RESEARCH ARTICLE

Putative copper- and zinc-binding motifs in *Streptococcus pneumoniae* identified by immobilized metal affinity chromatography and mass spectrometry

Xuesong Sun[1,**], Chuan-le Xiao[1*], Ruiguang Ge[2], Xingfeng Yin[1], Hui Li[1], Nan Li[1], Xiaoyan Yang[1], Ying Zhu[1], Xiang He[1] and Qing-Yu He[1]*

[1] Institute of Life and Health Engineering/National Engineering and Research Center of Genetic Medicine, Jinan University, Guangzhou, P. R. China
[2] School of Life Sciences, Sun Yat-Sen University, Guangzhou, P. R. China

Received: July 1, 2010
Revised: May 4, 2011
Accepted: May 11, 2011

The aim of metalloproteomics is to identify and characterize putative metal-binding proteins and metal-binding motifs. In this study, we performed a systematical metalloproteomic analysis on *Streptococcus pneumoniae* through the combined use of efficient immobilized metal affinity chromatography enrichment and high-accuracy linear ion trap-Orbitrap MS to identify metal-binding proteins and metal-binding peptides. In total, 232 and 166 putative metal-binding proteins were respectively isolated by Cu- and Zn-immobilized metal affinity chromatography columns, in which 133 proteins were present in both preparations. The putative metalloproteins are mainly involved in protein, nucleotide and carbon metabolisms, oxidation and cell cycle regulation. Based on the sequence of the putative Cu- and Zn-binding peptides, putative Cu-binding motifs were identified: H(X)mH (m = 0–11), C(X)$_2$C, C(X)nH (n = 2–4, 6, 9), H(X)iM (i = 0–10) and M(X)tM (t = 8 or 12), while putative Zn-binding motifs were identified as follows: H(X)mH (m = 1–12), H(X)iM (i = 0–12), M(X)tM (t = 0, 3 and 4), C(X)nH (n = 1, 2, 7, 10 and 11). Equilibrium dialysis and inductively coupled plasma-MS experiments confirmed that the artificially synthesized peptides harboring differential identified metal-binding motifs interacted directly with the metal ions. The metalloproteomic study presented here suggests that the comparably large size and diverse functions of the *S. pneumoniae* metalloproteome may play important roles in various biological processes and thus contribute to the bacterial pathologies.

Keywords:
Metalloproteome / Metal-binding motif / Microbiology / *Streptococcus pneumoniae*

1 Introduction

Metals, often at trace levels, play pivotal roles in the bacterial survival and infection in the hosts. It is estimated that around 30% of proteins requires a metal cofactor (thus named ''metalloproteins''), such as copper (Cu), zinc (Zn) and iron (Fe) to achieve their respective catalytic, regulatory and structural roles [1]. Some metalloproteins including superoxide dismutase C, ferritin and Hpn are essential in metal homeostasis and detoxification processes [2–4], while ferric uptake regulator, peroxide regulator and nickel response regulator control the intracellular concentrations

Correspondence: Professor Qing-Yu He, Institute of Life and Health Engineering, Jinan University, Guangzhou 510632, P. R. China
E-mail: tqyhe@jnu.edu.cn
Fax: +86-20-85227039

Abbreviations: GO, gene ontology; **ICP-MS,** inductively coupled plasma MS; **IDA,** immobilized iminodiacetic acid; **IMAC,** immobilized metal affinity chromatography; **LTQ,** linear ion trap; **SCX,** strong cation exchange

*These authors contributed equally to this work.

**Additional corresponding author: Dr. Xuesong Sun
E-mail: tsunxs@jnu.edu.cn

Colour Online: See the article online to view Figs. 2 and 4 in colour.

of metals, their import and export [5, 6]. Metalloenzymes contain metal ions in their active sites to bind substrates or help stabilizing proteins during catalytic reactions.

In the level of individual protein study, a variety of analytical techniques including ESI-MS, inductively coupled plasma MS (ICP-MS), high-throughput X-ray absorption spectroscopy, X-ray emission/fluorescence spectroscopy have been widely used to determine the metal-binding stoichiometry, metal-binding sites and metal-dependent structural changes of proteins [7–9]. At the same time, immobilized metal affinity chromatography (IMAC), with its immobilized metal ions to bind the surface-exposed amino acids of proteins/peptides in differential binding affinities, was commonly used to pre-fractionate metalloproteins for systematical metalloproteomic investigation [10–13]. The metal-binding proteins enriched with IMAC can be analyzed through LC-MS/MS (i.e. the hybrid linear ion trap (LTQ)-Orbitrap MS) with high-throughput protein sequence identifications [14–16]. Previously we have successfully employed this approach combining the IMAC metalloprotein separation with MS/MS protein identification to characterize the phosphoproteome of *Streptococcus pneumoniae* [17] and the metalloproteins either to be involved in nickel homeostasis [18] or to be the molecular targets of bismuth compounds in *Helicobacter pylori* [19].

Proteins containing Zn, Fe, Mn and Cu are the most abundant molecules in metalloproteins deposited in the protein data bank [20]. Moreover, Cu- and Zn-containing enzymes and proteins are prevalent in nature, playing important roles in various biological processes. Cu- and Zn-binding proteome has been investigated in hepatoma [9, 21], and Cu-binding proteome has been identified in *Escherichia coli* [22]. However, metal-binding protein profiles in Gram-positive bacteria have not been reported yet. In this study, we intend to characterize Cu- and Zn-binding proteins and motifs in the Gram-positive bacterium *S. pneumoniae*.

As an α-hemolytic, aerobic and encapsulated diplococcus *S. pneumoniae* commonly causes a large number of human diseases, including otitis media, sinusitis, bacterial meningitis, pneumonia and sepsis [23]. Several proteomic studies on this Gram-positive bacterium, including its proteome profile, drug resistance, pathogenicity, response to temperature alteration or iron restriction and regulatory networks of protein expression have been reported [17, 24–28]. In this work, we focused on the identification of Cu- and Zn-metalloproteomes and their respective binding motifs in *S. pneumoniae*. Eight peptides containing the identified metal-binding motifs were synthesized and used to confirm their respective metal-binding capabilities. Metal-binding sites are usually unique with respect to the number of ligands and spatial geometries [29]. As for Cu and Zn, they preferably bind to the site-specific ligands composed of residues such as cysteine, histidine and methionine according to the hard-soft acid-base theory [30]. The research results from this study provided insights for the prediction of proteins involved in metal homeostasis in bacteria through the analysis of the protein primary sequences.

2 Materials and methods

2.1 Cell culture and protein extraction

S. pneumoniae D39 was grown in Todd-Hewitt broth supplemented with 0.5% yeast extract (THY) in a controlled atmosphere chamber (37°C, 5% CO_2). Bacterial growth was monitored by the light absorbance at 600 nm using a Thermo model evolutionary 300 spectrometer. At OD_{600}~0.6, cells were harvested by centrifugation at 5000 × *g* for 20 min at 4°C. The harvested cells were washed three times with pre-chilled PBS buffer and then resuspended in an appropriate volume of lysis buffer (15 mM Tris-HCl, pH 8.0). The mixture was frozen–thawed for three cycles and then sonicated ten times each for 30 s. The lysate was centrifuged at 12 000 × *g* for 10 min at 4°C. Protein concentrations were determined using Bradford assay.

