The International Conference on Natural Sciences, Mathematics, Application, and Technology (ICON-SMART) Proceedings
e-ISSN: 2963-3915

Vol. 3, December 2023
pp. 1-7

In Silico Analysis of Copper Chaperone in Heavy Metal Tolerant Bacteria

Yohanis Irenius Mandik

[1]*Department of Chemistry, Faculty of Mathematics and Natural Sciences, Cenderawasih University*

Contact email: yimandik@gmail.com

Abstract. Transition metal ions, such as copper, iron, molybdenum, and cobalt, are essential in many biological processes. This is because these metals have different levels of oxidation state. Copper chaperone (CopZ) serves not only to manage copper translocation but also as an antioxidant. In this study, we conducted in silico analysis of the CopZ in a various bacteria that are tolerant to heavy metals. This study aimed to analyze the physicochemical properties and predict the structure of this protein. The alignment of the CopZ showed a conserved motif -CXXC- in all amino acid sequences aligned. CopZ of *Microbacterium laevaniformans*, a type of bacteria that is resistant to metals, nitrate, and low pH conditions, has been selected for structural prediction. Secondary structure analysis showed that CopZ from *M. laevaniformans* has 35.3% α-helix, 35.3% β-sheet, 16.2% turn, and 13.2% coil. In addition, the instability index was 35.63, which indicated that this protein is stable. The result of this study could be an initial framework to investigate further CopZ as a drug target candidate.

Keywords in-silico; CopZ; copper chaperone; structural prediction; -CXXC- motif.

INTRODUCTION

Transition metal ions, such as copper, iron, molybdenum, and cobalt, function as important cofactors in various biological processes in cells. This is because these metals can exist in various oxidation states in vivo. Copper is a redox-active metal used by almost all organisms and can fluctuate between oxidized (Cu^{2+}) and reduced (Cu^{+}) states. With this ability to change redox state, copper can coordinate with ligands such as carboxylic oxygen, imidazole nitrogen, thiolate, and sulfur thioether groups. However, copper's redox properties also make it toxic. The redox cycle between Cu(II) or Cu(I) can catalyze the production of toxic radicals, which in turn cause damage to lipids, proteins, DNA, and other biomolecules [1].

Under certain physiological cell conditions, the availability of free copper must be strongly controlled [2]. So that the transportation of copper into specific enzymes within the cell is managed by a group of proteins called "copper chaperones". These proteins directly transport copper in the cytoplasm to the specific site of utilization while preventing copper from improper interactions with other cellular components [3].

Several studies suggested that copper chaperones were associated with some human diseases. Mutations in central regulators of human cellular copper metabolism, the copper-transporting P-type ATPases, will disrupt the homeostatic copper balance, resulting in copper deficiency (Menkes disease), copper overload (Wilson disease), or idiopathic copper toxicosis [4-6]. Therefore further study of how copper chaperone is involved in those diseases is important.

The era of information technology brought a new chapter in the study of new drug discovery. Through computational studies, new drug discovery research can be shortened because it is neither time-consuming nor requires expensive chemicals. In addition, computational study is conducted based on the theoretical foundation as well as the availability of well-developed software that ca n support new drug discovery research. For that reason, we conducted in silico approach to investigate the physicochemical properties of some copper chaperones as well as the structural prediction of selected copper chaperones (CopZ) from heavy metal-tolerant bacteria. The result of this study is expected to be the cornerstone in further investigation about copper chaperone as a drug target candidate.

