Novel types of two-domain multi-copper oxidases: possible missing links in the evolution

Kensuke Nakamura[a,*], Takeshi Kawabata[a], Kei Yura[b], Nobuhiro Go[a,b]

[a] *Graduate School of Information Science, Nara Institute of Science and Technology, 8916-5, Takayama-cho, Ikoma, Nara 630-0101, Japan*
[b] *Quantum Bioinformatics Group, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, 8-1, Umemidai, Soraku-gun, Kyoto 619-0215, Japan*

Received 12 July 2003; revised 25 August 2003; accepted 26 August 2003

First published online 24 September 2003

Edited by Robert B. Russell

Abstract **An analysis of the genome sequence database revealed novel types of two-domain multi-copper oxidases. The two-domain proteins have the conspicuous combination of blue-copper and inter-domain trinuclear copper binding residues, which is common in ceruloplasmin and ascorbate oxidase but not in nitrite reductase, and therefore are considered to retain the characteristics of the plausible ancestral form of ceruloplasmin and ascorbate oxidase. A possible evolutionary relationship of these proteins is proposed.**

Key words: Copper binding protein; Cupredoxin; Cyanobacterium; Genome sequence; Phylogeny; Rhizobium

1. Introduction

Protein-bound metal ions exhibit various functions, such as catalysis, electron transfer, oxygen transfer and signal transmission. Despite their importance, very few bioinformatics resources have been dedicated to analyses of the interactions between metal ions and proteins, so far [1]. Copper is one of the essential metal elements. Multi-copper blue proteins[1] exploit the distinctive redox properties of copper [2,3]. They contain two types of copper binding sites, blue-copper binding (BCB) sites and inter-domain copper binding (IDCB) sites. Historically, blue-copper has been called a type 1 copper and the inter-domain coppers have been recognized as type 2 and type 3 coppers, based on their spectroscopic characteristics [3,4]. Amino acid residues that coordinate copper ions are well conserved in proteins; therefore the presence or absence of copper binding sites in each domain can be easily identified solely from the amino acid sequences (Fig. 1).

Multi-copper blue proteins consist of two, three or six homologous domains (Fig. 2). These domains are also homologous with cupredoxins, such as azurin and plastocyanin, which are mono-domain proteins with a blue copper [5]. Ceruloplasmin is a six-domain multi-copper blue protein [6]. Ascorbate oxidase and laccase are three-domain multi-copper blue proteins [7,8]. A wide variety of three-domain multi-copper blue proteins can be found in plants, fungi and bacteria [9,10]. In this letter, we use the name ascorbate oxidase to represent the three-domain multi-copper blue proteins. Nitrite reductase is a two-domain multi-copper blue protein, which has a BCB site in the first domain [11]. For its function, this two-domain protein forms a homo-trimer complex that structurally resembles the six-domain ceruloplasmin.

The IDCB sites of ascorbate oxidase and ceruloplasmin are trinuclear, consisting of one type 2 and two type 3 copper ions. The three-dimensional structure around these trinuclear copper binding sites, as well as their spatial proximity to the nearest BCB sites, is very similar between ceruloplasmin and ascorbate oxidase [6]. This strongly suggests that they most likely have evolved from a common ancestor. The IDCB site of nitrite reductase is mononuclear, with only a type 2 copper.

Ryden and Hunt [12], Murphy et al. [13] and Lindley et al. [4] have proposed mutually similar hypotheses regarding the course of domain evolution of multi-copper blue proteins. There is a consensus that a domain duplication occurred as the primary event, which explains the fact that all of the domains of multi-copper blue proteins can be classified into two groups, based on sequence similarity of the domains. These domain classes are called class IV and class V by Murphy et al. [13], and are color-coded pink and blue, respectively, in Fig. 2. Ryden and Hunt [12] explained the formation of a six-domain protein, ceruloplasmin, as a consequence of triplication of the two-domain hypothetical ancestor formed by the initial domain duplication, followed by several modifications of copper binding sites. They also suggested that three-domain multi-copper blue proteins, such as ascorbate oxidase, have evolved by a single domain addition to the two-domain protein. Alternatively, Murphy et al. suggested that ascorbate oxidase could have evolved from a six-domain protein by replacing domains 3 through 5 with a short linker [13]. In these hypotheses, the order of events for (1) the loss of BCB sites and (2) the acquisition of inter-domain sites was unclear. Also, the copper binding site organizations of the two-domain common ancestors have not been considered explicitly.

In either scenario of the evolutionary pathway, we now recognize that it requires the involvement of several types of two-domain ancestral proteins, as postulated in Fig. 2. The five two-domain hypothetical proteins are labeled with brack-

*Corresponding author. Fax: (81)-743-72 5391.
E-mail address: kenske@is.aist-nara.ac.jp (K. Nakamura).

