecular mechanisms by which these mutations impair the normal biological function of ATP7A/ATP7B are uncharacterised.

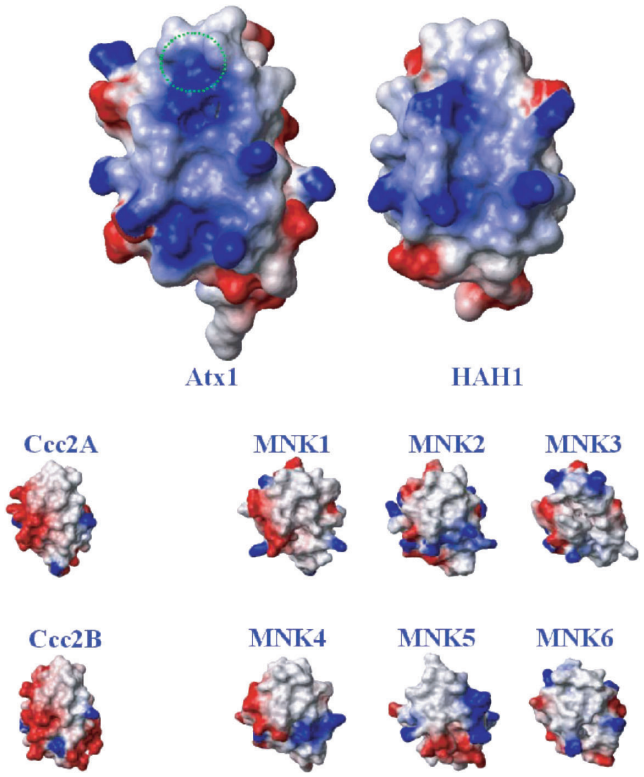

Figure 3. (Top) Electrostatic surface potential of yeast Atx1 and human HAH1. The protein face involved in the interaction with the partner is shown. A green circle highlights the metal-binding site. (Bottom) Electrostatic surface potential of the two soluble copper(I)-binding domains of yeast Ccc2 and of the six soluble copper(I)-binding domains of human ATP7A (Menkes protein). The protein face involved in the interaction with the partner is shown.

♦ Matrix Metalloproteinases: The Problem of Selectivity and Specificity

One of the major expectations from genome sequencing project has been in terms of a significant impact on how drugs are thought, developed and used. The underlying idea is that by gaining access to information about all the proteins that an organism can produce, it becomes possible to select for the most appropriate targets for pharmaceutical therapy (e.g. affecting a key metabolic process responsible for a disease, or against a pathogen by blocking one of its vital processes) or to identify proteins potentially interfering with the therapy (e.g. human proteins binding a drug aimed at a different protein). Clearly, in this frame the availability of the human genome is of utmost importance. When one aims at analyzing the role of protein within a pharmaceutical therapy, it is necessary to couple sequence information to structural information. The latter is needed to identify and characterize binding pockets and to evaluate e.g. the energetics of binding by a small compound.

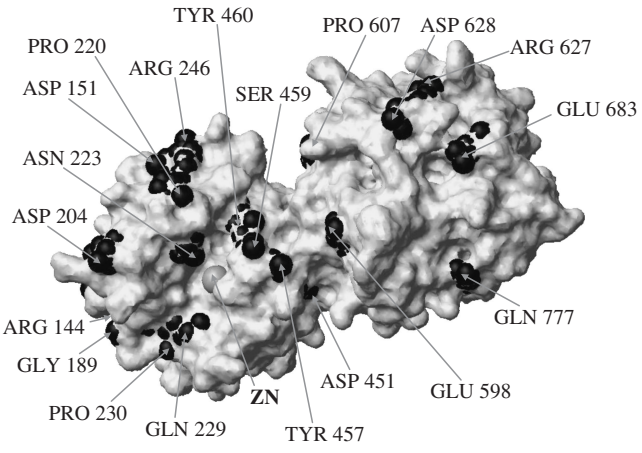

Figure 4. The location of residues showing conservation/variation patterns distinctive of the different MMP sub-families, and thus potentially important for tuning MMP specificity, is mapped onto the Van der Waals surface of MMP1.

