Structural Genomics of Proteins Involved in Copper Homeostasis

LUCIA BANCI* AND ANTONIO ROSATO

Magnetic Resonance Center, University of Florence, Via Luigi Sacconi 6, and Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019, Sesto Fiorentino, Italy

Received June 18, 2002

ABSTRACT

Genome sequencing projects have provided a wealth of data, most notably the primary sequences of all the proteins that a given organism can produce. The understanding of this information at the functional level is still in the beginning stages. Three-dimensional structural information is necessary to unravel at the atomic level the mechanisms by which a protein carries out its function, and such information can often be very useful to predict at least gross functional features, even in the absence of biochemical data. An exhaustive structural characterization of the proteins encoded in the genomes is thus highly desirable. To enhance the functional insights provided by genome-scale structural determination, we have prioritized our research to target specific processes of the cell, i.e., those responsible for controling metal homeostasis. In this Account, we present the results obtained by the Magnetic Resonance Center of the University of Florence on proteins involved in the homeostasis of copper. The general research strategy is presented, followed by a discussion focused on different key experimental aspects. An overview of the initial results and of their relevance to the understanding of molecular function and cellular processes is also given.

Introduction

After the first successful completion of the genome sequencing of a bacterium, that of *Haemophilus influenzae* in 1995,[1] the number of similar or even more ambitious projects being undertaken worldwide has been increasing steadily. In parallel, the time required to determine, e.g., the complete sequence of the genome of a bacterium has diminished considerably. With genome sequences coming by the dozen, the fundamental problem of attaching functional information to each of the proteins encoded in the various genomes becomes a crucial one. The possibility of inferring from experimental three-dimensional structural data functional information more reliable than that guessed from, e.g., sequence comparison methods alone has prompted worldwide efforts aimed at solving protein structures on a genome-wide scale (for an overview, see the special supplement of *Nature Structural Biology* on Structural Genomics, Vol. 7, No. 11, 1999). These efforts are often collectively referred to as "Structural Genomics" or "Structural Proteomics". It is to be noted that these efforts are aimed at very large scale production, which can be attained only by improving dramatically upon the time scale needed for all the steps, from protein expression/purification to sample/crystal preparation to structure determination. To fulfill this high-throughput (HTP) requirement, it is necessary to exploit parallel processing methods and automation as much as possible.

On the other hand, even for proteins for which direct functional information is available (e.g., from biochemical or genetics approaches), there is a high added value in the availability of structural information, which can provide a deeper understanding of the mechanisms by which a protein carries out its function. Structural information is therefore crucial also for the understanding of protein misfunction. In addition, experimental structures can be exploited to build computer models of the structures of other, uncharacterized proteins.[2] To do this on a genome scale, it is necessary to have experimental structures of different types of proteins, possibly from different types of organisms, so as to have enough structural templates for all of the proteins.[3] Traditionally, structural biologists have chiefly tackled the study of proteins for which detailed biochemical/biological information was available, by using a rich portfolio of biophysical techniques as well as approaches borrowed from molecular biology (most notably, site-directed mutagenesis). The main new twist in this scientific area to be gained from Structural Genomics is the concept of HTP that is inherent to the latter, and particularly the ideas of parallel processing methods and automation. Examples are the parallel analysis of series of homologous proteins from different organisms, or of libraries of computer-designed mutants.

Here we present our approach to Structural Genomics and its application to one of the projects presently being developed at the Magnetic Resonance Center (CERM) of the University of Florence.

The Philosophy and the Strategy

The motto of the Florence laboratory is, "From gene to function through the structure". We want to reconcile the HTP approach to structure determination with the research aimed at understanding the function and unraveling the physiological processes within the cell. We are, in fact, implementing different projects with the goal to provide new insights into the role of metal ions in prokaryota and eukaryota, by elucidating not only the structures of metal-binding proteins but also their network of interactions with other proteins and, possibly, with nucleic acids. In our projects, the criteria for target

Lucia Banci is Professor of Chemistry at the University of Florence, Italy. Her research interests are devoted to structural characterization of biological molecules through NMR spectroscopy, the methodological aspects of its application to proteins containing paramagnetic metal ions, computational simulations of proteins structure and dynamics, and genome browsing and functional genomics.

Antonio Rosato obtained his Ph.D. degree from the University of Florence, Italy, in 1998. He is now Associate Professor of Chemistry at the same university. His research interests focus on the mechanisms of folding/unfolding and the factors controlling the stability of metalloproteins, and on the theory and practice of NMR spectroscopy applied to paramagnetic metalloproteins.

* Address correspondence to this author at Magnetic Resonance Center, University of Florence. Tel.: +39 055 4574263. Fax: +39 055 4574253. E-mail: banci@cerm.unifi.it.

10.1021/ar010120r CCC: $25.00
Published on Web 01/10/2003

selection are thus largely determined by the function of the protein, at variance with other Structural Genomics projects, which are usually focused on specific organisms or on expanding the so-called fold library by determining new protein folds. Consequently, it can be said that our projects tend toward functional genomics.

