



Computationally characterizing and comprehensive analysis of zinc-binding sites in proteins [☆]

Zexian Liu ^a, Yongbo Wang ^b, Changhai Zhou ^a, Yu Xue ^b, Wei Zhao ^{a,*}, Haiyan Liu ^{a,*}

^a Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science & Technology of China, Hefei, Anhui 230027, China

^b Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

ARTICLE INFO

Article history:

Received 3 December 2012

Received in revised form 2 March 2013

Accepted 4 March 2013

Available online 14 March 2013

Keywords:

Zinc-binding

Geometric restriction

Training-independent

Cancer gene

Drug target

ABSTRACT

Zinc is one of the most essential metals utilized by organisms, and zinc-binding proteins play an important role in a variety of biological processes such as transcription regulation, cell metabolism and apoptosis. Thus, characterizing the precise zinc-binding sites is fundamental to an elucidation of the biological functions and molecular mechanisms of zinc-binding proteins. Using systematic analyses of structural characteristics, we observed that 4-residue and 3-residue zinc-binding sites have distinctly specific geometric features. Based on the results, we developed the novel computational program Geometric REstriction for Zinc-binding (GRE4Zn) to characterize the zinc-binding sites in protein structures, by restricting the distances between zinc and its coordinating atoms. The comparison between GRE4Zn and analogous tools revealed that it achieved a superior performance. A large-scale prediction for structurally characterized proteins was performed with this powerful predictor, and statistical analyses for the results indicated zinc-binding proteins have come to be significantly involved in more complicated biological processes in higher species than simpler species during the course of evolution. Further analyses suggested that zinc-binding proteins are preferentially implicated in a variety of diseases and highly enriched in known drug targets, and the prediction of zinc-binding sites can be helpful for the investigation of molecular mechanisms. In this regard, these prediction and analysis results should prove to be highly useful for further biomedical study and drug design. The online service of GRE4Zn is freely available at: <http://biocomp.ustc.edu.cn/gre4zn/>. This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

© 2013 Elsevier B.V. All rights reserved.

1. Introduction

Although the total net content of zinc in organisms is very low, it is still essential for survival. Free or loosely bound zinc ions function as an intracellular signal [1], and were shown to act as a second messenger recently [2]. However, the major role of the zinc ion is tightly coordinated with protein residues [3], and it is estimated that as much as 10% of the human proteome is made up of potential zinc-binding proteins [4]. A large number of studies have been conducted to describe the molecular mechanisms of zinc-binding [5,6]. Although zinc ions can be penta- or even hexa-coordinated, tetrahedral coordination is the predominant form for most of the zinc-binding sites [6]. The majority of the identified zinc-binding residues are made up of cysteine (C) and histidine (H) [3], while various other residues, such as glutamic acid (E), aspartic acid (D), serine, threonine, lysine and methionine can also coordinate with zinc ions [5]. In fact, cysteine,

histidine, aspartic acid and glutamic acid constitute almost all of the zinc-coordinating protein residues, while oxygen, nitrogen and sulfur donors in water molecules or other free ligands can also serve as coordinating moieties for zinc ions [5,6]. As a structural component that binds with amino acid (AA) residues, the zinc ion is critical for the functions of proteins, such as helping to stabilize the structure of “zinc-finger” transcription factors [3,7] and acting as the catalytic site in enzymes [3]. Thus, it is critical to identify zinc-binding sites in order to dissect the molecular functions and mechanisms in the proteins that contain them.

To date, a variety of experimental approaches, including X-ray diffraction, Nuclear Magnetic Resonance (NMR) and X-ray absorption fine structure (XAFS) techniques have been employed to identify zinc-binding residues in proteins [5]. However, since these experimental studies are both time- and labor-intensive, only a small proportion of potential zinc-binding proteins have characterized, even though genome-scale analysis have suggested that there are thousands of zinc-binding proteins [4]. Recently, a number of computational approaches have made contributions to this area to promote the discovery of zinc-binding proteins together with their binding sites. Since zinc ions only coordinate with restricted types of residues

[☆] This article is part of a Special Issue entitled: Computational Proteomics, Systems Biology & Clinical Implications. Guest Editor: Yudong Cai.

* Corresponding authors. Tel./fax: +86 551 63607450.

E-mail addresses: zhaowei@ustc.edu.cn (W. Zhao), hylu@ustc.edu.cn (H. Liu).

and the binding seems to follow certain specific patterns, a series of prediction studies were carried out based on sequence analysis [8–17]. A variety of algorithms were employed alone or in combination in these studies, including support vector machines (SVMs) [10,11,13,14,17], neural networks (NNs) [9,11,13,17], machine learning (ML) [12] and the homology-based method of PHI-Blast [8,15]. Since zinc-binding is heavily dependent on the three-dimensional conformations of protein residues, the structure-based predictions might be expected to achieve a better performance [5,6].

Previously, a handful of computational studies have contributed to the effort to predict zinc-binding sites based on protein structures [18–27]. Structural features such as secondary structure states (SS), solvent-accessible surface areas (SASAs), inter-residue distance matrices, geometrical features and residue properties were combined with various algorithms including NNs [18], SVMs [20,24], machine learning [24], random forest algorithm [21], and Bayesian classifier [22] in an effort to provide accurate predictions. Furthermore, the empirical Fold-X force field, Rosetta software and the fragment transformation method (FTM) were also employed to characterize zinc-binding structures [19,25,27] (Table 1). Recently, Zheng et al. presented a powerful computational framework which integrates various features including sequence, structure and network properties with the random forest algorithm to predict zinc-binding sites [28]. In addition, we previously developed a structure-based method (TEMSP or 3D TEmplate-based Metal Site Prediction) for predicting zinc-binding sites [26]. In these studies, complex classifiers, force field-based modeling and template-based calculation among these approaches afforded excellent results.

In this work, we systematically analyzed the structural features of zinc-binding sites from a well characterized and non-redundant dataset of 601 zinc-binding sites in 431 proteins. We observed that 4-residue (4-res) and 3-residue sites (3-res) have different sequential and structural features, such as sequence length distribution, AA preferences, SSs and geometrical distance. In particular, we found that the geometrical distance between zinc and the binding residue was specifically restricted in the 4-res and 3-res binding sites, respectively. Based on these observations, we developed a geometric restriction approach to characterize zinc-binding sites from protein structures. The geometrical distance ranges for the 4-res and 3-res binding were respectively calculated from the zinc-binding data and then were employed to predict potential zinc-binding residues. The

software program Geometric REstriction for Zinc-binding, <http://biocomp.ustc.edu.cn/gre4zn/> (GRE4Zn) was implemented, with a sensitivity of 97.68% and a precision of 98.93% under an Intersection over Union Ratio (IoUR) ≥ 0.5 , respectively, while the values were 95.56% and 96.58% under IoUR = 1.0. With this superior predictor, we performed a large-scale study on the structurally characterized proteins in PDB database [29]. Furthermore, we systematically analyzed the potential zinc-binding sites for *Escherichia coli*, *Saccharomyces cerevisiae* and *Homo sapiens*. The statistical analyses of gene ontology (GO) annotations for the results suggested that the enriched biological processes were different for different species. It was observed that zinc-binding proteins were involved in more complicated biological processes in higher organisms than lower organisms, which implies an ongoing evolution of the functions of zinc-binding proteins. Further analyses revealed that zinc-binding proteins are also significantly enriched in cancer genes and drug targets, and could serve as a useful resource for further biomedical investigation and drug design.

2. Materials and methods

2.1. Data preparation and analysis

The dataset of experimentally identified zinc-binding sites have been retrieved according to the methods described in our recent study [26]. The zinc-binding protein structures of less than 95% sequence identity were retrieved while redundant chains were removed. The “abnormal” zinc sites with multiple conformations, less than three coordinating atoms, metal atom occupancy lower than 0.5, or a B factor higher than 90, were discarded. Since the excess redundancy of a large number of homologous sites could lead to the overestimation of prediction accuracy, the zinc-binding proteins were clustered at a sequence identity cutoff of 30% with the CD-HIT program [30], and only one representative protein chain was retained from each cluster. Furthermore, 99 protein structures from the resulting 431 proteins were randomly selected as the testing dataset, while the others were considered as the training dataset.

