**Vol.2, No.2, 111-122 (2011)**
doi:10.4236/jbpc.2011.22014

**Journal of Biophysical Chemistry**

# MetalloPred: a tool for hierarchical prediction of metal ion binding proteins using cluster of neural networks and sequence derived features

**Pradeep Kumar Naik*, Piyush Ranjan, Pooja Kesari, Sankalp Jain**

Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat, India;
*Corresponding Author: pknaik73@rediffmail.com; pradeep.naik@juit.ac.in

Received 19 January 2011; revised 13 February 2011; accepted 21 February 2011.

## ABSTRACT

**Given a protein sequence, how can we identify whether it is a metalloprotein or not? If it is, which main functional class and subclasses it belongs to? This is an important biological question because they are closely related to the biological function of an uncharacterized protein. Particularly, with the avalanche of protein sequences generated in the post genomic era and since conventional techniques are time consuming and expensive, it is highly desirable to develop an automated method by which one can get a fast and accurate answer to these questions. Here, a top-down predictor, called MetalloPred, is developed which consists of 3 level of hierarchical classification using cascade of neural networks from sequence derived features. The 1st layer of the prediction engine is for identifying a query protein as metalloprotein or not; the 2nd layer for the main functional class; and the 3rd layer for the sub-functional class. The overall success rates for all the three layers are higher than 60% that were obtained through rigorous cross-validation tests on the very stringent benchmark datasets in which none of the proteins has 30% sequence identity with any other in the same class or subclass. MetalloPred achieved good prediction accuracies and could nicely complement experimental approaches for identification of metal binding proteins. MetalloPred is freely available to be used in-house as a standalone and is accessible at http://www.juit.ac.in/assets/Metallopred/.**

**Keywords:** Metalloprotein; Classification; Sequence Derived Parameters; Neural Networks

## 1. INTRODUCTION

Metalloprotein is a generic term for a protein that contains a metal ion cofactor. Metalloproteins have captivated chemists and biochemists, particularly since 1950s, when the first X-ray crystal structure of a protein, sperm whale myoglobin indicated the presence of an iron atom [1]. The metal ion is usually coordinated by nitrogen, oxygen or sulfur atoms belonging to amino acids in the polypeptide chain and/or a macro-cyclic ligand incorporated into the protein [2,3]. The presence of the metal ion allows metalloenzymes to perform functions such as redox reactions that cannot be performed by the limited set of functional groups found in amino acids [1]. Metalloproteins play important roles in structural stability and complex formation [4-8], gene expression regulation and alteration [9-12], DNA processing [13], signaling processes and cellular event [14], transport [11,15,16], metabolism control [15,17], antibody recognition [18] and other biological processes such as cellular respiration, photosynthesis, nitrogen fixation and antioxidant defense [19]. Approximately, 1/3 of structurally-determined proteins are metalloproteins [20]. Much effort has been devoted to understanding the structure and function of these proteins.

Traditionally the metalloproteins have been identified, based on experimental techniques such as absorbance spectroscopy [21], gel electrophoresis [22], metal-affinity columns and shift assay [23], chromatography [24], mass spectroscopy [22], NMR [9] and combined spectroscopic studies [25]. These techniques which require purified or semi-purified proteins of interest, do not facilitate identification of unknown proteins from a complex mixture, or require multi-step processes and very specialized equipment which limit their application ranges. Therefore, there is need to explore alternative methods for facilitating the identification of metalloproteins to complement these experimental methods. With the exponential growth of sequence data, an insurmountable task of characteriz-

ing these sequences with experimental methods is very cumbersome. It is thus desirable to explore automated computational methods for the annotation of novel protein sequences. Several sequenced-based computational methods have been explored based on similarity search, metal-binding sites sequence motifs [26,27] and multiple sequence alignments against known metalloproteins [28]. Because of the sequence, structural and functional diversity of metalloproteins [4-8,14-17], it is desirable to explore additional methods that predict metalloproteins directly from sequence or sequence-derived properties. For a newly-found protein sequence the most interesting thing people wish to know is about its biological function and hence the following questions are often asked: Is the query protein a metalloprotein or non-metalloprotein? If it is, which main functional class does it belong to? Or going further deeper, what about its sub-functional class? The present study was initiated in an attempt to develop a top–down approach to solve all these problems and make it accessible to the vast majority of experimental scientists by providing a user-friendly web-server

In this study, we have developed cascade of artificial neural network (NN) prediction systems for metalloproteins. The generalized classification obtained by the method suggests that MetalloPred could be useful as a starting point in initial screening and *ab initio* prediction of metalloproteins, and, in combination with comparative studies on completed genomic sequences, it could give further insight into the evolution of protein structure and function.

## 2. MATERIALS AND METHODS

### 2.1. Preparation of Dataset

All metalloproteins used in this study are collected from a comprehensive search of protein data bank (www.rcsb.org). A total of 14625 metalloprotein sequences were obtained and have been classified into calcium-binding (3466), magnesium-binding (2886), potassium-binding (173), sodium-binding (157), cobalt-binding (200), copper-binding (887), manganese-binding (968), molybdenum-binding (134), nickel-binding (147), vanadium-binding (11), zinc-binding (4861) and iron-binding (328). This data set was further refined by discarding protein sequences having length less than 20 amino acids, as they are very unlikely form a proper pocket to coordinate with metal ion. Some proteins were found to bind with more than one metal ion and have been discarded. With the aim of avoiding prejudiced learning in the networks, we scaled the sequences such that the inequality in the number of protein sequences in each class may be compromised. We reduced the proteins in each class with a similarity cutoff of 30% using BLASTClust [29]. A negative dataset consisting of 5738 protein sequences, representing non-class members is also selected from PDB database. These datasets are divided into training, testing and independent evaluation sets (**Table 1**).

### 2.2. Feature Extraction

Following three types of discrete feature vectors were constructed for each protein sequence.