Abbreviations: BCB, blue-copper binding; IDCB, inter-domain copper binding

[1] We use the term 'multi-copper blue protein' to refer to a group of proteins including nitrite reductase and multi-copper oxidases, such as ceruloplasmin, ascorbate oxidase and laccase.

0014-5793/03/$22.00
doi:10.1016/S0014-5793(03)01000-7

ets, as [**X**], [**Y**], [**A**], [**B**] and [**C**]. We explicitly pictured the evolutionary sequence of the two-domain proteins ([**X**]→[**Y**]→[**A**]→[**B**]), which would have led to the formation of multi-copper blue proteins.

In this letter, we describe our search for these hypothetical proteins in genome sequences and their identification. Our findings strongly support our proposed evolutionary pathways.

2. Materials and methods

The BLASTP program [14] was used to search the non-redundant (nr) set of protein sequences provided at NCBI [15]. Multiple sequence alignments were constructed using the T-Coffee program [16] followed by manual correction to remove the non-overlapped regions. Phylogenetic analysis was performed using the neighbor-joining method, which was followed by local rearrangements using the maximum-likelihood algorithm to predict the most likely tree. The tree topology was assessed with the re-sampling of the estimated log-likelihoods (RELL) bootstrap method. These analyses were carried out using the MOLPHY program [17]. Protein sequences are identified by their SwissProt entry names when possible. Otherwise, the NCBI accession number or NCBI sequence ID is indicated.

3. Results and discussion

3.1. Novel types of two-domain multi-copper blue proteins from whole genome sequences

We picked out the sequences of Cu-oxidase domains (PF00394) from the Pfam profile database [18]. Using them as queries, we ran BLAST searches against the nr protein sequence database. Among the obtained sequences, 11 proteins were found to possess the sequence pattern of copper binding sites characteristic of either [**A**], [**B**] or [**C**].