Since almost all zinc-binding residues are cysteines, histidines, glutamic acids and aspartic acids (CHEDs), in this study, only the sites with CHED residues were reserved for consideration. The dataset of 601 zinc-binding sites in 431 proteins was divided into two subsets according to the size of their ligands. The larger subset of “4 residues (4-res)” contains 473 sites in 317 proteins, in which four of the coordinating ligands are CHED residues, while the smaller subset of “3 residues (3-res)” has 128 sites in 123 proteins, in which the tetrahedral ligands were constituted with three CHED residues and another donor of oxygen, nitrogen or sulfur from either a water molecule or other free ligands. The distances between atoms were computed for the 4-res and 3-res datasets, respectively. For large-scale prediction, the protein structures determined by X-ray with a resolution of less than 3.0 Å were retrieved from the PDB database [29]. The protein structures were integrated with a sequence similarity threshold of 30% in organism-specific statistical analyses of *E. coli*, *S. cerevisiae* and *H. sapiens*. From these large-scale analyses, 377, 411 and 1027 structure chains were predicted to contain zinc-binding sites among 1263, 726 and 3087 protein structures, respectively, from the PDB database.

The SS and SASA analyses were carried out by STRIDE [31]. The enrichment analyses of the annotations of SCOP structural classifications, GO, cancer genes and drug targets were performed with a hypergeometric distribution [31]. The comparison of the SCOP structural classifications for 4-res and 3-res zinc-binding was performed with Yates' Chi-square (χ^2) test with the 2×2 contingency table [31]. The GO annotation file for PDB was downloaded from the GOA database at the EBI (<http://www.ebi.ac.uk/goa>) [32], while the SCOP structural classification of proteins data was downloaded from the SCOP database (<http://scop.mrc-lmb.cam.ac.uk/scop/>) [33]. In

Table 1
Summary of a number of previous studies on the prediction of zinc-binding.

	Software	PMID	Algorithm
<i>Sequence-based predictions</i>			
Andreini et al. [8]		14962940	PHI-BLAST
Lin et al. [9]		15912584	NN
Passerini et al. [11]	MLP	16927295	SVM, NN
Lin et al. [10]	SVMProt	17254297	SVM
Passerini et al. [12]	Zinc Finder	17280606	ML
Shu et al. [14]	PredZinc	18245129	SVM
Lippi et al. [13]	MetalDetector	18635571	SVM, NN
Andreini et al. [15]		19697929	PHI-BLAST
Bertini et al. [16]		20443034	HMM
Passerini et al. [17]	MetalDetector v2.0	21576237	SVM, NN
<i>Structure-based predictions</i>			
Sodhi et al. [18]	MetSite	15313626	NN
Schymkowitz et al. [19]		16006526	Fold-X
Babor et al. [20]	CHED	17657805	SVM
Goyal and Mande [23]		17847089	Template-based
Ebert and Altman [22]	FEATURE	18042678	Bayesian
Bordner [21]		18940825	Random forest
Levy et al. [24]	SeqCHED	19173310	SVM, ML
Wang et al. [25]	Rosetta	20054832	Rosetta
Zhao et al. [26]	TEMSP	21414989	Template-based
Lu et al. [27]		22723976	FTM
Zheng et al. [28]	ZincIdentifier	23166753	Random forest

addition, the protein structures in this study were visualized in Pymol (<http://www.pymol.org/>) [34].

2.2. The geometric restriction approach (GRE)

Previous studies reviewed that zinc-coordinating atoms (zcAtom) are different in C, H, E and D residues: i.e. an SG atom for C, ND1 or NE2 atom for H, OE1 or OE2 atom for E, and OD1 or OD2 atom for D [5,6]. Although one residue may coordinate two zinc ions with different zcAtoms, in this study, for one residue it was considered that only one zcAtom coordinated.

We predict zinc-binding residues from the protein structure based simply on geometric restriction, and the prediction results of the 4-res and 3-res types were first computed separately and subsequently merged together. The prediction effort was conducted as follows.

Given a certain protein structure, the CHED residues in the protein were exhaustively combined to retrieve quadruplets or triplets. From the dataset, it was observed that the longest distance among the C α atoms for zinc-binding residues is approximately 12 Å, while exhaustive tests of the distance limitation for the C α atoms from 14 Å to 20 Å did not result in any difference in the prediction. Thus, only CHED residues within 13 Å were considered in this study. An extensive search of the tetrahedral coordination model (TCM) was carried out for the zcAtoms.

For the 4-res type, the center of the four zcAtoms was considered as the site of a potential zinc ion:

$$\text{zinc}_{(x,y,z)} = \frac{1}{4} \sum_{1 \leq i \leq 4} \text{zcAtom}_i(x,y,z) \quad (1)$$

For the 3-res type, a virtual atom (vAtom) was added to simulate the fourth coordinating atom from a water molecule or other free ligand. Since the TCM was a tetrahedron with four sides while the distance between the zcAtoms is approximately 3.3 Å [5,6], for an ideal TCM, the distance between the zinc and plane of the three zcAtom is 0.67 Å, while the distance between the vAtom and the plane of the three zcAtom is 2.69 Å. Furthermore, the zinc and vAtom could both be on the same side of the plane of the three zcAtom. Given the vector $\text{norV}_{(x,y,z)}$ as the normal vector for the plane of the three zcAtom, we considered the positions of the zinc ion and the virtual atom to be as follows:

$$\text{Zinc}_{(x,y,z)} = \frac{1}{3} \sum_{1 \leq i \leq 3} \text{zcAtom}_i(x,y,z) \pm 0.67 * \text{norV}_{(x,y,z)} \quad (2)$$

$$\text{vAtom}_{(x,y,z)} = \frac{1}{3} \sum_{1 \leq i \leq 3} \text{zcAtom}_i(x,y,z) \pm 2.69 * \text{norV}_{(x,y,z)}. \quad (3)$$

Based on the rationale computed from the training datasets (Table 4), the proper TCM for which the distances between the zinc ion and zcAtoms, the zinc ion and non-coordinating atoms and the vAtom and non-coordinating atoms that were within the defined ranges were preserved.

Finally, the prediction results of both the 4-res and 3-res types were merged. If two residues combinations among the quadruplets and triplets shared 3 residues, the results with less TCMs were dropped.

2.3. Performance evaluation

The performance evaluation was performed as described in our previous study [26]. The IoUR was used to quantify the prediction accuracy of the residues. Sensitivity (*Sn*) and precision (*Pr*) were employed to evaluate the prediction performance. The measurements were defined as shown below:

$$\text{IoUR} = \frac{N(\text{predicted ligand residues} \cap \text{actual ligand residues})}{N(\text{predicted ligand residues} \cup \text{actual ligand residues})}$$

$$\text{Sn} = \frac{TP}{TP + FN} \quad \text{and} \quad \text{Pr} = \frac{TP}{TP + FP}$$

2.4. Implementation of the online service

The GRE4Zn prediction server was implemented in PHP + Java. One is able to submit the protein structure file in PDB format to the server, and the calculation may be restrained to certain selected chains by specifying the chain codes. The prediction results provided after submission, and the user is then able to view the structure file of the predicted binding residues.

3. Results

3.1. Systematic analyses reveal the differences between 4-res and 3-res zinc-binding

Since the zinc ion is critical for protein function, the accurate prediction of zinc-binding sites would help advance the understanding of the biological functions and molecular mechanisms of zinc-binding proteins. A number of studies which have contributed to this area are summarized in Table 1, of which ten were sequence-based and the other eleven were structure-based. Since a variety of residues such as C, D, E and H were employed to coordinate with the zinc ion, it is a complex matter to accurately predict all the *bona fide* residues. A number of approaches and algorithms have been employed to characterize the zinc-binding residues in proteins and a variety of sequential or structural features considered.