1) *Amino acid composition*: given the sequence of a protein, its amino acid composition was computed and then used to generate a set of 20 features representing composition of 20 standard amino acids in the protein sequences that include A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, and Y. These features have been widely used in predicting different structural classes

**Table 1.** Number of proteins used for training and validation of MetalloPred.

| Class and subclass | Number of proteins | Training set | Test set | Independent set |
|---|---|---|---|---|
| Non-metal binding | 5738 | 828 | 207 | 3322 |
| Metal binding | 5371 | 680 | 170 | 3017 |
| Alkali earth metal binding (S1) | 2193 | 319 | 80 | 955 |
| Calcium | 1282 | 572 | 143 | 567 |
| Magnesium | 911 | 418 | 105 | 388 |
| Alkali metal binding (S2) | 83 | 56 | 14 | 13 |
| Potassium | 70 | 48 | 12 | 10 |
| Sodium | 13 | 8 | 2 | 3 |
| Transition metal binding (S3) | 3095 | 305 | 76 | 2049 |
| Cobalt | 55 | 40 | 10 | 5 |
| Copper | 306 | 221 | 55 | 30 |
| Iron | 158 | 66 | 16 | 76 |
| Manganese | 278 | 103 | 26 | 149 |
| Molybdenum | 49 | 35 | 9 | 5 |
| Nickel | 54 | 39 | 10 | 5 |
| Vanadium | 8 | 4 | 2 | 2 |
| Zinc | 2187 | 326 | 82 | 1779 |

 

[30-32] and subcellular localization [33-36] of proteins. The formula used to calculate amino acid composition is:

$$AAcomp(i) = \frac{AA(i)}{\sum_{j=1}^{20} AA(j)}$$

where AA($i$) = Frequency of $i^{th}$ amino acid.

2) *Physicochemical properties*: twelve sequence derived properties for each protein sequence was calculated using EMBOSS (EBI) package [37]. The parameters include molecular weight, total charge, isoelectric point, mole percentages of tiny (A, C, G, S, T); small (A, C, D, G, N, P, S, T, V); aliphatic (I, L, V); aromatic (F,H,W,Y); non-polar (A, C, F, G, I, L, M, P, V, W, Y); polar (D, E, H, K, N, Q, R, S, T), charged (D, E, H, K, R); acidic (D, E) and basic (H, K, R) amino acids.

3) *Pseudo amino acid composition* (*PseAA*)

This class of descriptor consists of a set of 37 features, 20 of which are weighted amino acid compositions and rest 17 are correlation factors calculated among amino acids for each protein sequence [38].

A protein sequence $P$ with $L$ amino acid resides can be represented as:

$$P = R_1 R_2 R_3 R_4 \cdots R_L \quad (1)$$

where $R_1$ represents the 1st residue of the protein $P$, $R_2$ the 2nd residue, and so forth. According to the simplest discrete model, the amino acid composition of the protein P based on the equation (1) can be expressed as:

$$P = \begin{bmatrix} f_1 & f_2 & \cdots & f_{20} \end{bmatrix}^T \quad (2)$$

where $f_u (u = 1, 2, \cdots, 20)$ are the normalized occurrence of frequencies for the 20 native amino acids in P, and T the transposing operator. The additional 17 features are a series of rank-different correlation factors along a protein chain and were calculated as follows.

A protein sequence P consisting of $L$ amino acid resides can be represented as:

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & p_{20} & p_{20+1} & \cdots & p_{20+\lambda} \end{bmatrix}^T, (\lambda < L) \quad (3)$$

where $20 + \lambda$ components are given by

$$p_u = \begin{cases} \dfrac{f_u}{\sum_{i=1}^{20} f_i + w\sum_{k=1}^{\lambda} \tau_k}, (1 \le u \le 20) \\ \dfrac{w\tau_{u-20}}{\sum_{i=1}^{20} f_i + w\sum_{k=1}^{\lambda} \tau_k}, (20+1 \le u \le 20+\lambda) \end{cases} \quad (4)$$

where $w$ is the weight factor and $\tau_k$ is the $k$-th tier correlation factor that reflects the sequence order correlation between all the $k$-th most contiguous residues as formulated by

$$\tau_k = \frac{1}{L-k} \sum_{i=1}^{L-k} J_{i,i+k}, (k < L) \quad (5)$$

with

$$J_{i,i+k} = \frac{1}{\Gamma} \sum_{g=1}^{\Gamma} \left[ \Phi_\xi (R_{i+k}) - \Phi_\xi (R_i) \right]^2 \quad (6)$$

where $\Phi_\xi(R_i)$ is the $\xi$-th function of the amino acid $R_i$, and $\Gamma$ the total number of the functions considered. $\Phi_1(R_i)$, $\Phi_2(R_i)$ and $\Phi_3(R_i)$ represented respectively the hydrophobicity value [39], hydrophilicity value [40], and side chain mass of amino acid $R_i$ (**Table 2**); while $\Phi_1(R_{i+k})$, $\Phi_2(R_{i+k})$ and $\Phi_3(R_{i+k})$ are the corresponding values for the amino acid $R_{i+k}$. Therefore, the total number of functions considered is $\Gamma = 3$.

It can be seen from equation (3) that the first 20 components, *i.e.*, $p_1$, $p_2$, $\cdots$, $p_{20}$ are associated with the conventional AA composition of protein, while the remaining components $p_{20+1}$, $\cdots$, $p_{20+\lambda}$ are the correlation factors that reflect the 1st tier, 2nd tier, $\cdots$, and the $\lambda^{th}$ tier sequence order correlation patterns. It is through these additional λ factors that the important sequence-order information is incorporated.

## 2.3. System Architecture and Component of NN Topology

The overall classification system consists of three layers of successive multilayer feed forward (acyclic) artificial NNs (**Figure 1**), each one with a single hidden layer at which the computation takes place. Some com-

**Table 2.** Hydrophobicity, hydrophilicity and mass of side chain scales for 20 amino acids used in calculating pseudo amino acid composition (Pse*AA*).

| Amino acid | Hydrophobicity[a] | Hydrophilicity[b] | Mass of side chain |
|---|---|---|---|
| A | 0.62 | –0.5 | 15 |
| C | 0.29 | –1 | 47 |
| D | –0.9 | 3 | 59 |
| E | –0.74 | 3 | 73 |
| F | 1.19 | –2.5 | 91 |
| G | 0.48 | 0 | 1 |
| H | –0.4 | –0.5 | 82 |
| I | 1.38 | –1.8 | 57 |
| K | –1.5 | 3 | 73 |
| L | 1.06 | –1.8 | 57 |
| M | 0.64 | –1.3 | 75 |
| N | –0.78 | 0.2 | 58 |
| P | 0.12 | 0 | 42 |
| Q | –0.85 | 0.2 | 72 |
| R | –2.53 | 3 | 101 |
| S | –0.18 | 0.3 | 31 |
| T | –0.05 | –0.4 | 45 |
| V | 1.08 | –1.5 | 43 |
| W | 0.81 | –3.4 | 130 |
| Y | 0.26 | –2.3 | 107 |

[a]Hydrophobicity values are from reference [39], [b]Hydrophilicity values are from reference [40].