2.2 Proteins binding onto IMAC columns

IMAC columns were prepared with 1 mL of immobilized iminodiacetic acid (IDA, Pierce) loaded in poly-prep columns (Bio-Rad). After washing with 10 mL water, 2 mL of 50 mM $CuSO_4$ or $ZnCl_2$ were loaded into the respective IDA columns and then the excessive unbound metal ions were washed out with 5 mL water. An IDA column without loading metal ions was used as a control in all experiments. Protein extracts were dialyzed with an IMAC binding buffer (20 mM phosphate buffer containing 0.5 M NaCl and 3 M urea, pH 7.4), filtered with 0.45-μm cellulose membranes and then loaded onto the Cu- or Zn-IMAC columns pre-equilibrated with binding buffer. Unbound proteins were washed off consecutively with 10 mL binding buffer and 10 mL binding buffer supplemented with 5 mM imidazole. Metal-binding proteins were eluted with phosphate buffer containing 50 mM EDTA. All fractions eluted from columns were concentrated with YM-3 centrifuge columns (Millipore). The experiments were repeated three times.

2.3 In-solution trypsin digestion

Whole cell protein extracts from *S. pneumoniae* D39 (15 mg) or metal-binding proteins (100 μg) isolated with IMAC were treated with 10 mM DTT (37°C, 3 h) and then 20 mM iodoacetamide (25°C, 1 h in dark). Proteins were precipitated with four volumes of pre-chilled acetone, centrifuged at 12 000 × *g* for 30 min and washed twice with ethanol. The resulting pellet was re-dissolved in 25 mM Tris-HCl buffer (pH 7.6) and then digested with sequencing grade trypsin (1:25 w/w; Promega, Madison, WI) at 37°C for 20 h.

2.4 Peptide binding onto IMAC columns

Digested peptides from whole cell lysates were diluted tenfold with IMAC binding buffer, loaded onto the Cu- or Zn-IMAC column (1 mL), and separated according to the method described in Section 2.2. The resulting metal-binding peptides were acidified with 1/20 volumes of 20% TFA, and then purified using a Sep-Pak C18 cartridge (0.7 mL, Waters) according to the protocol provided by the manufacturer. The peptides eluted with 10, 15, 20, 25, 35 and 40% ACN were combined, and then dried down by a Speed-Vac and stored at −80°C until subsequent MS analysis. Experiments were performed in triplicate.

2.5 Peptide analysis by strong cation exchange-RPLC-MS/MS

Typtic digestion products of metal-binding proteins were redissolved in 30 μL of Buffer A (5% ACN/0.1% formic acid in water) and separated with a strong cation exchange (Biobasic-SCX, Thermo Electron, San Jose, CA) and RPLC (Thermo Electron) coupled with a nanospray LTQ-Orbitrap MS (Thermo Electron) according to the previously described protocol [17]. Briefly, the peptide mixtures were loaded onto the SCX column (100 × 0.32 mm; 5 μm particle size) pre-equilibrated with 5% ACN/0.1% formic acid in water using an auto-sampler of a Surveyor HPLC system (Thermo Finnigan, San Jose, CA), and then eluted with three concentrations of NH_4Cl (0, 0.05 and 1 M) prepared in Buffer A. Each elution step was maintained for 30 min with a constant flow-rate of 6 μL/min.

Each fraction eluted from SCX column and the metal-binding peptides obtained in Section 2.4 were loaded into a C18 column (100 μm id, 10 cm length, 5 μm-size resin (Michrom Bioresources, Auburn, CA) pre-equilibrated with Buffer A with a flow rate of 250 nL/min. Peptides were eluted with a 0–35% ACN gradient plus 0.1% formic acid in water over 120 min and analyzed online with a LTQ-Orbitrap MS via a data-dependent TOP10 method. The parameters used for the MS analysis were: spray voltage, 1.85 kV; no sheath and auxiliary gas flow; ion transfer tube temperature, 200°C; 35% normalized collision energy using for MS2; ion selection thresholds, 1000 counts for MS2 and activation $q = 0.25$ and activation time of 30 ms during MS2 acquisitions. The mass spectrometers were operated in positive ion mode with a data-dependent automatic switch between MS and MS/MS acquisition modes [31].

2.6 Database analysis

All the raw data files were processed using BioWorks 3.3.1 (Thermo Finnigan) and the derived peak lists were searched with the SEQUEST engine (Thermo Fisher Scientific, Waltham, MA, version 28.13) against the forward and reversed NCBI *S. pneumoniae* D39 protein database. The following search criteria were employed: full tryptic specificity required; two missed cleavages allowed; carbamidomethylation set as fixed modification, whereas oxidation (M) as variable modifications. Peptides identified by SEQUEST engine were filtered according to their charge state, cross-correlation score (Xcorr: +1 peptides, ≥ 1.8; +2 peptides, ≥ 2.1; +3 peptides, ≥ 3.1; +4 peptides, ≥ 4.1), normalized difference in correlation score (ΔCn: ≥ 0.1) and the tryptic nature of each peptide. The parameters used were conservative to minimize the false-positive rate to lower than 1%. At least two unique peptides are required for protein identifications.

2.7 Bioinformatics analysis

The information of the function and localization of the identified metalloproteins were obtained from the gene ontology (GO) database of *S. pneumoniae* D39 as described previously and Uniprot database [17, 32]. The bioinformatics toolbox of MATLAB was used to obtain one level of upward regulated GO terms of metalloproteins and the interaction network was constructed by all GO terms (Gene Ontology Consortium, http://www.geneontology.org/; Basic GO Usage, http://bioconductor.org/docs/vignettes.html). Each node was labeled by the gene name, locus number or GO term number.

2.8 Equilibrium dialysis and ICP-MS

Eight artificial peptides containing the identified Cu- or Zn-binding motifs and one unrelated peptide were synthesized (Table 3). The binding between Cu or Zn with the synthesized peptides were monitored with equilibrium dialysis plus ICP-MS according to a previously described method [33]. Peptides (200 μL, 100 μM) were incubated with 1 mM $CuSO_4$ or $ZnSO_4$ in binding buffer (20 mM HEPES buffer, pH 7.4) containing 100 mM NaCl at room temperature for 30 min, and then dialyzed against 2 L binding buffer twice. Metal concentrations were determined by ICP-MS (US EPA 6020A-2007). Experiments were repeated in triplicate.

3 Results

3.1 Identification of metal-binding proteins

To identify the Cu- and Zn-metalloproteomes of *S. pneumoniae*, total cellular proteins were separated by Cu- and Zn-IMAC columns, respectively, and the enriched Cu- and Zn-binding proteins were analyzed by LC-MS/MS. As proteins were not retained by the control column without prior metal loading, the proteins bound to the Cu- or Zn-IMAC were solely attributed to the metal-binding

capabilities. As shown in Fig. 1, most of the proteins non-specifically bound with Cu-IMAC were washed off from the Zn-IMAC column by the binding buffer; there may be quite a few proteins left and eluted by the binding buffer containing 5 mM imidazole as revealed in Lane 6. Totally, 232 and 166 proteins were isolated by Cu- and Zn-IMAC columns, respectively. It was found that 133 proteins were present in both columns (sheet 5 of Supporting Information Table 1), indicating that there exist some common binding features of the two metal ions. The detailed information about the identified proteins eluted from Cu- and Zn-IMAC with their protein NCBI accession numbers and gene names is listed in Supporting Information Tables S1.