In this study, since there were instances of both 4-res and 3-res types of zinc-binding, we systematically analyzed the two types individually. The analyses of the distribution of the zinc-binding residues suggested that the sequence distances between the N-terminal and C-terminal coordinating residues were diverse (Fig. 1A), especially in the case of 3-res type. Previous studies showed that a variety of features, including AA types, Ss and SASAs could be employed to predict zinc-binding sites [18], and we present a series of such analyses in this study. The distributions patterns of the four types of residues in the zinc-binding sites are presented in Fig. 1B. It was observed that in the 4-res type dataset, cysteines constituted more than 70% of the binding residues, while more than half of the binding residues were histidines in the 3-res. Interestingly, these results were consistent with the distribution of AA types with structural and catalytic functions in recent systematical analyses reported by Andreini et al. [35]. However, there is no distinction between the distribution of AA types in the 4-res and 3-res types of zinc-binding in their study [35]. The Ss of these residues were also analyzed and are presented in Fig. 1C. It was observed more than half of the residues in the 4-res zinc-binding type were in the turn regions of proteins, while for the 3-res type, approximately a quarter of the residues were in such regions. Furthermore, SASAs are an important structural feature which indicates the proportion of exposure to solvent, so they were analyzed and are presented in Fig. 1D. It was observed that less than half of the surface area was exposed to solvent for most of the residues. It is interesting that although the 4-res and 3-res types of zinc-binding have a number of feature differences, they have a similar pattern of surface area distribution.

In addition, we statistically analyzed the SCOP structural classifications for the zinc-binding proteins and the results are presented in Tables 2 and 3. It was evident that small proteins were enriched in the 4-res zinc-binding proteins, while alpha + beta proteins and alpha/beta proteins were over-represented in 3-res zinc-binding proteins (Table 2). Furthermore, in contrast with the 4-res zinc-binding proteins, there were more all beta proteins in the 3-res zinc-binding proteins (Table 3).

Taken together, it is suggested that there are obvious differences between the 4-res and 3-res types of zinc-binding. Therefore, we

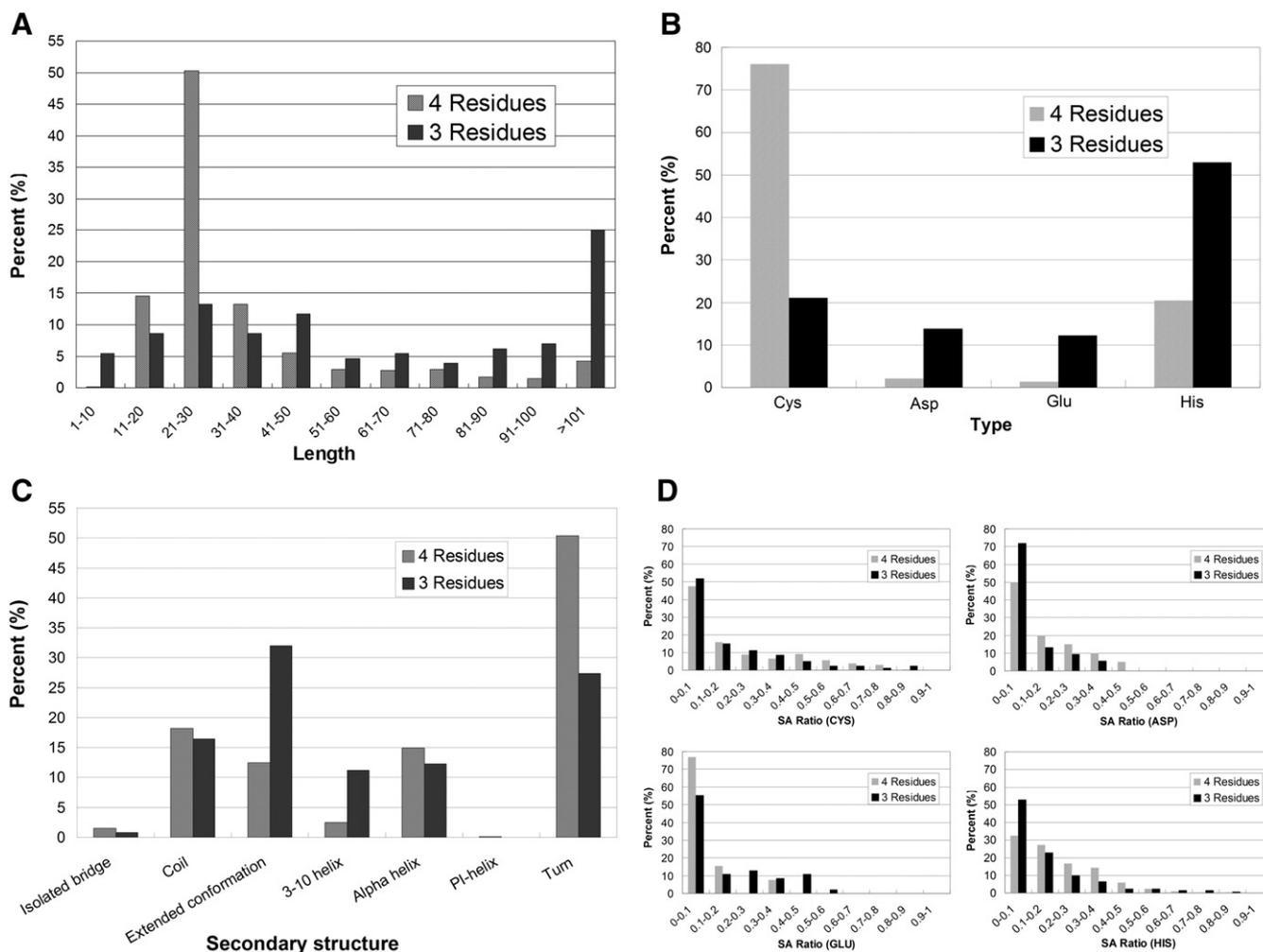


Fig. 1. The structural features of 4-res and 3-res zinc-binding. (A) The distribution of the sequence separations between zinc-binding residues. (B) The distribution of amino acid types of zinc-binding residues in the 4-res and 3-res types of zinc-binding sites. (C) The distribution of the secondary structure types of zinc-binding residues in the 4-res and 3-res. (D) The distribution pattern of the SASA ratio of zinc-binding residues in the 4-res and 3-res types of zinc-binding sites.

conducted structure-based prediction of zinc-binding sites, and the computational analyses of the 4-res and 3-res types were carried out separately in this study.

3.2. The geometric restriction approach for zinc-binding site prediction

Although the above analyses did not indicate any distinct features among the sequential and structural properties, it was observed that the geometric features of zinc-binding were well defined. The distances of the zinc ion and coordinating atoms were calculated for

the 4-res and 3-res types of zinc-binding in the datasets separately, and the results are presented in Fig. 2. Fig. 2A provides a model for the 4-res zinc-binding of *E. coli* DNA polymerase III (PDB code: 1A5T), while the diphtheria toxin repressor (PDB code: 1B10) from *Corynebacterium diphtheriae* is shown in Fig. 2B. The distribution patterns of the distance between the zinc ion and its coordinating atoms for the 4-res and 3-res types are presented in Fig. 2C and D, respectively. We found that there were different distance between the

Table 2

The enriched SCOP classifications for zinc-binding proteins (p -value < 0.01).

Type	Classification	Zinc-binding		Proteome		E-ratio ^c	p -value
		Num. ^a	Per. ^b	Num.	Per.		
4-res	Small proteins	135	82.82%	4,223	4.91%	16.86	1.70E−147
3-res	Alpha + beta proteins	36	50.00%	25,537	29.71%	1.68	2.37E−04
	Alpha/beta proteins	36	50.00%	28,306	32.93%	1.52	1.99E−03

^a The number of proteins annotated.

^b The proportion of proteins annotated.

^c Enrichment ratio: the proportion in zinc-binding divided by the one in SCOP database.

Table 3

The comparison of SCOP classifications between the 4-res and 3-res zinc-binding proteins.

Classification	4-res		3-res		E-ratio ^c	X^2	p -value
	Num. ^a	Per. ^b	Num.	Per.			
Small proteins	135	82.82%	2	2.78%	29.82	128.34	9.45E−30
Alpha + beta proteins	27	16.56%	36	50.00%	0.33	26.78	2.29E−07
Alpha/beta proteins	30	18.40%	36	50.00%	0.37	23.14	1.50E−06
All beta proteins	18	11.04%	19	26.39%	0.42	7.75	5.38E−03

^a The number of proteins annotated.