The general strategy of our project can be schematized as below:

1. Genome browsing in bacteria and eukaryotes
2. Phylogenetic comparative analysis
3. Protein selection and expression
4. Structure determination by NMR and/or X-ray
5. Modeling
6. Protein−protein interaction and function

Here, we present an overview of the results obtained by applying the concept of HTP structure determination of proteins involved in copper homeostasis. The work carried out not only provides a structural characterization of these proteins but also adds a functional dimension, mainly by characterizing the interaction with their substrate, copper(I) and/or copper(II), and with their physiological partners.

Copper homeostasis is the result of a very complex ensemble and interaction of a number of cellular processes. Several aspects of the metal homeostasis need to be tightly controlled. First, the concentration of free copper ions should be negligible, as the copper ions are highly toxic due to their redox properties, e.g., by catalyzing the formation of radicals which can damage the cell. The copper ions in the cell are indeed sequestered from solution, being bound to proteins. On the other hand, newly produced copper-binding proteins need to uptake copper ions to achieve their active form. As the free copper concentration in the cell is very low ($<10^{-18}$ M),[4] a system permitting rapid and efficient metal transfer and preventing nonspecific reactions involving copper must be available. Thus, the function of some of the proteins involved in copper homeostasis is that of "chaperoning" the copper ion to copper-dependent proteins through very selective protein−protein interactions. The unraveling of this delicate interplay has started only recently.[5−8] The focus of our project is on proteins involved at any level in copper homeostasis of both prokaryota and eukaryota. On one hand, this definition includes essentially all copper-binding proteins in any organism. However, we have devoted special attention to proteins actively involved in the mechanisms of copper uptake and release, transport, and resistance. These proteins can, in fact, be regarded as the keepers of copper homeostasis in organisms, while proteins requiring copper in order to perform their function are more akin to passive users of the system. This approach requires that, for each target protein selected (see below), quite extensive literature searches be performed to acquire all the available information on its physiological role.

Genome Browsing. Because our project goes across all organisms, step 1 is carried out using all the genetic information available, i.e., taking into account all available genome sequences, as well as the entries in all gene banks. To search for copper-binding proteins (as mentioned above, any copper-binding protein is intrinsically of relevance to the present project), gene annotations in the genomes are scanned systematically. The sequences of the proteins which are potentially copper-binding are then used as input to the program BLAST[9] to identify homologues in the gene data banks. These sequence alignments usually permit the identification of the metal-binding site(s) as the region most conserved. The metal-binding consensus sequence can then be used to perform PHI-BLAST[10] searches, in which more distant sequence relationships can be looked for, with the restraint that the consensus sequence identified be conserved. With this procedure, for any protein identified from genome browsing as described above, a family of protein sequences with the same consensus sequence for metal binding and at least distant sequence similarities (i.e., sequence identity greater than 10%) is thus built. Of course, each family can contain more than one of the proteins independently identified by browsing genome annotations. It is important to note that, for our purposes, we chopped multidomain proteins into single domains and used the latter as if they were individual proteins in all our bioinformatic analyses. Due to our structural approach, i.e., in solution through NMR spectroscopy, we also limited our selection criteria to proteins or protein domains of less than 200 amino acids. For each family, all the sequences were then used to build sequence alignments and phylogenetic trees, to separate the proteins into different clusters. The subgroups identified may actually reflect functional differentiations. A number of homologues from each subgroup was selected for the subsequent steps. The use of homologues from different organisms improves the chances of success in getting good expression of soluble and folded proteins for each protein subgroup.[11] For each subgroup, one experimental structure is sufficient to generate reliable structural models based on homology modeling for all the proteins in it.[2] Some degree of redundancy may, however, be desirable and is useful in deriving functional and mechanistic insights at the atomic level.

Protein Expression. Protein cloning and expression has been carried out using pET vectors and BL21(DE3) or BL21(DE3)pLysS *Escherichia coli* strains. For protein purification, a His_6 tag was fused at either the N- or the C-terminus of the target protein, and a cleavage site for Factor Xα was introduced between the tag and the target protein, in order to be able to remove the tail. When working with metalloproteins, this step is particularly important, in order to guarantee that there is no interference of the His_6 tag with respect to metal binding. Proteins were purified by affinity chromatography. A second step of affinity chromatography *after* cleaving the His_6 tag was particularly effective in getting rid of impurities that coeluted with the tagged protein in the first step. We have found that, in several instances, performing the purification under reducing conditions and/or in the absence of O_2 significantly improved the yield and the behavior (most notably foldedness) in solution of the proteins. To rationalize this, it is important to keep in mind that copper ions,

particularly in the case of copper(I), are often bound to cysteines by coordinating to their Sγ atoms. Since the purification procedure involves depletion of the proteins from any metal ion through treatment with a large excess of EDTA, the thiols in the metal-binding region may, under oxidizing conditions, form intra- or intermolecular disulfide bridges. The formation of intermolecular disulfide bridges can then result in lowered protein solubility. Use of excess reductant (e.g., DTT) or removal of O_2 at all stages of the purification can avoid this problem by keeping the thiols in the reduced form, thus preventing the formation of disulfide bridges. When using NMR spectroscopy for structure determination, protein labeling with stable isotopes such as ^{15}N and ^{13}C is highly advisable.[12] To do this, cells are grown in minimal medium using isotopically enriched nitrogen and carbon sources, or in an isotopically enriched complex medium. We have not observed a general trend for the dependence of the protein yield on the particular growth media used. Therefore, we routinely assayed the yield in both media on small-scale cultures (typically 100 mL) by SDS−PAGE electrophoresis and then performed large-scale cultures in the medium that showed the best performance.

