# DCCP and DICP: Construction and Analyses of Databases for Copper- and Iron-Chelating Proteins

Hao Wu[1]*, Yan Yang[1,2]*, Sheng-Juan Jiang[1], Ling-Ling Chen[1]#, Hai-Xia Gao[1], Qing-Shan Fu[1,3], Feng Li[1,2], Bin-Guang Ma[1], and Hong-Yu Zhang[1]#

[1] *Shandong Provincial Research Center for Bioinformatic Engineering and Technique, Shandong University of Technology, Zibo 255049, China;* [2] *Isotope Research Laboratory, Sichuan Agriculture University, Yaan 625014, China;* [3] *College of Life Sciences, Shandong Normal University, Jinan 250014, China.*

**Copper and iron play important roles in a variety of biological processes, especially when being chelated with proteins. The proteins involved in the metal binding, transporting and metabolism have aroused much interest. To facilitate the study on this topic, we constructed two databases (DCCP and DICP) containing the known copper- and iron-chelating proteins, which are freely available from the website http://sdbi.sdut.edu.cn/en. Users can conveniently search and browse all of the entries in the databases. Based on the two databases, bioinformatic analyses were performed, which provided some novel insights into metalloproteins.**

**Key words: copper and iron, database, primary sequence analysis, secondary structure prediction, SCOP classification**

## Introduction

Metalloproteins account for one third of the natural proteins, which perform important biological functions in living organisms. As is well known, both copper and iron participate in a variety of metabolic processes, such as oxygen transport, ATP synthesis, cellular respiration, metal homeostasis, and antioxidant defense (*1, 2*). Despite the essential role of metal ions in many cellular processes, excessive free metal ions in the cytoplasm are highly toxic (*3, 4*). When the intracellular free copper or iron ions reach a high level, they will compete with other metal ions for important biological ligands or active sites of enzymes. Furthermore, excessive free copper or iron ions will generate reactive oxygen species that degrade DNA, proteins, and lipids (*4, 5*). That is to say, both deficiency and excess of metal ions are disadvantageous, which will result in a number of fatal diseases, such as Wilson and Menkes disease, Parkinson disease, Alzheimer disease, and prion diseases (*2, 6–9*).

To use metal ions safely, living organisms have developed highly specialized systems of proteins to recruit, deliver, and eliminate them. On the other hand, due to the hydrophilic property of metal ions, they cannot permeate through the bio-membranes without carriers. To solve the problem, many metal-chelating proteins have come into being within the millions of years' evolution, which can specially bind metal ions, assist them to traverse the membrane and hand them to metal-dependent proteins, finally complete the compartmentalization of metal ions and maintain the homeostasis of the metal (*10–13*).

In recent years, more and more attention was paid to metalloproteins. Although there are some metalloprotein databases relevant to this topic, such as MDB (Metalloprotein Database and Browser, http://metallo.scripps.edu/, last updated on Nov. 20, 2003) and PROMISE (The Prosthetic centers and Metal Ions in Protein Active Sites, http://metallo.scripps.edu/PROMISE/MAIN.html, last updated on Mar. 1, 1999), they are out of date now and lack an overall analysis on metalloproteins. This aroused our interest to construct two databases for copper-chelating proteins (DCCP) and iron-chelating proteins (DICP), as well as perform analyses on these proteins.

* **These authors contributed equally to this work.**
# **Corresponding authors.**
**E-mail: zhanghy@sdut.edu.cn;**
      **llchen@sdut.edu.cn**

# Results and Discussion

## Database construction

Currently, DCCP involves two types of proteins, one with the primary sequences, called DCCP_1D, which contains 5,088 entries, and the other with the experimentally determined 3D structures, called DCCP_3D, containing 480 entries of proteins. In DCCP_3D, only 390 proteins contain copper ions in the structures, and the others without copper ions are fragments of some copper-chelating proteins, such as prion protein. In addition, it is found that four proteins in DCCP_3D are correlated with bleomycin antibiotics (Number: 182, 183, 377, and 417), indicating that copper proteins also have the function related to antibiotics besides catalysis and electron transfer. All the proteins were selected in a wide range of species, from the low organisms such as bacteria to the high organisms such as human. Most of the records were obtained from PDB (Protein Data Bank), GenBank (Release 141.0), SWISS-PROT, and PIR (Protein Information Resource). Furthermore, several protein sequences in DCCP_1D were obtained by EST fragments linkage.