METHODS

Sequence retrieval and primary structure analysis

Sequences of copper chaperone and/or heavy metal transport/detoxification protein were retrieved in FASTA format from NCBI (http://ncbi.nlm.nih.gov). A total of 11 protein sequences were selected to be investigated. Table 1 showed the protein used in this study. All sequences then aligned with COBALT (http://www.ncbi.nlm.nih.gov/tools/cobalt/). In addition, the physicochemical properties, such as molecular weight, theoretical isoelectric point (pI), extinction coefficient (EC), instability index (II), aliphatic index (AI), and grand average hydropathy (GRAVY) were computed using the Expasy's ProtParam tool [7] (http://web.expasy.org/protparam/). Moreover, the secondary structure composition of all copper chaperones was analyzed by SOPMA (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_sopma.html)

Functional analysis and homology modeling

The presence of disulfide bridges (SS bonds) in CopZ is predicted by DiANNA (DiAminoacid Neural Network Application) [8] (http://clavius.bc.edu/~clotelab/DiANNA/). The modeling of three-dimensional structure of CopZ of *M. laevaniformans* was conducted by Phyre^2 [9] (http://www.sbg.bio.ic.ac.uk/~phyre2) and Swiss-Model (http://swissmodel.expasy.org/workspace/), and subsequently visualized by YASARA (www.yasara.org). The modeled structure was evaluated by PROCHECK (http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/) and validated by QMEAN (http://swissmodel.expasy.org/qmean/) and ModFOLD (http://www.reading.ac.uk/bioinf/ModFOLD/).

TABLE 1. Protein used in this study

No	NCBI reference sequence	Definition	Organism source
1	WP_005049374.1	Heavy metal transport/ detoxification protein	*Microbacterium laevaniformans*
2	WP_019157648.1	copper chaperone	*Brevibacterium* sp. JC43
3	WP_017883414.1	copper chaperone	*Leucobacter* sp. UCD-THU
4	YP_003155379.1	copper chaperone	*Brachybacterium faecium* DSM 4810
5	WP_005189537.1	copper chaperone	*Gordonia amarae*
6	WP_006213359.1	copper chaperone	*Kocuria palustris*
7	YP_004079619.1	copper chaperone	*Mycobacterium gilvum* Spyr1
8	WP_020018251.1	copper chaperone	*Promicromonospora sukumoe*
9	YP_004574732.1	copper chaperone	*Microlunatus phosphovorus* NM-1
10	WP_007924932.1	copper chaperone	*Janibacter hoylei*
11	YP_003148870.1	copper chaperone	*Kytococcus sedentarius* DSM 20547

RESULTS AND DISCUSSION

The alignment of CopZ (presented in Figure 1) showed that all CopZs from heavy metal tolerant bacteria have –CXXC– motif, a metal-binding sequence motif [10], near the N-terminal residue of the protein. All CopZ had sequence –CGHC– for this motif.

Based on data presented in Table 2, it was shown that most CopZ are hydrophobic due to the presence of high non-polar residues content (residues A, C, F, G, I, L, M, P, V, and Y). The table also showed that all copper chaperones have no Trp residue. Moreover, all copper chaperones have at least two Cys residues, indicating the possibility of disulfide bridge (SS bonds).

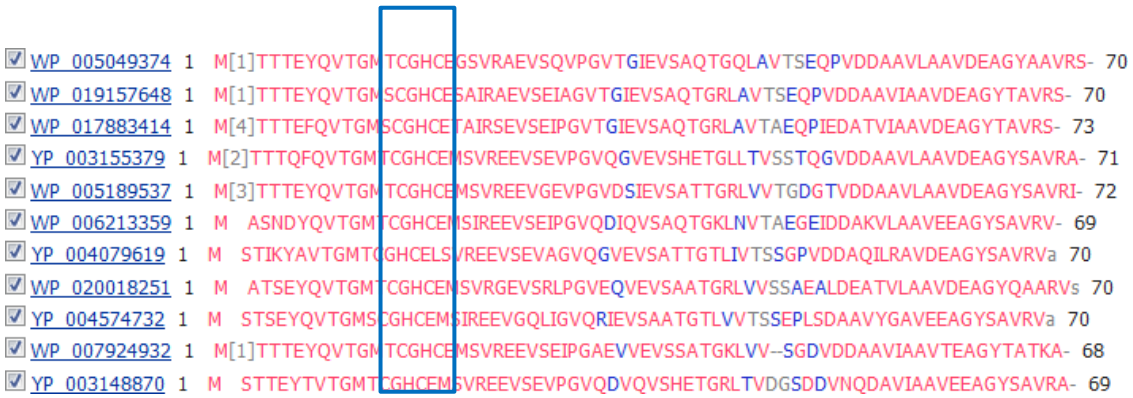

FIGURE 1. Sequence alignment of CopZ (blue-box indicated –CXXC- motif)