Structure Determination. Structure determination by NMR has been done following standard methods,[12,13] and thus it is not necessary to describe it here. For structure determination by NMR, the "foldedness" of proteins was checked through $^{1}H-^{15}N$ HSQC spectra. Usually proteins were dissolved in phosphate or acetate buffer. pH and ionic strength were systematically varied to achieve optimal foldedness, as judged by $^{1}H-^{15}N$ HSQC spectra.

The use of NMR spectroscopy for structure determination with respect to X-ray is particularly suitable for copper-transporting proteins, which in most cases bind the metal ion through the sulfur atoms of Cys residues. In the metal-free form, the cysteines are easily oxidized and form disulfide bridges. On the other hand, the copper-bound form experiences partial loss of copper during crystallization, leading to a disordered copper site in the solid-state structure, which is the result of the presence of mixtures of oxidized and reduced apo-protein, as well as metalated protein.[14−16] Indeed, at present a few structures of proteins involved in copper homeostasis have been determined also through X-ray.[14,15,17−20] However, either the copper ion had been replaced with nonphysiological metal ions or the apo form was oxidized, i.e., with a disulfide bridge. In both cases, the structures could not provide information on the structural properties of the metal site in a biologically relevant state. Indeed, the first structure ever reported for a protein involved in copper homeostasis (i.e., the fourth metal-binding domain of the soluble portion of the so-called Menkes protein) has been solved through NMR.[21] The metal bound was Ag(I), thus not relevant from the physiological point of view, but still this structure revealed for the first time the fold of this class of proteins.

Modeling. Step 5 above has been performed, as mentioned, by homology modeling methods, using standard procedures,[2] with the only exception that particular care was taken in modeling the metal-binding site (see ref 22 for details). Analysis of the modeling results can provide, in addition to common structural features, interesting functional hints, for instance, on the factors determining metal ion affinity and selectivity, or on the factors responsible for interaction with physiological partners.[22]

Functional Analysis. Finally, we deemed it particularly important in a project aimed also at the analysis of functional details of the proteins investigated to characterize the interaction of protein partners along the copper delivery/binding/removal pathway. Indeed, we feel that the picture of copper homeostasis can be complete only upon understanding the atomic details of how the copper ion(s) is passed from one transporter to another and to the final protein targets (the passive users). Therefore, we set out to characterize protein−protein interactions at the atomic level by means of NMR spectroscopy, an extraordinary and probably unique tool in this respect.[23] Protein−protein interaction studies typically were performed by titrating a stable isotope-enriched protein with its unlabeled (putative) partner and analyzing (i) chemical shift variations and the appearance of new peaks along the titration; (ii) variation of ^{15}N R_1 and R_2 relaxation rates; (iii) variation of intramolecular NOEs; and (iv) the occurrence of intermolecular NOEs. These measurements are performed with the first protein labeled and the second unlabeled, and vice versa, so that information on both partners is obtained. Also, different combinations of apo- and holo-proteins are used. The latter experiments are of importance to understand the flow of the metal ions along a chain of transporters. The above data provide information on the three-dimensional structure of protein−protein adducts at the atomic level.[23]

Results

Target Selection and Expression. From our bioinformatic analyses, and also on the basis of literature data, we were able (as of April 2002) to identify 350 proteins as being either copper-binding or involved in copper homeostasis (Figure 1). As already mentioned, we excluded from this selection the proteins constituting final "users" of the copper ion. This ensemble included proteins quite well characterized from the biochemical and genetic point of view as well as proteins with no functional annotation at all. The proteins were divided into 20 groups on the basis of sequence alignments and phylogenetic trees. For each group, a number of targets between one and four was then selected on the basis of features such as predicted stability, protein size, p*I*, etc. When possible, targets were selected from more than one organism. Proteins from eukaryota and prokaryota with similar sequences were grouped into a single family; for these families, targets were selected from both eukaryota and prokaryota.

In total, 39 targets were selected for expression (Figure 1), of which 20 could be produced in *E. coli* in good quantities, as monitored by SDS−PAGE electrophoresis of whole cells. The targets showing good expression could be purified to at least 90% purity by a combination of

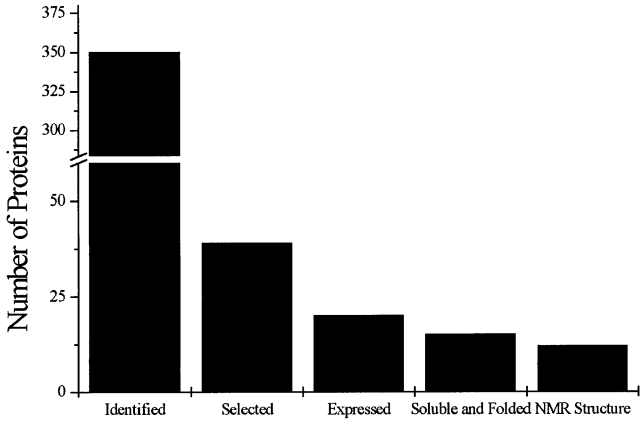

FIGURE 1. Summary of the results obtained at various stages of the project described herein.