^b The proportion of proteins annotated.

^c The enrichment ratio: the proportion of the 4-res zinc-binding proteins divided by the one in the 3-res zinc-binding proteins.

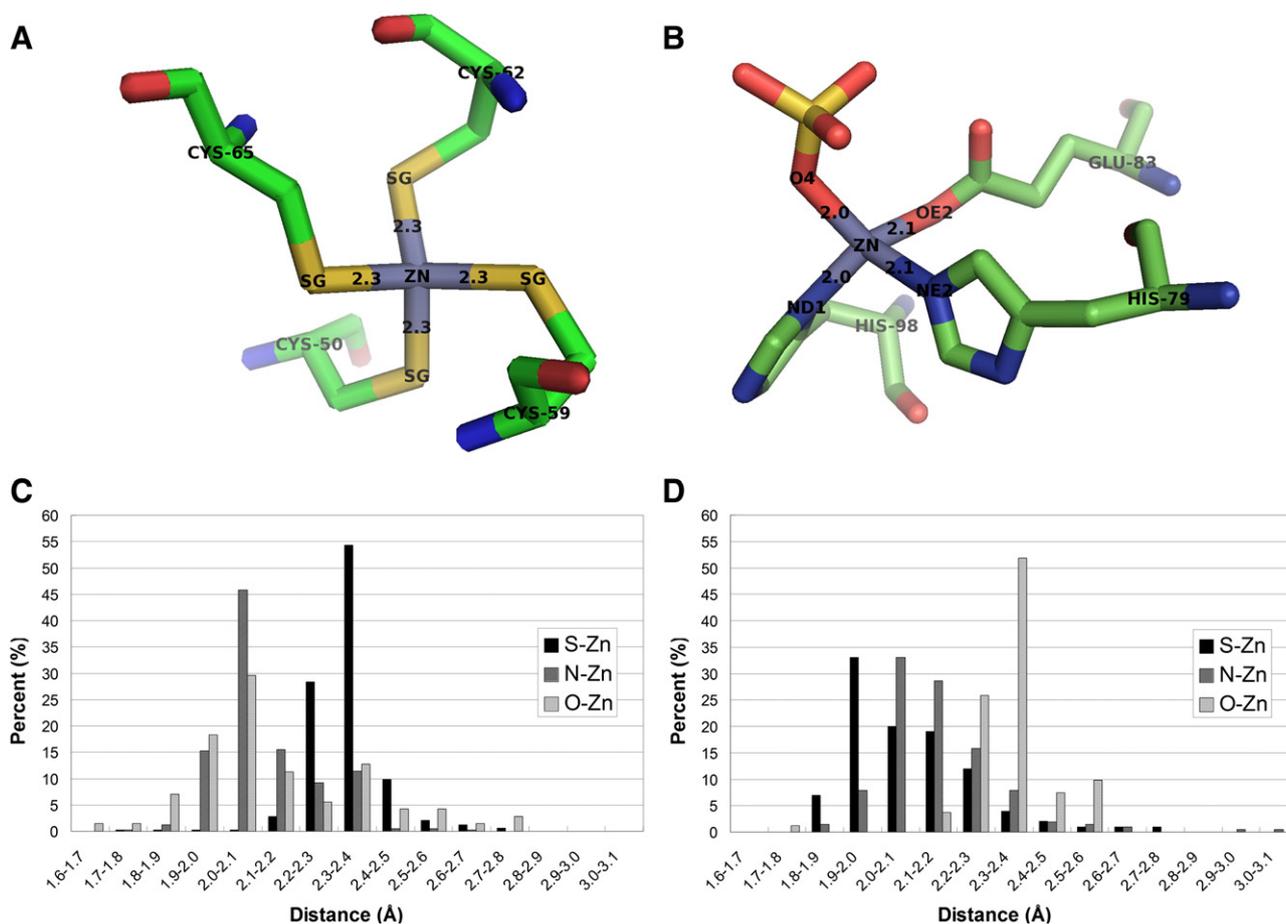


Fig. 2. The distribution pattern of the distances between the zinc ion and coordinating atoms. (A) An example of 4-res zinc-binding (PDB code: 1A5T). (B) An example of 3-res zinc-binding (PDB code: 1B10). (C) The distribution pattern of the distances between the zinc ion and coordinating atoms in 4-res zinc-binding. (D) The distribution pattern of the distances between the zinc ion and coordinating atoms in 3-res zinc-binding.

zinc ion and various atoms. For example, the majority of the distances between sulfur (S) and zinc (Zn) were found to be approximately 2.3–2.4 Å, while the distances for nitrogen (N) and the zinc ion were 2.0–2.1 Å. Furthermore, the distance preferences for the same type of atoms from the 4-res and 3-res types were also different. For instance, the distances between oxygen (O) and zinc in the 4-res type were extraordinarily diverse, while the distances for more than half of the 3-res type were 2.3–2.4 Å. Table 4 presents the distance limitations we derived from the training dataset, which is stable with a different dataset, such as the testing dataset (data not shown). It is obvious that although the coordination sphere is flexible, there were considerable preferences and restraints on zinc-binding.

Based on these observations, we proposed a geometric restriction (GRE) approach to predict zinc-binding residues from the protein

structure. The geometric features of zinc-coordination were employed to characterize the zinc-binding residues. Using the GRE approach, we developed the predictor program GRE4Zn. The positions of the zinc ion were simply modeled using Eqs. (1) and (2), while the distances between the zinc ion and its coordinate atoms were checked to guarantee the formation of the coordination bond. Subsequently, the distances between the zinc ion and non-coordinate atoms were examined to ensure the correct coordination. It was assumed that the distance between the zinc ion and non-coordinate atoms must be greater than the zinc ion and its coordinating atoms. For the 3-res type, the distance between the virtual atom and other atoms were checked to avoid clash. All the distance ranges were directly calculated from the datasets, and all these procedures were implemented into GRE4Zn.

3.3. Performance evaluation and comparison

To evaluate the prediction performance of GRE4Zn, the measurements in our previous study of TEMSP [26] were employed. Since there were several zinc-binding residues at every site, all of the binding residues should be determined. However, predictions with only one mismatch were still valuable. In our previous study, IoUR was employed, which quantifies the prediction accuracy by calculating the correct proportion over the sum of actual and predicted residues [26].

In this study, we adopted the IoUR measurement to calculate the performance under the conditions $\text{IoUR} \geq 0.5$ and $\text{IoUR} = 1.0$, which indicated that at least one half and all of the residues were correctly predicted, respectively. The details are presented in Table 5. Since

Table 4

The geometric distances employed in GRE4Zn for zinc-binding prediction. “–” indicates that there is no limitation for these atoms.

	Zn-S (Å)	Zn-O (Å)	Zn-N (Å)	Zn-OX ^a (Å)	Zn-C (Å)
<i>The distance between zinc and binding atoms</i>					
4-res	1.60–3.05	1.70–2.55	1.78–2.55	–	–
3-res	2.00–2.65	1.80–2.55	1.80–2.55	1.80–2.55	–
<i>The distance between zinc and non-binding atoms</i>					
4-res	>3.05	>2.55	>2.55	–	–
3-res	–	>2.55	>2.55	–	>2.55

^a OX presents the virtual atom for zinc-coordination.

Table 5
The prediction performance of GRE4Zn and a comparison with TEMSP and CHED.

	IoUR ≥ 0.5		IoUR = 1.0	
	<i>Sn</i> (%)	<i>Pr</i> (%)	<i>Sn</i> (%)	<i>Pr</i> (%)
GRE4Zn (4-res) ^a	97.68	98.93	95.56	96.58
GRE4Zn (3-res) ^a	88.28	86.26	77.34	75.57
GRE4Zn (Merged) ^a	97.19	91.88	91.35	85.78
GRE4Zn (Merged) ^b	97.01	87.84	89.55	81.08
TEMSP ^b	86.00	95.90	73.50	82.00
CHED ^b	82.84	52.61	37.31	23.70

^a Training dataset.