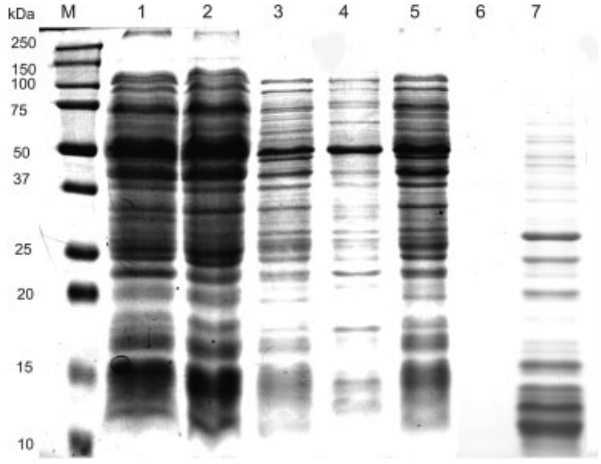

Figure 1. SDS-PAGE gel of fractions eluted from Cu-IMAC (lanes 2–4) and Zn-IMAC (lanes 5–7). Lane 1: whole cell lysates; Lanes 2 and 5: fractions washed with binding buffer; Lanes 3 and 6: fractions washed with binding buffer plus 5 mM imidazole; Lanes 4 and 7: fractions eluted with 50 mM EDTA. The gel was stained with Commassie blue.

3.2 Classification of metal-binding proteins

Information on the identified metal-binding proteins is obtained through GO analysis. Among the identified *S. pneumoniae* metal-binding proteins, only 70 and 50% of the proteins were annotated for their functions in biological processes and cellular localizations, respectively, albeit with extensive and intensive literature searching. High percentage of proteins without annotation information is due to the absence of functional studies on these kinds of proteins. These hypothetical proteins may play roles as metalloenzymes, metal transporters or metallochaperons in bacterial cells. Efforts are needed to verify their respective biological roles and functions in the bacterium.

As shown in Fig. 2, both the putative Cu- and Zn-binding proteins are present throughout the whole cell, with ~40% from the cytoplasm and 5% from the membrane

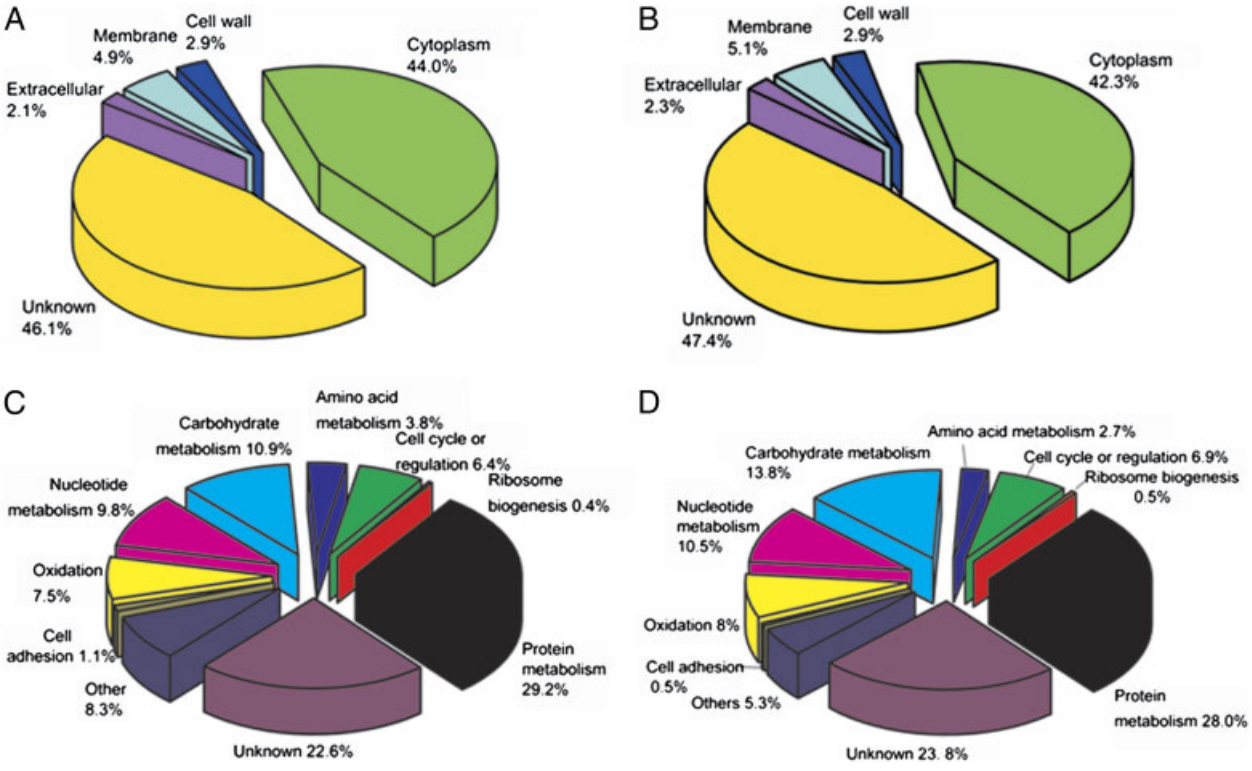

Figure 2. Cellular location and functional categories of identified Cu-binding proteins and Zn-binding proteins of *S. pneumoniae* D39. Fractions of identified Cu-binding (A) and Zn-binding (B) proteins in different cellular compartments; percentages of identified Cu-binding (C) and Zn-binding (D) proteins in different molecular functions.

(Supporting Information Table S2). As regards the respective functions in biological processes, the Cu- and Zn-binding proteins have similar distributions (Fig. 2): ~30% in protein metabolism, ~10% in nucleotide metabolism, ~12% in carbon metabolism, ~3% in amino acid and peptide metabolism and ~6% in cell cycle and cell division.

After database searching, six candidate Cu-binding proteins involved in nucleotide metabolism can also bind other divalent cations. For example, uracil phosphoribosyltransferase and adenylosuccinate synthetase bind Mg^{2+}; ribose-phosphate pyrophosphokinase cysteinyl transfer RNA (tRNA) synthetase bind Zn^{2+}; phosphopentomutase and carbamoyl-phosphate synthase bind Mn^{2+}.

3.3 Identification of metal-binding motifs

To identify the potential metal-binding motifs, the metal-binding capabilities of proteolytic peptides derived from the whole bacterial cell were examined. A physiologically relevant buffer condition was used during the IMAC experiments to mimic the in vivo condition as much as possible. Totally, 115 and 90 peptides were eluted from Cu-IMAC and Zn-IMAC columns, respectively. Among them, 36 Cu- and 35 Zn-binding peptides contain putative metal-binding motifs (Tables 1 and 2), six of which, DFHVVAETGIHAR, HGIEGVVVIGGDGSYHGAMR, HVPVYIQEDMVGHK, HYAHIDAPGHADYVK, HYILAEDYHQDYLR and TLHSGQHFAQGIVADLK are shared in both fractions. The detailed information of identified Cu- and Zn-binding peptides is presented in Supporting Information Table S3. Two representative ESI mass spectra of the metal-binding peptides extracted from the Cu- and Zn-IMAC columns are shown in Fig. 3.

