Classification of mononuclear zinc metal sites in protein structures

SAMUEL KARLIN[†] AND ZHAN-YANG ZHU

Department of Mathematics, Stanford University, Stanford, CA 94305-2125

Contributed by Samuel Karlin, October 3, 1997

ABSTRACT Our study of the extended metal environment, particularly of the second shell, focuses in this paper on zinc sites. Key findings include: (i) The second shell of mononuclear zinc centers is generally more polar than hydrophobic and prominently features charged residues engaged in an abundance of hydrogen bonding with histidine ligands. Histidine–acidic or histidine–tyrosine clusters commonly overlap the environment of zinc ions. (ii) Histidine tautomeric metal bonding patterns in ligating zinc ions are mixed. For example, carboxypeptidase A, thermolysin, and sonic hedgehog possess the same ligand group (two histidines, one unidentate acidic ligand, and a bound water), but their histidine tautomeric geometries markedly differ such that the carboxypeptidase A makes only N^{δ1} contacts, thermolysin makes only N^{ε2} contacts, and sonic hedgehog uses one of each. Thus the presence of a similar ligand cohort does not necessarily imply the same topology or function at the active site. (iii) Two close histidine ligands HX_mH, $m \leq 5$, rarely both coordinate a single metal ion in the N^{δ1} tautomeric conformation, presumably to avoid steric conflicts. Mononuclear zinc sites can be classified into six types depending on the ligand composition and geometry. Implications of the results are discussed in terms of divergent and convergent evolution.

This paper highlights similarities and dissimilarities of the extended metal environment for Zn metal sites. For definitions, methods, and corresponding results on Cu, Fe, and Mn metal centers, see ref. 1. In a protein structure with one or more metal cofactors, we defined and analyzed the extended metal environments, emphasizing three layers of interactions: the metal core, the ligand group, and the second shell, which consists of all residues within 3.5-Å distance to some ligand (see legend to Table 1). We further identified distinctive residue clusters of the protein structures often overlapping the metal environment. For example, histidine–cysteine–methionine clusters are common in the neighborhood of type I copper sites (1); histidine–tyrosine clusters are prominent in many cases about iron metal sites (1).

The Classification of Zinc Metal Sites

Zinc serves structural, chemical, and regulatory roles in biological systems and is an essential ingredient of the active site in many protease and phosphatase proteins (3–5). Lipscomb and Sträter (5) review families, motifs, and enzymology of protein structures containing one or more zinc metal atoms. Table 1 exhibits the ligand groups and second-shell residues of a representative collection of available protein structures containing one or more mononuclear Zn²⁺ ions. The principal ligands coordinating Zn ions comprise combinations of H (histidine), acidic (D and/or E), and C (cysteine) residues, water ligands, and sometimes the residues Y, N, S, and T. The ligand composition and geometry suggest six natural classes (Table 1): *class I*, ligand group involving at least three histidine

residues, which share an elongated zinc binding motif HXXX-HXXGXXH (E and G are not ligands) (5); *class II*, ligand arrays that feature a proximal histidine pair HXH and a third histidine rather distant in the sequence; *class III*, combinations of H and C ligands; *class IV*, two separated H ligands, an acidic unidentate plus a bound water molecule; *class V*, predominantly acidic ligands; and *class VI*, other ligand compositions.

Histidine Ligand Tautomeric Conformation. See also ref. 1 for a discussion of histidine tautomeric patterns for copper and iron proteins. The tautomeric preferences of histidine ligands of a mononuclear zinc ion are as follows: (i) A ligand group of three H abiding by the sequence motif HX₃HX₅H invariably invokes N^{ε2} contacts. (ii) A ligand group containing three H residues in the arrangement HXH and a distant H adopts for the proximal histidines the conformation N^{ε2}, whereas the distant ligand acquires about equally the N^{ε2} or N^{δ1} conformation, with the exception of the metallo-β-lactamase (1bmc) structure. (iii) A ligand group including precisely two H, HX_mH, $m \leq 4$, bonds to the metal via the N^{ε2} contact. (iv) A ligand group of two H, from HX_mH, $m \geq 5$ adopts about equally the tautomeric alternatives: both N^{ε2}, both N^{δ1}, or mixed with one N^{ε2} and one N^{δ1}. (v) The ligand group with a single H generally favors ligation via N^{δ1}. There are variants on the foregoing coordination geometry: the two zinc histidine ligands of lysozyme (1lba) both ligate via N^{δ1} and similarly for carboxypeptidase A (2ctc). The tumor suppressor p53 protein also adopts the N^{δ1} atom for its histidine contact. The bond lengths of histidine ligands to the Zn ion are generally in the tight range of 1.95–2.10 Å, with the closest atom mostly attained by the N^{ε2} nitrogen attachment.

Second Shell. A prominent feature of the second shell of a mononuclear zinc environment is that of charged residues, which generally engage ligand residues via hydrogen bonds. The second shell tends to contain more polar residues than nonpolar residues and often by a factor of at least 2. The second shell involves more buried than exposed residues (about 1/3 are solvent exposed).

There are two separated Zn ions in each chain of the tramtrack protein, whose second shell is largely exposed. The ligands for Zn-A171 and Zn-A172 show exactly the same ligand group and identical spacings in the primary sequence. These correspond to two tandem zinc finger motifs where, apart from the two C residues and two H residues, there is hardly any conservation in the intervening sequences. It is tantalizing that these two zinc metal centers are dramatically deviant in the second shell. The first zinc atom has a preferred polar second shell, whereas the second zinc atom in its second shell is devoid of polar residues but contains a surfeit of hydrophobic residues (Table 1).