 

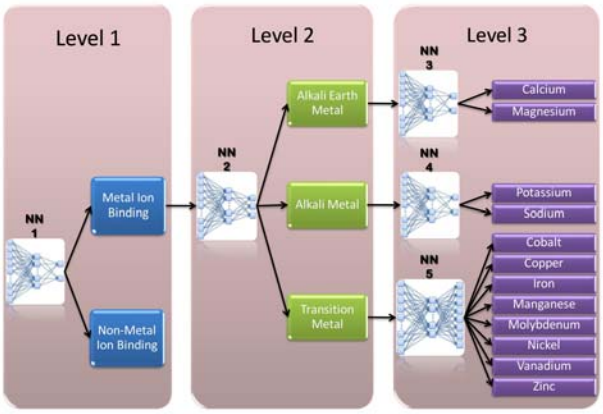


**Figure 1.** A schematic drawing to classify metalloproteins and non-metalloproteins as well as the three main functional classes of metalloproteins and their subclasses.

mon features shared by all NNs are the following:

1) There is full connectivity as every node in each network layer is connected to every other node in the adjacent forward layer.

2) There are a small number of nodes in the hidden layer responsible for the actual learning process carried out by each component network.

3) The activation function on each node is a nonlinear, sigmoid logistic function of the weighted sum of all synaptic weights (plus a constant bias).

NN1 is binary classifiers which classify an input protein sequence as a metalloprotein or non-metalloprotein. If the input protein sequence is classified as a metalloprotein then it is processed by NN2 which gets classified into one of the three main classes of metal binding proteins (alkali earth metal, alkali metal or transition metal). Each class consists of an independent NN [(alkali earth metal (NN3), alkali metal (NN4) and transition metal (NN5)] for classification of input protein sequence specifically into metal ion it binds. We have used three categories of sequence derived features such as physicochemical properties, amino acid composition and pseudo amino acid composition for training of NNs. Using these parameters independently and with combination we have developed seven neural network clusters: $NN_{prop}$, $NN_{AAcomp}$, $NN_{pseAA}$, $NN_{pseAA+prop}$, $NN_{AAcomp+prop}$, $NN_{pseAA+AAcomp}$ and $NN_{pseAA+AAcomp+prop}$. Before the learning process, all network synaptic weights are initialized to small random values which have been optimized to final weights during the learning process based on backpropagation algorithm [41].

An important issue in the design of a NN classification system is the network's generalization, that is, its ability to give correct predictions when it is presented with unseen examples. With a small number of training samples and a relatively large number of synaptic weights, there is always the possibility that the network's free parameters will adapt to the special features of the training data (over-fitting). A straightforward way to overcome this problem is to use a sufficient number of training examples (usually more than 30 times the number of adjustable network parameters). However, the protein classes are unbiased and it is not possible to have these many numbers. Therefore to control the over fitting in our application, we have employed non-convergent criteria (early stopping method); the training process is stopped before the finishing of optimization procedure. We follow the common method which is to withhold and use part of the training data (20%) as an internal validation set. Training is stopped at the point at which the classification error on the holdout subset begins to rise.

In the prediction phase, just like the forward pass in learning, network weights are globally fixed (those obtained after the convergence of the training process) and the NN is presented with an unknown example for classification. In the same hierarchical manner, the input signal propagates once in the forward direction and the output value constitutes the network's decision based on the already studied training examples. The prediction accuracy of the models has been validated using self test, jackknife test and independent data set. For jackknife test we randomized the test set for 100 times and recorded average performance accuracy.

## 3. RESULTS AND DISCUSSION

To assess the performance of the MetalloPred, we applied several tests. We created a new independent test set with well-characterized protein sequences from all level of classes and sub-classes (**Table 1**) to evaluate the performance of the new integrated system. In addition we have also performed sub-sampling test (self test) and jackknife test for examining the accuracy of MetalloPred. All these validation tests are commonly used for performance evaluation of a predictor. Jackknife test is deemed the most rigorous and objective [30] and hence has been increasingly adopted by investigators in examining the quality of various prediction methods [42-44]. A direct comparison with results from previous metal binding protein prediction studies may not be most appropriate because of the differences in the protein classes predicted, datasets, protein descriptors, prediction methods and parameters.

### 3.1. Performance of 1st Layer of Neural Network

The performance and validation results of NN1 are given in **Table 3**. An overall accuracy of 99.74% and

 

**Table 3.** Performance accuracy and validation results of 1st layer of MetalloPred based on combination of pseudo amino acid composition, amino acid composition and physicochemical properties.

| Class | Train | Test | Validation of NN1 (%) | | |
|---|---|---|---|---|---|
| | set (%) | set (%) | Independent set | Self test | Jackknife test |
| Metal binding | 99.85 | 89.41 | 81.38 | 78.71 | 73.92 |
| Non-metal binding | 99.64 | 86.57 | 85.47 | 82.27 | 74.44 |
| Average | 99.74 | 87.99 | 83.42 | 80.49 | 74.18 |

87.99% for the training and test set data using combination of sequence derived features such as pseudoamino acid composition, amino acid composition and physicochemical properties. While considering the validation techniques by using an independent data set, self test and jackknife test, the overall accuracy of the 1st layer of MetalloPred is 83.42%, 80.49% and 74.18% respectively. The details of the performance accuracy and validation results based on different types of sequence derived feature have been represented in supplementary **Table 1**.

### 3.2. Performance of 2nd Layer of Neural Network

The overall success rate in identifying the metalloproteins among their three main functional classes is 99.25% (using training set) and 81.91% (using test set) (**Table 4**). Similarly the overall performance accuracy based on three types of validation tests has been found to be 75.16% (using independent data set), 73.24% (using self test) and 64.23% (using jackknife test). The corresponding results by Metallopred on the data set for three major classes of metalloproteins using different types of sequence derived features are given in supplementary **Table 2**.

### 3.3. Performance of 3rd Layer of Neural Network

The performance accuracy and validation results of NNs in identifying subclasses of alkali earth metal (NN3), alkali metal (NN4) and transition metal (NN5) binding proteins using the combination of all sequence derived features has been given in **Table 5**. The corresponding results by MetalloPred on the detection of calcium and magnesium metal binding proteins are 91.72% (training set), 91.07% (test set), 77.46% (independent data set), 81.1% (self test) and 75.51% (jackknife test) on the data set 'S1'. Similarly for the data set 'S2' the performance accuracy for the detection of potassium and sodium binding proteins are 97.95% (training set), 96.4% (test set), 79.2% (independent data set), 91.57% (self test) and 83.33% (jackknife test). The overall accuracy of detection of cobalt-binding, copper-binding, iron-binding, manganese-binding, molybdenum-binding, nickel-binding, vanadium-binding and zinc-binding is 98.88% (training set), 95.39% (test set), 84.06% (independent data set), 71.39% (self test) and 67.98% (jackknife test) using the data set 'S3'. The details of the performance accuracy have been represented in supplementary **Table 3**.