In terms of the specificity of metal-binding [29], Cu preferably binds to the side groups of amino acid residues including cysteine, histidine and methionine according to the hard-soft acid-base theory [30]. Sequence prediction for the Cu-binding peptides revealed the following putative metal-binding motifs: H(*X*)*m*H ($m = 0$–11), C(*X*)$_2$C, C(*X*)*n*H ($n = 2$–4, 6, 9), H(*X*)iM ($i = 0$–10) and M(*X*)tM ($t = 8$ or 12). Cu-binding peptides originated from dihydroorotase, cell wall surface anchor family protein, alanyl-tRNA synthetase, ATP-dependent Clp protease, glucose-6-phosphate isomerase, dTDP-glucose 4, 6-dehydratase and translation elongation factor Tu have at least one histidine-dominant motif such as HH, H(*X*)2H, H(*X*)3H or H(*X*)5H. In total, 18 proteins, including ATP-binding protein of sugar ABC transporter and cell division protein FtsH, identified from the Cu-peptide isolation method, were not detected by the Cu-protein isolation method (Supporting Information Table S4).

Similar to putative Cu-binding motifs, putative Zn-binding motifs in *S. pneumoniae* were H(*X*)*m*H ($m = 1$–12), H(*X*)*i*M ($i = 0$–12), M(*X*)*t*M ($t = 0$, 3 and 4), C(*X*)*n*H ($n = 1$, 2, 7, 10 and 11). However, the motif C(*X*)*r*C was not found in Zn-binding peptides of *S. pneumonaie*. After comparison, we found that 16 proteins, including heat shock protein GrpE and oligopeptide-binding protein AmiA, interfered from the identified Zn-binding peptides were not detected in the metalloprotein preparation (Supporting Information Table S4). The missed detection of these proteins in metalloprotein preparation processes may be attributed to the buried binding sites under the protein surface.

3.4 Statistics of amino acids in metal-binding motifs

The types of amino acids neighboring the potential metal-binding ligands (histidine, cysteine and methionine) in the metal-binding motifs were analyzed statistically. As shown in Supporting Information Fig. S1, there are eight amino acids showing high frequency in both Cu- and Zn-binding peptides. Among them, glutamic acid and alanine are present with the highest frequency in Cu- and Zn-binding motifs, respectively. Interestingly, both of the two acidic amino acids, glutamic acid and aspartic acid, show high occurrence frequency in the metal-binding motifs, suggesting that acidic residues with negative charge under physiological conditions contribute much to the binding between metal ions and metal-binding motifs. In addition to the above two amino acids, other residues presenting with comparably high frequency in metal-binding sites are neutral amino acids, including Gly, Ala, Val, Ile, Thr and Ser.

3.5 Interaction network of metalloproteins

Protein–protein interaction networks were constructed with GO analysis to understand the protein functions involved in the regulation of cellular processes. As shown in Fig. 4A, Cu-binding protein interaction map consisted of two minor and two major networks, covering 29 and 26 proteins, respectively dTDP-4-dehydrorhamnose reductase (rfbD) was the root of one large network in which the downstream target proteins were involved in tRNA aminoacylation, rRNA binding and DNA replication.

Similar to the Cu-binding protein interaction map, the Zn-binding protein interaction map included one major network and two minor networks (Fig. 4B). In the major network, L-lactate dehydrogenase (Ldh) was the root and its downstream targets were involved in nucleotide binding and transcription process, and D-alanine-poly(phosphoribitol) ligase subunit 2 (dltC), dTDP-glucose 4,6-dehydratase (rfbB) and lactate oxidase (lctO) were the partial nodes with high connectivity between each other.

Interestingly, all of the downstream proteins of rfbB in both Cu- and Zn-binding protein maps were involved in the oxidation reduction process, including the putative UDP-glucose 6-dehydrogenase (cps2K), pyruvate oxidase (spxB), 6-phosphogluconate dehydrogenase (gnd) and reduced form of NADH oxidase (nox). Further investigations are needed to elucidate and confirm the biological significance of the present protein–protein interaction maps.

Table 1. Putative Cu-binding peptides from the digest of total *S. pneumoniae* cell lysate

Peptide sequence	Protein name	Motif	*m/z* (measured)	*m/z* (calculated)	Charge
HYAHIDAPGHADYVK	Translation elongation factor Tu	H(X)$_2$H; H(X)$_8$H; H(X)$_5$H	1693.815	1693.808	4
AIEIHTVGAGQMTDLDK	Conserved hypothetical protein TIGR00488	H(X)$_6$M	1798.907	1798.900	3
FDGATMSNNTHVSGDDHIDVTNPMK	Cell wall surface anchor family protein	H(X)$_5$H; H(X)$_{12}$M; H(X)$_6$M;	2703.187	2703.177	3
IGATTIYVTHDQTEAMTLADR	Sugar ABC transporter, ATP-binding protein	H(X)$_5$M	2307.134	2307.129	3
VDTVVISTQHDPEATNEQIHQDVIDK	*S*-adenosylmethionine synthetase	H(X)$_9$H	2931.446	2931.433	3
MPEAVEEESHALSYDEVK	Cell division protein FtsH	M(X)$_8$H	2062.928	2062.927	3
TLHSGQHFAQGIVADLK	Alanyl-tRNA synthetase	H(X)$_3$H	1821.966	1821.961	3
ITSHELLEMDVVDK	Acetyl-CoA carboxylase, carboxyl transferase, α subunit	H(X)$_4$M	1628.824	1628.820	3
VPFVEHNIMTSPLTR	Regulatory protein Spx	H(X)$_2$M	1740.914	1740.910	3
SGVITVIATDHAPHHVDEK	Dihydroorotase	H(X)$_2$H; H(X)$_3$H; H(X)$_0$H	2026.040	2026.035	3
DLYQQHHNVILPDEVLK	ATP-dependent Clp protease, ATP-binding subunit	H(X)$_0$H	2061.082	2061.076	3
HLEADMEDGVLVIR	ATP-dependent Clp protease, ATP-binding subunit	H(X)$_4$M	1596.809	1596.805	2
HGIEGVVVIGGDGSYHGAMR	6-Phosphofructokinase	H(X)$_2$M	2010.986	2010.981	3
LTQSHYILAEEVHSIPTK	Conserved hypothetical protein	H(X)$_7$H	2066.097	2066.092	3
IVSHSVQDAALGEGEGCLSVDR	Peptide deformylase	H(X)$_{12}$C	2299.104	2299.098	3
HVPVYIQEDMVGHK	Ribosomal protein S19	H(X)$_{11}$H; H(X)$_8$M; M(X)$_2$H	1651.832	1651.826	3
KGPFVDEHLMK	Ribosomal protein S19	H(X)$_1$M	1300.678	1300.672	3
GYPTDMDHIISPEGMGK	Conserved hypothetical protein	H(X)$_6$M; M(X)$_1$H; M(X)$_8$M	1847.835	1847.830	2
HFDMAETVELPK	Cell division protein FtsZ	H(X)$_2$M	1416.686	1416.683	2
ENDLLLITADHGNDPTYAGTDHTR	Phosphopentomutase	H(X)$_{10}$H	2639.237	2639.233	3
HYILAEDYHQDYLR	Peptide methionine sulfoxide reductase msrA/msrB 1	H(X)$_7$H	1835.877	1835.871	3
LMAQHQTEEPTVAQMEELK	Tagatose 1,6-diphosphate aldolase	H(X)$_9$M; M(X)$_2$H; M(X)$_{12}$M	2213.064	2213.058	3
HALLGAGIESSHSYER	Conserved hypothetical protein	H(X)$_{10}$H	1726.859	1726.851	3
IGADAVIAEGMEAGGHIGK	Trans-2-enoyl-ACP reductase II	M(X)$_4$H	1795.906	1795.901	3
TIDGGLSCQYEHQFVITK	Methionine aminopeptidase, type I	C(X)$_3$H	2096.017	2096.012	2
AAIDFLNHHFANLQTK	Glucose-6-phosphate isomerase	H(X)$_0$H	1839.954	1839.950	3
DFHVVAETGIHAR	Phosphocarrier protein HPr	H(X)$_7$H	1451.743	1451.739	3
NIGIMAHVDAGK	Translation elongation factor G	M(X)$_1$H	1225.641	1225.636	2
GMVPSGASTGEHEAVELR	Phosphopyruvate hydratase	M(X)$_9$H	1826.879	1826.870	2
FHHVSTDEVYGDLPLR	dTDP-glucose 4,6-dehydratase	H(X)$_0$H	1884.934	1884.924	3
ILVSNPDMHQTLTETLK	Conserved hypothetical protein	M(X)$_0$H	1940.020	1940.016	2
GENDLSHQPISFGGMLQK	Cell division protein FtsA	H(X)$_7$M	1957.948	1957.944	3
IIGVDGTVANCGVHGK	Alcohol dehydrogenase, zinc-containing	C(X)$_2$H	1596.824	1596.816	2
TTICGTDLHIIK	Alcohol dehydrogenase, zinc-containing	C(X)$_4$H	1371.734	1371.730	2
IPDVMYVVDPHK	Ribosomal protein S2	M(X)$_5$H	1412.728	1412.724	2
LVQASNYPQCQLCLENEGYHGR	Galactose-1-phosphate uridylyltransferase	C(X)$_2$C; C(X)$_9$H; C(X)$_6$H	2636.209	2636.198	3