Diverse Statistically Significant 3D Residue Clusters in Zinc Protein Structures. For background on statistically significant residue clusters, see the preceding paper (1). Many of the zinc structures possess histidine {H} and/or cysteine–histidine {CH} and/or histidine–acidic {HED} residue clusters generally overlapping the zinc metal environments (Table 2). Aminopeptidase (1lcp) contains a significant mixed charge cluster {KRED} at the interface just reaching to the zinc-490.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1997 by The National Academy of Sciences 0027-8424/97/9414231-6\$2.00/0
PNAS is available online at <http://www.pnas.org>.

[†]To whom reprint requests should be addressed. e-mail: fd.zgg@forsythe.stanford.edu.

Table 1. Single Zn ions in protein three-dimensional (3D) structures

Ligands	Bond Length	Bond Atom	Second Shell: residues, hydrophobic Φ vs hydrophilic Π, secondary structure, side-chain solvent accessibility		
Class I (H X ₁ , H X ₂ , H, Y):					
1. 1atl (Zn-A401): Atrolysin C (<i>Crotalus atrox</i>): Two chains (A & B), 200aa each chain; structural class: α+β; (SLE: 2-(thiomethylene-4-methylpentanoic acid)					
H-A142	2.02	N ₆₂	V-A138, T-A139, I-A165, R-A167	Φ={1I, 1V, 1L, 2M}	
H-A146	2.07	N ₆₂	E-A143, M-A150, M-A166	Π={1R, 1E, 1T} Others={1G, 1P, 1Wat, 1 CH ₃ , 1 SLE}	
H-A152	2.05	N ₆₂	M-A166, Wat-953	3α, 0β, 7 coil	
Y-C302	2.17, 2.14	O, OXT	E-A143, R-A167, P-A168, G-A169, L-A170, SLE-C301, Methyl Group-C302	7 buried, 3 exposed	
2. 1mnc: Neutrophil Collagenase (human); Single chain, 158aa; α+β; (PLH: Methylamino-Phenylalanyl-Leucyl-Hydroxamate)					
Zn	H-218 1.98	N ₆₂	E-219, L-235, M-236, Y-237	Φ={1I, 3L, 1M, 1A, 2Y}	
-281	H-222 2.04	N ₆₂	Q-186, E-219, L-226	Π={1E, 1Q, 1N}	
	H-228 2.05	N ₆₂	M-236	Others={1G, 1P, 2Wat}	
	PLH-280 2.84, 2.29, 2.11	N ₁ , O ₁ , O ₂	G-179, I-180, L-181, A-182, E-219, P-238, N-239, Y-240, Wat-58, Wat-64	1α, 1β, 11 coil 7 buried, 6 exposed	
Zn	H-168 1.86	N ₆₂	Y-164, R-166, S-172, F-174	Φ={2F, 1A, 1Y}	
-280	D-170 1.92	O ₆₂	S-172, F-185, Wat-43	Π={1R, 1E, 1S} Others={4Wat}	
	H-183 1.83	N ₆₂	S-172, E-219, Wat-46	1α, 2β, 4 coil	
	H-196 1.96	N ₆₁	F-174, A-182, Wat-21, Wat-42	3 buried, 4 exposed	
3. 1sat (Zn-472): Serratia Protease (<i>Serratia marcescens</i>): Single chain, 468aa; α+β					
H-176	2.07	N ₆₂	Q-172, T-173, E-177, L-213, M-214	Φ={3L, 1M}	
H-180	2.13	N ₆₂	L-138, E-177, L-184, M-214, Wat-511	Π={2R, 1Q, 1E, 1T}	
H-186	1.99	N ₆₂	R-41, M-214, Wat-516, Wat-617	Others={4Wat}	
Y-216	2.77	OH	R-209, Wat-612	3α, 0β, 6 coil	
Wat-510	1.84	O	E-177	6 buried, 3 exposed	
4. 1iae (Ni-999): Astacin with Zn replaced by Nickel (<i>Astacus astacus</i> L.): Single chain, 200aa; α/β					
H-92	2.07	N ₆₂	Q-88, T-89, E-93, I-146	Φ={1I, 1F, 1M}	
H-96	2.18	N ₆₂	E-93, F-100, M-147	Π={1E, 1H, 1S, 2T}	
H-102	2.09	N ₆₂	H-104, T-105, M-147, Wat-('331, 426, '437)	Others={1G, 5Wat}	
Y-149	2.27	OH	H-104, S-153, Wat-328, Wat-330	3α, 0β, 6 coil	
Wat-300	2.30	O	E-93	9 buried, 0 exposed	
Wat-600	2.22	O			
Class II (3H-HXH, distant H):					