TABLE 2. Amino acid composition of CopZ (in mole percentage) by using Protparam tool

Amino acid residue	CopZ protein										
	WP_005049374.1	WP_019157648.1	WP_017883414.1	YP_003155379.1	WP_005189537.1	WP_006213359.1	YP_004079619.1	WP_020018251.1	YP_004574732.1	WP_007924932.1	YP_003148870.1
Ala (A)	15.7	17.1	13.2	11.6	11.8	10.6	10.4	15.2	10.4	13.2	8.8
Arg (R)	2.9	4.3	4.4	2.9	4.4	3.0	4.5	6.1	4.5	1.5	4.4
Asn (N)	0.0	0.0	0.0	0.0	0.0	3.0	0.0	0.0	0.0	0.0	1.5
Asp (D)	4.3	4.3	2.9	4.3	7.4	6.1	4.5	3.0	1.5	4.4	7.4
Cys (C)	2.9	2.9	2.9	2.9	2.9	3.0	3.0	3.0	3.0	2.9	2.9
Gln (Q)	7.1	4.3	4.4	5.8	1.5	6.1	3.0	4.5	4.5	1.5	4.4
Glu (E)	8.6	10.0	11.8	10.1	10.3	12.1	9.0	12.1	11.9	11.8	11.8
Gly (G)	10.0	8.6	8.8	10.1	11.8	9.1	10.4	9.1	10.4	8.8	8.8
His (H)	1.4	1.4	1.5	2.9	1.5	1.5	1.5	1.5	1.5	1.5	2.9
Ile (I)	1.4	5.7	7.4	0.0	1.5	6.1	4.5	0.0	4.5	2.9	1.5
Leu (L)	2.9	1.4	1.5	4.3	2.9	3.0	4.5	6.1	4.5	1.5	1.5
Lys (K)	0.0	0.0	0.0	0.0	0.0	3.0	1.5	0.0	0.0	2.9	0.0
Met (M)	2.9	2.9	1.5	2.9	2.9	3.0	1.5	3.0	3.0	4.4	2.9
Phe (F)	0.0	0.0	1.5	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Pro (P)	2.9	1.4	2.9	1.4	1.5	1.5	1.5	1.5	1.5	1.5	1.5
Ser (S)	7.1	8.6	7.4	8.7	5.9	7.6	10.4	9.1	13.4	8.8	8.8
Thr (T)	11.4	11.4	14.7	11.6	13.2	6.1	10.4	7.6	7.5	13.2	10.3
Trp (W)	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Tyr (Y)	2.9	2.9	1.5	1.4	2.9	3.0	3.0	3.0	4.5	2.9	2.9
Val (V)	15.7	12.9	11.8	17.4	17.6	12.1	16.4	15.2	13.4	16.2	17.6

Table 3 summarized of physicochemical properties of CopZ from some heavy metal tolerant bacteria. Based on data in Table 3, it is shown that the average molecular weight of copper chaperones calculated is 6844.5 – 7174.7 Da. Isoelectric point (pI) is the pH at which the surface of protein is covered with charge but net charge of the protein is zero. At pI proteins are stable and compact. The computed pI value of all proteins indicates that these copper chaperones are acidic which have pI< 7. The computed isolelectric point (pI) will be useful for developing buffer systems for purification by isoelectric focusing method. Expasy's ProtParam computed the extinction coefficient in