3.6 Binding between metal-binding motifs and metal ions

Equilibrium dialysis and ICP-MS were used to validate the direct interaction between putative metal-binding peptides and metal ions. Eight peptides containing either putative Cu- or Zn-binding motifs identified in this study were synthesized. In parallel, an unrelated peptide without the identified putative metal-binding motif was synthesized as a control. These peptides contained different kinds of

Table 2. Putative Zn-binding peptides from the digest of total *S. pneumoniae* cell lysate

Peptide sequence	Protein name	Motif	*m/z* (measured)	*m/z* (calculated)	Charge
HYAHIDAPGHADYVK	Translation elongation factor Tu	H(*X*)2H; H(*X*)$_8$H; H(*X*)$_5$H	1693.817	1693.808	3
VHMAYIGHPVAGDEVYGPR	Ribosomal large subunit Pseudouridine synthase D	H(*X*)$_5$H; H(*X*)$_0$M; M(*X*)$_4$H	2068.017	2068.007	3
IALEHNIAWQEIQGTGHR	Acetoin dehydrogenase complex, E2 component,	H(*X*)$_{11}$H	2073.067	2073.062	3
SVDNLSVGFHEDHGVGAPK	β-Galactosidase precursor, putative	H(*X*)$_2$H	1964.953	1964.946	3
TLHSGQHFAQGIVADLK	Alanyl-tRNA synthetase	H(*X*)$_3$H	1821.969	1821.961	3
FMHHYNFPQYSVGETGR	Polyribonucleotide nucleotidyltransferase	H(*X*)$_0$H; M(*X*)$_0$H; M(*X*)$_1$H	2069.940	2069.929	3
ATISMIQSMVEHR	Conserved hypothetical protein	M(*X*)$_6$H; M(*X*)$_2$H; M(*X*)$_3$M	1502.751	1502.745	2
IVEQFHVAHISTGDMFR	Adenylate kinase	H(*X*)$_2$H; H(*X*)$_8$M; H(*X*)$_5$M	2002.991	2002.980	3
NPIVFAFHPSAQESSAHAAR	Alcohol dehydrogenase, iron-containing	H(*X*)$_8$H	2137.067	2137.057	3
GEEYHQLTSDHSLVNELLK	Phosphatase, putative	H(*X*)5H	2212.095	2212.088	3
DLYQQHHNVILPDEVLK	ATP-dependent Clp protease, ATP-binding subunit	H(*X*)$_0$H	2061.084	2061.076	3
HGIEGVVVIGGDGSYHGAMR	6-Phosphofructokinase	H(*X*)$_2$M	2010.990	2010.981	3
VVFHSLSSDHMQEVVK	ATP-dependent Clp protease, ATP-binding subunit	H(*X*)$_5$H; H(*X*)$_6$M; H(*X*)$_0$M	1841.929	1841.921	3
HVPVYIQEDMVGHK	Ribosomal protein S19	H(*X*)$_{11}$H; H(*X*)$_8$M; M(*X*)$_2$H	1651.835	1651.826	3
EGYEMSFDTMVLTGDN AANPHGIPAANK	Proline dipeptidase PepQ	M(*X*)$_{10}$H; M(*X*)$_4$M	2982.334	2982.324	3
IYYTHSNHPGGLK	Ribosomal protein L13	H(*X*)$_2$H	1486.751	1486.744	3
DGHEIPVISGSVPPHLTR	Glutamyl aminopeptidase PepA	H(*X*)$_{11}$H	1911.018	1911.008	3
NVVEWGIANPHEAV MMASFNPAK	*N*-acetylglucosamine-6-phosphate deacetylase	H(*X*)$_3$M; H(*X*)$_4$M; M(*X*)$_0$M	2544.213	2544.201	3
HYILAEDYHQDYLR	Peptide methionine sulfoxide reductase msrA/msrB 1	H(*X*)$_7$H	1835.881	1835.871	3
HENMIAVLAVDEVK	ATP synthase F1, ε subunit	H(*X*)$_2$M	1567.820	1567.815	2
TNHHAFETGLAYHTATMVR	cmp-binding-factor 1	H(*X*)$_0$H; H(*X*)$_9$H; H(*X*)$_8$H; H(*X*)$_{12}$M; H(*X*)$_3$M	2157.040	2157.029	4
DLMHHLIELYK	Phosphotransferase system, lactose-specific IIA component	H(*X*)$_0$H; M(*X*)$_0$H; M(*X*)$_1$H	1427.737	1427.735	3
FVAPHEVEYMQSQVTAADELIR	Glucose-6-phosphate isomerase	H(*X*)$_4$M	2549.245	2549.234	3
AHAGLDIGDTAIGMHVK	Conserved hypothetical protein TIGR01440	H(*X*)$_{12}$H; H(*X*)$_{11}$M; M(*X*)$_0$H	1705.876	1705.869	3
GIHLAVQGCEHVNR	Conserved hypothetical protein TIGR01440	C(*X*)$_1$H; H(*X*)$_5$C; H(*X*)$_7$H	1589.804	1589.797	3
DAFEHILCGASMVQVGTTLHK	Dihydroorotate dehydrogenase A	C(*X*)$_{11}$H; H(*X*)$_2$C; H(*X*)$_6$M; M(*X*)$_7$H	2330.135	2330.127	3
DFHVVAETGIHAR	Phosphocarrier protein HPr	H(*X*)$_7$H	1451.746	1451.739	3
HSLHIGDQLQVK	Iron-dependent transcriptional regulator	H(*X*)$_2$H	1374.756	1374.749	3
THEYCTNNQPNNHS DHVDPYPYLAK	Autolysin/*N*-acetylmuramoyl-L-alanine amidase	C(*X*)$_7$H; C(*X*)$_{10}$H; H(*X*)$_2$C; H(*X*)$_{10}$H; H(*X*)$_2$H	3014.328	3014.312	4
SLTLTSEHDNLFMEEIAK	Conserved hypothetical protein	H(*X*)4M	2094.013	2094.006	3
VHVHTEDPGLVMQEGLK	Conserved hypothetical protein	H(*X*)$_1$H; H(*X*)$_9$M; H(*X*)$_7$M	1888.967	1888.959	3
HYGYTDLHLLVGNDGLR	β-*N*-acetylhexosaminidase	H(*X*)$_6$H	1942.983	1942.977	3
GLAMVQESLIHALK	Heat shock protein GrpE	M(*X*)$_6$H	1525.846	1525.841	2
SAIIDYDIYHGHQIVSK	Conserved hypothetical protein	H(*X*)$_1$H	1959.009	1958.997	2
DMEAILSHVEEVK	DegV family protein	M(*X*)$_5$H	1499.746	1499.741	2