5. 1add (Zn-400): Adenosine Deaminase (mouse): Single chain, 349aa; α/β					
H-15	2.24	N ₆₂	E-260, N-293	Φ={1L, 1A}	
H-17	2.03	N ₆₂	IDA-353 (1DA: 1-Deaza-Adenosine)	Π={2R, 1D, 1E, 1N, 1H, 1S}	
H-214	2.05	N ₆₂	R-101, R-156, D-181, L-182, A-183	Others={1DA}	
D-295	2.27	O ₆₁	H-238, S-265, IDA-353	1α, 6β, 2 coil	
Wat-461	1.80	O	H-238, IDA-353	9 buried, 0 exposed	
6. 2cba (Zn-262): Carbonic Anhydrase II (human): Single chain, 258aa; all β					
H-94	2.10	N ₆₂	F-66, N-67, Q-92, L-118, Wat-359	Φ={2F, 1L, 1A} Π={1Q, 2N, 2E, 1T} Others={1Wat}	
H-96	2.12	N ₆₂	E-106, A-116, N-244	0α, 7β, 3 coil	
H-119	2.11	N ₆₁	F-93, E-106, E-117	8 buried, 2 exposed	
Wat-263	2.05	O	T-199		
7. 1ton (Zn-200): Tonin (Rat); Single chain, 228aa; all β					
H-57	2.04	N ₆₂	Y-94, D-102, Wat-92	Φ={1I, 1L}; Π={1D, 1S, 1Y}	
H-97	2.05	N ₆₂	Y-94, I-95, P-95A, L-95B	Others={1P, 1Wat}; 0α, 0β, 6 coil	
H-99	2.05	N ₆₂	D-102, S-214	3 buried, 3 exposed	
8. 1bmc (Zn-228): Metallo-β-Lactamase (<i>Bacillus cereus</i>): Single chain, 213aa; α/β					
H-86	2.47	N ₆₂	E-91, D-152	Φ={1A, 1C} Π={1R, 2D, 1N}	
H-88	2.38	N ₆₁	D-183	1α, 0β, 5 coil	
H-149	2.37	N ₆₂	N-153, C-168, A-184	5 buried, 1 exposed	
Class III (combination of C, H & Acidic residues):					
9. 1clc (Zn-594): Endoglucanase (<i>Clostridium thermocellum</i>): Single chain, 541aa; α+β					
C-155	2.35	S _Y	G-195, Y-250, R-289, Wat-617	Φ={1L, 1Y} Π={1R, 1D, 1E, 1Y}	
C-173	2.20	S _Y	P-282	Others={1G, 1P, 1Wat}	
H-174	2.26	N ₆₁	D-177, E-285, R-289	1α, 0β, 7 coil	
H-197	2.29	N ₆₂	L-153, Y-202, Wat-639	7 buried, 1 exposed	
10. 1tup (Zn-A951): Tumor Suppressor P53 (human): Three identical chains (A, B & C), 195aa each chain; all β					
C-A176	2.33	S _Y	E-A180, Wat-2123	Φ={1V, 1L, 1M, 1C},	
H-A179	1.99	N ₆₁	R-A175, C-A182	Π={1R, 1E, 1N} Others={2Wat}	
C-A238	2.28	S _Y	L-A194, M-A246, V-A274, Wat-2139	1α, 1β, 5 coil	
C-A242	2.23	S _Y	N-A239	4 buried, 3 exposed	
11. 1gpc (Zn-1): Core GP32 (Bacteriophage T4): Single chain, 218aa; all β					
H-64	2.69	N ₆₂	V-89, V-109, R-111	Φ={2V, 1Y}, Π={1R, 2S, 1T}	
C-77	2.24	S _Y	S-79, T-80	Others={1Wat}	
C-87	2.14	S _Y	Wat-29	3α, 0β, 4 coil	
C-90	2.21	S _Y	Y-84, S-94	4 buried, 3 exposed	
12. 1thb: Beta 3 Alcohol Dehydrogenase (human): Two identical chains (A&B), 374aa each chain; α/β; (PYZ: 4-iodopyrazole)					
Zn-	C-A97 2.38	S _Y	K-A99, K-A113	Φ={none} Π={4K, 1N}	
A375	C-A100 2.28	S _Y	K-A104	Others={1G, 1P, 1Wat}	
	C-A103 2.26	S _Y	G-A98, K-A323, Wat-A482	0α, 0β, 7 coil	
	C-A111 2.20	S _Y	P-A95, K-A113, N-A114	1 buried, 6 exposed	
Zn-	C-A46 2.26	S _Y	D-A49	Φ={1F}, Π={1D, 4T, 1S}	
	H-A67 2.10	N ₆₂	D-A49, T-A143, T-A145, Wat-A463	Others={1 Wat}	
	A376	C-A174 2.16	S _Y	F-A93, S-A177, T-A178	4α, 0β, 3 coil
		PYZ 2.30	N ₁	T-A48, F-A93	7 buried, 0 exposed
13. 8atc (Zn-B154): Aspartate Carbamoyltransferase (<i>E. coli</i>) Four chains: A & C (identical) 310aa, α/β; B & D (identical): 146aa; α+β					
C-B109	2.06	S _Y	H-B117, Wat-B162	Φ={none}, Π={2N, 1H, 1S}	
C-B114	2.05	S _Y	N-B111, S-B116, H-B117	Others={2Wat}	
C-B138	2.04	S _Y	Wat-B178	2α, 0β, 2 coil	
C-B141	2.06	S _Y	N-A132, N-B111	2 buried, 2 exposed	