280 nm because proteins absorb strongly in this wavelength while other substances in protein solutions do not. The extinction coefficient of Copper chaperones at 280 nm ranges from 1615 to 4595 M^{-1} cm^{-1} to the concentration of Cys, Trp, and Tyr. The highest extinction coefficient belongs to copper chaperone of *Microlunatus phosphovorus* NM-1, since it has the most residue Cys and Tyr more than other copper chaperones. Copper chaperones from *Leucobacter* sp. UCD-THU and *Brachybacterium faecium* DSM 4810 have the lowest EC, because they have the least Cys and Tyr residue among all copper chaperones studied. Interestingly, based on ProtParam's analysis, it has been shown that 8 copper chaperones have the same value of Extinction Coefficient (EC), i.e. 3105 M^{-1} cm^{-1}, while 2 copper chaperones have the same value, i.e. 1615 M^{-1} cm^{-1}. It was because they have Cys and Tyr residue almost the same. CopZ from *Microlunatus phosphovorus* NM-1 had the highest EC value, because it had the highest amount of Cys and Tyr residues. The computed protein concentration and extinction coefficients support the quantitative study of protein–protein and protein-ligand interactions in solution. On the basis of instability index, Expasy's ProtParam classifies *Brevibacterium* sp. JC43, *Leucobacter* sp. UCD-THU, and *Microlunatus phosphovorus* NM-1 as unstable (instability index > 40) compare to other copper chaperone. The aliphatic index (AI) which is defined as the relative volume of a protein occupied by aliphatic side chains (A, V, I and L) is regarded as a positive factor for the increase of thermal stability of globular proteins (Ikai, 1980). The high aliphatic index of all copper chaperones infers that these proteins may be stable for a wide range of temperature. The Grand Average hydropathy (GRAVY) value of protein is the sum of hydropathy values of all the amino acids, divided by the number of residues in the sequence (Kyte and Dolittle, 1982). GRAVY value of CopZ are ranging from -0.256 to 0.157. The very low GRAVY index of all copper chaperones infers that these copper chaperones could result in a better interaction with water.

TABLE 3. Physicochemical properties of CopZ

No	Protein	Length	MW (Da)	EC	pI	II	AI	GRAVY
1	WP_005049374.1	70	7072.7	3105	4.05	30.25	78	0.107
2	WP_019157648.1	70	7133.8	3105	4.18	40.83	82.29	0.154
3	WP_017883414.1	68	7013.7	1615	4.22	51.8	81.76	0.084
4	YP_003155379.1	69	7070.7	1615	4.18	15.26	78.99	0.106
5	WP_005189537.1	68	6935.6	3105	4.01	16.81	80.15	0.097
6	WP_006213359.1	66	7001.7	3105	4.19	23.47	81.21	-0.239
7	YP_004079619.1	67	6872.6	3105	4.44	17.57	92.99	0.209
8	WP_020018251.1	66	6844.5	3105	4.39	18.52	82.73	0.036
9	YP_004574732.1	67	6968.7	4595	4.33	41.97	84.33	0.096
10	WP_007924932.1	68	6941.7	3105	4.12	26.33	77.35	0.157
11	YP_003148870.1	68	7174.7	3105	4.11	28.97	71.47	-0.256

Notes: MW, molecular weight; EC, extinction coefficient; pI, isoelectric point; II, instability index; AI, aliphatic index; GRAVY, Grand Average hydropathy.

Table 4 shows the summary of secondary structure predicted by SOPMA. The table infers that CopZ from *M. laevaniformans*, *Brevibacterium* sp. JC43, and *Janibacter hoylei* are mostly α-helices because they have rich alanine content (as shown in Table 2). The result also shows that all CopZ have no 3_{10}-helix; π-helix, β-bridge, ambigous states (As); Other states (Os). However, in average all CopZ have balance combination of α-helix, β-strand, and coil.

Table 4. Secondary structure composition of CopZ by SOPMA

No.	Protein	Hh	Gg	Ii	Bb	Ee	Tt	Cc	As	Os
1	WP_005049374.1	35.71	0	0	0	21.43	11.43	31.43	0	0
2	WP_019157648.1	31.43	0	0	0	28.57	11.43	28.57	0	0
3	WP_017883414.1	30.88	0	0	0	25.00	11.76	32.35	0	0
4	YP_003155379.1	30.43	0	0	0	28.99	11.59	28.99	0	0
5	WP_005189537.1	32.35	0	0	0	23.53	10.29	33.82	0	0
6	WP_006213359.1	30.30	0	0	0	25.76	10.61	33.33	0	0
7	YP_004079619.1	29.85	0	0	0	26.87	11.94	31.34	0	0
8	WP_020018251.1	31.82	0	0	0	25.76	12.12	30.30	0	0
9	YP_004574732.1	31.34	0	0	0	26.87	11.94	29.85	0	0
10	WP_007924932.1	35.29	0	0	0	20.59	14.71	29.41	0	0
11	YP_003148870.1	29.41	0	0	0	25.00	11.76	33.82	0	0