^b Testing dataset.

the 4-res and 3-res types of zinc-binding were predicted separately, the performance was also calculated independently. For the 4-res type, The *Sn* and *Pr* values of GRE4Zn were 97.68% and 98.93%, respectively under IoUR ≥ 0.5 , while the values were 95.56% and 96.58% under IoUR = 1.0 (Table 5). For the 3-res type, GRE4Zn achieved a performance of 88.28% (*Sn*) and 86.26% (*Pr*) when IoUR ≥ 0.5 , while the performance was 77.34% (*Sn*) and 75.57% (*Pr*) under IoUR = 1.0 (Table 5). For the 3-res zinc-binding, only three coordinating atoms were from the three residues, while the fourth coordinating atom was from a water molecule or other free ligand. Since the fourth coordinating atom was not as restrained as the protein residues and should therefore be simulated as a virtual coordinating atom during prediction, it is more complicated to correctly predict the 3-res zinc-binding sites. Such a characterization is consistent with the result that the prediction of the 4-res type was more accurate than the 3-res type. When the prediction results from the 4-res and 3-res types were merged, GRE4Zn yields a performance of 97.19% (*Sn*) and 91.88% (*Pr*) under IoUR ≥ 0.5 , while the *Sn* and *Pr* values are 91.35% and 85.78% under IoUR = 1.0, respectively (Table 5).

To further evaluate the performance of GRE4Zn, comparisons were conducted with the testing datasets for TEMSP and CHED,

which respectively are a predictor from our previous study and a popular structure- based predictor of metal binding sites [20,26]. The results are shown in Table 5. It was observed that GRE4Zn achieved a comparable performance with TEMSP under IoUR ≥ 0.5 (Table 5). Although the *Pr* of GRE4Zn was lower (87.84% vs. 95.90%), the *Sn* was much better than TEMSP (97.01% vs. 86.00%) (Table 5). Furthermore, GRE4Zn was better than TEMSP under IoUR = 1.0, since the *Pr* of the two were comparable (81.08% vs. 82.00%) but the *Sn* of GRE4Zn was much better (89.55% vs. 73.50%) (Table 5). The comparison between GRE4Zn and CHED indicated that GRE4Zn was much better. Under IoUR ≥ 0.5 , both the *Pr* and *Sn* of GRE4Zn were higher than CHED (*Pr*: 95.90% vs. 52.61%; *Sn*: 86.00% vs. 82.84%). The comparison under IoUR = 1.0 was consistent with this result (*Pr*: 82.00% vs. 23.70%; *Sn*: 73.50% vs. 37.31%). Taken together, it was demonstrated that GRE4Zn achieved a superior performance in the prediction of zinc-binding sites.

3.4. Large-scale analysis reveals the functional importance of zinc-binding

Since previous studies have suggested that 10% of the human proteome is made up of potential zinc-binding proteins [4], it might be of great value to have a systematic view based on large-scale computational analyses. With the superior GRE4Zn program, large-scale studies of structurally characterized proteins in the PDB database were performed. For example, UPF1, a human RNA-dependent ATPase and 5'-3' RNA helicase, is critical for nonsense-mediated decay [36], the structure of which was characterized in a previous study (PDB: 2WJY) [36]. It was observed that there were three zinc ion binding sites for UPF1, i.e. Cys123-Cys145-His155, Cys137-Ser140-His159 and Cys186-Cys213 (Fig. 3A) [36]. It seems that the first two are 3-res zinc-binding sites, while the third was only partially characterized. However, GRE4Zn suggested that the potential binding model for Cys123-Cys145-His155 was Cys123-Cys126-Cys145-His155 (Fig. 3B). Furthermore, GRE4Zn suggested Cys165 was to be the fourth binding residue

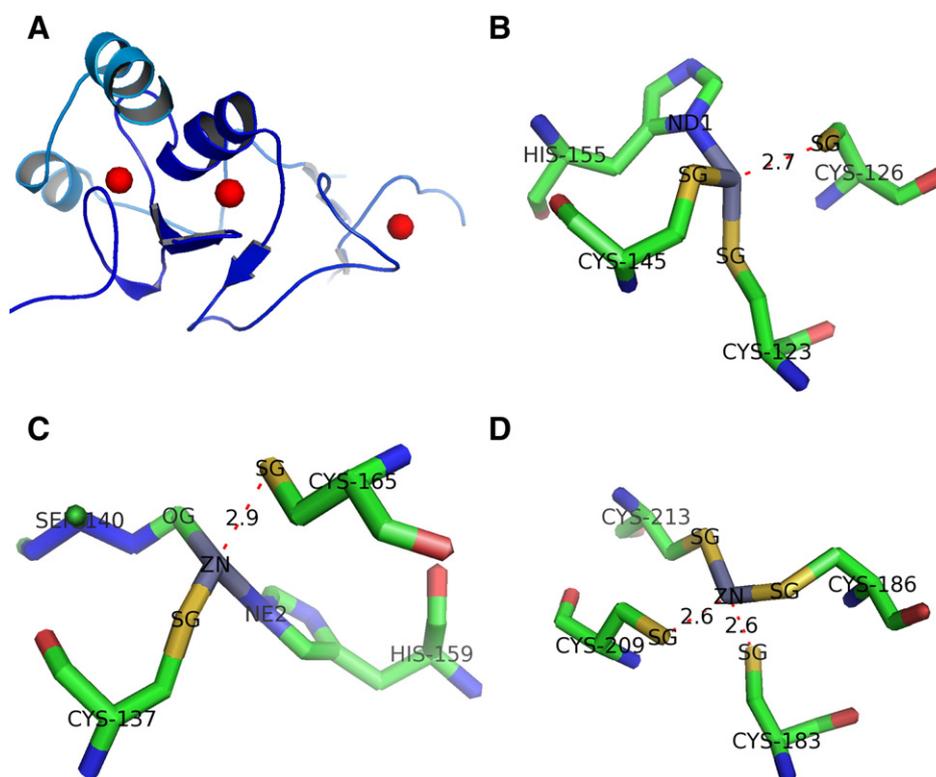


Fig. 3. An example (PDB ID: 2WJY) of zinc-binding prediction using GRE4Zn. (A) An overview of the three zinc ions (red solid spheres) in the protein structure. (B) The predicted 3-res binding model for Cys123-His145-His155. (C) The predicted 4-res binding model for Cys137-Ser140-His159. (D) The predicted 4-res binding model for Cys186-Cys213.

Table 6The most enriched biological processes in the potential zinc-binding proteins in the PDB database for *E. coli*, *S. cerevisiae* and *H. sapiens*, respectively (p -value $< 10^{-7}$).

Description of GO term	Zinc-binding		Proteome		E-ratio ^c	p-value
	Num. ^a	Per. ^b	Num.	Per.		
<i>The most enriched biological processes for zinc-binding protein for Escherichia coli</i>						
Iron ion transport (GO:0006826)	18	4.95%	18	0.45%	10.93	1.36E–19
Carbohydrate metabolic process (GO:0005975)	49	13.46%	147	3.69%	3.64	5.24E–17
Metabolic process (GO:0008152)	76	20.88%	323	8.12%	2.57	3.90E–16
Cellular iron ion homeostasis (GO:0006879)	20	5.49%	32	0.80%	6.83	8.28E–14
Oxidation-reduction process (GO:0055114)	61	16.76%	258	6.48%	2.59	4.37E–13
L-Lyxose metabolic process (GO:0019324)	10	2.75%	10	0.25%	10.93	3.66E–11
L-Arabinose catabolic process to xylulose 5-phosphate (GO:0019569)	9	2.47%	9	0.23%	10.93	4.09E–10
Arabinose catabolic process (GO:0019568)	9	2.47%	9	0.23%	10.93	4.09E–10
TRNA 3'-end processing (GO:0042780)	11	3.02%	16	0.40%	7.52	9.31E–09
Lipopolysaccharide biosynthetic process (GO:0009103)	24	6.59%	74	1.86%	3.55	1.36E–08
Proteolysis (GO:0006508)	26	7.14%	88	2.21%	3.23	2.98E–08
<i>The most enriched biological processes for zinc-binding protein for Saccharomyces cerevisiae</i>						
Transcription from RNA polymerase II promoter (GO:0006366)	220	55.70%	702	35.60%	1.56	4.40E–20
Transcription, DNA-dependent (GO:0006351)	223	56.46%	752	38.13%	1.48	1.01E–16
Glutamine biosynthetic process (GO:0006542)	20	5.06%	20	1.01%	4.99	7.30E–15
Maintenance of transcriptional fidelity during DNA-dependent transcription elongation from RNA polymerase II promoter (GO:0001193)	40	10.13%	62	3.14%	3.22	9.84E–15
Transcription-coupled nucleotide-excision repair (GO:0006283)	40	10.13%	64	3.25%	3.12	4.82E–14
Nitrogen compound metabolic process (GO:0006807)	20	5.06%	22	1.12%	4.54	1.13E–12
<i>The most enriched biological processes for zinc-binding protein for Homo sapiens</i>						
Regulation of ARF GTPase activity (GO:0032312)	16	1.74%	16	0.29%	6.10	2.46E–13
Apoptosis (GO:0006915)	73	7.96%	203	3.63%	2.19	4.41E–12
Protein deacetylation (GO:0006476)	14	1.53%	14	0.25%	6.10	9.40E–12
Central nervous system development (GO:0007417)	13	1.42%	14	0.25%	5.66	6.90E–10
Proteolysis (GO:0006508)	60	6.54%	170	3.04%	2.15	9.18E–10
Superoxide metabolic process (GO:0006801)	10	1.09%	10	0.18%	6.10	1.35E–08