Table 1. (continued)

Ligands	Bond Length	Bond Atom	Second Shell: residues, hydrophobic Φ vs hydrophilic Π , secondary structure, side-chain solvent accessibility
14. 1lba (Zn-151): Lysozyme Mutant A6K, Del (2-5) (Bacteriophage T7); Single chain, 146aa; α/β			
<u>H</u> -17	2.52	N ₆₁	<u>H</u> -68, <u>V</u> -79
<u>H</u> -122	2.67	N ₆₁	A-126, <u>Wat</u> -202
C-130	2.15	S ₇	<u>H</u> -123
<u>Wat</u> -199	2.83	O	<u>Y</u> -46
Φ =(1V, 1A) Π =(2H, 1 ⁺ Y) Others={1Wat}			
0 α , 1 β , 4 coil 3 buried, 2 exposed			
15. 2drp: Tramtrack Protein (<i>Drosophila melanogaster</i>); Two identical chains (A&D) 65aa each chain; small protein			
Zn- <u>C</u> -A113	2.28	S ₇	<u>V</u> -A115
<u>C</u> -A116	2.27	S ₇	R-A118
A171 <u>H</u> -A129	1.99	N ₆₂	<u>Y</u> -A120, <u>N</u> -A125, <u>F</u> -A126, <u>T</u> -A132, <u>S</u> -A133
<u>H</u> -A134	1.96	N ₆₂	V-A115, <u>Y</u> -A130
Φ =(1F, 1V, 1Y), Π =(1R, 1N, 1S, 1T, 1 ⁺ Y), Others={none}			
5 α , 1 β , 2 coil 2 buried, 6 exposed			
Zn- <u>C</u> -A143	2.29	S ₇	P-A144, <u>F</u> -A145, <u>F</u> -A147
<u>C</u> -A146	2.28	S ₇	
A172 <u>H</u> -A159	2.02	N ₆₂	<u>F</u> -A150, <u>M</u> -A156, <u>I</u> -A162, <u>I</u> -A163
<u>H</u> -A164	1.99	N ₆₂	V-A160
Φ =(2I, 3F, 1V, 1M), Π =(none) Others={1P}			
4 α , 1 β , 3 coil 1 buried, 7 exposed			
16. 1ctt (Zn-296): Cytidine Deaminase (<i>E. coli</i>); Single chain, 294aa; α/β; (DHZ: 3,4-dihydrozebularine)			
<u>H</u> -102	2.02	N ₆₁	<u>Q</u> -105, <u>S</u> -106, <u>H</u> -131, <u>DH</u> Z-295
C-129	2.42	S ₇	<u>H</u> -131, R-133
<u>C</u> -132	2.11	S ₇	<u>E</u> -104, <u>F</u> -135, <u>M</u> -136
<u>Wat</u> -700	1.84	O	<u>E</u> -104, <u>DH</u> Z-295
Φ =(1F, 1M), Π =(1R, 1E, 1Q, 1H, 1S) Others={1DHZ}			
7 α , 0 β , 0 coil 5 buried, 2 exposed			
Class IV (2H, 1 D or E, 1 water; the D or E is an unidentate):			
17. 1glc (Zn-4): Glycerol Kinase (Bacteriophage T4); Two chains, F: 161aa, G: 489aa; α/β			
<u>H</u> -F75	2.44	N ₆₂	F-F71, <u>A</u> -F76, <u>F</u> -F91, <u>T</u> -F95, <u>V</u> -F96
<u>H</u> -F90	2.21	N ₆₂	<u>G</u> -F92, <u>P</u> -F137
E-G478	3.15, 1.97	O ₆₁ , O ₆₂	F-F71, V-F96, E-G475, <u>T</u> -G477, <u>Y</u> -G481, R-G482
<u>Wat</u> -1	2.07	O	V-F96
Φ =(2F, 1V, 1A, 1Y) Π =(1R, 1E, 2T), Others={1G, 1P}			
2 α , 1 β , 8 coil 6 buried, 5 exposed			
18. 1vhh (Zn-400): Sonic Hedgehog (<i>Mus musculus</i>); Single chain, 152aa; α/β			
<u>H</u> -141	2.06	N ₆₂	<u>G</u> -128, <u>W</u> -129, H-135, <u>S</u> -139, <u>A</u> -146
<u>D</u> -148	1.97, 2.81	O ₆₁ , O ₆₂	<u>T</u> -126, <u>E</u> -127, <u>Wat</u> -315
<u>H</u> -183	2.08	N ₆₁	<u>E</u> -54, <u>E</u> -177, <u>Wat</u> -212
<u>Wat</u> -201	2.05	O	<u>E</u> -177
Φ =(1W, 1A), Π =(3E, 1H, 1S, 1T) Others={1G, 2Wat}			
0 α , 4 β , 5 coil 4 buried, 5 exposed			
19. 2ctc (Zn-308): Carboxypeptidase (<i>Bos taurus</i>); Single chain, 307aa; α/β			
<u>H</u> -69	2.08	N ₆₁	<u>R</u> -127, <u>D</u> -142, LOF-309, <u>Wat</u> -3
<u>H</u> -72	2.31, 2.25	O ₆₁ , O ₆₂	<u>T</u> -75, <u>Q</u> -76, <u>S</u> -197, <u>Wat</u> -(4, '53, '80)