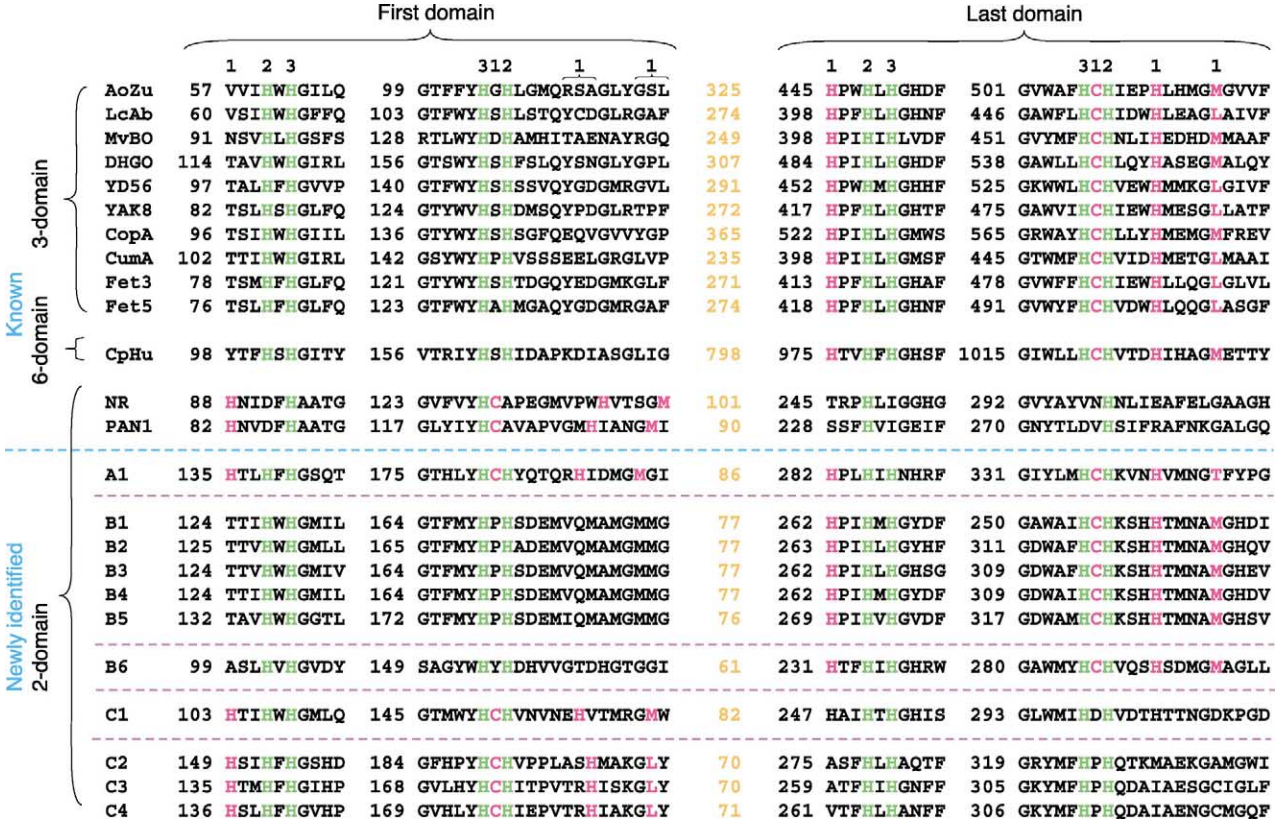

Fig. 1. Sequence alignment around the copper binding sites in multi-copper blue proteins. The sequences in the top half (above the blue broken line) are of conventional multi-copper blue proteins, and those in the bottom half (below the blue broken line) are of the newly identified two-domain multi-copper blue proteins. The novel two-domain proteins were clustered into five groups, according to sequence similarity, and are separated by purple dotted lines. The numbers 1, 2, and 3, at the top of the alignments, indicate the consensus positions of the type 1, type 2 and type 3 copper binding residues, respectively. The residues colored in red are of BCB sites. The residues colored in green are type 2/3 IDCB sites. Two sequence fragments, from both the first and last domains of the molecules, are indicated. In the cases of the three-domain and six-domain multi-copper oxidases, the last domain is the third and sixth domain, respectively. Numbers on the left side of each sequence fragment show the first residue position of the fragment. The yellow number in the middle column is the number of residues between the second and third fragments. The sequence ID (SwissProt ID or NCBI accession number), the common name (if any), and the origin of each sequence are listed below. AoZu: ASO_CUCPM, ascorbate oxidase from *Cucurbita pepo* var. *melopepo*; LcAb: LAC1_AGABI, laccase from *Agaricus bisporus*; MvBO: BLRO_MYRVE, bilirubin oxidase from *Myrothecium verrucaria*; DHGO: Q00292, dihydrogeodin oxidase from *Aspergillus terreus*; YD56: YD56_YEAST, YD56 from *Saccharomyces cerevisiae*; YAK8: YAK8_SCHPO, YAK8 from *Schizosaccharomyces pombe*; CopA: CPA1_PSESM, CopA from *Pseudomonas syringae*; CumA: NP_743195, CumA from *Pseudomonas putida*; Fet3: FET3_YEAST, ferrireductase from *Saccharomyces cerevisiae*; Fet5: P43561, ferrireductase from *Saccharomyces cerevisiae*; CpHu: CERU_HUMAN, ceruloplasmin from *Homo sapiens*; NR: NIR_ACHCY, nitrite reductase from *Achromobacter cycloclastes*; PAN1: ANIA_NEIGO, PAN1 from *Neisseria gonorrhoeae*; A1: Q9HQF4 from *Halobacterium* strain sp. NRC-1; B1: Q92S43 from *Sinorhizobium meliloti*; B2: Q8U8U7 from *Agrobacterium tumefaciens* str. C58; B3: ZP_00029340 from *Burkholderia fungorum*; B4: NP_768850 from *Bradyrhizobium japonicum* USDA 110; B5: ZP_00052601 from *Magnetospirillum magnetotacticum*; B6: Q93HV5 from *Streptomyces griseus*; C1: ZP_00002680 from *Nitrosomonas europaea*; C2: NP_711736 from *Leptospira interrogans* serovar lai str; C3: ZP_00073469 from *Trichodesmium erythraeum* IMS 101; C4: NP_487982 from *Nostoc* sp. PCC 7120.