A possible approach is that of searching the human genome sequence for all proteins that bear sequence similarity to the protein intended as pharmaceutical target (either from a pathogen or from a human metabolic pathway that one wants to regulate). This provides a collection of human proteins that should also bear structural similarity to the target, and can be done using the tools described in the first two sections. An example can be that of matrix metalloproteinases (MMPs), a family of zinc proteases which are involved in the degradation and remodelling of the extracellular matrix (ECM). This process is of crucial importance for the development and morphogenesis of tissues. MMPs have an important role in tumor progression, and because of this have been extensively investigated as targets for drug therapy.[31,32] By analysing the human genome, 23 different MMPs are identified, together with six additional isoforms.[33] Phylogenetic analysis shows that this ensemble can be subdivided in a few sub-families; this subdivision correlates well with the organization of the proteins into domains (with few exceptions). In addition, the subdivision correlates with potential differences in the mechanisms of regulation of the catalytic activity, suggesting that it reflects a true physiological differentiation.[33]

The availability of solid state and/or solution structures for a few MMPs allows the prediction of the 3-D structure also for those MMPs lacking this information. In this case the method of choice for building structural models is homology modelling.[34] It is important to note that most MMP structures available only contain the catalytic domain, while only two structures are available for full-length MMPs.[33] This can limit the applicability of homology modelling methods: indeed while it is possible to build good (see[34] for a definition of what good means) models for the catalytic domains of all 23+6 MMPs above, the same can be done for only a dozen of proteins in the case of full-length structure prediction. The analysis of surface properties, of the shape, size and amino acid composition of the catalytic pocket provides hints on the selectivity towards different substrates or different inhibitors (Figure 4).[33] The availability of structural models can be exploited in several ways, e.g. to help in the interpretation of experimental screening data or to design mutant proteins with improved properties.[35] However, the main interest is possibly that of in silico screening of libraries of small compounds to look for candidate drugs.

♦ Perspectives in Bioinorganic Chemistry

The above selected topics represent only a very minor part of the whole field of bioinorganic chemistry but provide hints for a new methodological approach. What matters is that any selection of a bioinorganic research starts from a genomic approach, i.e. studies of the occurrence among living organisms (at the moment the number of complete genome sequences available for prokaryotes and eukaryotes are 182). Furthermore, the analysis of the genomic context in bacteria (i.e. the organization of genes within operons) provides hints to understand the metabolic process in which the protein is involved. Only after genomic analyses, the research project can be planned.

One should consider that metalloproteins containing copper, zinc and iron can easily be 20% of the whole human genome. If other metal ions are considered (e.g. calcium, magnesium, sodium, potassium, etc.), then the field seems very broad. Indeed, bioinorganic chemistry has a big challenge ahead!

References

1 R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J.-F. Tomb, B. A. Dougherty, J. M. Merrick, K. McKenney, G. Sutton, W. FitzHugh, C. Fields, J. D. Gocayne, J. Scott, R. Shirley, L.-I. Liu, A. Glodek, J. M. Kelley, J. F. Weidman, C. A. Phillips, T. Spriggs, E. Hedblom, M. D. Cotton, T. R. Utterback, M. C. Hanna, D. T. Nguyen, D. M. Saudek, R. C. Brandon, L. D. Fine, J. L. Fritchman, J. L. Fuhrmann, N. S. M. Geoghagen, C. L. Gnehm, L. A. McDonald, K. V. Small, C. M. Fraser, H. O. Smith, and J. C. Venter, *Science*, **269**, 496 (1995).

2 J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira,

X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. G. Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R.-R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Y. Wang, A. Wang, X. Wang, J. Wang, M.-H. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. C. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y.-H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C. Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigó, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y.-H. Chiang, M. Coyne, C. Dahlke, A. D. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh, and X. Zhu, *Science*, **291**, 1304 (2001).

3 J. C. Venter, K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith, *Science*, **304**, 66 (2004).

4 S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman, *J. Mol. Biol.*, **215**, 403 (1990).

5 S. F. Altschul, T. L. Madden, A. Schaeffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman, *Nucleic Acids Res.*, **25**, 3389 (1997).

6 I. Bertini and A. Rosato, *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 3601 (2003).

7 C. Andreini, I. Bertini, and A. Rosato, *Bioinformatics*, **20**, 1373 (2004).

8 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.*, **28**, 235 (2000).

9 J. M. Castagnetto, S. W. Hennessy, V. A. Roberts, E. D. Getzoff, J. A. Tainer, and M. E. Piquet, *Nucleic Acids Res.*, **30**, 379 (2002).