Notes: Alpha helix (Hh); 3_{10} helix (Gg); π-helix (Ii); β-bridge (Bb); Extended strand (Ee); β-turn (Tt); Random coil (Cc); Ambigous states (As); Other states (Os).

Copper chaperone from *M. laevaniformans* was selected to be modeled since this bacterium was reported to be resistant to metals, nitrate, and low pH conditions. This protein was modeled using Phyre2 server, Swiss-Model, and Geno3D. The three modeling servers were analyzed and compared using PROCHECK, QMEAN, and ModFOLD. The result of structure prediction is illustrated in Figure 2. The –CGHC– motif was illustrated in a ball and stick model. The tool DiANNA recognized the presence of 2 Cysteines in the CopZ *M. laevaniformans* sequence and predicted one probable SS bond pattern of pairs (as discussed in the primary structure analysis) in the protein, which is located in position 14-17 (as shown by CGHC motif in Figure 1 and 2).

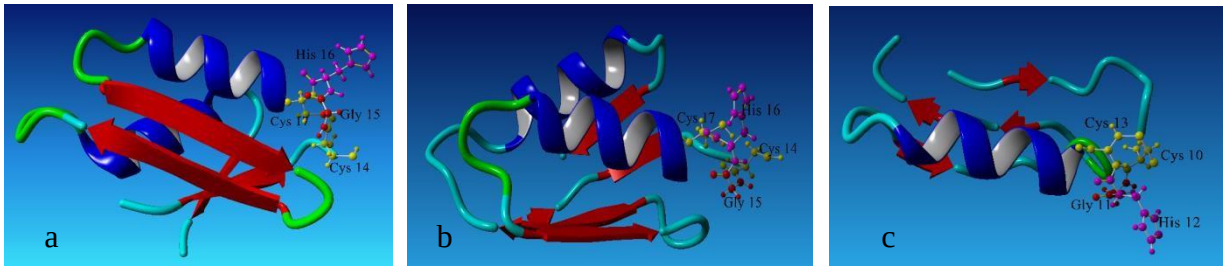

FIGURE 2. Three-dimensional structure prediction of copper chaperone *M. laevaniformans* by using: (a) Phyre[2]; (b) Swiss-Model; (c) Geno3D. Blue ribbon indicated α-helix structure, red ribbon indicated β-strand structure, green and light-blue strand indicated turn and coil structure, respectively. CGHC motif is shown in ball and stick model; Cys residue in yellow, Gly residue in red, and His residue in magenta.

Table 5 summarized the result of comparison analysis of three modeling server for CopZ *M. laevaniformans*. Based on evaluation analysis, the Ramachandran plot of Phyre [2] and Swiss-Model were higher than Geno3D result. A good quality model would be expected to have over 90% in the most favored (RFR) regions (A, B, L). The percentage of residues in the favored regions is one of the better guides to predict the quality of a modeled structure stereochemically [11]. Although Phyre[2] and Swiss-Model have RFR scores of 85.5% and 88.9%, these were still within the standard acceptable limits for the 3D structures modeled. In the contrary, the Geno3D model only had 63.9%. In addition, based on QMEAN analysis, it was shown that modeled structures from Phyre2 and Swiss-Model have QMEAN and Z-score score higher than Geno3D. This QMEAN analysis suggested that the more positive the

score revealed the more qualified the structure [12]. Moreover, ModFOLD analysis also suggested that the modeled structure conducted by Phyre2 and Swiss-Model servers was better than Geno3D. According to 'confidence and P-value' they have very low scores that indicated them as 'certified' structures. A certified structure has $p < 0.001$ and shows less than a 1/1000 chance that the model is incorrect [13]. It was also shown from ModFOLD analysis that the global model quality score of the modeled structure conducted by Phyre2 and Swiss-Model servers was higher than Geno3D. However, all servers have scores greater than 0.4 which generally indicates all models are more complete, confident models, and highly similar to the native structure.