^a The number of proteins annotated.^b The proportion of proteins annotated.^c Enrichment ratio: the proportion in zinc-binding proteins divided by the one in the proteome.

for Cys137-Ser140-His159, even though serine was not considered in GRE4Zn (Fig. 3C). In addition, Cys-183-Cys186-Cys209-Cys-213 was suggested as the zinc-binding sites for Cys186-Cys213 (Fig. 3D). These results are consistent with those in a recent comprehensive study by Andreini et al. [37].

To further characterize the functional roles, organism-specific statistical analyses were performed for the functional annotations of potential zinc-binding proteins. For *E. coli*, the most enriched were basic biological processes such as Iron ion transport/homeostasis (GO:0006826, GO:0006879) and metabolic processes (GO:0005975, GO:0008152) (Table 6). It was observed that the transcription-related processes (GO:0006366, GO:0006351, GO:0001193, GO:0006283) were over-represented (Table 6) in the potential zinc-binding proteins of *S. cerevisiae*. The statistical results for *H. sapiens* suggested that certain complex biological processes, including the regulation of ARF GTPase activity (GO:0032312), apoptosis (GO:0006915) and central nervous system development (GO:0007417) were enriched (Table 6). From the statistical results of the GO terms, it was observed that the most enriched biological processes of potential zinc-binding proteins were different in different species. Furthermore, the results indicated that zinc-binding proteins might have even more complex functions and be involved in more complicated biological processes that enable these proteins to play expanded roles in higher organisms.

Since zinc plays critical biological and physiological roles in organisms, it is implicated in human health [1,38]. The functional analyses of zinc-binding in cancer were conducted by means of a large-scale analysis. The human cancer genes from the Cancer Gene Census of the database Catalogue Of Somatic Mutations In Cancer (COSMIC) [39] were retrieved and then analyzed by GRE4Zn. Utilizing the hypergeometric distribution, it was observed zinc-binding proteins were enriched in cancer genes (Table 7). From the large-scale prediction, it was observed that various cancer genes, such as BRCA1, RB1, CDH1 and DICER1, are zinc-binding proteins. For example, the breast

and ovarian cancer susceptibility gene BRCA1 encodes RING domain which is stabilized by the zinc ion and exhibits ubiquitin ligase activity, while the disruption of zinc-binding by a cancer-predisposing mutation abolished BRCA1 activity [40].

Furthermore, statistical analyses of the KEGG pathway were performed (Table 8) using DAVID [41]. It was observed that zinc-binding proteins were enriched in "Pathways in cancer", "Bladder cancer" and "Small cell lung cancer". In the KEGG "Pathways in cancer", approximately 30 proteins were predicted to be zinc-binding by GRE4Zn (Table 8), such as the tumor suppressor p53, retinoblastoma-associated protein and transforming growth factor beta-1. From the results of this large-scale prediction, approximately 9 proteins in the KEGG pathway of "Bladder cancer", including tumor suppressor p53, matrix metalloproteinase 1 and thrombospondin 1, were also suggested to contain zinc-binding sites. Furthermore, approximately 12 proteins involved in the KEGG pathway of "Small cell lung cancer" were predicted to be zinc-binding, including tumor suppressor p53, TNF receptor-associated factor 2 and X-linked inhibitor of apoptosis. Taken together,

Table 7

The statistical analyses of the enrichment in the zinc-binding proteins in cancer genes and drug targets.

	Zinc-binding		Human PDB		E-ratio ^d	p-value
	Total ^a	Annotated ^b	Total ^c	Annotated		
Cancer genes	1027	26	19651	189	2.63	6.13E–06
Drug targets	1027	50	19651	244	3.92	5.23E–17

^a Total, the number of large-scale predicted zinc-binding proteins.^b Annotated, the number of predicted zinc-binding proteins annotated as cancer genes/drug targets.^c Total, the total number of structurally characterized proteins.^d E-ratio, short for enrichment ratio, the proportion of annotated proteins in the predicted zinc-binding proteins divided by that of in all the human protein structures.

Table 8
The statistical analyses of the over-represented KEGG pathway.

KEGG term	Zinc-binding		p-value	Benjamini ^a
	Num.	Per.		
Purine metabolism	21	4.7%	2.3E–05	3.4E–03
Complement and coagulation cascades	13	2.9%	6.6E–05	5.0E–03
Pathways in cancer	30	6.7%	5.4E–04	2.7E–02
Bladder cancer	9	2%	5.9E–04	2.2E–02
Small cell lung cancer	12	2.7%	1.6E–03	4.9E–02

^a Benjamini, the Benjamini-Hochberg correction, p-value < 0.05.

these analyses indicate that zinc-binding proteins are heavily involved in cancer.

To further dissect the functional importance of zinc-binding, large-scale analyses were performed with GRE4Zn on human drug targets retrieved from the Potential Drug Target Database (PDTP) [42]. The statistical analysis suggested that zinc-binding proteins were enriched in the drug targets (Table 7). For example, previous studies suggested that zinc is important for bone formation and mineralization [38], while bone morphogenetic protein 7 (BMP7) induces cartilage and bone formation [43]. Recently, the TGF-beta family member protein secreted from bone stromal cells BMP7 was found to be critical in tumor dormancy and the recurrence of prostate cancer [44], and was proposed to be a potential drug target. The prediction by GRE4Zn suggested there is a 4-res zinc-binding site of Cys38-Cys71-Cys104-Cys138 in BMP7 (PDB ID: 1BMP) (Fig. 4A). This prediction might provide helpful insight into the molecular mechanism by which BMP7 is involved in cancer signaling. The degradation of extracellular matrix is critical for invasion and metastasis of cancer cells [45,46]. As one of the serine proteinase systems involved in

extracellular matrix degradation, urokinase-type plasminogen activator (uPA) is a potential target for cancer therapy [45,46]. Interestingly, uPA is inhibited by the zinc ion, and this inhibition suppresses the invasion of human urological cancer cells [47]. With GRE4Zn, uPA was predicted to contain the 3-res zinc-binding site Cys42-His57-Cys58 (PDB ID: 1C5W) (Fig. 4B). Since a disulfide bond in Cys42–Cys58 was identified in the protein structure [48], which is critical for protein stability, it is suggested that the zinc ion might inhibit uPA through the formation of zinc-binding sites, which would influence the disulfide bond. Taken together, it is suggested that the prediction of zinc-binding will prove to be very helpful to achieving a better understanding of the molecular mechanism of drug targets.