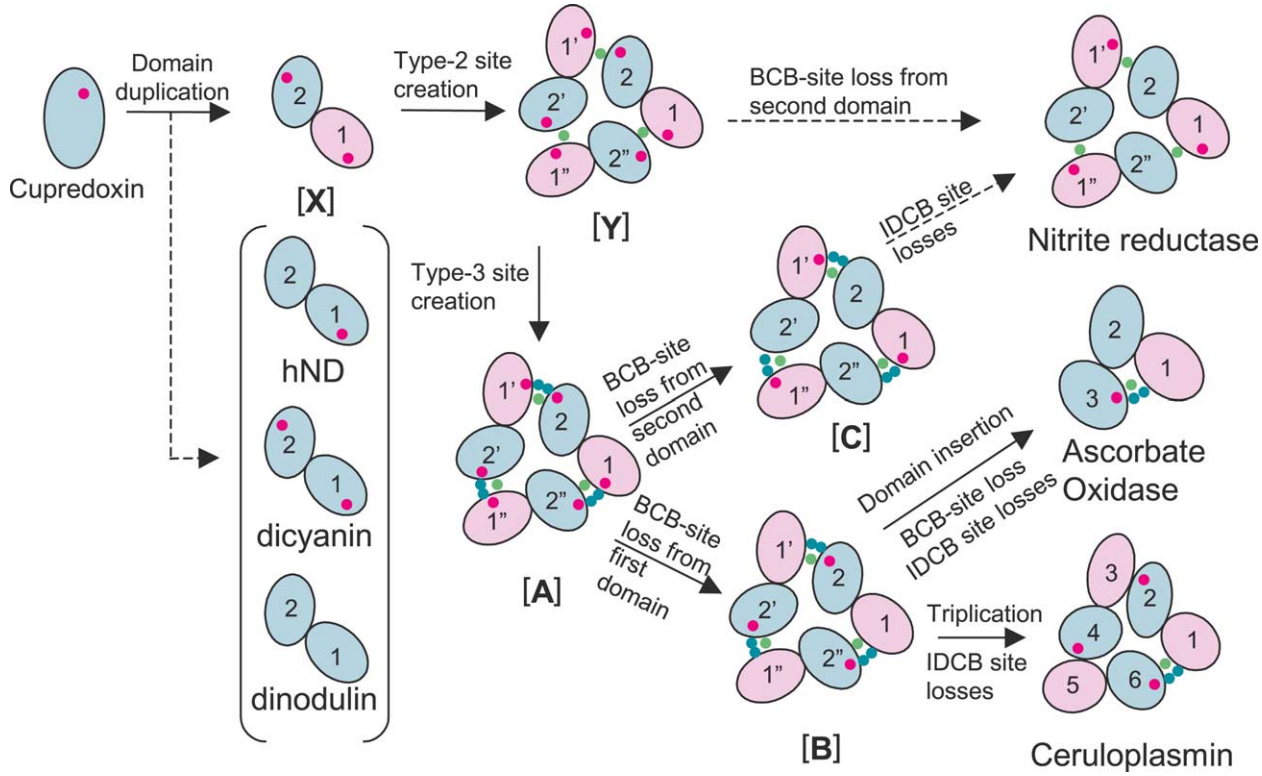

Fig. 2. Schematic representation of domain organizations of the multi-copper blue proteins and their hypothetical common ancestors. An oval shape indicates a cupredoxin or multi-copper blue protein domain. Pink and blue colors show the domain class (classes IV and V, respectively). Red, green, and blue dots indicate type 1, type 2, and type 3 coppers, respectively. Solid arrows indicate postulated pathways and dotted arrows indicate alternative pathways. Conventional multi-copper blue proteins are shown in the right-most column. Five hypothetical proteins are shown with bracketed labels. Three of them ([**X**], [**Y**] and [**A**]) have BCB sites in both domains, whereas [**B**] and [**C**] have a BCB site only in the first and second domain, respectively. The hypothetical proteins [**A**], [**B**] and [**C**] have trinuclear IDCB sites, whereas [**Y**] has mononuclear IDCB sites, and [**X**] has no IDCB site. Doubled BCB domain proteins are shown in parentheses underneath [**X**].

Sequence alignments around the copper binding sites of the novel two-domain multi-copper blue proteins are shown in the bottom half of Fig. 1. One sequence (A1: Q9HQF4) from an archaeon, *Halobacterium* sp. NRC-1, has the copper binding site pattern of the hypothetical ancestor [**A**]. Six sequences (B1: Q92S43, B2: Q8U8U7, B3: ZP_00029340, B4: NP_768850, B5: ZP_00052601 and B6: Q93HV5) share the copper binding site pattern of [**B**], and four sequences (C1: ZP_00002680, C2: NP_711736, C3: ZP_00073469 and C4: NP_487982) have the copper binding site pattern of [**C**]. All of the domains of these proteins have four IDCB histidines (residues shown in green in Fig. 1), which are enough to hold three copper ions at the inter-domain site. The proteins with patterns [**A**] and [**B**] have BCB sites (residues shown in red) in the second domain, and the proteins with patterns [**B**] and [**C**] have BCB sites in the first domain. An analysis of the sequence similarity revealed that the first and second domains of the 11 novel two-domain proteins can be classified into two groups, corresponding to the class IV and class V domains of multi-copper blue proteins. This finding is consistent with the expectation that these 11 proteins are the direct descendants of the hypothetical ancestors [**A**], [**B**] and [**C**], as depicted in Fig. 2.

Since the monomeric two-domain multi-copper blue proteins cannot form IDCB sites, we assume that these proteins aggregate to form homo-trimers, like the case of nitrite reductase. When the two-domain protein in the first group (A1) forms a homo-trimer as we assumed, and is fully loaded with copper ions, the resulting complex can hold as many as 15 copper ions per unit.

These two-domain multi-copper blue proteins have not been previously recognized. In the process of automated or semi-automated sequence annotation, the differences between the copper binding residue locations in these two-domain multi-copper blue proteins have been overlooked. The sequence similarity with the Cu-oxidase profile in each of the two domains is easily detected. However, these proteins have been inadequately annotated, as either a nitrite reductase or a fragment of other multi-copper oxidases, such as ascorbate oxidase.