affinity chromatography and, in some instances, size exclusion chromatography. $^{1}H-^{15}N$ HSQC[24] spectra for all of these were acquired in order to check the foldedness of the protein. Whenever feasible, we first checked the apo-protein and then proceeded on to investigating metal-binding properties as well as the NMR features of the holo-protein. Two of the expressed proteins could only be purified as holo-proteins. Metal-binding properties were checked by atomic absorption, EXAFS, UV−visible spectroscopy, and, when appropriate, EPR. Atomic absorption and EXAFS are exquisitely complementary in characterizing metal-binding properties, as the first technique provides information about the stoichiometry of binding, while the second provides information about the donor atoms to the metal. All proteins showing good $^{1}H-^{15}N$ HSQC spectra were retained for structure determination by NMR. A $^{1}H-^{15}N$ HSQC spectrum was classified as good when it showed dispersed signals with line widths in keeping with what was expected on the basis of protein size, and having a number of peaks between 80% and 120% of what was expected (i.e., we retained systems with multiple conformations in solution, provided that this phenomenon appeared to be limited to a relatively small part of the protein). For those systems which did not show a good $^{1}H-^{15}N$ HSQC spectrum, different buffer conditions (pH, ionic strength) were explored to improve foldedness. In five cases, we could not obtain good $^{1}H-^{15}N$ HSQC spectra under any of the experimental conditions assayed. Finally, the number of proteins for which until now we could obtain good $^{1}H-^{15}N$ HSQC spectra, and which were thus subjected to structure determination, was 15 (Figure 1). For one system with $^{1}H-^{15}N$ HSQC spectra typical of a largely unstructured protein (CopAa), it was possible to improve the foldedness by introducing a single mutation, designed through the analysis of the energetics, calculated using statistical potentials,[25] of a structural model of the protein. The mutation was Ser46 to Val; it is to be noted that sequence alignments indicate that, in this position, a hydrophobic residue occurs much more commonly than a hydrophilic one. The solution structure was then solved, and the Val46 residue was shown to form a number of energetically favorable interactions with the hydrophobic side chains of surrounding residues (Ile8, Met10, Val39, Leu41, Val48).

Protein Structures and Interactions. Twelve structures have been obtained for seven different systems (Table 1; some examples are shown in Figure 2). For three systems, both the apo- and holo-protein were structurally characterized. The systems characterized include small soluble proteins (copper chaperones), which bind copper and carry it either to its final target protein or to other proteins, which in turn transfer copper in other compartments/organelles or outside the cell. An example of the latter systems is constituted by the soluble domain of the copper-tranporting ATPases, which are membrane-bound proteins. In most cases, these proteins adopt the same fold as copper chaperones. It is the so-called "ferredoxin-like", $\beta\alpha\beta\beta\alpha\beta$ fold, with the secondary structure elements connected by loops. At the top of Figure 2, one example of a soluble chaperone (CopZ) and of a soluble domain of a membrane-bound ATPase (CopA), both adopting the $\beta\alpha\beta\beta\alpha\beta$ fold, are shown.

The metal-binding site, which has the common CXXC consensus sequence, is highly solvent-exposed; the two Cys ligands are located at the end of the first α helix and in the external loop between the first β strand and the first α helix. The design of this metal-binding site at the surface of the protein is optimal for its function, i.e., copper transfer, at variance with other catalytic or electron-transfer sites, which are usually more buried and more protected inside the protein frame. The average S−Cu−S angle observed in the solution structures for both chaperones and the soluble domains of ATPases is $122° \pm 30°$, suggesting that copper is three-coordinate.[26,27] EXAFS measurements indicated that the three coordination is achieved by three sulfur donor atoms.[5,28] For both classes of proteins, most of the sequences located in the available genomes and all of the systems structurally characterized feature a Met two positions before the first copper-binding Cys. The involvement of this Met in metal coordination has been, however, excluded as its side chain points toward the hydrophobic core of the protein, stabilizing the metal-binding loop through hydrophobic contacts.[21,22,29] The coordination sphere of copper(I) may be completed by the sulfur of an exogenous ligand, like DTT, which is added to keep the copper and the protein in a reduced state. The metal-binding loop may have different properties in the copper-free (apo), with respect to the copper-loaded (holo), form (Figure 3). This appears to be true for copper chaperones but not for ATPases.[27] Indeed, while in the soluble domains of ATPases the loop does not change its conformation upon copper binding and is highly ordered also in the absence of copper, in chaperones the copper-binding loop experiences a significant structural change (Figure 3) and a sizably increased disorder in solution upon copper release. This suggests that, in the chaperone proteins, the loop carrying its metal cargo is flexible,[26] in order to adapt its conformation to optimize metal transfer, while in the receiving proteins, the loop is well organized also in the absence of the metal[27] to be ready to accept it.