Table 5. Comparison of predicted structure validation

Modeling server	PROCHECK (Ramachandran plot analysis)			QMEAN		ModFOLD	
	RFR (%)	RAR (%)	ROR (%)	QMEAN score	Z-score	Confidence and P-value	Global model quality score
Phyre[2]	85.5	14.5	0.0	0.771	0.35	1.412×10^{-4}	0.8429
Geno3D	63.9	36.1	0.0	0.457	-0.91	4.774×10^{-3}	0.5042
Swiss-Model	88.9	9.3	1.9	0.797	0.55	1.703×10^{-4}	0.8249

Notes: RFR: residues in favoured region; RAR: residues in allowed region; ROR: residues in outlier region

CONCLUSION

All CopZs were hydrophobic and contain –CGHC– motif. The physicochemical properties of all CopZs would provide adequate essential data about the proteins and their properties. Secondary structure analysis also predicted that all CopZs have mixed structures of α-helix, β-strand, turn, and random-coiled. Moreover, based on evaluation and validation analysis, it can be concluded that Phyre[2] and the Swiss model were good quality servers to perform structure homology and modeling.

REFERENCES

1. Halliwell, B. and Gutteridge, J.M. 1985. The importance of free radicals and catalytic metal ions in human disease. Mol. Asp. Med. 8, 89–193.
2. Rae, T.D., Schmidt, P.J., Pufahl, R.A., Culotta, V.C., and O'Halloran, T.V. "Undetectable Free Intracellular Copper: the Requirement of a Copper Chaperone for Superoxide Dismutase" *Science* **1999**, *284*, 805-808.
3. Rosenzweig. A.C. 2001. Copper Delivery by Metallochaperone Proteins, Accounts of Chemical Research, 34 (2): 119–128.
4. Hamza, I., Schaefer, M., Klomp, L.W. and Gitlin, J.D. 1999. Interaction of the copper chaperone HAH1 with the Wilson disease protein is essential for copper homeostasis, Proceedings of the National Academy of Sciences of the USA, 96(23): 13363-13368.
5. Wernimont, A.K., Yatsunyk, L.A. and Rosenzweig, A.C. 2004. Protein and Its Copper Chaperone Binding of Copper(I) by the Wilson Disease, Journal of Biological Chemistry, 279:12269 -12276.
6. deBie, P., Muller, P., Wijmenga, C., and Klomp, L.W.J. 2007. Molecular pathogenesis of Wilson and Menkes disease: correlation of mutations with molecular defects and disease phenotypes. Journal of Medical Genetics, 44(11): 673–688.
7. Gasteiger E., Hoogland C., Gattiker A., Duvaud S., Wilkins M.R., Appel R.D., and Bairoch A. (2005). *Protein Identification and Analysis Tools on the ExPASy Server;*(In) John M. Walker (ed): The Proteomics Protocols Handbook, Humana Press. pp. 571-607.
8. Ferre, F. and Clote, P. 2005. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res.*1 (33): W230-232.
9. Kelley, L.A. and Sternberg, M.J.E. 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*. 4(3): 363-371.
10. Harrison, M.D., Jones, C.E., Solioz, M. and Dameron, C.T. 2000. Intracellular copper routing: the role of copper chaperones, TIBS 25:29-32.
11. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. Proteins, **12**, 345-364.

12. Benkert, P., Tosatto, S.C.E. and Schomburg, D. 2008. "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics, 71(1): 261 -277.
13. Morris, A.L., MacArthur, M.W., Hutchinson, E.G., and Thornton, J.M. 1992. Stereochemical quality of protein structure coordinates. Proteins, **12**, 345-364.