4. Discussion

The zinc ion is essential for living organisms and zinc-binding is critical for protein function [1–3,5]. The precise identification of zinc-coordinating sites is a key step forward in understanding the biological roles and molecular mechanisms of zinc-binding proteins. Although a great number of studies have contributed to this area, the zinc metallome is still far from being understood [4]. Since experimentally identifying zinc-binding sites is highly labor-intensive, it is of great importance to develop computational approaches to provide the prediction information that is needed for further investigation. Recently, a series of computational studies on zinc-binding were carried out [8–27], and a number of software packages have been developed (Table 1). Although the predictors might be different in sequence-based and structure-based approaches and the algorithms employed are quite diverse, all of the predictors do need to train their model with the dataset. Since training procedures might introduce the

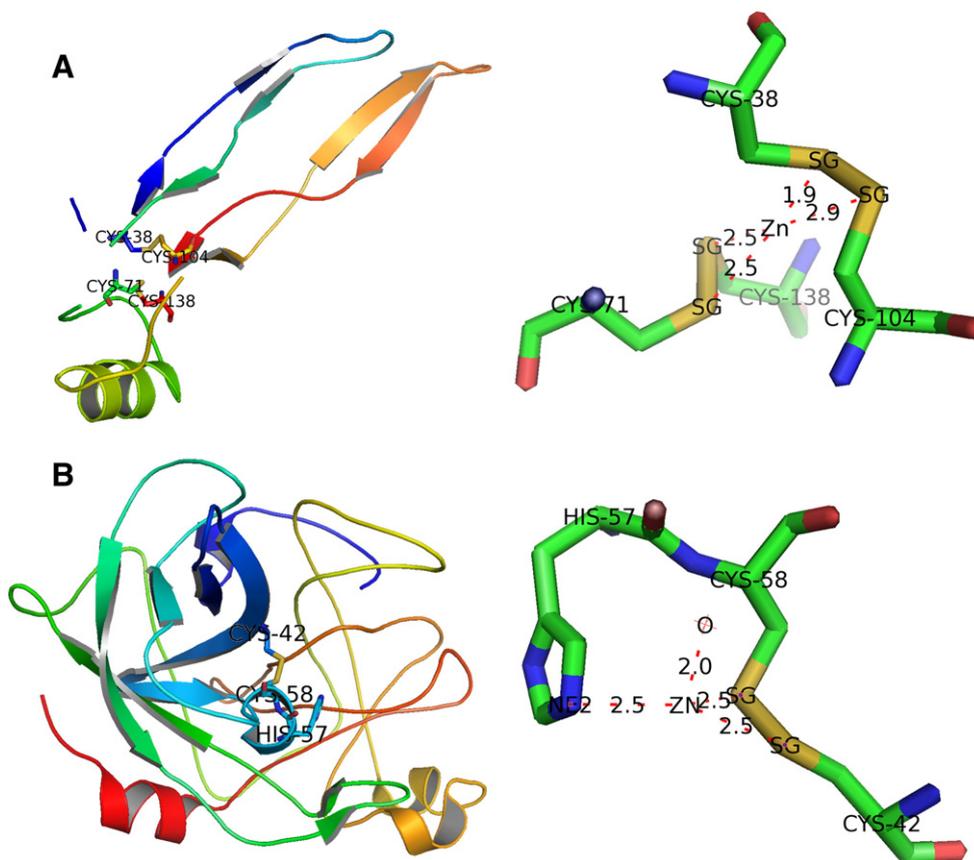


Fig. 4. Prediction of zinc-binding sites in BMP7 and uPA. (A) BMP7. (B) uPA.

problem of over-fitting problem into the prediction effort, it is of great importance to employ a training-independent approach.

In this study, we found that there were a number of differences between the 4-res and 3-res types of zinc-binding. Further analyses of the 4-res and 3-res types of zinc-binding suggested that the geometric features of zinc-binding were sharply constrained. Based on these observations, we proposed the approach of geometric restriction presented in this study, which is both simple and training-independent. The GRE4Zn software was developed, and the performance was demonstrated to be highly promising on the basis of both evaluation and comparison with other studies. Furthermore, since other metals such as iron, copper, calcium are able to bind with protein residues through coordination in a like manner as zinc, it is evident that the coordination geometry is also critical for the binding of these metals [49]. Thus, we suggest that this GRE method may be applied for other metals. However, since the coordination patterns differ for different metals [23,49], the details of the geometrical relationships should be investigated thoroughly and the method should be adapted to different metal bindings.

With the superior program GRE4Zn we have carried out large-scale studies for structurally characterized proteins. The results show that there is a dynamic variation in the functions of zinc-binding across species. Further analyses suggest that zinc-binding proteins are heavily implicated in disease and cancers, so the prediction of zinc-binding sites is useful for obtaining insights into cancer therapy and molecular mechanisms of drug targets. Taken together, we anticipate that computational predictions and analysis, substantiated by experimental investigation, will help advancing the understanding of the molecular functions and mechanisms of the binding of zinc as well as other metals.

Acknowledgements

The authors thank Jichao Wang, Zhangyan Dai, Peng Xiong and Dr. Quan Chen for their helpful suggestions during the studies. This work was supported, in whole or in part, by the National Natural Science Foundation of China (21173203, 81272578 and 31171263), the National Basic Research Program (973 project) (2012CB910101 and 2013CB933900), and the Innovation Foundation of USTC for Young Scientists (WK2070000028). Pacific Edit reviewed the manuscript prior to submission.