For the current data sets in which none of the protein sequence has ≥30% sequence identity to any other in a same class or subclass, the overall success rates by the MetalloPred in identifying the main functional classes of metalloproteins and their subclasses is very high. In an earlier study, contribution of individual feature property to protein classification is investigated by separately conducting classification by the use of each feature property [45-47].

The same method was employed here. An analysis on the classification of the group of all metal binding proteins seems to suggest that, in order of prominence, the hydrophobicity and hydrophilicity play a more prominent role than other feature properties. Hydrophobicity has been shown to be important for metal-protein interactions such that metal binding sites usually appear in clusters with hydrophobic environment. High-affinity metal binding sites in some proteins are located at sequence segments with specific amino acid composition, and specific sequence motifs have been used for predicting metal-binding proteins [48-50]. It was also found that polarity and solvent accessibility of the binding site influences the functional properties of metal-binding proteins. Therefore, our prediction results are consistent with these experimental findings. Overall MetalloPred is a very powerful predictor in identifying metalloproteins, their main classes, and their subclasses.

## 4. CONCLUSIONS

From a practical point of view, the most important aspect of a prediction model is its ability to make correct predictions. Till date most of the available methods use the 3-D structure of the protein to predict and classify metal ion binding proteins. This is a very tedious job and requires much costlier endeavors. The sequence of a protein is an important determinant for the detailed molecular function of proteins and would consequently also be useful for prediction of metal ion binding protein and

 

**Table 4.** Performance accuracy and validation results of 2nd layer of MetalloPred based on combination of pseudo amino acid composition, amino acid composition and physicochemical properties.

| Class | Train | Test | Validation of NN2 (%) | | |
|---|---|---|---|---|---|
| | set (%) | set (%) | Independent set | Self test | Jackknife test |
| Alkali earth metal | 99.06 | 82.50 | 73.66 | 74.86 | 69.76 |
| Alkali metal | 100.00 | 84.29 | 76.15 | 72.86 | 62.86 |
| Transition metal | 98.69 | 78.95 | 75.68 | 72.01 | 60.06 |
| Average | 99.25 | 81.91 | 75.16 | 73.24 | 64.23 |

**Table 5.** Performance accuracy and validation results of 3rd layer of MetalloPred based on combination of pseudo amino acid composition, amino acid composition and physicochemical properties.

| Class and | Train | Test | Validation of NNs (%) | | |
|---|---|---|---|---|---|
| subclass | Set (%) | Set (%) | Independent set | Self test | Jackknife test |
| (a) Alkali earth metal binding (NN3) | | | | | |
| Calcium | 93.01 | 88.81 | 73.32 | 78.53 | 77.48 |
| Magnesium | 90.43 | 93.33 | 81.60 | 83.67 | 73.54 |
| Average | 91.72 | 91.07 | 77.46 | 81.10 | 75.51 |
| (b) Alkali metal binding (NN4) | | | | | |
| Potassium | 98.40 | 97.80 | 80.00 | 93.15 | 86.67 |
| Sodium | 97.50 | 95.00 | 78.40 | 90.00 | 80.00 |
| Average | 97.95 | 96.40 | 79.20 | 91.57 | 83.33 |
| (c) Transition metal binding (NN5) | | | | | |
| Cobalt | 100.0 | 100.0 | 80.00 | 66.00 | 62.00 |
| Copper | 93.21 | 81.82 | 80.00 | 70.36 | 68.77 |
| Iron | 100.0 | 100.0 | 84.74 | 70.98 | 60.00 |
| Manganese | 100.0 | 96.15 | 84.90 | 62.64 | 61.78 |
| Molybdenum | 100.0 | 100.0 | 80.00 | 68.64 | 69.55 |
| Nickel | 100.0 | 90.00 | 80.00 | 68.78 | 62.45 |
| Vanadium | 100.0 | 100.0 | 96.20 | 92.50 | 90.00 |
| Zinc | 97.85 | 95.12 | 86.66 | 71.27 | 69.26 |
| Average | 98.88 | 95.39 | 84.06 | 71.39 | 67.98 |

their classes. Additionally much encouraging results have been predicted using the sequence derived parameters technique. Therefore, a much accurate and reliable method is to predict the metal ion binding proteins and metal ion binding protein classes based on both strategies. Cascade of neural networks used in this study appears to be a potentially useful tool for the prediction of metal-binding proteins of different classes. The prediction accuracy may be further enhanced with the further expansion of our knowledge about metal-binding proteins, particularly for those small metal-binding classes, more refined representation of the structural and physicochemical properties of proteins and the improvement of prediction algorithms such as the better treatment of imbalanced dataset in the next version of our prediction tool.

## REFERENCES

[1] Finkelstein, J. (2009) Metalloproteins. *Nature*, **460**, 813-813. doi:10.1038/460813a

[2] Rosette, M. and Malone, R. (2002) Bioinorganic chemistry a short course. John Wiley and Sons, New York.

[3] Gregolinski, J., Starynowicz, P., Hua, K.T., Lunkley, J.L., Muller, G. and Lisowski, J. (2008) Helical lanthanide (III) complexes with chiral nonaaza macrocycle. *Journal of American Chemical Society*, **130**, 17761-17773. doi:10.1021/ja805033j

[4] Wintz, H., Fox, T., Wu, Y.Y., Feng, V., Chen, W., Chang, H.S., Zhu, T. and Vulpe, C. (2003) Expression profiles of Arabidopsis thaliana in mineral deficiencies reveal novel transporters involved in metal homeostasis. *Journal of Biological Chemistry*, **278**, 47644-47653. doi:10.1074/jbc.M309338200

[5] Cox, E.H. and McLendon, G.L. (2000) Zinc-dependent protein folding. *Current Opinion in Chemical Biology*, **4**, 162-165. doi:10.1016/S1367-5931(99)00070-8

[6] Michel, S.L. and Berg, J.M. (2002) Building a metal binding domain, one half at a time. *Chemical Biology*, **9**, 667-668. doi:10.1016/S1074-5521(02)00160-6

[7] Guntinas, M.B., Bordin, G. and Rodriguez, A.R. (2002)

 

Identification, characterization and determination of metal-binding proteins by liquid chromatography. *Analytical and Bioanalytical Chemistry*, **374**, 369-378. doi:10.1007/s00216-002-1508-3

[8] Yang, W., Lee, H.W., Hellinga, H. and Yang, J.J. (2002) Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins*, **47**, 344-356. doi:10.1002/prot.10093

[9] Jensen, M.R., Petersen, G., Lauritzen, C., Pedersen J. and Led, J.J. (2005) Metal binding sites in proteins: Identification and characterization by paramagnetic NMR relaxation. *Biochemistry*, **44**, 11014-11023. doi:10.1021/bi0508136

[10] Wu, H., Yang, Y., Jiang, S.J., Chen, L.L., Gao, H.X., Fu, Q.S., Li, F., Ma, B.G. and Zhang, H.Y. (2005) DCCP and DICP: Construction and analyses of databases for copper- and iron-chelating proteins. *Genomics*, *Proteomics and Bioinformatics*, **3**, 52-57.