<u>H</u> -196	2.13	N ₆₁	<u>I</u> -68, <u>T</u> -75, <u>T</u> -268, <u>E</u> -270, <u>L</u> -271, <u>Wat</u> -20
<u>Wat</u> -29	2.00	O	<u>LOF</u> (L-Phenyl Lactate)-309
Φ =(1I, 1L), Π =(1R, 1D, 1E, 1Q, 1S, 2T) Others={5Wat, 1LOF}			
2 α , 2 β , 5 coil 9 buried, 0 exposed			
20. 8thn (Zn-321): Thermolysin (<i>Bacillus thermoproteolyticus</i>); Single chain (E), 316aa; α/β			
<u>H</u> -E142	1.93	N ₆₂	<u>D</u> -E138, <u>V</u> -E139, <u>S</u> -E169, <u>D</u> -E170
<u>H</u> -E146	1.97	N ₆₂	<u>E</u> -E143, <u>D</u> -E150, <u>N</u> -E165, <u>S</u> -E169, <u>Wat</u> -(334, 362, '538)
<u>E</u> -E166	2.79, 1.88	O ₆₁ , O ₆₂	<u>Y</u> -E157, <u>V</u> -E230, <u>H</u> -E231, <u>Wat</u> -395
<u>Wat</u> -392	2.16	O	<u>Y</u> -E157, <u>H</u> -E231
Φ =(2V, 1Y), Π =(1E, 3D, 1N, 1H, 1S), Others={4Wat}			
7 α , 0 β , 3 coil 7 buried, 3 exposed			
Class V (mostly acidic residues):			
21. 1frp (Zn-5): Fructose-1,6-bisphosphatase (pig kidney); Two identical chains (A&B), 321aa each chain; α/β			
<u>D</u> -A118	2.39	O ₆₂	E-A97, <u>P</u> -A119, <u>T</u> -A134, <u>Wat</u> -251
<u>D</u> -A121	2.36	O ₆₁	<u>M</u> -A248, <u>V</u> -A249
<u>E</u> -A280	2.42	O ₆₂	<u>L</u> -A275, <u>R</u> -A276, <u>M</u> -A284
FDP-1	2.52	O ₁	<u>S</u> -A123, <u>S</u> -A124, <u>N</u> -A212, <u>Y</u> -A215, <u>Y</u> -A244, <u>G</u> -A246, <u>M</u> -A248, <u>Y</u> -A264, <u>K</u> -A274, <u>R</u> -A276, <u>R</u> -B243, <u>Wat</u> -(155, '202, '256, '258, '299, '301)
Fructose-2,6-bisphosphate			
<u>Wat</u> -7	2.47	O	
Φ =(1V, 1L, 2M) Π =(2R, 1K, 1N, 1E, 2S, 1T, 3 ⁺ Y) Others={1G, 1P, 7Wat}			
3 α , 3 β , 11 coil 11 buried, 6 exposed			
22. 4enl (Zn-438): Enolase (Baker's yeast); Single chain, 436aa; α/β			
<u>D</u> -246	2.15	O ₆₂	<u>E</u> -168, <u>A</u> -248, <u>K</u> -396
<u>E</u> -295	1.95	O ₆₂	<u>P</u> -297, <u>L</u> -343, <u>K</u> -396, <u>Wat</u> -133
<u>D</u> -320	2.06	O ₆₂	<u>D</u> -296, <u>T</u> -323, <u>L</u> -343, <u>K</u> -345, <u>Wat</u> -(21, '204, '228)
<u>Wat</u> -3	2.36	O	<u>E</u> -168, <u>K</u> -396
<u>Wat</u> -5	2.03	O	<u>D</u> -296
Φ =(1L, 1A) Π =(2K, 1D, 1E, 1T) Others={1P, 4Wat}			
0 α , 3 β , 5 coil 8 buried, 0 exposed			
Others:			
23. 1thj (Zn-A214): Carbonic Anhydrase (<i>Methanosarcina thermophila</i>); Trimer (A, B & C), 213aa each chain; α/β (mainly β)			
<u>H</u> -A81	2.06	N ₆₁	<u>R</u> -A59, <u>D</u> -A61, <u>E</u> -A62
<u>H</u> -A122	2.05	N ₆₂	<u>A</u> -A82, <u>A</u> -A125
<u>H</u> -C117	2.15	N ₆₂	<u>Q</u> -A120, <u>Q</u> -C75, <u>D</u> -C76
<u>Wat</u> -215	1.99	O	<u>Q</u> -C75
Φ =(2A), Π =(1R, 2D, 1E, 2Q) Others={none}			
0 α , 3 β , 5 coil 8 buried, 0 exposed			
24. 1lcp (Zn-A490): Leucine Aminopeptidase (Bovine); Two identical chains (A&B), 484aa each chain; α/β			
<u>L</u> -A170	2.80	O	<u>L</u> -A166, <u>M</u> -A274
<u>M</u> -A171	2.85	O	<u>A</u> -A167
<u>T</u> -A173	2.67	O	<u>R</u> -A169, <u>P</u> -A174, <u>Wat</u> -189
<u>R</u> -A271	2.71	O	<u>E</u> -A172, <u>D</u> -A268, <u>M</u> -A274, <u>Wat</u> -46
<u>Wat</u> -45	2.72	O	<u>M</u> -A178
Φ =(1L, 2M, 1A) Π =(1R, 1D, 1E) Others={1P, 2Wat}			
4 α , 0 β , 4 coil 3 buried, 5 exposed			