References

- [1] H. Haase, L. Rink, Functional significance of zinc-related signaling pathways in immune cells, *Annu. Rev. Nutr.* 29 (2009) 133–152.
- [2] S. Yamasaki, K. Sakata-Sogawa, A. Hasegawa, T. Suzuki, K. Kabu, E. Sato, T. Kurosaki, S. Yamashita, M. Tokunaga, K. Nishida, T. Hirano, Zinc is a novel intracellular second messenger, *J. Cell Biol.* 177 (2007) 637–645.
- [3] J.E. Coleman, Zinc proteins: enzymes, storage proteins, transcription factors, and replication proteins, *Annu. Rev. Biochem.* 61 (1992) 897–946.
- [4] C. Andreini, L. Banci, I. Bertini, A. Rosato, Counting the zinc-proteins encoded in the human genome, *J. Proteome Res.* 5 (2006) 196–201.
- [5] D.S. Auld, Zinc coordination sphere in biochemical zinc sites, *Biometals* 14 (2001) 271–313.
- [6] K. Patel, A. Kumar, S. Durani, Analysis of the structural consensus of the zinc coordination centers of metalloprotein structures, *Biochim. Biophys. Acta* 1774 (2007) 1247–1253.
- [7] J.M. Berg, Zinc finger domains: hypotheses and current knowledge, *Annu. Rev. Biophys. Chem.* 19 (1990) 405–421.
- [8] C. Andreini, I. Bertini, A. Rosato, A hint to search for metalloproteins in gene banks, *Bioinformatics* 20 (2004) 1373–1380.
- [9] C.T. Lin, K.L. Lin, C.H. Yang, I.F. Chung, C.D. Huang, Y.S. Yang, Protein metal binding residue prediction based on neural networks, *Int. J. Neural Syst.* 15 (2005) 71–84.
- [10] H.H. Lin, L.Y. Han, H.L. Zhang, C.J. Zheng, B. Xie, Z.W. Cao, Y.Z. Chen, Prediction of the functional class of metal-binding proteins from sequence derived physico-chemical properties by support vector machine approach, *BMC Bioinforma.* 7 (Suppl. 5) (2006) S13.
- [11] A. Passerini, M. Punta, A. Ceroni, B. Rost, P. Frasconi, Identifying cysteines and histidines in transition-metal-binding sites using support vector machines and neural networks, *Proteins* 65 (2006) 305–316.
- [12] A. Passerini, C. Andreini, S. Menchetti, A. Rosato, P. Frasconi, Predicting zinc binding at the proteome level, *BMC Bioinforma.* 8 (2007) 39.
- [13] M. Lippi, A. Passerini, M. Punta, B. Rost, P. Frasconi, MetalDetector: a web server for predicting metal-binding sites and disulfide bridges in proteins from sequence, *Bioinformatics* 24 (2008) 2094–2095.
- [14] N. Shu, T. Zhou, S. Hovmöller, Prediction of zinc-binding sites in proteins from sequence, *Bioinformatics* 24 (2008) 775–782.
- [15] C. Andreini, I. Bertini, A. Rosato, Metalloproteomes: a bioinformatic approach, *Acc. Chem. Res.* 42 (2009) 1471–1479.
- [16] I. Bertini, L. Decaria, A. Rosato, The annotation of full zinc proteomes, *J. Biol. Inorg. Chem.* 15 (2010) 1071–1078.
- [17] A. Passerini, M. Lippi, P. Frasconi, MetalDetector v2.0: predicting the geometry of metal binding sites from protein sequence, *Nucleic Acids Res.* 39 (2011) W288–W292.
- [18] J.S. Sodhi, K. Bryson, L.J. McGuffin, J.J. Ward, L. Wernisch, D.T. Jones, Predicting metal-binding site residues in low-resolution structural models, *J. Mol. Biol.* 342 (2004) 307–320.
- [19] J.W. Schymkowitz, F. Rousseau, I.C. Martins, J. Ferkinghoff-Borg, F. Stricher, L. Serrano, Prediction of water and metal binding sites and their affinities by using the Fold-X force field, *Proc. Natl. Acad. Sci. U. S. A.* 102 (2005) 10147–10152.
- [20] M. Babor, S. Gerzon, B. Raveh, V. Sobolev, M. Edelman, Prediction of transition metal-binding sites from apo protein structures, *Proteins* 70 (2008) 208–217.
- [21] A.J. Bordner, Predicting small ligand binding sites in proteins using backbone structure, *Bioinformatics* 24 (2008) 2865–2871.
- [22] J.C. Ebert, R.B. Altman, Robust recognition of zinc binding sites in proteins, *Protein Sci.* 17 (2008) 54–65.
- [23] K. Goyal, S.C. Mande, Exploiting 3D structural templates for detection of metal-binding sites in protein structures, *Proteins* 70 (2008) 1206–1218.
- [24] R. Levy, M. Edelman, V. Sobolev, Prediction of 3D metal binding sites from translated gene sequences based on remote-homology templates, *Proteins* 76 (2009) 365–374.
- [25] C. Wang, R. Vernon, O. Lange, M. Tyka, D. Baker, Prediction of structures of zinc-binding proteins through explicit modeling of metal coordination geometry, *Protein Sci.* 19 (2010) 494–506.
- [26] W. Zhao, M. Xu, Z. Liang, B. Ding, L. Niu, H. Liu, M. Teng, Structure-based de novo prediction of zinc-binding sites in proteins of unknown function, *Bioinformatics* 27 (2011) 1262–1268.
- [27] C.H. Lu, Y.F. Lin, J.J. Lin, C.S. Yu, Prediction of metal ion-binding sites in proteins using the fragment transformation method, *PLoS One* 7 (2012) e39252.
- [28] C. Zheng, M. Wang, K. Takemoto, T. Akutsu, Z. Zhang, J. Song, An integrative computational framework based on a two-step random forest algorithm improves prediction of zinc-binding sites in proteins, *PLoS One* 7 (2012) e49716.
- [29] P.W. Rose, B. Beran, C. Bi, W.F. Bluhm, D. Dimitropoulos, D.S. Goodsell, A. Pric, M. Quesada, G.B. Quinn, J.D. Westbrook, J. Young, B. Yukich, C. Zardecki, H.M. Berman, P.E. Bourne, The RCSB protein data bank: redesigned web site and web services, *Nucleic Acids Res.* 39 (2011) D392–D401.
- [30] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT suite: a web server for clustering and comparing biological sequences, *Bioinformatics* 26 (2010) 680–682.
- [31] Z. Liu, J. Cao, Q. Ma, X. Gao, J. Ren, Y. Xue, GPS-YNO2: computational prediction of tyrosine nitration sites in proteins, *Mol. Biosyst.* 7 (2011) 1197–1204.
- [32] D. Barrell, E. Dimmer, R.P. Huntley, D. Binns, C. O'Donovan, R. Apweiler, The GOA database in 2009—an integrated gene ontology annotation resource, *Nucleic Acids Res.* 37 (2009) D396–D403.
- [33] A. Andreeva, D. Howorth, J.M. Chandonia, S.E. Brenner, T.J. Hubbard, C. Chothia, A.G. Murzin, Data growth and its impact on the SCOP database: new developments, *Nucleic Acids Res.* 36 (2008) D419–D425.
- [34] L.L.C. Schrodinger, The PyMOL Molecular Graphics System, Version 1.3r1, 2010.
- [35] C. Andreini, I. Bertini, G. Cavallaro, Minimal functional sites allow a classification of zinc sites in proteins, *PLoS One* 6 (2011) e26325.
- [36] M. Clerici, A. Mourao, I. Gutsche, N.H. Gehring, M.W. Hentze, A. Kulozik, J. Kadlec, M. Sattler, S. Cusack, Unusual bipartite mode of interaction between the nonsense-mediated decay factors, UPF1 and UPF2, *EMBO J.* 28 (2009) 2293–2306.
- [37] C. Andreini, G. Cavallaro, S. Lorenzini, A. Rosato, MetalPDB: a database of metal sites in biological macromolecular structures, *Nucleic Acids Res.* 41 (2013) D312–D319.
- [38] C.T. Chasapis, A.C. Loutsidou, C.A. Spiliopoulou, M.E. Stefanidou, Zinc and human health: an update, *Arch. Toxicol.* 86 (2012) 521–534.
- [39] S.A. Forbes, G. Tang, N. Bindal, S. Bamford, E. Dawson, C. Cole, C.Y. Kok, M. Jia, R. Ewing, A. Menzies, J.W. Teague, M.R. Stratton, P.A. Futreal, COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer, *Nucleic Acids Res.* 38 (2010) D652–D657.
- [40] P.S. Brzovic, J.E. Meza, M.C. King, R.E. Kleivit, BRCA1 RING domain cancer-predisposing mutations. Structural consequences and effects on protein-protein interactions, *J. Biol. Chem.* 276 (2001) 41399–41406.
- [41] W. da Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* 4 (2009) 44–57.
- [42] Z. Gao, H. Li, H. Zhang, X. Liu, L. Kang, X. Luo, W. Zhu, K. Chen, X. Wang, H. Jiang, PDTD: a web-accessible protein database for drug target identification, *BMC Bioinforma.* 9 (2008) 104.
- [43] D.L. Griffith, P.C. Keck, T.K. Sampath, D.C. Rueger, W.D. Carlson, Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor beta superfamily, *Proc. Natl. Acad. Sci. U. S. A.* 93 (1996) 878–883.
- [44] A. Kobayashi, H. Okuda, F. Xing, P.R. Pandey, M. Watabe, S. Hirota, S.K. Pai, W. Liu, K. Fukuda, C. Chambers, A. Wilber, K. Watabe, Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone, *J. Exp. Med.* 208 (2011) 2641–2655.
- [45] A.H. Mekaw, D.L. Morris, M.H. Pourgholami, Urokinase plasminogen activator system as a potential target for cancer therapy, *Future Oncol.* 5 (2009) 1487–1499.

- [46] R. Hildenbrand, H. Allgayer, A. Marx, P. Stroebel, Modulators of the urokinase-type plasminogen activation system for cancer, *Expert Opin. Investig. Drugs* 19 (2010) 641–652.
- [47] K. Ishii, S. Usui, Y. Sugimura, H. Yamamoto, K. Yoshikawa, K. Hirano, Inhibition of aminopeptidase N (AP-N) and urokinase-type plasminogen activator (uPA) by zinc suppresses the invasion activity in human urological cancer cells, *Biol. Pharm. Bull.* 24 (2001) 226–230.
- [48] B.A. Katz, R. Mackman, C. Luong, K. Radika, A. Martelli, P.A. Sprengeler, J. Wang, H. Chan, L. Wong, Structural basis for selectivity of a small molecule, S1-binding, submicromolar inhibitor of urokinase-type plasminogen activator, *Chem. Biol.* 7 (2000) 299–312.
- [49] T. Dudev, C. Lim, Metal binding affinity and selectivity in metalloproteins: insights from computational studies, *Annu. Rev. Biophys.* 37 (2008) 97–116.