[11] Hantke, K. (2001) Iron and metal regulation in bacteria. *Current Opinion in Microbiology*, **4**, 172-177. doi:10.1016/S1369-5274(00)00184-3

[12] Bouton, C.M. and Pevsner, J. (2000) Effects of lead on gene expression. *Neurotoxicology*, **21**, 1045-1055.

[13] Feng, M., Patel, D., Dervan, J.J., Ceska, T., Suck, D., Haq, I. and Sayers, J.R. (2004) Roles of divalent metal ions in flap endonuclease-substrate interactions. *Nature Structural and Molecular Biology*, **11**, 450-456. doi:10.1038/nsmb754

[14] Carafoli, E. (2002) Calcium signaling: A tale for all seasons. *Proceeding National Academic of Science USA*, **99**, 1115-1122. doi:10.1073/pnas.032427999

[15] Harris, E.D. (2000) Cellular copper transport and metabolism. *Annual Review of Nutrition*. **20**, 291-310. doi:10.1146/annurev.nutr.20.1.291

[16] O'Halloran, T.V. and Culotta, V.C. (2000) Metallochaperones: An intracellular shuttle service for metal ions. *Journal of Biological Chemistry*, **275**, 25057-25060. doi:10.1074/jbc.R000006200

[17] Vallee, B.L. and Auld, D.S. (1990) Active-site zinc ligands and activated $H_2O$ of zinc enzymes. *Proceeding National Academic of Science USA*, **87**, 220-224. doi:10.1073/pnas.87.1.220

[18] Zhou, T., Hamer, D.H., Hendrickson, W.A., Sattentau, Q.J. and Kwong, P.D. (2005) Interfacial metal and antibody recognition. *Proceeding National Academic of Science U.S.A.*, **102**, 14575-14580. doi:10.1073/pnas.0507267102

[19] Lieu, P.T., Heiskala, M., Peterson, P.A. and Yang, Y. (2001) The roles of iron in health and disease. *Molecular Aspects of Medicine*, **22**, 1-87. doi:10.1016/S0098-2997(00)00006-6

[20] Barondeau, D.P. and Getzoff, E.D. (2004) Structural insights into protein metal ion partnerships. *Current Opinion of Structural Biology*, **14**, 765-774. doi:10.1016/j.sbi.2004.10.012

[21] Reed, G.H. and Poyner, R.R. (2000) Mn2+ as a probe of divalent metal ion binding and function in enzymes and other proteins. *Metal Ions Biological Systems*, **37**, 183-207.

[22] Binet, M.R., Ma, R., McLeod, C.W. and Poole, R.K. (2003) Detection and characterization of zinc- and cadmium-binding proteins in Escherichia coli by gel electrophoresis and laser ablation inductively coupled plasma-mass spectrometry. *Analytical Biochemistry*, **318**, 30-38. doi:10.1016/S0003-2697(03)00190-8

[23] Herald, V.L., Heazlewood, J.L., Day, D.A. and Millar, A.H. (2003) Proteomic identification of divalent metal cation binding proteins in plant mitochondria. *FEBS Letter*, **537**, 96-100. doi:10.1016/S0014-5793(03)00101-7

[24] Papoyan, A. and Kochian, L.V. (2004) Identification of Thlaspi caerulescens genes that may be involved in heavy metal hyperaccumulation and tolerance: Characterization of a novel heavy metal transporting ATPase. *Plant Physiology*, **136**, 3814-3823. doi:10.1104/pp.104.044503

[25] Schnepf, R., Haehnel, W., Weighardt, K. and Hildebrandt, P. (2004) Spectroscopic identification of different types of copper centers generated in synthetic four-helix bundle proteins. *Journal of American Chemical Society*, **126**, 14389-14399. doi:10.1021/ja0484294

[26] Etterna, T.J., Huynen, M.A., De Vos, W.M. and Van der Oost, J. (2003) TRASH: A novel metal-binding domain predicted to be involved in heavy-metal sensing, trafficking and resistance. *Trends in Biochemical Science*, **28**, 170-173. doi:10.1016/S0968-0004(03)00037-9

[27] Rigden, D.J. and Galperin, M.Y. (2004) The DxDxDG motif for calcium binding: Multiple structural contexts and implications for evolution. *Journal of Molecular Biology*, **343**, 971-984. doi:10.1016/j.jmb.2004.08.077

[28] Andreini, C., Banci, L., Bertini, I. and Rosato, A. (2006) Counting the zinc-proteins encoded in the human genome. *Journal of Proteome Research*, **5**, 196-201. doi:10.1021/pr050361j

[29] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403-410.

[30] Chou, K.C. and Zhang, C.T. (1995) Prediction of protein structural classes. *Critical Review in Biochemistry and Molecular Biology*, **30**, 275-349. doi:10.3109/10409239509083488

[31] Klein, P. (1986) Prediction of protein structural class by discriminant analysis. *Biochimica Biophysica Acta*, **874**, 205-215. doi:10.1016/0167-4838(86)90119-6

[32] Anfisen, C.B. (1973) Principles that govern the folding of protein chains. *Science*, **181**, 223-230. doi:10.1126/science.181.4096.223

[33] Chou, K.C. (2000) Review: Prediction of protein structural classes and subcellular locations. *Current Protein and Peptide Science*, **1**, 171-208. doi:10.2174/1389203003381379

[34] Hua, S. and Sun, Z. (2001) Support vector machine approach for protein subcellular localization prediction. *Bioinformatics*, **17**, 721-728. doi:10.1093/bioinformatics/17.8.721

[35] Nakai, K. (2000) Protein sorting signals and prediction of subcellular localization. *Advances in Protein Chemistry*, **54**, 277-344. doi:10.1016/S0065-3233(00)54009-1

[36] Zhou, G.P. and Doctor, K. (2003) Subcellular location prediction of apoptosis proteins. *Proteins*, **50**, 44-48. doi:10.1002/prot.10251

[37] Rice, P., Longden, I. and Bleasby, A. (2000) EMBOSS: The European molecular biology open software suite. *Trends in Genetics*, **16**, 276-277. doi:10.1016/S0168-9525(00)02024-2

 

[38] Chou, K.C. (2001) Prediction of protein cellular attributes using pseudo amino acid composition. *Proteins*, *Structure*, *Function and Genetics*, **43**, 246-255. doi:10.1002/prot.1035

[39] Tanford, C. (1962) Contribution of hydrophobic interactions to the stability of the globular conformation of proteins. *Journal of American Chemical Society*, **84**, 4240-4247. doi:10.1021/ja00881a009

[40] Hopp, T.P. and Woods, K.R. (1981) Prediction of protein antigenic determinants from amino acid sequences. *Proceeding of National Academic of Science USA*, **78**, 3824-3828. doi:10.1073/pnas.78.6.3824

[41] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning internal representations by error propagation. In: D.E. Rumelhart, J.L. McClelland, PDP Research Group Editors, *Parallel distributed processing*: *Explorations in the microstructure of cognition*. Foundations, MIT Press, Cambridge, 318-362.