metal-binding motifs, including H(*X*)*m*H, C(*X*)*n*H and M(*X*)*i*H. The results showed that the control peptide was unable to bind either Cu or Zn (Table 3). Synthesized peptides, HYAHIDAPGHADYVK, HVPVYIQEDMVGHK and HALLGAGIESSHSYER bound more than one molar equivalent of Cu per molecule as Cu binding requires only

Table 3. Cu^{2+} or Zn^{2+}- peptide stoichiometries for the peptides as determined with equilibrium dialysis and ICP-MS

	Peptide sequence	Cu (or Zn)/peptide	Motifs
Cu-binding peptides	HYAHIDAPGHADYVK	1.32±0.1	$H(X)_2H$; $H(X)_8H$; $H(X)_5H$
	HVPVYIQEDMVGHK	1.31±0.05	$H(X)_{11}H$; $H(X)_8M$; $M(X)_2H$
	HALLGAGIESSHSYER	1.21±0.09	$H(X)_{10}H$
	TIDGGLSCQYEHQFVITK	0.85±0.08	$C(X)_3H$
	VILLDVQGIEEVNR (control)	0.02±0.00	N/A
Zn-binding peptides	VHMAYIGHPVAGDEVYGPR	0.92±0.10	$H(X)_5H$; $H(X)_0M$; $M(X)_4H$
	GIHLAVQGCEHVNR	0.85±0.12	$C(X)_1H$; $H(X)_5C$; $H(X)_7H$
	HYGYTDLHLLVGNDGLR	0.91±0.08	$H(X)_6H$
	ATISMIQSMVEHR	0.98±0.08	$M(X)_6H$; $M(X)_2H$; $M(X)_3M$
	VILLDVQGIEEVNR (control)	0.06±0.00	N/A

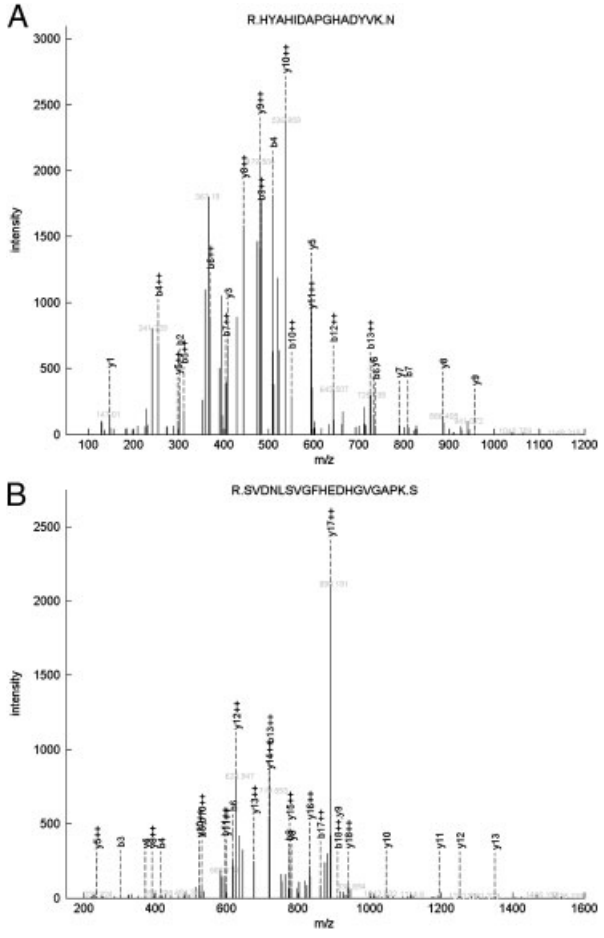

Figure 3. Representative MS/MS spectra of the identified Cu-binding peptide (R.HYAHIDAPGHADYVK.N) of translation elongation factor Tu and Zn-binding peptide (R.SVDNLSVGFHED HGVGAPK.S) of β-galactosidase precursor.

one histidine, whereas other peptides bound Cu and Zn in about 1:1 stoichiometry (Table 3). These results confirmed the putative metal-binding motifs identified in this study.

4 Discussion

Metalloproteins with the intrinsic metal atoms play catalytic, regulatory and structural roles in bacteria. So far, the metalloproteome in *Streptococcus* species has not yet been reported. The current study focused on the pathogenic bacterium *S. pneumoniae*, using efficient IMAC peptide enrichment and high-accuracy LTQ-Orbitrap MS analysis to systematically characterize metalloproteins and metal-binding motifs.

In total the *S. pneumoniae* D39 genome encodes 1914 proteins. The identified Cu- and Zn-binding proteins from *S. pneumoniae* D39 account for 12.1 and 8.7% of the total proteins, respectively. As shown in Fig. 4, many important node proteins (such as LysS or NrdE) interact with several upstream and downstream target proteins. The interaction network indicated that metalloproteins are involved in several important biological processes including protein, carbohydrate, nucleotide metabolism and cell cycle. It was noted that some proteins involved in oxidation reduction process were also identified as metal-binding proteins.

Interestingly, 40 ribosomal proteins including 16 proteins from 30S subunit and 24 from 50S subunit were identified as putative metal-binding proteins. The high percentage of ribosomal proteins was also found in the study on the hepatocellular metalloproteins [21]. Another study showed that the smallest protein in large prokaryotic ribosomal subunit L36 from *Thermus thermophilius* is a Zn ribbon protein [34]. The findings seem to indicate that transition metal ions may function in ribosomal assembly. To date, the role of transition metal ions during ribosome biogenesis has not yet been evaluated besides their contribution in the stabilization of the tertiary rRNA structure [35], and thus future biological investigations are required.

Some cytoplasmic proteins without predicted metal-binding domains were also identified in this study, which may be due to their binding with other metalloproteins. For example, the heat shock protein GrpE, a typical chaperone protein without typical metal-binding capabilities, was identified in this study [36].