DICP also includes two types of proteins, DICP_1D and DICP_3D. The former contains 20,461 entries and the latter 1,195 entries, of which 108 are fragments of iron-chelating proteins, such as apoferritins. Different from DCCP, the entry number in DICP is given according to its function. What's more, users can conveniently BLAST the query sequences against all the entries in this database.

Similar to MDB and PROMISE, DCCP and DICP provide convenient searching tools. Users can make a quick or advanced search by using PDB ID, GenBank accession number, common name of protein, species name, valence of ion, cellular localization, or protein function. The most important feature of MDB and PROMISE is that they offer quantitative information on geometrical parameters of metal-binding sites in protein structures available from PDB, while DCCP also presents the possible binding sites of copper ions with ligands. The new feature of DCCP is that it offers the 3D structures of 2,777 primary sequences in DCCP_1D, which are modeled according to the homologous structures (similarity >30%) in PDB by the module MODELER in Insight II package. The predicted structures can be taken as alternatives for these structure-unknown proteins, which can provide detailed information for studying the structure-function relationships of copper-chelating proteins. Another novel feature is that the available cellular location of copper- and iron-chelating proteins is provided, which will facilitate the study on the functions of these proteins.

## Database analyses

To get some new insights into metalloproteins, we performed the primary sequence comparison, secondary structure prediction, and SCOP structure classification for copper-chelating proteins, iron-chelating proteins, and general proteins.

Based on DCCP_3D, we selected 390 copper-containing proteins to do the analyses. Some proteins have more than one chain, so the total sequence number is 574. To compare with ion-containing proteins, the same amount of sequences without redundancy (two proteins holding the same function, coming from the same species and with the sequence similarity higher than 99% are considered to be redundancy) was selected from DICP_3D (Some data are redundant in DICP_3D. For instance, the proteins that contain heme as cofactor hold nearly one third of all proteins). Furthermore, the same amount of general proteins was selected from PDB, which contained no metals, as a control to compare with copper- and iron-chelating proteins. The data in DCCP_1D and DICP_1D were excluded because of the deficiency of some necessary information.