[42] Zhou, G.P. (1998) An intriguing controversy over protein structural class prediction. *Journal of Protein Chemistry*, **17**, 729-738. doi:10.1023/A:1020713915365

[43] Chou, K.C. and Cai, Y.D. (2002) Using functional domain composition and support vector machines for prediction of protein subcellular location. *Journal of Biological Chemistry*, **277**, 45765-45769. doi:10.1074/jbc.M204161200

[44] Huang, Y. and Li, Y. (2004) Prediction of protein subcellular locations using fuzzy k-NN method. *Bioinformatics*, **20**, 21-28. doi:10.1093/bioinformatics/btg366

[45] Lin, H.H., Han, L.Y., Zhang, H.L., Zheng, C.J., Xie, B. and Chen, Y.Z. (2006) Prediction of the functional class of lipid-binding proteins from sequence derived properties irrespective of sequence similarity. *Journal of Lipid Research*, **47**, 824-831. doi:10.1194/jlr.M500530-JLR200

[46] Fierro-Monti, I. and Mathews, M.B. (2000) Proteins binding to duplexed RNA: One motif, multiple functions. *Trends in Biochemical Science*, **25**, 241-246. doi:10.1016/S0968-0004(00)01580-2

[47] Perez-Canadillas, J.M. and Varani, G. (2001) Recent advances in RNA-protein recognition. *Current Opinion in Structural Biology*, **11**, 53-58. doi:10.1016/S0959-440X(00)00164-0

[48] Hunt, J.A., Ahmed, M. and Fierke, C.A. (1999) Metal binding specificity in carbonic anhydrase is influenced by conserved hydrophobic core residues. *Biochemistry*, **38**, 9054-9062. doi:10.1021/bi9900166

[49] Rapisarda, V.A., Chehin, R.N., De Las Rivas, J., Rodriguez-Montelongo, L., Farias, R.N. and Massa, E.M. (2002) Evidence for Cu(I)-thiolate ligation and prediction of a putative copper-binding site in the Escherichia coli NADH dehydrogenase-2. *Archives Biochemistry and Biophysics*, **405**, 87-94. doi:10.1016/S0003-9861(02)00277-1

[50] Abbott, J.J., Pei, J., Ford, J.L., Qi, Y., Grishin, V.N., Pitcher, L.A., Phillips, M.A. and Grishin, N.V. (2001) Structure prediction and active site analysis of the metal binding determinants in gamma-glutamylcysteine synthetase. *Journal of Biological Chemistry*, **276**, 42099-42107. doi:10.1074/jbc.M104672200

 

**Supplementary Table 1.** The summary of the performance accuracy and validation results of 1st layer (NN1) of MetalloPred based on different sequence derived features.

| Class | Training set | Test set | Validation of NN1 (%) | | |
|---|---|---|---|---|---|
| | (%) | (%) | Independ-ent set | Self test | Jackknife test |
| 1. Pseudo Amino Acid Composition | | | | | |
| A | 100.00 | 73.53 | 70.98 | 70.14 | 67.01 |
| B | 100.00 | 75.77 | 73.93 | 72.71 | 63.24 |
| Average | 100.00 | 74.65 | 72.45 | 71.42 | 65.12 |
| 2. Amino Acid Composition | | | | | |
| A | 85.44 | 63.53 | 64.09 | 68.86 | 64.04 |
| B | 87.08 | 62.32 | 65.94 | 65.51 | 62.46 |
| Average | 86.26 | 62.93 | 65.01 | 67.18 | 63.25 |
| 3. Physicochemical Properties | | | | | |
| A | 84.41 | 68.82 | 63.34 | 57.22 | 55.16 |
| B | 88.41 | 69.08 | 69.77 | 62.87 | 58.70 |
| Average | 86.41 | 68.95 | 66.55 | 60.04 | 56.93 |
| 4. Pseudo Amino Acid Composition + Amino Acid Composition | | | | | |
| A | 98.68 | 78.82 | 74.75 | 63.91 | 61.67 |
| B | 98.31 | 84.88 | 79.95 | 67.97 | 65.99 |
| Average | 98.49 | 81.85 | 77.35 | 65.94 | 63.83 |
| 5. Pseudo Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 99.71 | 83.82 | 79.58 | 69.35 | 65.89 |
| B | 99.76 | 84.73 | 78.05 | 72.66 | 68.03 |
| Average | 99.73 | 84.27 | 78.81 | 71.00 | 66.96 |
| 6. Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 98.68 | 66.47 | 61.91 | 62.73 | 56.37 |
| B | 97.22 | 70.39 | 62.60 | 60.49 | 59.37 |
| Average | 97.95 | 68.43 | 62.25 | 61.61 | 57.87 |
| 7. Pseudo Amino Acid Composition + Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 99.85 | 89.41 | 81.38 | 78.71 | 73.92 |
| B | 99.64 | 86.57 | 85.47 | 82.27 | 74.44 |
| Average | 99.74 | 87.99 | 83.42 | 80.49 | 74.18 |

A: Metal binding; B: Non-metal binding

**Supplementary Table 2.** The summary of the performance accuracy and validation results of 2nd layer (NN2) of MetalloPred based on different sequence derived features.

| Class | Training | Test | Validation of NN1 (%) | | |
|---|---|---|---|---|---|
| | Set (%) | Set (%) | Indepen-dent set | Self test | Jackknife test |
| 1. Pseudo Amino Acid Composition | | | | | |
| A | 93.42 | 60.00 | 64.19 | 68.26 | 59.13 |
| B | 98.21 | 85.71 | 76.92 | 92.86 | 84.29 |
| C | 95.74 | 71.05 | 48.56 | 71.32 | 62.24 |
| Average | 95.79 | 72.25 | 63.22 | 77.48 | 68.55 |
| 2. Amino Acid Composition | | | | | |
| A | 99.06 | 65.00 | 61.15 | 67.77 | 59.21 |
| B | 100.00 | 57.14 | 76.92 | 90.00 | 78.57 |
| C | 98.69 | 68.42 | 45.00 | 68.36 | 60.33 |
| Average | 99.25 | 63.52 | 61.02 | 75.38 | 66.04 |
| 3. Physicochemical Properties | | | | | |
| A | 93.10 | 71.25 | 63.04 | 66.40 | 49.11 |
| B | 83.93 | 64.29 | 69.23 | 72.86 | 61.43 |
| C | 89.18 | 59.21 | 58.79 | 53.40 | 51.74 |
| Average | 88.74 | 64.92 | 63.69 | 64.22 | 54.09 |
| 4. Pseudo Amino Acid Composition + Amino Acid Composition | | | | | |
| A | 98.12 | 52.50 | 72.67 | 75.69 | 59.94 |