A ligand residue or a residue in the second shell is shown by the one-letter amino acid code followed by chain identifier and primary sequence position (Protein Data Bank residue number). Bond lengths are in Å. Residues are underlined when the residue is in an α -helix; in italic when the residue is in a β -strand; or in ordinary font when the residue is in a coil location. Boldface letters indicate that the residue side-chain atoms are buried (side-chain solvent accessibility is less than 10%). A residue in second shells is designated by the symbol * when one of its side-chain atoms forms a hydrogen bond with a side-chain atom of a ligand; the symbol † when one of its main-chain atoms forms a hydrogen bond with a side-chain atom of a ligand; the symbol ‡ when one of its side-chain atoms forms a hydrogen bond with a main-chain atom of ligand. Exclusive main chain and main-chain hydrogen bonds are not indicated. *Data set*: a representative set of protein structures was based on the list of Hobohm and Sander (2) version of December, 1996, with pairwise sequence identity less than 25%. Total number of proteins in the list is 443, and the number of those with metal, heme, or iron-sulfur linkage is 129 (29%). The structure data set was augmented with several recent protein structures known to contain metal centers.

Table 2. Significant residue clusters in protein 3D structures with single zinc ion center

Cluster Type	Residues	Comments
2. 1mnc: Neutrophil Collagenase (human); Single chain, 158aa (PLH: Methylamino-Phenylalanyl-Leucyl-Hydroxamate)		
{HED}	<u>H-168</u> , Zn-282, <u>H-183</u> , D-170, <u>H-196</u> , R-166, D-167, E-200, <u>F-174</u> , S-172, P-173, <u>A-182</u> , PLH-280, Zn-281, <u>H-218</u> , H-222, H-228, <u>E-219</u> , G-179, <u>L-235</u> , <u>D-254</u> , <u>L-181</u> , N-178, Ca-283, <u>D-198</u> , <u>I-180</u> , <u>E-201</u> , D-175	Overlaps the environment of two mononuclear zinc centers;
{HY}	<u>Y-240</u> , PLH-280, Zn-281, <u>H-218</u> , H-222, H-228, P-238, Y-237	About Zn-281
{HM}	Zn-281, <u>H-218</u> , H-222, H-228, PLH-280, <u>E-219</u> , M-236	
3. 1sat: Serratia Protease (Serratia Marcescens); Single chain, 468aa;		
{HM}	<u>M-214</u> , <u>H-180</u> , Zn-472, H-186, <u>H-176</u>	About Zn-472
{HY}	H-229, <u>D-168</u> , Y-230, <u>I-280</u> , Y-169, R-209, <u>Q-172</u> , Y-216, Zn-472, H-186, <u>H-176</u> , <u>H-180</u>	About Zn-472
4. 1iae: Astacin with Zn replaced by Nickel (<i>Astacus Astacus</i> L.); Single chain, 200aa;		
{H}	<u>H-104</u> , <u>H-102</u> , Ni-999, <u>H-92</u> , <u>H-96</u>	About Ni-999
{CH}	C-64, <u>C-84</u> , <u>E-93</u> , H-92, Ni-999, H-102, <u>H-96</u>	About Ni-999
{HY}	<u>H-96</u> , Ni-999, <u>H-92</u> , Y-149, H-102, <u>H-104</u> , Y-141, H-148	About Ni-999
{HED}	H-181, <u>E-36</u> , H-87, <u>I-91</u> , M-182, <u>D-186</u> , S-145, M-147, <u>H-96</u> , Ni-999, <u>H-92</u> , Y-149, H-102, H-104, Y-141, H-148, <u>E-103</u>	About Ni-999
{HEDY}	<u>H-96</u> , Ni-999, <u>H-92</u> , Y-149, H-102, H-104, Y-141, H-148, <u>E-103</u> , R-106, D-108, Y-136	About Ni-999
5. 1add: Adenosine Deaminase (mouse); Single chain, 349aa; (1DA: 1-Deaza-Adenosine)		
{H}	Zn-400, <u>H-17</u> , <u>H-214</u> , <u>H-15</u> , D-295, 1DA-353, <u>H-238</u>	About zinc
{HEDY}	<u>H-238</u> , Y-240, <u>S-266</u> , H-278, M-315, <u>E-217</u> , 1DA-353, D-296, D-19, <u>H-17</u> , Zn-400, <u>H-214</u> , <u>H-15</u> , D-295, <u>E-260</u> , R-235, <u>E-99</u> , Y-97	
6. 2cba: Carbonic Anhydrase II (human); Single chain, 258aa;		
{HY}	H-3, H-4, W-5, H-64, N-62, P-201, P-20, T-200, Y-7, N-67, <u>H-94</u> , Zn-262, <u>H-119</u> , <u>H-96</u> , <u>E-117</u> , Q-92, H-107, Y-194	About zinc center
7. 1ton: Tonin (Rat); Single chain, 228aa;		