### *Primary sequence comparison*

As is well known, proteins are very complicated biomolecules, in which twenty amino acids can be arranged in different orders. Even if a little difference in the sequence can cause an obvious change of the protein's function. So we firstly compared the amino acid compositions of copper-chelating proteins, iron-chelating proteins, and the same amount of general proteins. The twenty amino acids are classified into three groups, charged, polar no-charged, and non-polar, respectively. Figure 1 illustrates the average proportion of each amino acid of the three kinds of proteins. Independent-samples T test (P<0.001) was preformed to compare the amino acid content of copper- or iron-chelating proteins and general proteins, respectively (data not shown). The results show that the average contents of His, Arg, Thr, Gly, Leu, Val, Phe, Pro, and Met in copper-chelating proteins are significantly different from those of general proteins, suggesting that some of which may play impor-
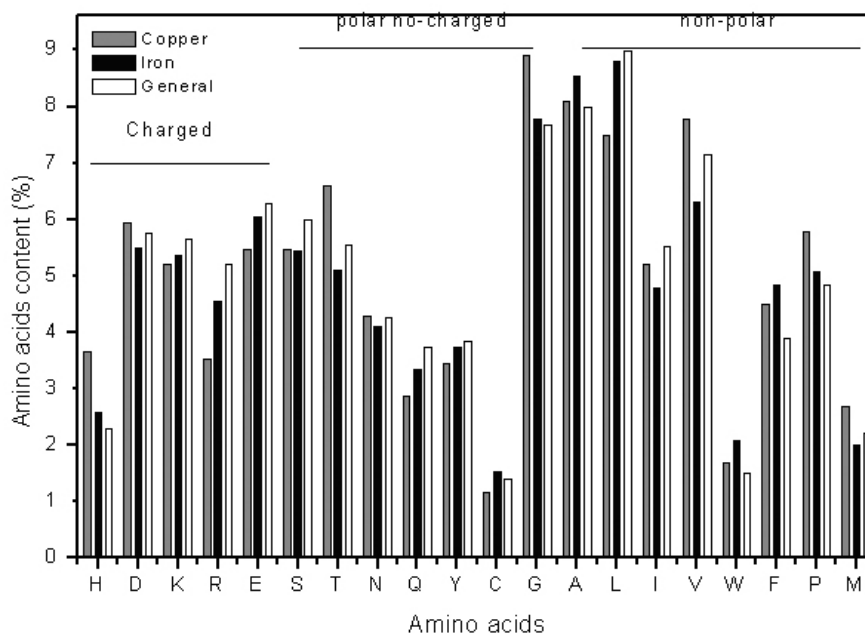
**Fig. 1** Plot for the twenty amino acid contents of proteins in DCCP_3D, DICP_3D, and the same amount of general proteins selected from PDB. The gray columns indicate copper-chelating proteins, black columns indicate iron-chelating proteins, and white columns indicate general proteins. Independent-samples T test (P<0.001) shows that the average contents of His, Arg, Thr, Gly, Leu, Val, Phe, Pro, and Met in copper-chelating proteins are significantly different from those of general proteins, and the contents of Arg, Ser, Thr, Ala, Ile, Val, Trp, and Phe in iron-chelating proteins are different from those of general proteins.

tant roles in copper-chelating proteins. In fact, His, Gly, and Met are the most important Cu(II)-binding amino acid residues (14). Furthermore, it is interesting to note that the content of Gly is the highest in copper-chelating proteins, which is prone to form β-sheet (15). For iron-chelating proteins, the average contents of Arg, Ser, Thr, Ala, Ile, Val, Trp, and Phe are significantly different from those of general proteins, indicating that the preference of amino acids in iron proteins is different from that in copper proteins and general proteins. While for the charged (H, D, K, R, E), polar no-charged (S, T, N, Q, Y, C, G), and non-polar (A, L, I, V, W, F, P, M) amino acid residues, the three kinds of proteins have similar content (Table 1), which is confirmed by the independent-samples T test (P<0.001).

**Table 1 Contents of Three Amino Acid Groups**

| Group | Charged (%) | Polar no-charged (%) | Non-polar (%) |
|---|---|---|---|
| Copper-chelating proteins | 23.8 | 32.8 | 43.2 |
| Iron-chelating proteins | 24.0 | 31.0 | 42.4 |
| General proteins | 25.2 | 32.5 | 42.1 |

Charged: H, D, K, R, E; Polar no-charged: S, T, N, Q, Y, C, G; Non-polar: A, L, I, V, W, F, P, M.

### Secondary structure prediction

Since metalloproteins account for about one third of the natural proteins, it is highly necessary to know the prediction accuracies of the existing software for secondary structure prediction for these proteins. Based on the selected proteins, we evaluated the popular software for secondary structure prediction, such as PredictProtein (including PHD and PROF), NNPREDICT, PSIPRED, and JPRED. The evaluation criterion was based on $Q_3$ index, which is defined as the number of amino acids predicted correctly divided by the number of all amino acids. The standard PDB secondary-structure annotations include eight types of protein secondary structure elements (H, I, G, E, B, S, T, -), while the software only predict three states,

$\alpha$-helix, $\beta$-sheet, and loop, respectively. The eight states are reduced to three states according to several criteria (*16*). Table 2 lists the prediction accuracies based on five widely-used criteria. It can be seen that the prediction accuracies are very similar for different conversion criteria. The last column lists the prediction accuracies for general proteins, which were selected from references (*16*). The predicted results for copper-chelating proteins are as good as those for general proteins, suggesting that the existing software is appropriate to predict the secondary structures of metalloproteins.