 

| | | | | | |
|---|---|---|---|---|---|
| B | 96.43 | 64.29 | 68.46 | 65.71 | 52.86 |
| C | 97.38 | 63.16 | 60.26 | 54.23 | 54.19 |
| Average | 97.31 | 59.98 | 67.13 | 65.21 | 55.66 |
| 5. Pseudo Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 99.06 | 71.25 | 66.34 | 69.05 | 64.78 |
| B | 100.00 | 85.71 | 84.62 | 77.14 | 70.00 |
| C | 98.69 | 77.11 | 71.24 | 66.41 | 62.62 |
| Average | 99.25 | 78.02 | 74.07 | 70.87 | 65.80 |
| 6. Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 94.04 | 68.75 | 67.33 | 68.42 | 57.92 |
| B | 89.29 | 71.43 | 69.23 | 87.14 | 78.57 |
| C | 93.77 | 71.05 | 44.61 | 67.69 | 59.75 |
| Average | 92.37 | 70.41 | 60.39 | 74.42 | 65.41 |
| 7. Pseudo Amino Acid Composition + Amino Acid Composition + Physicochemical Properties | | | | | |
| A | 99.06 | 82.50 | 73.66 | 74.86 | 69.76 |
| B | 100.00 | 84.29 | 76.15 | 72.86 | 62.86 |
| C | 98.69 | 78.95 | 75.68 | 72.01 | 60.06 |
| Average | 99.25 | 81.91 | 75.16 | 73.24 | 64.23 |

A: Alkali earth metal binding; B: Alkali metal binding; C: Transition metal binding.

**Supplementary Table 3.** The summary of the performance accuracy and validation results of 3rd layer of MetalloPred based on different sequence derived features.

| Classes | Training | Test | Validation of NN3 (%) | | |
|---|---|---|---|---|---|
| and subclass | Set (%) | Set (%) | Indpend-ent set | Self test | Jackknife test |
| (a) Alkali earth metal binding proteins (NN3) | | | | | |
| 1. Pseudo Amino Acid Composition | | | | | |
| Calcium | 95.98 | 86.01 | 63.67 | 72.59 | 63.92 |
| Magnesium | 94.74 | 78.10 | 64.95 | 66.16 | 57.55 |
| Average | 95.36 | 82.05 | 64.31 | 69.37 | 60.73 |
| 2. Amino Acid Composition | | | | | |
| Calcium | 92.48 | 90.91 | 72.84 | 73.29 | 63.36 |
| Magnesium | 90.43 | 92.38 | 63.66 | 64.05 | 54.49 |
| Average | 91.45 | 91.64 | 68.25 | 68.67 | 58.92 |
| 3. Physicochemical Properties | | | | | |
| Calcium | 89.51 | 86.71 | 66.84 | 69.09 | 52.59 |
| Magnesium | 87.56 | 86.67 | 62.11 | 66.73 | 49.33 |
| Average | 88.53 | 86.69 | 64.47 | 67.91 | 50.96 |
| 4. Pseudo Amino Acid Composition + Amino Acid Composition | | | | | |
| Calcium | 94.76 | 90.21 | 74.25 | 81.12 | 64.76 |
| Magnesium | 93.54 | 93.33 | 75.52 | 77.82 | 62.91 |
| Average | 94.15 | 91.77 | 74.88 | 79.47 | 63.83 |
| 5. Pseudo Amino Acid Composition + Physicochemical Properties | | | | | |
| Calcium | 97.20 | 88.11 | 55.56 | 62.52 | 56.22 |
| Magnesium | 95.22 | 79.05 | 57.47 | 72.08 | 64.63 |
| Average | 96.21 | 83.58 | 56.51 | 67.30 | 60.42 |
| 6. Amino Acid Composition + Physicochemical Properties | | | | | |
| Calcium | 94.23 | 88.81 | 72.49 | 73.71 | 63.08 |
| Magnesium | 92.82 | 91.43 | 57.22 | 65.01 | 56.60 |
| Average | 93.52 | 90.12 | 64.85 | 69.36 | 59.84 |
| 7. Pseudo Amino Acid Composition + Amino Acid Composition + Physicochemical Properties | | | | | |
| Calcium | 93.01 | 88.81 | 73.32 | 78.53 | 77.48 |
| Magnesium | 90.43 | 93.33 | 81.60 | 83.67 | 73.54 |
| Average | 91.72 | 91.07 | 77.46 | 81.10 | 75.51 |
| (b) Alkali metal binding proteins (NN4) | | | | | |
| 1. Pseudo Amino Acid Composition | | | | | |
| Potassium | 94.00 | 88.00 | 70.00 | 91.67 | 83.33 |
| Sodium | 87.50 | 85.00 | 66.67 | 60.00 | 58.00 |
| Average | 90.75 | 86.50 | 68.33 | 75.83 | 70.66 |
| 2. Amino Acid Composition | | | | | |
| Potassium | 92.00 | 84.00 | 80.00 | 91.67 | 81.67 |
| Sodium | 87.50 | 70.00 | 59.33 | 54.00 | 52.18 |