{H}	H-57, Zn-200, H-97, H-99,	About zinc center
{CHY}	H-57, Zn-200, H-97, Y-94, A-56, R-90, Y-59, Y-64, N-35, S-60, E-39, Y-40, C-42, C-58	About zinc; the two cysteines form a disulphide bridge
8. 1bmc: Metallo- β -Lactamase (<i>Bacillus cereus</i>); Single chain, 213aa;		
{CH}	C-168, <u>H-149</u> , Zn-228, H-88, <u>H-86</u>	About zinc
{HED}	D-183, H-88, Zn-228, <u>H-149</u> , <u>H-86</u> , T-85, <u>R-91</u> , <u>D-90</u> , D-56, <u>I-83</u> , <u>L-54</u> , <u>H-28</u> , <u>E-30</u> , <u>G-43</u> , D-81, <u>H-106</u> , D-126	About zinc
9. 1cle: Endoglucanase (<i>Clostridium Thermocellum</i>); Single chain, 541aa;		
{CH}	C-173, Zn-594, <u>H-197</u> , <u>H-174</u> , P-282, <u>H-170</u> , C-155	About zinc
{CHY}	<u>H-197</u> , Zn-594, C-173, Y-202, C-155, <u>H-174</u>	About zinc
10. 1tup: Tumor Suppressor P53 (human); Three identical chains (A, B & C), 195aa each chain;		
{CH}	H-A178, H-A179, Zn-A951, C-A242, C-A238, C-A176, P-A177, C-A182	About zinc
{CHM}	H-A178, H-A179, Zn-A951, C-A242, C-A238, C-A176, L-A194, Y-A236, T-A253, M-A133, V-A272, C-A135, C-A141	About zinc
11. 1gpc: Core GP32 (Bacteriophage T4); Single chain, 218aa;		
{CH}	<u>H-64</u> , Zn-1, C-87, C-90, C-77	Complete ligand array
12. 1htb: Beta 3 Alcohol Dehydrogenase (human); Two identical chains (A&B), 374aa each chain;		
{C}	C-A97, Zn-A375, C-A111, C-A103, C-A100	All ligands coordinating zinc
13. 8ate: Aspartate Carbamoyltransferase (<i>E. Coli</i>) Four chains: A & C (identical) 310aa, B & D (identical) 146aa		
{C}	C-B109, Zn-B154, C-B138, C-B114, C-E141	Complete ligand group
{CH}	H-B117, C-B114, Zn-B154, C-B138, C-B109, C-B141	About zinc
14. 1lba: Lysozyme Mutant A6K, Del (2-5) (Bacteriophage T7); Single chain, 146aa;		
{CH}	Zn-151, C-130, <u>H-17</u> , H-122, H-68, C-18	About Zn-151
16. 1ctt: Cytidine Deaminase (<i>E. Coli</i>); Single chain, 294aa;		
{CHM}	C-132, Zn-296, H-102, <u>E-104</u> , C-129, H-131, <u>M-136</u>	About zinc
18. 1vhm: Sonic Hedgehog (<i>Mus Musculus</i>); Single chain, 152aa;		
{HED}	H-181, E-177, <u>H-183</u> , Zn-400, <u>D-148</u> , <u>H-141</u> , S-139, H-135, E-127, D-130, Y-142, G-133, H-134, D-132, E-91	About zinc
20. 8tln: Thermolysin (<i>Bacillus Thermoproteolyticus</i>); Single chain (E), 316aa;		
{HED}	D-E150, W-E115, H-E146, Zn-321, <u>E-E166</u> , <u>H-E142</u> , Y-E157, <u>N-E165</u> , D-E170, R-E203, <u>D-E138</u> , Ca-317, E-E187, D-E185, Ca-318, <u>E-E177</u> , E-E190 (The ligand and second shell residues of Ca ions were not underlined)	Traverses zinc and calcium environments
{ED}	D-E191, E-E190, Ca-318, <u>E-E177</u> , D-E185, Ca-317, <u>D-E138</u> , E-E187	About two proximal calcium ions
21. 1frp: Fructose-1,6-biophosphatase (pig kidney); Two identical chains (A&B), 321aa each chain;		
{HEDY}	E-A97, E-A98, R-A276, <u>E-A280</u> , Zn-5, <u>D-A121</u> , <u>D-A118</u> , FDP-1, Y-A244, Y-A264, Y-A215	About zinc
23. 1thj: Carbonic Anhydrase (<i>Methanosarcina thermophila</i>); Trimer, 213 aa each chain;		
{H}	<u>H-A81</u> , Zn-A214, <u>H-A122</u> , <u>H-C117</u>	residue ligands of the zinc ion
{HED}	E-B13, T-A212, E-A211, <u>G-A208</u> , E-A207, H-A204, P-B18, <u>T-B17</u> , W-B19, Y-A200, E-B88, E-B89, N-87, B-91, P-B92, <u>I-B97</u> , E-B104, E-B94, D-B95, N-B96, I-B93, E-B84, Y-A201, <u>E-B62</u> , <u>H-B81</u> , Zn-B214, <u>H-B122</u> , <u>H-A117</u> , D-A76	interface of chain A & B and residue ligands of the zinc metal ion
24. 1lcp: Leucine Aminopeptidase (Bovine); Two identical chains (A&B), 484aa each chain;		
{KRED}	E-B80, R-B396, E-A404, R-A408, D-B458, K-B457, E-B459, S-B370, R-B470, P-B471, D-B395, <u>T-B474</u> , <u>T-B393</u> , R-B473, <u>E-B477</u> , R-B481, <u>Q-B164</u> , <u>R-B168</u> , E-B172, R-B271	Just reaches the environment of the zinc metal ion