**Table 2 Secondary Structure Prediction Accuracies of Copper-Chelating Proteins in DCCP_3D Based on Five Secondary-Structure-Predicting Software**

| Software | Copper-chelating proteins | | Average accuracy for |
|---|---|---|---|
| | conversion*[1,2] | conversion*[3,4,5] | General proteins |
| PHD | 74.8 | 76.4 | 71.9–73.5 |
| PROF | 76.9 | 78.3 | >78 |
| NNPREDICT | 60.6 | 62.3 | >65 |
| PSIPRED | 79.1 | 80.3 | 78–80.6 |
| JPRED | 74.6 | 76.6 | 72.9–74.8 |

* Conversion criteria from 8 types to 3 types: [1]H, G, I—H; E, B—E; Other—C. [2]H, G—H; E, B—E; Other—C. [3]H—H; E—E; Other—C. [4]H, G, I—H; E—E; Other—C. [5]H, G—H; E—E; Other—C.

### SCOP classification

Finally, SCOP classification based on the structure domain level for three kinds of proteins were analysed (*17*). Table 3 enumerates the number of structure domains in eleven SCOP classes of the three types of proteins. It is revealed that the numbers of domains in all beta proteins and the membrane and cell surface proteins of copper-chelating proteins are twice more than that of general proteins, while that of all alpha proteins and the alpha and beta proteins (a/b) are opposite. Therefore, in the currently structure-known metalloproteins, copper-chelating proteins prefer to form $\beta$-sheet. The fact is consistent with that Gly is the most abundant amino acid residue in copper proteins, which has a high propensity to form $\beta$-sheet (*15*). However, for iron-chelating proteins, the number of domains in all alpha proteins is nearly treble of the general proteins and sextuple of the copper-chelating proteins, and the membrane and cell surface proteins are about decuple of the general proteins. The reason is that some of the iron-chelating proteins, such as the ferritin-like, heme-dependent peroxidases, and heme oxygenase, are categorized into all alpha proteins in the SCOP. Furthermore, some iron-chelating proteins play an important role in the process of photosynthesis, so the number of domains in membrane and cell surface proteins and peptides is much higher than that of general proteins. The different functions of copper and iron proteins make them select different structure domains. It should be pointed that the statistical result is based on the currently available copper- and iron-chelating proteins with known structures, which might be updated when more data are available.

In conclusion, the primary sequence comparison, secondary structure prediction, and SCOP structure classification of three kinds of proteins indicate that metalloproteins are different from normal proteins in some aspects, such as the amino acid composition and the SCOP structure classification, which are relevant to their functions. However, some common features indeed exist in these proteins, which results in the secondary structure predicting accuracies for metalloproteins. As many diseases, especially neurodegenerative diseases, are associated with the imbalance in metabolism of metal ions or mutations of metalloproteins (*18*), the research on metalloproteins will definitely offer opportunities for the discovery of novel therapeutic and diagnostic agents for a series of diseases. With the rapid development of proteomics and experimental technologies, more and more structures of copper- and iron-chelating proteins will be available. DCCP and DICP will be updated semiyearly. New identified structures will be added in time and new analyses will be carried out. It can be expected that our databases will facilitate the research on metal-chelating proteins.