Table 1. Available Structures of Proteins Involved in Copper Homeostasis[a]

protein	organism	PDB code	technique	ref
Atx1	yeast	1FES (apo), 1FDS (holo)	NMR	26
Ccc2 (1st domain)	yeast	1FVS (apo), 1FVQ (holo)	NMR	27
CopAa (1st domain)	*Bacillus subtilis*	pending (apo + holo)	NMR	
CopAb (2nd domain)	*Bacillus subtilis*	1JWW (apo), 1KQK (holo)	NMR	35
CopZ	*Bacillus subtilis*	pending (apo), 1KOV (holo)	NMR	36
Cu_7 metallothionein	yeast	1FMY (holo)	NMR	37
CopC	*Pseudomonas syringae*	1M42 (apo)	NMR	31
Menkes protein (4th domain)	*Homo sapiens*	1AW0 (apo), 2AW0 (contains one Ag^+ ion)	NMR	21
CopZ	*Enterococcus hirae*	1CPZ (apo)	NMR	38
Cu_7 metallothionein	yeast	1AQR (holo)	NMR	39
CCS (2nd domain)	*Homo sapiens*	1DO5 (apo)	X-ray	14
CCS	yeast	1QUP (apo)	X-ray	14
CCS(2nd domain)	yeast	1EJ8 (apo)	X-ray	19
CCS in complex with SOD	yeast	1JK9 (apo, contains one Zn^{2+} ion)	X-ray	20

[a] Only structures relative to proteins in a physiologically relevant state are shown. The structures above the break have been solved within the project described in this Account. All other structures have been solved by different research groups worldwide.

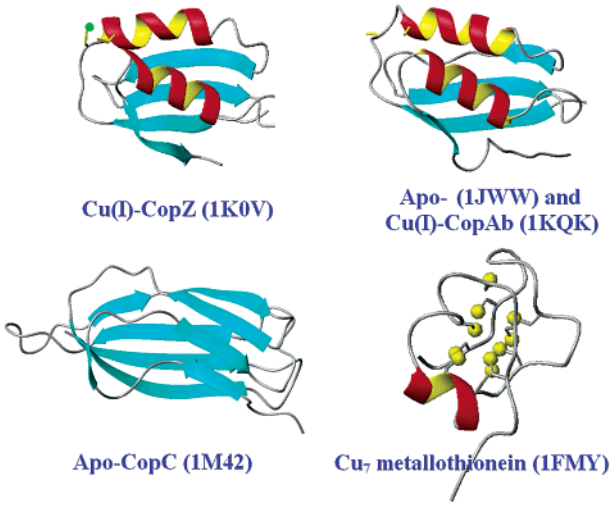

FIGURE 2. Ribbon display of some structures solved within the project described herein. The name of the protein, its metalation state, and the PDB code of the structure are also displayed. The corresponding publications are (starting from the upper left corner, and going clockwise) refs 36, 35, 37, and 31.

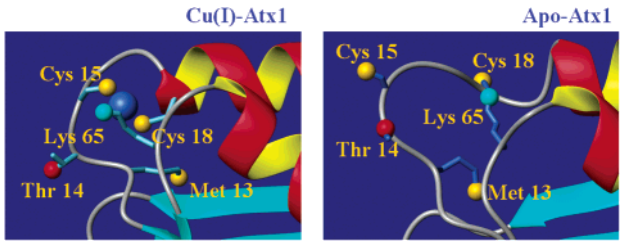

FIGURE 3. Variations in the structure of the metal-binding site upon changes in the metalation state of Atx1. Variations of the protein dynamics in solution are also observed.[26]

Electrostatic interactions are the major driving forces for protein−protein recognition and interaction in these systems. Models of the complex between soluble chaperones and the soluble domain of ATPases have been derived on the basis of NMR spectral changes upon protein interactions (Figure 4).[30] The NMR data thus provide experimental evidence for the molecular mechanism of copper transfer hypothesized on the basis of the X-ray structure of the metal-loaded human copper chaperone, which forms a homodimer in the crystal.[18] The

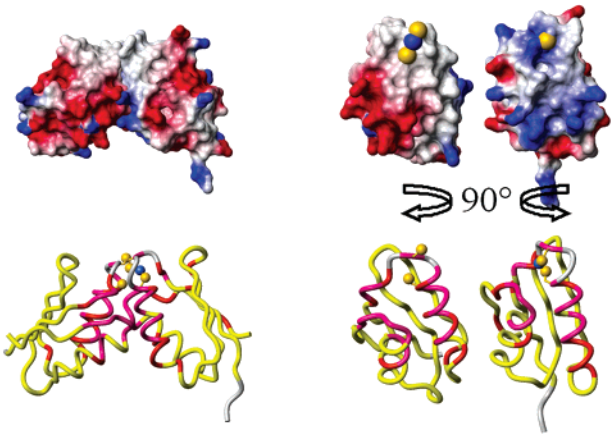

FIGURE 4. (Bottom) Three-dimensional model of the adduct and interaction surfaces as defined by NMR chemical shift mapping studies for Atx1 (right) and Ccc2 (left).[30] Chemical shift variations upon binding to the partner are color-coded from orange (smaller) to magenta (larger). Regions not displaying variations are in yellow, while regions for which no data are available are in gray. The side chains of the metal-binding cysteines are also displayed, with the donor sulfur atom shown as a gold sphere. The Cu^+ ion in Atx1 is shown as a blue sphere. (Top) Electrostatic surfaces of Atx1 (right) and Ccc2 (left). The orientation of the proteins is the same as in the bottom part of the figure.