3.2. *Phylogenetic relationships of the conventional and novel multi-copper blue proteins*

To investigate the evolutionary relationship between these new two-domain proteins and other conventional multi-domain proteins, we carried out a phylogenetic analysis. To construct a multiple sequence alignment of proteins with different numbers of domains, we used partial sequences for large proteins. For ascorbate oxidase and ceruloplasmin, only the first and last domains were used. The procedure described in Section 2 provided a multiple alignment of 175 amino acids without gaps. A phylogenetic tree of the new two-domain proteins together with ceruloplasmin, ascorbate oxidase, and nitrite reductase, using the maximum-likelihood

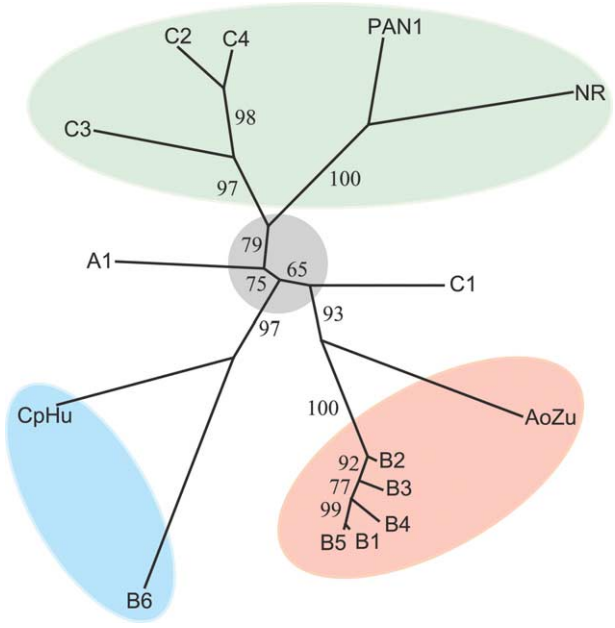

Fig. 3. A rootless phylogenetic tree constructed from a sequence alignment of the conserved regions of multi-copper blue proteins. The dendrogram was generated using the neighbor-joining method followed by local rearrangements using the maximum-likelihood algorithm [17]. Numbers at branches are RELL bootstrap values. Sequences are identified by the codes used in Fig. 1. The gray background circle indicates the low-reliability region, and the colored background ovals indicate the suggested evolutionary relationships.

method, is shown in Fig. 3. An attempt to determine the root position by making a multiple alignment with each domain of these proteins was not successful, due to the low resolution caused by the shortness of the multiple alignment (ca. 50 amino acids) and the low sequence identity among them. Indeed, even in the case of Fig. 3, the distant relationship (region within gray shaded circle) has low bootstrap values and is not reliable. Nevertheless, we can draw several conclusions from the phylogenetic tree. Among the new proteins with a BCB site only in the second domain (type [**B**]), five proteins (B1 through B5), except B6, were close to each other. Also, among the proteins with a BCB site only in the first domain (type [**C**]), three proteins (C2, C3 and C4), except C1, were close to each other. Therefore, the novel two-domain multi-copper blue proteins are categorized into five groups, consisting of (1) A1, (2) B1–B5, (3) B6, (4) C1 and (5) C2–C4, as separated by purple lines in Fig. 1. In the phylogenetic tree, ascorbate oxidase shares the closest common ancestor with the second group (B1–B5) (sequence identity ~35%). Also, ceruloplasmin shares the closest common ancestor with the third group (B6), and nitrite reductase shares it with the fifth group (C2–C4). These suggested relationships are consistent with our hypothesis that ascorbate oxidase and ceruloplasmin evolved from the type [**B**] two-domain protein, as depicted in Fig. 2. The relationship between nitrite reductase and the type [**C**] proteins is rather ambiguous. It is also possible that nitrite reductase has directly descended from a type [**Y**] protein.

We hypothesized that types [**B**] and [**C**] share the type [**A**] protein as a common ancestor. However, the evolutionary pathways from [**A**] to [**B**] and [**C**] still remain ambiguous. For example, it is possible that two independent BCB site losses occurred from [**A**], to generate the two distinct groups of type [**B**] proteins (groups B1–B5 and B6), each of which led to the three- and six-domain multi-copper oxidases. Also, the separation of two groups of type [**C**] proteins on the phylogenetic tree is not consistent with our parsimonious scenario