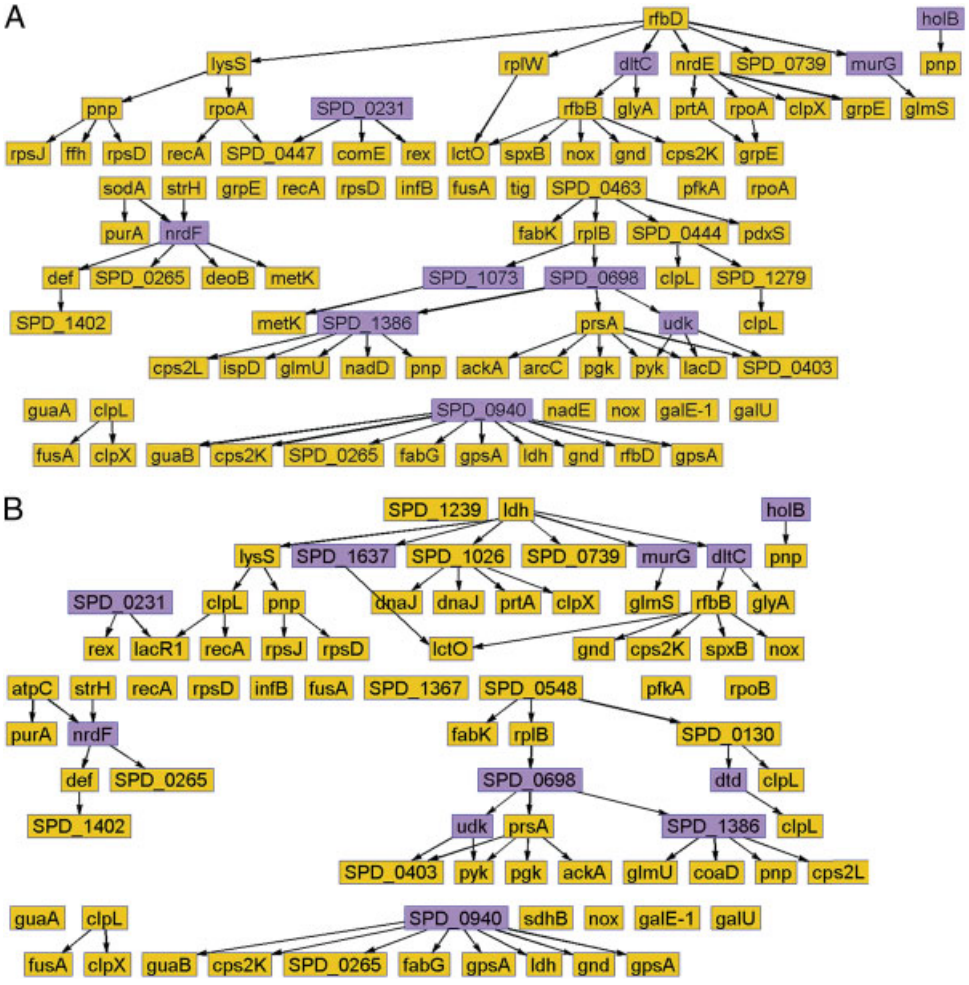

Figure 4. Interaction network maps of the identified Cu-binding (A) and Zn-binding (B) proteins in *S. pneumoniae* D39. Proteins in yellow represent metalloproteins identified in this study; proteins in purple are proteins interacted with the identified metalloproteins in D39 strain.

Zn is important for maintaining protein structure and stability. The proportion of Zn proteome in total proteins encoded by the genomes in eukaryota (8.8%), bacteria (5%) and archaea (6%) has been reported [37]. Several Zn-metalloproteins including alcohol dehydrogenase identified in *S. pneumoniae* have been reported as zinc-finger proteins in the previous studies [38]. Totally 11 proteins of *S. pneumoniae* in protein data-bank contain Zn, including glutamyl aminopeptidase (pepA), penicillin binding protein 2B (PBP-2B), penicillin binding protein 1A (PBP-1A), AdcAII, alkaline amylopullulanase, phosphate transport system regulatory protein (PhoU), pneumococcal histidine triad A protein (phtA), peptidoglycan deacetylase (PgdA), choline binding protein, peptide deformylase and protein surface antigen (PsaA). However, only three proteins (pepA, peptide deformylase, PsaA) were detected in the eluates from Zn-IMAC. Several lipoproteins predicted to contain Zn, such as the zinc transporter AdcAII, were not isolated through the Zn-IMAC, either due to their low abundance, limited solubility or Zn sites in a kinetically labile state. In addition, Zn trafficking membrane proteins were absent from the detection in this study.

Previous studies have demonstrated that metal ions can bind to ligands including imidazole nitrogen from histidine residues, carboxylate oxygen from acidic residues such as aspartic acid/glutamic acid, thiol group from cysteine or thio ether from methionine residue and amide nitrogen and carbonyl oxygen from the peptide backbone [29, 39]. Based on the frequency of residues in the identified metal-binding peptides, we concluded that the ligand binding affinity for a residue to both Cu and Zn is in the order of histidine > methionine > cysteine (Tables 1 and 2).

Cu prefers four-, five- or six-coordinate geometry that is commonly square planar or octahedral with weak axial coordination, whereas Zn has a preference for four-coordinate tetrahedral geometry. Some putative Cu-binding motifs (i.e. HXH, HXXH and MXXH) in *S. pneumoniae* have also been found in hepatoma cell lines by Sarkar and coworkers using similar separation and identification methods [9]. In comparison, Cu-binding motifs $C(X)_2C$ and $C(X)_nH$ ($n = 2$–9) were only detected in *S. pneumoniae*, with the distance between two histidine ligands being shorter than 12 amino acids, whereas the distance can be as far as 24 residues (i.e. enolase) in hepatoma, indicating that metal

binding motifs may be species-specific. The direct interactions between metal ions and peptides containing differential kinds of putative metal-binding motifs identified in this study were verified by equilibrium dialysis and ICP-MS (Table 3).

In summary, in this work we globally characterized the putative metalloproteome and metal-binding motifs of the pathogenic bacterium *S. pneumoniae*. The functional classification and interaction network of the identified metalloproteins indicated that metalloproteins are involved in many important processes including protein, nucleotide and carbon metabolism. Compared to the known mammalian metalloproteomes, *S. pneumoniae* has a bigger size of metal-binding proteins. These results imply that the metalloproteins are closely related to many important biological processes. Further investigations on the molecular functions of the metalloproteins in bacteria are undergoing.

This work was partially supported by the 2007 Chang-Jiang Scholars Program, ''211'' Projects, National Natural Science Foundation of China (20871057, to Q.-Y.H; 31000373, to X.S.; 20801061, to R.G.), Guangdong Natural Science Foundation (9251063201000006, to Q.-Y.H; 10451063201005247, to X.S.), Fundamental Research Funds for the Central Universities (21610101, to Q.-Y.H, and 21611201, to X.S.).

The authors have declared no conflict of interest.

5 References

[1] Tainer, J. A., Roberts, V. A., Getzoff, E. D., Metal-binding sites in proteins. *Curr. Opin. Biotechnol.* 1991, *2*, 582–591.

[2] Ge, R., Sun, H., Bioinorganic chemistry of bismuth and antimony: target sites of metallodrugs. *Acc. Chem. Res.* 2007, *40*, 267–274.

[3] Vašák, M., Hasler, D. W., Metallothioneins: new functional and structural insights. *Curr. Opin. Chem. Biol.* 2000, *4*, 177–183.

[4] Chasteen, N. D., Ferritin. Uptake, storage, and release of iron. *Met. Ions Biol. Syst.* 1998, *35*, 479–514.

[5] Andrews, S. C., Robinson,A. K., Rodriguez-Quinones,F., Bacterial iron homeostasis. *FEMS Microbiol. Rev.* 2003, *27*, 215–237.

[6] Dosanjh, N. S., Michel, S. L., Microbial nickel metalloregulation: NikRs for nickel ions. *Curr. Opin. Chem. Biol.* 2006, *10*, 123–130.