resulting picture indicates that the two partners interact through two regions of opposite electrostatic potential; the two CXXC metal-binding domains result close enough to ease the metal transfer. Loop 5 of one protein is in contact with helix α1 of the partner, and the two α1 helices make an angle of about 45° to optimize the contacts between the two metal-binding sites. In such a complex, several electrostatic interactions are established. The spatial arrangement of the two proteins in the complex may facilitate the metal transfer through the formation of a series of two- and three-coordinate metal-bridged intermediates, which involve the cysteine ligands of both proteins as transient copper ligands. This process is facilitated by the geometric arrangement of the metal-binding site on the protein surface and is driven by the tendency of copper(I) to saturate its coordination sphere. Indeed, in the isolate protein, bicoordinate copper(I) completes its coordination sphere by binding to an exogenous molecule, as discussed above.[5,28] The latter can

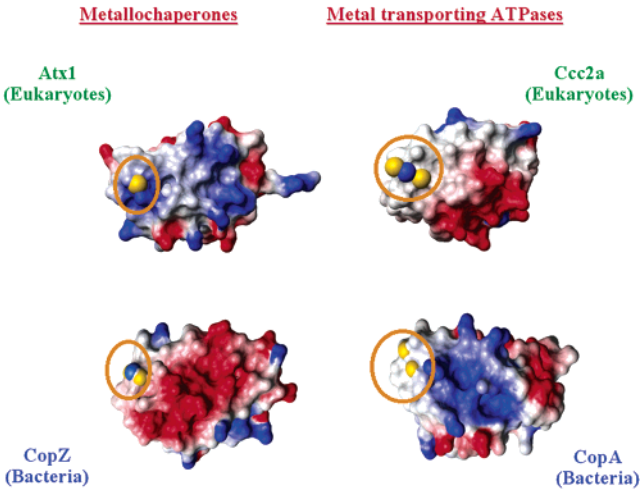

FIGURE 5. Electrostatic surfaces of soluble copper(I) transporters (metallochaperones, left column) and of the soluble domain of their membrane-bound physiological partners (metal-transporting ATPases, right column) from eukaryota (top) and bacteria (bottom). The proteins are oriented such that the region where interaction with the partner occurs faces the reader (see also Figure 4). The metal-binding site in each protein is highlighted by the orange circle. There is an evolutionary relationship between these domains; however, note the reversal in electrostatic charge at the interaction region between corresponding proteins from eukaryota and bacteria.[22]

be easily displaced by the incoming copper-recipient protein. The largest spectral and structural changes observed in the recipient protein for the first (in sequence) metal-binding cysteine with respect to the second cysteine suggest that it is the former which establishes first a bond with the copper ion.[30]

Modeling. Once the three-dimensional structures of proteins in various sequence subclasses have been experimentally determined, structural models can be derived for the other protein sequences identified in gene data banks and belonging to these subclasses. For instance, low-energy, reliable structural models can be obtained for all the sequences homologous to those of Atx1 and Ccc2, which all share the common $\beta\alpha\beta\beta\alpha\beta$ fold.[22] All these proteins show a similar hydrophobic core, involving essentially only parts of the secondary structure elements which contain highly conserved residues. Sequence variability is much higher in protein regions far from the metal-binding site. The completely conserved residues are only the two copper-binding cysteines, while highly conserved residues are located mainly in the loops around the metal-binding site. Specifically, a Gly-Met pair occurring one residue before the first metal-binding cysteine is almost always conserved. Other sequence variations appear to be associated with belonging to a specific subgroup, suggesting that they may be of mechanistic importance. The loop containing the metal-binding site (loop 1) is stabilized by interactions with the two other loops (3 and 5), which are close in space and form an end of this roughly cylindrical fold. These loops contain highly conserved amino acids, which are also involved in protein−protein interactions. In loop 5, a Lys is completely conserved in eukaryotic chaperones, while in bacteria as well as in the soluble domains of copper ATPases, a conserved aromatic residue (Tyr or Phe) is nearly always present. The protein surface is characterized by residues that are less conserved than those in the hydrophobic core. Nevertheless, very similar electrostatic potential surfaces are observed within any given subgroup, while charge reversals may be observed for proteins of different classes (e.g., homologous proteins from bacteria with respect to eukaryota; see Figure 5). Some of the sequence, structural, and electrostatic potential variations between eukaryotic and bacterial proteins might reflect a different interaction specificity in the complete and complex pathway of copper homeostasis.[22]

Conclusions

In the present Account, we have described our approach to Structural Genomics, which has the peculiarity of having a particular focus on the function of the investigated proteins. For this reason, the selection of targets is based on potential functional properties and is performed across the available genomes of all organisms. We have applied this approach to proteins involved in copper homeostasis, and we have experimentally characterized a subset of the sequences identified in the data banks. Information on the structural and functional properties of these proteins has been highlighted, which is often found to be generally relevant to all proteins in the same subclasses.