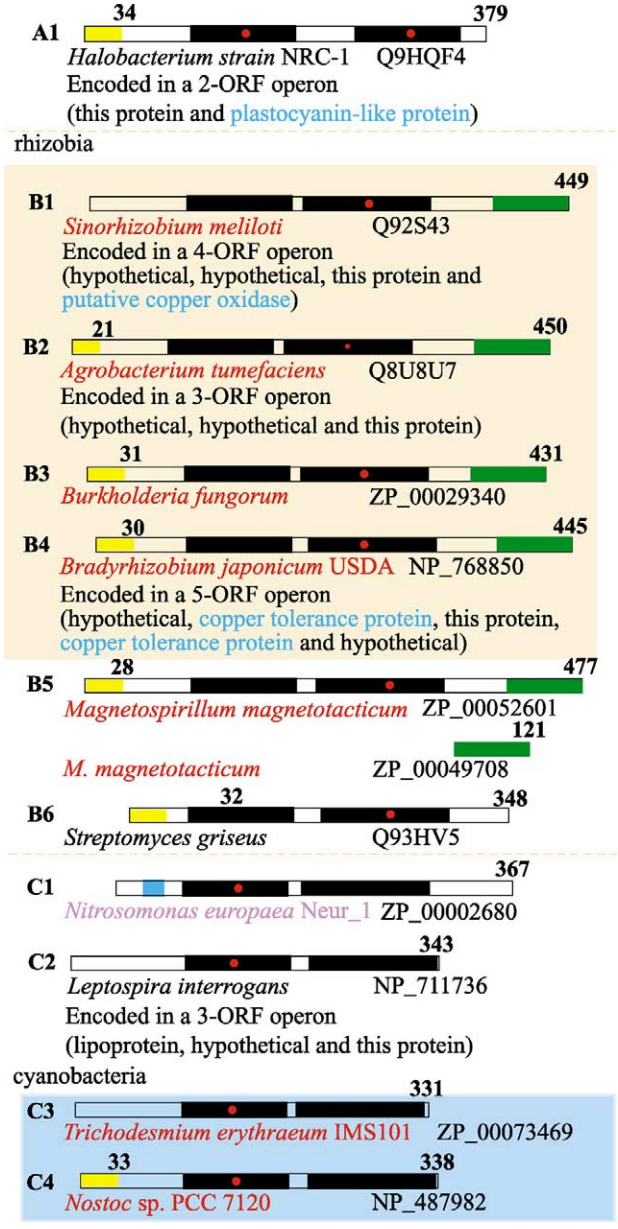

Fig. 4. Sequence analyses of novel multi-copper oxidases. Each sequence is described by a rectangle with the name of the species from which the protein is derived, an accession number, and an operon structure, if available. A black box in a rectangle indicates a multi-copper oxidase domain, a yellow box indicates a putative signal peptide, a blue box indicates a transmembrane region, and a green box indicates a conserved domain found only in the six proteins shown. Functions of the green domains are unknown. A red circle in a black box is a type 1 copper. Red species names indicate that these species fix nitrogen molecules and a pink name refers to a species that oxidizes ammonia to nitrate. The brown background covering B1 to B4 indicates that these proteins are derived from rhizobia, a root nodule-forming species. The blue background covering C3 and C4 indicates that these proteins are derived from nitrogen-fixing cyanobacteria.

shown in Fig. 2. This may be a consequence of the low reliability of the tree in the distant region (gray shaded area in Fig. 3), or it may suggest independent BCB site loss events for each group of type [**C**] proteins.

The evolutionary pathway which led to the acquisition of the IDCB sites ([**X**]→[**Y**]) is intriguing. It is hard to imagine that a doubled cupredoxin [**X**] acquired IDCB sites by simultaneous mutations of at least four distant locations on the sequence. There is a cupredoxin (amicyanin from *Paracoccus denitrificans*) that possesses two histidines at sequence positions that correspond to the IDCB site. If this kind of cupredoxin duplicated, then the product would be ready to form an IDCB site by trimer formation.

3.3. BCB domain duplication events

Recently Nersissian et al. [19,20] isolated a duplicated cupredoxin protein called dicyanin (Q9M510) from tomato. From a BLAST analysis, they also found two other kinds of duplicated cupredoxin proteins. One of them is a two-domain protein with no blue copper, named dinodulin (Q9MAK3, Q9LFI4 and Q9M135), from *Arabidopsis thaliana*. The other is the duplicated cupredoxin protein hND, which lacks blue copper in the first domain, from *Halobacterium* sp. NRC-1 (Q9HPH3), the same origin as A1. Unlike the multi-copper blue proteins, these duplicated cupredoxins lack IDCB sites (Fig. 2).

A phylogenetic analysis of the domains of duplicated cupredoxins together with the domains of multi-copper blue proteins (dendrogram not shown) showed that the domains of dicyanin and the domains of dinodulins formed clusters independent of the domain cluster of the multi-copper blue proteins. This observation suggests that the duplications of cupredoxin to form dicyanin and dinodulin were separate events from the one that led to the foundation of the multi-copper blue protein family.