10 Z. Zhang, A. A. Schaffer, W. Miller, T. L. Madden, D. J. Lipman, E. V. Koonin, and S. F. Altschul, *Nucleic Acids Res.*, **26**, 3986 (1998).

11 I. Bertini, C. Luchinat, and M. S. Viezzoli, in "Zinc Enzymes," ed. by I. Bertini, C. Luchinat, W. Maret, and M. Zeppezauer, Birkhauser, Boston (1986).

12 L. Banci and M. Piccioli, in "Encyclopedia of Magnetic Resonance," ed. by D. M. Grant and R. K. Harris, Encycl. of Nuclear Magn. Reson. (1996).

13 S. Puig and D. J. Thiele, *Curr. Opin. Chem. Biol.*, **6**, 171 (2002).

14 A. C. Rosenzweig, *Acc. Chem. Res.*, **34**, 119 (2001).

15 L. Banci and A. Rosato, *Acc. Chem. Res.*, **36**, 215 (2003).

16 L. A. Finney and T. V. O'Halloran, *Science*, **300**, 931 (2003).

17 C. Askwith and J. Kaplan, *Trends Biochem. Sci.*, **23**, 135 (1998).

18 K. Jordan, D. A. Natale, E. V. Koonin, and M. Y. Galperin, *J. Mol. Evol.*, **53**, 622 (2001).

19 F. Arnesano, L. Banci, I. Bertini, S. Ciofi-Baffoni, E. Molteni, D. L. Huffman, and T. V. O'Halloran, *Genome Res.*, **12**, 255 (2002).

20 J. Gitschier, B. Moffat, D. Reilly, W. I. Wood, and W. J. Fairbrother, *Nat. Struct. Biol.*, **5**, 47 (1998).

21 C. E. Jones, N. L. Daly, P. A. Cobine, D. J. Craik, and C. T. Dameron, *J. Struct. Biol.*, **143**, 209 (2003).

22 L. Banci, I. Bertini, R. Del Conte, M. D'Onofrio, and A. Rosato, *Biochemistry*, **43**, 3396 (2004).

23 M. J. Petris, J. F. Mercer, J. G. Culvenor, P. Lockhart, and J. Camakaris, *EMBO J.*, **15**, 6084 (1996).

24 M. Schaefer, R. G. Hopkins, M. L. Failla, and J. D. Gitlin, *Am. J. Physiol.*, **276**, G639 (1999).

25 F. Arnesano, L. Banci, I. Bertini, D. L. Huffman, and T. V. O'Halloran, *Biochemistry*, **40**, 1528 (2001).

26 F. Arnesano, L. Banci, I. Bertini, F. Cantini, S. Ciofi-Baffoni, D. L. Huffman, and T. V. O'Halloran, *J. Biol. Chem.*, **276**, 41365 (2001).

27 J. M. Walker, D. Huster, M. Ralle, C. T. Morgan, N. J. Blackburn, and S. Lutsenko, *J. Biol. Chem.*, **279**, 15376 (2004).

28 R. Tsivkovskii, B. C. MacArthur, and S. Lutsenko, *J. Biol. Chem.*, **276**, 2234 (2001).

29 D. Huster and S. Lutsenko, *J. Biol. Chem.*, **2789**, 32212 (2003).

30 D. M. Danks, P. E. Campbell, B. J. Stevens, V. Mayne, and E. Cartwright, *Pediatrics*, **50**, 188 (1972).

31 L. M. Coussens, B. Fingleton, and L. M. Matrisian, *Science*,

295, 2387 (2002).
32 C. M. Overall and C. Lopéz-Otín, *Nat. Rev. Cancer*, **2**, 657 (2002).
33 C. Andreini, L. Banci, I. Bertini, C. Luchinat, and A. Rosato, *J. Proteome Res.*, **3**, 21 (2004).
34 M. A. Marti-Renom, A. C. Stuart, A. Fiser, R. Sanchez, F. Melo, and A. Sali, *Annu. Rev. Biophys. Biomol. Struct.*, **29**, 291 (2000).
35 L. Banci, I. Bertini, A. Ciulli, M. Fragai, C. Luchinat, and B. Terni, *J. Mol. Catal. A: Chem.*, **204–205**, 401 (2003).