Thanks to the availability of complete genome sequences for a number of organisms, genetic information on the role of different proteins in copper homeostasis is becoming available at increasing speed. This provides completely new clues for the identification of targets relevant to the project, reducing the need for, e.g., identification of known metal-binding motifs in the primary sequence at the level of target selection. As a consequence, the targets selected will bear less and less sequence homology to proteins of known structure, possibly resulting in the identification of new folds or in the association of new functions to known folds. A notable example is the recently solved structure of CopC from *Pseudomonas syringae*.[31] The methodology developed in this project will be instrumental in defining key issues such as the bases for metal selectivity or the mechanisms for partner recognition and interaction, which are best experimentally investigated by a combination of molecular biology and structural approaches. The identification and characterization of new metal-binding motifs is another critical point. Here, given the intrinsic relatively low affinity of metal-binding sites in chaperones (which reflects a functional requirement), high-throughput approaches may be prone to errors or fail to give results altogether (e.g., note the difficulty of crystallizing holo-Cu(I) chaperones in a functionally relevant form). It is noteworthy that, when the physiologically relevant metal ion is paramagnetic, special NMR methods can be used to structurally characterize the binding site.[32−34] The case of CopC from *P. syringae* constitutes an excellent example

also in this respect.[31] Finally, in most instances, an important challenge in the functional investigation of systems involved in metal homeostasis is to establish the directionality of metal transfer(s). The NMR investigation of adducts as described in this Account is a well-suited technique to address this problem, especially when complemented by other techniques.

We thank Prof. Ivano Bertini for many fruitful discussions over the course of the project. Financial support from the EC (contract HPRI-1999-CT-00009) to the PARABIO Research Infrastructure, which has provided NMR time, is gratefully acknowledged. Finally, we thank our co-workers and collaborators, whose names are listed in the references, for their contribution to the research described in this Account.

References

(1) Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M.; et al. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science* **1995**, *269*, 496−512.

(2) Marti-Renom, M. A.; Stuart, A. C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291−325.

(3) Vitkup, D.; Melamud, E.; Moult, J.; Sander, C. Completeness in structural genomics. *Nat. Struct. Biol.* **2001**, *8*, 559−566.

(4) Rae, T.; Schmidt, P. J.; Pufahl, R. A.; Culotta, V. C.; O'Halloran, T. V. Undetectable intracellular free copper: the requirement of a copper chaperone for superoxide dismutase. *Science* **1999**, *284*, 805−808.

(5) Pufahl, R. A.; Singer, C. P.; Peariso, K. L.; Lin, S.-J.; Schmidt, P. J.; Fahrni, C. J.; Cizewski Culotta, V.; Penner-Hahn, J. E.; O'Halloran, T. V. Metal ion chaperone function of the soluble Cu(I) receptor Atx1. *Science* **1997**, *278*, 853−856.

(6) Valentine, J. S.; Gralla, E. B. Delivering copper inside yeast and human cells. *Science* **1997**, *278*, 817−818.

(7) Pena, M. M. O.; Lee, J.; Thiele, D. J. A delicate balance: homeostatic control of copper uptake and distribution. *J. Nutr.* **1999**, *129*, 1251−1260.

(8) O'Halloran, T. V.; Culotta, V. C. Metallochaperones: An Intracellular Shuttle Service for Metal Ions. *J. Biol. Chem.* **2000**, *275*, 25057−25060.

(9) Altschul, S. F.; Madden, T. L.; Schaeffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389−3402.

(10) Zhang, Z.; Schaffer, A. A.; Miller, W.; Madden, T. L.; Lipman, D. J.; Koonin, E. V.; Altschul, S. F. Protein sequence similarity searches using patterns as seeds. *Nucleic Acids Res.* **1998**, *26*, 3986−3990.

(11) Montelione, G. T. Structural genomics: an approach to the protein folding problem. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13488−13489.

(12) Cavanagh, J.; Fairbrother, W. J.; Palmer, A. G., III; Skelton, N. J. *Protein NMR Spectroscopy. Principles and practice*; Academic Press: San Diego, 1996.

(13) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; Wiley: New York, 1986.

(14) Lamb, A. L.; Wernimont, A. K.; Pufahl, R. A.; Culotta, V. C.; O'Halloran, T. V.; Rosenzweig, A. C. Crystal structure of the copper chaperone for superoxide dismutase. *Nat. Struct. Biol.* **1999**, *6*, 724−729.

(15) Rosenzweig, A. C.; Huffman, D. L.; Hou, M. Y.; Wernimont, A. K.; Pufahl, R. A.; O'Halloran, T. V. Crystal structure of the Atx1 metallochaperone protein at 1.02 A resolution. *Structure Folding Des.* **1999**, *7*, 605−617.

(16) Rosenzweig, A. C. Copper delivery by metallochaperone proteins. *Acc. Chem. Res.* **2001**, *34*, 119−128.

(17) Lamb, A. L.; Wernimont, A. K.; Pufahl, R. A.; O'Halloran, T. V.; Rosenzweig, A. C. Crystal structure of the second domain of the human copper chaperone for superoxide dismutase. *Biochemistry* **2000**, *39*, 1589−1595.

(18) Wernimont, A. K.; Huffman, D. L.; Lamb, A. L.; O'Halloran, T. V.; Rosenzweig, A. C. Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins. *Nat. Struct. Biol.* **2000**, *7*, 766−771.