**Table 3 SCOP Classifications Results**

| SCOP classification (domain) | Copper-chelating | Iron-chelating | General |
|---|---|---|---|
| All alpha proteins | 57 | 356 | 120 |
| All beta proteins | 777 | 148 | 369 |
| Alpha and beta proteins (a/b) | 27 | 127 | 225 |
| Alpha and beta proteins (a+b) | 139 | 198 | 205 |
| Multi-domain proteins (alpha and beta) | 8 | 13 | 13 |
| Membrane and cell surface proteins and peptides | 93 | 130 | 13 |
| Small proteins | 19 | 44 | 28 |
| Coiled coil proteins | 0 | 0 | 16 |
| Low resolution protein structures | 2 | 6 | 1 |
| Peptides | 0 | 0 | 0 |
| Designed proteins | 0 | 0 | 0 |

SCOP classifications of copper-chelating proteins in DCCP_3D and iron-chelating proteins in DICP_3D as well as general proteins selected from PDB.

## Materials and Methods

The construction of DCCP and DICP was based on the well-known web technologies: a fast database management system (MySQL; http://www.mysql.com), a stable web server (Apache; http://www.apache.org), and a powerful web scripting language (PHP; http://www.php.net) on the basis of an open source operation system—Linux. With these free and powerful tools, we created the two databases with interactive web interfaces. Each entry has a unique database identification number, amino acid sequence, protein function, and other available information. Further information can be found at the website http://sdbi.sdut.edu.cn/en.

## Acknowledgements

## References

1. Chan, S., *et al.* 1998. The role of copper, molybdenum, selenium, and zinc in nutrition and health. *Clin. Lab. Med.* 18: 673-685.

2. Lieu, P.T., *et al.* 2001. The roles of iron in health and disease. *Mol. Aspects. Med.* 22: 1-87.

3. Lippard, S.J. 1999. Free copper ions in the cells? *Science* 284: 748-749.

4. Emerit, J., *et al.* 2001. Iron metabolism, free radicals, and oxidative injury. *Biomed. Pharmacother.* 55: 333-339.

5. Gaetke, L.M. and Chow, C.K. 2003. Copper toxicity, oxidative stress, and antioxidant nutrients. *Toxicology* 189: 147-163.

6. Waggoner, D.J., *et al.* 1999. The role of copper in neurodegenerative disease. *Neurobiol. Dis.* 6: 221-230.

7. Bush, A.I. 2000. Metals and neuroscience. *Curr. Opin. Chem. Biol.* 4: 184-191.

8. Lehmann, S. 2002. Metal ions and prion diseases. *Curr. Opin. Chem. Biol.* 6: 187-192.

9. Yantiri, F. and Andersen, J.K. 1999. The role of iron in Parkinson disease and 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine toxicity. *IUBMB Life* 48: 139-141.

10. Harris, E.D. 2000. Cellular copper transport and metabolism. *Annu. Rev. Nutr.* 20: 291-310.

11. O'Halloran, T.V. and Culotta, V.C. 2000. Metallochaperones, an intracellular shuttle service for metal ions. *J. Biol. Chem.* 275: 25057-25060.

12. Harrison, M.D., *et al.* 1999. Copper chaperones: function, structure and copper-binding properties. *J. Biol. Inorg. Chem.* 4: 145-153.

13. Papanikolaou, G. and Pantopoulos, K. 2005. Iron metabolism and toxicity. *Toxicol. Appl. Pharmacol.* 202: 199-211.

14. Ji, H.F. and Zhang, H.Y. 2004. A theoretical study on Cu(II)-binding modes and antioxidant activity of mammalian normal prion protein. *Chem. Res. Toxicol.* 17: 471-475.

15. Street, A.G. and Mayo, S.L. 1999. Intrinsic $\beta$-sheet propensities result from van der Waals interactions be-

tween side chains and the local backbone. *Proc. Natl. Acad. Sci. USA* 96: 9074-9076.

16. Cuff, J.A. and Barton, G.J. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins* 34: 508-519.

17. Murzin, A.G., *et al.* 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247: 536-540.

18. Wong, B.S., *et al.* 2001. Oxidative impairment in scrapie-infected mice is associated with brain metals perturbations and altered antioxidant activities. *J. Neurochem.* 79: 689-698.