3.4. Possible functional annotation of the newly identified two-domain multi-copper blue proteins

The proteins that we have identified lacked adequate annotation for biological functions in the sequence databases. We used conventional function prediction methods to obtain information about the biological functions of the proteins. The results are summarized in Fig. 4.

Most of the proteins classified in the multi-copper blue protein family are enzymes that catalyze the oxidation/reduction of small organic/inorganic substrates [21]. The IDCB sites of the novel multi-copper blue protein are trinuclear. Trinuclear IDCB sites are found in multi-copper oxidases, such as ceruloplasmin and ascorbate oxidases, whereas the IDCB sites of nitrite reductase are mononuclear. The catalytic activity of the trinuclear site is the reduction of oxygen to water, and that of the mononuclear site is the reduction of nitrite to nitric oxide. The similarity of the IDCB sites suggests that the molecular functions of the novel proteins should be similar to those of ascorbate oxidase or ceruloplasmin, rather than to that of nitrite reductase.

Proteins B1 to B5 have the same domain organizations (Fig. 4) and were clustered in the phylogenetic tree (Fig. 3). Analyses by SOSUI [22] and PSORT [23] predicted that these sequences have signal peptides at the N-terminus and are localized to the periplasmic space of the organism. These sequences are rich in methionine. *CopC*, one of the multi-copper oxidases, has a methionine-rich region that binds excess copper ions [24,25].

We have analyzed the genomic context of these newly identified proteins to gain insight into their function. B1 is encoded in a putative operon with four genes, which also contains *CopC*. B4 is encoded in a putative operon with five genes, which contains two more copper tolerance proteins [26]. These copper-containing proteins, including *CopC*, are assumed to be factors for copper resistance. The presumptive function of the proteins is to secrete copper ions in the periplasm [27]. Proteins encoded in the same operon are transcribed at the same time, and often have relevant biological functions [28]. Proteins B1 to B5 are, therefore, probably periplasmic proteins related to copper ion transport. A genomic link between a multi-copper oxidase (*CopC*) and a copper resistance protein (*CopA*) was suggested previously [24].

Seven out of 11 organisms with the novel multi-copper blue protein are capable of nitrogen fixation. B1, B2, B3 and B4 are all derived from rhizobia, bacteria that have established symbiosis with plants and fix nitrogen [29]. B5 is derived from *Magnetospirillum magnetotacticum*, which was recently discovered to fix nitrogen [30]. C3 and C4 are derived from cyanobacteria that fix nitrogen [31]. In addition to these, C1 is derived from *Nitrosomonas europaea*, which oxidizes ammonia to nitrate [32]. The distribution of our newly found proteins among those organisms and the fact that other organisms whose entire genomes have been sequenced lack the proteins suggest that the proteins may have functions related to nitrogen fixation. One of the best characterized proteins in multi-copper blue proteins is nitrite reductase, whose substrate/product is nitrite/nitric oxide. The catalytic functions of nitrite reductase rely on two types of copper ions and an aspartic acid residue located close to a type 2 copper ion [33]. This aspartic acid residue is replaced with histidine in the novel multi-copper blue protein. From these observations, we suggest that the novel proteins function differently from nitrite reductase.

The proteins in the multi-copper oxidase family are famous as moonlighting proteins, which are able to change their functions in response to environmental changes [34]. New functions of the multi-copper oxidase family may emerge from the novel proteins we have identified. The biological functions of these proteins should be revealed by future experiments.

3.5. Conclusions

The eleven novel multi-copper blue proteins reported here are unprecedented in their distribution of trinuclear IDCB sites in consecutive domains, and thus deserve recognition as new classes of multi-copper blue proteins. They share the conspicuous copper binding site pattern with ascorbate oxidase and ceruloplasmin, and fit in perfectly as a common ancestral form of these multi-copper blue proteins. A sequence analysis indicated that there were several independent occasions of cupredoxin domain duplication, and one of them led to the foundation of the multi-copper blue protein family.

Acknowledgements: We thank Dr. Gautam Basu for helpful discussions. Our research was supported by the Special Coordination Funds Promoting Science and Technology, and a Grant-in-Aid for Scientific Research on Priority Areas (C) 'Genome Information Science' from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