(19) Hall, L. T.; Sanchez, R. J.; Holloway, S. P.; Zhu, H.; Stine, J. E.; Lyons, T. J.; Demeler, B.; Schirf, V.; Hansen, J. C.; Nersissian, A. M.; Valentine, J. S.; Hart, P. J. X-ray crystallographic and analytical ultracentrifugation analyses of truncated and full-length yeast copper chaperones for SOD (LYS7): a dimer-dimer model of LYS7−SOD association and copper delivery. *Biochemistry* **2000**, *39*, 3611−3623.

(20) Lamb, A. L.; Torres, A. S.; O'Halloran, T. V.; Rosenzweig, A. C. Heterodimeric structure of superoxide dismutase in complex with its metallochaperone. *Nat. Struct. Biol.* **2001**, *8*, 751−755.

(21) Gitschier, J.; Moffat, B.; Reilly, D.; Wood, W. I.; Fairbrother, W. J. Solution structure of the fourth metal-binding domain from the Menkes copper-transporting ATPase. *Nat. Struct. Biol.* **1998**, *5*, 47−54.

(22) Arnesano, F.; Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Molteni, E.; Huffman, D. L.; O'Halloran, T. V. Metallochaperones and metal transporting ATPases: a comparative analysis of sequences and structures *Genome Res.* **2002**, *12*, 255−271.

(23) Zuiderweg, E. R. Mapping protein−protein interactions in solution by NMR spectroscopy. *Biochemistry* **2002**, *41*, 1−7.

(24) Bodenhausen, G.; Ruben, D. J. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett.* **1980**, *69*, 185−188.

(25) Sippl, M. J. Recognition of Errors in the Three-Dimensional Structures *Proteins: Struct., Funct. Genet.* **1993**, *17*, 355−362.

(26) Arnesano, F.; Banci, L.; Bertini, I.; Huffman, D. L.; O'Halloran, T. V. Solution Structure of the Cu(I) and Apo forms of the Yeast Metallochaperone, Atx1. *Biochemistry* **2001**, *40*, 1528−1539.

(27) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Huffman, D. L.; O'Halloran, T. V. Solution structure of the yeast copper transporter domain Ccc2a in the apo and Cu(I)-loaded states. *J. Biol. Chem.* **2001**, *276*, 8415−8426.

(28) Banci, L.; Bertini, I.; Del Conte, R.; Mangani, S.; Meyer-Klaucke, W. X-ray absorption spectroscopy study of CopZ, a copper chaperone in *Bacillus subtilis*. *Biochemistry* **2003**, in press.

(29) Rosenzweig, A. C.; Huffman, D. L.; Hou, M. Y.; Wernimont, A. K.; Pufahl, R. A.; O'Halloran, T. V. Crystal structure of the Atx1 metallochaperone protein at 1.02 A resolution. *Structure Folding Des.* **1999**, *7*, 605−617.

(30) Arnesano, F.; Banci, L.; Bertini, I.; Cantini, F.; Ciofi-Baffoni, S.; Huffman, D. L.; O'Halloran, T. V. Characterization of the binding interface between the copper chaperone Atx1 and the first cytosolic domain of Ccc2 ATPase. *J. Biol. Chem.* **2001**, *276*, 41365−41376.

(31) Arnesano, F.; Banci, L.; Bertini, I.; Thompsett, A. Solution structure of CopC: a cupredoxin-like protein involved in copper homeostasis. *Structure* **2002**, *10*, 1337−1347.

(32) Bertini, I.; Luchinat, C.; Rosato, A. The solution structure of paramagnetic metalloproteins. *Prog. Biophys. Mol. Biol.* **1996**, *66*, 43−80.

(33) Bertini, I.; Luchinat, C.; Parigi, G. *Solution NMR of Paramagnetic Molecules*; Elsevier: Amsterdam, 2001.

(34) Bertini, I.; Luchinat, C.; Piccioli, M. Paramagnetic Probes in Metalloproteins. Turning Limitations into Advantages. *Methods Enzymol.* **2001**, *339*, 314−340.

(35) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; D'Onofrio, M.; Gonnelli, L.; Marhuenda-Egea, F. C.; Ruiz-Dueñas, F. J. NMR characterization of the N-Terminal domain of a potential copper translocating P-type ATPase from bacillus subtilis. *J. Mol. Biol.* **2002**, *317*, 415−429.

(36) Banci, L.; Bertini, I.; Del Conte, R.; Markey, J.; Ruiz-Dueñas, F. J. Copper trafficking: the solution structure of *Bacillus subtilis* CopZ. *Biochemistry* **2001**, *40*, 15660−15668.

(37) Bertini, I.; Hartmann, H. J.; Klein, T.; Liu, G.; Luchinat, C.; Weser, U. High-resolution solution structure of the protein part of Cu7 metallothionein. *Eur. J. Biochem.* **2000**, *267*, 1008−1018.

(38) Wimmer, R.; Herrmann, T.; Solioz, M.; Wüthrich, K. NMR structure and metal interactions of the CopZ copper chaperone. *J. Biol. Chem.* **1999**, *274*, 22597−22603.

(39) Peterson, C. W.; Narula, S. S.; Armitage, I. M. 3D solution structure of copper and silver-substituted yeast metallothioneins. *FEBS Lett.* **1996**, *379*, 85−93.

AR010120R