[7] Gao, Y., Chen, C., Chai, Z., Advanced nuclear analytical techniques for metalloproteomics. *J. Anal. At. Spectrom.* 2007, *22*, 856–866.

[8] Solomon, E. I., Spectroscopic methods in bioinorganic chemistry: blue to green to red copper sites. *Inorg. Chem.* 2006, *45*, 8012–8025.

[9] She, Y. M., Narindrasorasak, S., Yang, S., Spitale, N. et al., Identification of metal-binding proteins in human hepatoma lines by immobilized metal affinity chromatography and mass spectrometry. *Mol. Cell. Proteomics* 2003, *2*, 1306–1318.

[10] Porath, J., Carlsson, J., Olsson, I., Belfrage, G., Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 1975, *258*, 598–599.

[11] Shi, W., Chance, M. R., Metallomics and metalloproteomics. *Cell. Mol. Life Sci.* 2008, *65*, 3040–3048.

[12] Sun, X., Chiu, J. F., He, Q. Y., Application of immobilized metal affinity chromatography in proteomics. *Exp. Rev. Proteomics* 2005, *2*, 649–657.

[13] Sun, X., Chiu, J. F., He, Q. Y., Fractionation of proteins by immobilized metal affinity chromatography. *Methods Mol. Biol.* 2008, *424*, 205–212.

[14] Makarov, A., Denisov, E., Lange, O., Horning, S., Dynamic range of mass accuracy in LTQ Orbitrap hybrid mass spectrometer. *J. Am. Soc. Mass Spectrom.* 2006, *17*, 977–982.

[15] Makarov, A., Denisov, E., Kholomeev, A., Balschun, W. et al., Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 2006, *78*, 2113–2120.

[16] Perry, R. H., Cooks, R. G., Noll, R. J., Orbitrap mass spectrometry: instrumentation, ion motion and applications. *Mass Spectrom. Rev.* 2008, *27*, 661–699.

[17] Sun, X., Ge, F., Xiao, C. L., Yin, X. F. et al., Phosphoproteomic analysis reveals the multiple roles of phosphorylation in pathogenic bacterium *Streptococcus pneumoniae*. *J. Proteome Res.* 2010, *9*, 275–282.

[18] Sun, X., Ge, R., Chiu, J. F., Sun, H. et al., Identification of proteins related to nickel homeostasis in *Helicobater pylori* by immobilized metal affinity chromatography and two-dimensional gel electrophoresis. *Metal Based Drugs* 2008, *2008*, 289490.

[19] Ge, R., Sun, X., Gu, Q., Watt, R. M. et al., A proteomic approach for the identification of bismuth-binding proteins in *Helicobacter pylori*. *J. Biol. Inorg. Chem.* 2007, *12*, 831–842.

[20] Shi, W., Zhan, C., Ignatov, A., Manjasetty, B. A. et al., Metalloproteomics: high-throughput structural and functional annotation of proteins in structural genomics. *Structure* 2005, *13*, 1473–1486.

[21] Smith, S. D., She, Y. M., Roberts, E. A., Sarkar, B., Using immobilized metal affinity chromatography, two-dimensional electrophoresis and mass spectrometry to identify hepatocellular proteins with copper-binding ability. *J. Proteome Res.* 2004, *3*, 834–840.

[22] Sevcenco, A. M., Krijger, G. C., Pinkse, M. W., Verhaert, P. D. et al., Development of a generic approach to native metalloproteomics: application to the quantitative identification of soluble copper proteins in *Escherichia coli*. *J. Biol. Inorg. Chem.* 2009, *14*, 631–640.

[23] Lim, W. S., Macfarlane, J. T., Boswell, T. C., Harrison, T. G. et al., Study of community acquired pneumonia aetiology (SCAPA) in adults admitted to hospital: implications for management guidelines. *Thorax* 2001, *56*, 296–301.

[24] Cash, P., Argo, E., Ford, L., Lawrie, L. et al., A proteomic analysis of erythromycin resistance in *Streptococcus pneumoniae*. *Electrophoresis* 1999, *20*, 2259–2268.

[25] Lee, M. R., Bae, S. M., Kim, T. S., Lee, K. J., Proteomic analysis of protein expression in *Streptococcus pneumoniae* in response to temperature shift. *J. Microbiol.* 2006, *44*, 375–382.

[26] Soualhine, H., Brochu, V., Menard, F., Papadopoulou, B. et al., A proteomic analysis of penicillin resistance in *Streptococcus pneumoniae* reveals a novel role for PstS, a subunit of the phosphate ABC transporter. *Mol. Microbiol.* 2005, *58*, 1430–1440.

[27] Lee, K. J., Bae, S. M., Lee, M. R., Yeon, S. M. et al., Proteomic analysis of growth phase-dependent proteins of *Streptococcus pneumoniae*. *Proteomics* 2006, *6*, 1274–1282.

[28] Nanduri, B., Shah, P., Ramkumar, M., Allen, E. B. et al., Quantitative analysis of *Streptococcus pneumoniae* TIGR4 response to in vitro iron restriction by 2-D LC ESI MS/MS. *Proteomics* 2008, *8*, 2104–2114.

[29] Karlin, S., Zhu, Z. Y., Karlin, K. D., The extended environment of mononuclear metal centers in protein structures. *Proc. Natl. Acad. Sci. USA* 1997, *94*, 14225–14230.

[30] Pearson, R. G., Hard and soft acids and bases. *J. Am. Chem. Soc.* 1963, *85*, 3533–3539.

[31] Schroeder, M. J., Shabanowitz, J., Schwartz, J. C., Hunt, D. F. et al., A neutral loss activation method for improved phosphopeptide sequence analysis by quadrupole ion trap mass spectrometry. *Anal. Chem.* 2004, *76*, 3590–3598.

[32] Drabkin, H. J., Hollenbeck, C., Hill, D. P., Blake, J. A., Ontological visualization of protein-protein interactions. *BMC Bioinformatics* 2005, *6*, 29.

[33] Sun, X., Ge, R., Chiu, J. F., Sun, H. et al., Lipoprotein MtsA of MtsABC in *Streptococcus pyogenes* primarily binds ferrous ion with bicarbonate as a synergistic anion. *FEBS Lett.* 2008, *582*, 1351–1354.

[34] Kou, W., Kolla, H. S., Ortiz-Acevedo, A., Haines, D. C. et al., Modulation of zinc- and cobalt-binding affinities through changes in the stability of the zinc ribbon protein L36. *J. Biol. Inorg. Chem.* 2005, *10*, 167–180.

[35] Kaczanowska, M., Ryden-Aulin, M., Ribosome biogenesis and the translation process in *Escherichia coli*. *Microbiol. Mol. Biol. Rev.* 2007, *71*, 477–494.

[36] Heiss, K., Junkes, C., Guerreiro, N., Swamy, M. et al., Subproteomic analysis of metal-interacting proteins in human B cells. *Proteomics* 2005, *5*, 3614–3622.

[37] Andreini, C., Banci, L., Bertini, I., Rosato, A., Zinc through the three domains of life. *J. Proteome Res.* 2006, *5*, 3173–3178.

[38] Maret, W., Exploring the zinc proteome. *J. Anal. Atom. Spectrom.* 2004, *19*, 15–19.

[39] Holm, R. H., Kennepohl, P., Solomon, E. I., Structural and functional aspects of metal sites in biology. *Chem. Rev.* 1996, *96*, 2239–2314.