References

[1] Bertini, I. and Rosato, A. (2003) Proc. Natl. Acad. Sci. USA 100, 3601–3604.
[2] Messerschmidt, A. (Ed.) (1997) Multi-copper Oxidases, World Scientific, Singapore.
[3] Solomon, E.I., Sundaram, U.M. and Machonkin, T.E. (1996) Chem. Rev. 96, 2563–2605.
[4] Lindley, P., Card, G., Zaitseva, I. and Zaitsev, V. (1999) Perspect. Bioinorg. Chem. 4, 51–89.
[5] Messerschmidt, A., Huber, R., Poulos, T. and Wieghardt, K. (Eds.) (2001) Metalloproteins, John Wiley and Sons, Chichester.
[6] Zaitseva, I., Zaitsev, V., Card, G., Moshkov, K., Bax, B., Ralph, A. and Lindley, P. (1996) J. Biol. Inorg. Chem. 1, 15–23.
[7] Ducros, V., Brzozowski, A.M., Wilson, K.S., Ostergaard, P., Schneider, P., Svendson, A. and Davies, G.J. (2001) Acta. Crystallogr. D 57, 333–336.
[8] Messerschmidt, A., Luecke, H. and Huber, R. (1993) J. Mol. Biol. 230, 997–1014.
[9] Shimizu, A., Kwon, J.-H., Sasaki, T., Satoh, T., Sakurai, N., Sakurai, T., Yamaguchi, S. and Samejima, T. (1999) Biochemistry 38, 3034–3042.
[10] Alexandre, G. and Zhulin, I.B. (2000) Trends Biotechnol. 18, 41–42.
[11] Adman, E.T., Godden, J.W. and Turley, S. (1995) J. Biol. Chem. 270, 27458–27474.
[12] Ryden, L.G. and Hunt, L.T. (1993) J. Mol. Evol. 36, 41–66.
[13] Murphy, M.E.P., Lindley, P.F. and Adman, E.T. (1997) Protein Sci. 6, 761–770.
[14] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[15] National Center for Biotechnology Information, NIH, Bethesda, MD.
[16] Notredame, C., Higgins, D.G. and Heringa, J. (2000) J. Mol. Biol. 302, 205–217.
[17] Hasegawa, M., Kishino, H. and Saitou, N. (1991) J. Mol. Evol. 32, 443–445.
[18] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L.L. (2002) Nucleic Acids Res. 30, 276–280.
[19] Nersissian, A.M. and Shipp, E.L. (2002) Adv. Protein Chem. 60, 271–340.
[20] Hart, P.J., Nersissian, A.M., Herrmann, R.G., Nalbandyan, R.M., Valentine, J.S. and Eisenberg, D. (1996) Protein Sci. 5, 2175–2183.
[21] Hellman, N.E. and Gitlin, J.D. (2002) Annu. Rev. Nutr. 22, 439–458.
[22] Hirokawa, T., Boon-Chieng, S. and Mitaku, S. (1998) Bioinformatics 14, 378–379.
[23] Nakai, K. (2000) Adv. Protein Chem. 54, 277–344.
[24] Arnesano, F., Banci, L., Bertini, I., Mangani, S. and Thompsett, A.R. (2003) Proc. Natl. Acad. Sci. USA 100, 3814–3819.
[25] Barnett, M.J., Fisher, R.F., Jones, T., Komp, C., Abola, A.P., Barloy-Hubler, F., Bowser, L., Capela, D., Galibert, F., Gouzy, J., Gurjal, M., Hong, A., Huizar, L., Hyman, R.W., Kahn, D., Kahn, M.L., Kalman, S., Keating, D.H., Palm, C., Peck, M.C., Surzycki, R., Wells, D.H., Yeh, K.C., Davis, R.W., Federspiel, N.A. and Long, S.R. (2001) Proc. Natl. Acad. Sci. USA 98, 9883–9888.
[26] Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. and Tabata, S. (2002) DNA Res. 9, 189–197.
[27] Huynen, M., Snel, B., Lathe, W.III. and Bork, P. (2000) Genome Res. 10, 1204–1210.
[28] Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Trends Biochem. Sci. 23, 324–328.
[29] Weidner, S., Puhler, A. and Kuster, H. (2003) Curr. Opin. Biotechnol. 14, 200–215.
[30] Bazylinski, D.A., Dean, A.J., Schuler, D., Phillips, E.J. and Lovley, D.R. (2000) Environ. Microbiol. 2, 266–273.
[31] Stal, L.J. (2000) in: The Ecology of Cyanobacteria: Their Diversity in Time and Space (Whitton, B.A. and Potts, M., Eds.), pp. 61–120, Kluwer Academic, Dordrecht.
[32] Chain, P., Lamerdin, J., Larimer, F., Regala, W., Lao, V., Land, M., Hauser, L., Hooper, A., Klotz, M., Norton, J., Sayavedra-Soto, L., Arciero, D., Hommes, N., Whittaker, M. and Arp, D. (2003) J. Bacteriol. 185, 2759–2773.
[33] Ellis, M.J., Dodd, F.E., Sawers, G., Eady, R.R. and Hasnain, S.S. (2003) J. Mol. Biol. 328, 429–438.
[34] Bielli, P. and Calabrese, L. (2002) Cell. Mol. Life Sci. 59, 1413–1427.