 

| | | | | | |
|---|---|---|---|---|---|
| Average | 89.75 | 77.00 | 69.66 | 72.83 | 66.92 |
| 3. Physicochemical Properties | | | | | |
| Potassium | 97.92 | 91.67 | 70.00 | 70.00 | 55.00 |
| Sodium | 92.00 | 80.00 | 63.33 | 60.00 | 52.48 |
| Average | 94.96 | 85.83 | 66.66 | 65.00 | 53.74 |
| 4. Pseudo Amino Acid Composition + Amino Acid Composition | | | | | |
| Potassium | 96.00 | 91.67 | 60.00 | 59.67 | 54.00 |
| Sodium | 87.50 | 84.40 | 63.33 | 53.00 | 51.00 |
| Average | 91.75 | 88.03 | 61.66 | 56.33 | 52.50 |
| 5. Pseudo Amino Acid Composition + Physicochemical Properties | | | | | |
| Potassium | 97.30 | 94.00 | 70.00 | 69.33 | 68.33 |
| Sodium | 93.75 | 87.00 | 67.33 | 52.07 | 51.20 |
| Average | 95.52 | 90.50 | 68.66 | 60.70 | 59.76 |
| 6. Amino Acid Composition + Physicochemical Properties | | | | | |
| Potassium | 97.92 | 86.00 | 64.00 | 65.00 | 66.67 |
| Sodium | 92.60 | 75.00 | 63.67 | 50.00 | 50.00 |
| Average | 95.26 | 80.50 | 63.83 | 57.50 | 58.33 |
| 7. Pseudo Amino Acid Composition + Amino Acid Composition + Physicochemical Properties | | | | | |
| Potassium | 98.40 | 97.80 | 80.00 | 93.15 | 86.67 |
| Sodium | 97.50 | 95.00 | 78.40 | 90.00 | 80.00 |
| Average | 97.95 | 96.40 | 79.20 | 91.57 | 83.33 |
| (c) Transition metal binding proteins (NN5) | | | | | |
| 1. Pseudo Amino Acid Composition | | | | | |
| Cobalt | 92.50 | 100.0 | 60.0 | 88.0 | 78.0 |
| Copper | 98.19 | 80.00 | 53.33 | 78.26 | 69.20 |
| Iron | 100.0 | 87.50 | 60.53 | 62.20 | 53.66 |
| Manganese | 100.0 | 100.0 | 66.44 | 76.74 | 68.22 |
| Molybdenum | 100.0 | 100.0 | 60.0 | 84.09 | 72.73 |
| Nickel | 97.44 | 90.0 | 60.0 | 93.88 | 83.67 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 98.77 | 93.90 | 59.13 | 60.29 | 51.23 |
| Average | 96.28 | 93.92 | 64.34 | 80.12 | 71.46 |
| 2. Amino Acid Composition | | | | | |
| Cobalt | 100.0 | 100.0 | 80.0 | 88.0 | 80.0 |
| Copper | 98.64 | 87.27 | 66.67 | 71.38 | 59.06 |
| Iron | 98.48 | 100.0 | 48.68 | 67.07 | 58.54 |
| Manganese | 100.0 | 100.0 | 65.77 | 70.54 | 61.24 |
| Molybdenum | 100.0 | 100.0 | 60.0 | 93.18 | 84.09 |
| Nickel | 100.0 | 90.0 | 80.0 | 79.59 | 67.35 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 99.39 | 90.24 | 58.62 | 60.05 | 52.45 |
| Average | 99.56 | 95.94 | 69.33 | 78.41 | 69.53 |
| 3. Physicochemical Properties | | | | | |
| Cobalt | 100.0 | 100.0 | 40.0 | 72.0 | 58.0 |
| Copper | 85.97 | 85.45 | 50.00 | 57.10 | 53.33 |
| Iron | 89.39 | 93.75 | 52.63 | 65.85 | 57.56 |
| Manganese | 94.17 | 92.31 | 51.68 | 56.59 | 51.09 |
| Molybdenum | 100.0 | 100.0 | 60.0 | 79.55 | 63.64 |
| Nickel | 100.0 | 90.0 | 40.0 | 56.94 | 56.73 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 93.87 | 90.24 | 42.72 | 56.81 | 51.13 |
| Average | 95.42 | 93.97 | 52.43 | 66.54 | 58.93 |
| 4. Pseudo Amino Acid Composition + Amino Acid Composition | | | | | |
| Cobalt | 100.0 | 100.0 | 70.0 | 64.0 | 56.0 |
| Copper | 98.64 | 92.73 | 76.67 | 68.41 | 66.81 |
| Iron | 100.0 | 93.75 | 78.42 | 69.27 | 58.29 |
| Manganese | 100.0 | 100.0 | 78.12 | 65.58 | 54.73 |
| Molybdenum | 100.0 | 100.0 | 70.0 | 67.27 | 58.18 |
| Nickel | 100.0 | 90.0 | 70.0 | 68.57 | 66.33 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 99.39 | 92.68 | 73.05 | 60.69 | 56.72 |
| Average | 99.75 | 96.14 | 76.30 | 69.54 | 63.38 |
| 5. Pseudo Amino Acid Composition + Physicochemical Properties | | | | | |
| Cobalt | 100.0 | 100.0 | 80.0 | 84.0 | 76.0 |
| Copper | 98.19 | 88.09 | 78.0 | 65.29 | 59.49 |
| Iron | 100.0 | 100.0 | 73.68 | 75.61 | 70.73 |

 

| | | | | | |
|---|---|---|---|---|---|
| Manganese | 100.0 | 96.15 | 71.68 | 63.57 | 56.59 |
| Molybdenum | 100.0 | 100.0 | 80.0 | 81.82 | 77.27 |
| Nickel | 100.0 | 90.0 | 80.0 | 71.43 | 63.27 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 98.77 | 90.24 | 60.48 | 64.46 | 56.86 |
| Average | 99.62 | 95.56 | 78.0 | 75.77 | 70.03 |
| 6. Amino Acid Composition + Physicochemical Properties | | | | | |
| Cobalt | 100.0 | 100.0 | 60.0 | 62.0 | 52.0 |
| Copper | 98.19 | 89.09 | 46.67 | 70.29 | 59.06 |
| Iron | 100.0 | 100.0 | 78.95 | 80.49 | 71.95 |
| Manganese | 100.0 | 100.0 | 61.74 | 67.44 | 55.81 |
| Molybdenum | 100.0 | 100.0 | 80.0 | 84.09 | 75.0 |
| Nickel | 100.0 | 90.0 | 60.0 | 55.10 | 44.90 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 98.77 | 91.46 | 71.65 | 63.24 | 54.41 |
| Average | 99.62 | 96.32 | 68.93 | 71.27 | 62.58 |
| 7. Pseudo Amino Acid Composition + Amino Acid Composition + Physicochemical Properties | | | | | |
| Cobalt | 100.0 | 100.0 | 80.0 | 66.0 | 62.0 |
| Copper | 93.21 | 81.82 | 80.0 | 70.36 | 68.77 |
| Iron | 100.0 | 100.0 | 84.74 | 70.98 | 60.0 |
| Manganese | 100.0 | 96.15 | 84.90 | 62.64 | 61.78 |
| Molybdenum | 100.0 | 100.0 | 80.0 | 68.64 | 69.55 |
| Nickel | 100.0 | 90.0 | 80.0 | 68.78 | 62.45 |
| Vanadium | 100.0 | 100.0 | 100.0 | 100.0 | 100.0 |
| Zinc | 97.85 | 95.12 | 86.66 | 71.27 | 69.26 |
| Average | 98.88 | 95.39 | 84.06 | 71.39 | 67.98 |

 