The same sequential numbers of proteins are used as in Table 1. There is no statistically significant residue cluster (at 1% level) in atrolysin (1. 1atl), tramtrack zinc finger (15. 2drp), glycerol kinase (17. 1glc), carboxypeptidase (19. 2ctc), and enolase (22. 4enl). Procedures for identifying diverse residue clusters are described in the *Methods* sections of Karlin and Zhu (6) and Zhu and Karlin (7); see also the companion paper by Karlin *et al.* (1). A residue is described in two parts: part one, one-letter amino acid code and part two, a chain identifier (if any) and a residue number used in the Protein Data Bank file of the structure. The part one is underlined when the residue is in an α -helix; in italic when the residue is in a β -strand, and in ordinary font when the residue is in a coil. Boldface letters indicate that the residue side-chain atoms are buried (side-chain solvent accessibility less than 10%). Part two of a residue is doubly underlined when the residue is a ligand to the proximal metal ion or singly underlined when the residue is in the second shell of the proximal metal ion. Unless stated otherwise, a cluster is confined to one of identical chains and has analogs in the other chains.

Table 3. Properties of four class IV proteins

	Carboxypeptidase A	Thermolysin	Sonic hedgehog	Glycerol kinase
Ligands	H, H, acidic unidentate, bound water			
	In coils and β -strands. H residues are distant in sequence.	All ligands in α -helices. Classic zinc protease motif: HEX ₂ H	All buried in strands and coils; not a protease but autoprotoleolytic	Two buried H 15 residues apart in chain F; one H in coil, one H in β -strand; E exposed in coil
His ligand contact	2 N ^{δ1}	2 N ^{ε2}	1 N ^{δ1} , 1 N ^{ε2}	2 N ^{ε2}
His bond lengths	Moderate, 2.08–2.13 Å	Tight, 1.93–1.97 Å	Moderate, 2.06–2.08 Å	Long, 2.44–2.21 Å
2nd shell	In buried coils and β -strands	Mostly in α -helices, partly exposed	Many in β -strands, partly exposed	Partly exposed; mixed in secondary structures
Special residues of 2nd shell	D-142 H-bonds to ligand H-69	D-170 H-bonds to ligand H-142	E-54 H-bonds to ligand H-183	T-95 H-bonds to H-75
Significant clusters	None	{HED} about Zn ²⁺ ; {ED} about dicalcium	{HED} about Zn ²⁺ ; {ED} cluster covers the zinc site	None

Class I. These are characterized as mononuclear zinc proteins having a histidine ligand group following the motif HEX₂HX₂GX₂H (5). Several of these structures have an additional Y ligand. We propose the more precise motif H^(a)EX₂H^(b)X₂G(L/M/F)XH^(c), where the displayed E invariably occurs in the second shell and establishes a hydrogen bond of its carboxylate with the N^{ε2} nitrogen of H^(b). One of (L/M/F) is pervasively in the second shell close to H^(b). The G of the motif is external to the second shell. The second shell is about equally hydrophobic and polar, with R and E manifest among the polar residues. The first two histidine ligands of the motif HX₃HX₅H are in a common α -helix and the third histidine is in the following coil, giving a stable base for zinc ions to interact with.

Class II. These are examples of zinc metals involving three histidine ligands with two in the sequence order HXH and the third H more than 20 positions away. In these cases histidines predominantly ligate the zinc ion via N^{ε2}. The metallo- β -lactamase (8), which has its zinc coordination group involving H-86, H-88, H-149, is an anomaly in that the proximal histidines entail ligation by N^{ε2} for H-86, but N^{δ1} for H-88. The second shell of the class II zinc environments mostly favor polar (including several charged) residues (Table 1). In all these structures a significant {H}, {HED}, or {HY} cluster overlaps the extended zinc environment. In the protein structures of adenosine deaminase and carbonic anhydrase II, the paired one-apart histidine ligands occur in the same β -strand. However, this is not the case for tonin and metallo- β -lactamase, where both proximal histidines occur in a common coil.

Class III. A striking number of zinc protein structures have cysteine pairs coordinating Zn positioned in the order CX₂C, reminiscent of cytochrome heme bonding forms following the motif CX₂CH and tandem zinc finger segments following the motif CX₂CX₁₂HX₂₋₃H. Other examples of CX₂₋₄C occur frequently in the environment of iron-sulfur clusters (9). Class III two-apart cysteine ligands occur in two structural forms. The first of two cysteines and the two intervening residues are part of a turn. This motif is seen in the zinc ligand groups of tumor suppressor p53, aspartate carbamoyltransferase, and the tramtrack zinc finger protein. The other motif places the second cysteine (or histidine) ligand in a short helix, while the first cysteine ligand occurs as the N-cap residue of the helix. The second motif is seen in core GP32, β -3 alcohol dehydrogenase, and cytidine deaminase.

Mononuclear zinc ligand groups of distributed histidines (HX_mH, $m > 4$), pairs of cysteine ligands, and a possible acidic residue have the two histidines generally contact the metal in the mixed conformation of one N^{ε2} and one N^{δ1} tautomeric geometries. The tramtrack protein coordinates its two separate zinc ions in the N^{ε2} contact. Cytidine deaminase has a single histidine ligand (N^{δ1} contact), an acidic ligand, and two close cysteines CXHC. (here H is in the second shell). Core GP32 uses a single histidine (N^{δ1} contact) and three cysteines. Similarly, tumor suppressor p53 coordinates Zn-A with a single

histidine (N^{δ1} contact) and three cysteines. The mononuclear zinc ligand groups composed of only H and C residues are mostly part of a cysteine {C} and/or a cysteine-histidine {CH} cluster (Table 2). The second shell features predominantly charged residues often interacting with ligands.

Class IV. These are mononuclear zinc ion proteins coordinated by two histidines, one acidic unidentate residue, and a bound water. Four prominent current structures are of this kind: carboxypeptidase A (2ctc), thermolysin (8tln), sonic hedgehog (1vhh), and glycerol kinase (1glc). Carboxypeptidase A and thermolysin are stated to share structural similarity about the mononuclear zinc metal center (5, 10). However, our analysis, as shown in Table 3, indicates that the coordination topologies and other second-shell features of the foregoing four structures are innately different.

The zinc ligand group in 1glc traverses both chains, with zinc at the interface. The second shell of 1glc involves E-475G and R-482G, which mutually salt bridge. The second shell in 1vhh features E-54, which hydrogen bonds with the ligand H-183, and H-135, which interacts with the ligand H-141. Another similarity relates a second glutamate residue—E-270 in 2ctc, E-143 in 8tln, and E-177 in 1vhh—all bound to the zinc-bound water; and also a positively charged residue—R-127 in 2ctc, H-231 in 8tln, and H-135 in 1vhh—provides an essential general acid/base flexible addendum.

Thermolysin (8tln) features a {HED} cluster about the Zn²⁺ ion (Table 2). Around the dicalcium core of 8tln there is a significant acidic {ED} cluster featuring 3 E and 3 D residues, which putatively augment stability of the thermolysin structure (6). Thermolysin contains four calcium ions, all apparently contributing in structural capacities that are considered responsible for the high thermal stability of this protein. Two are enclosed by an acidic charge cluster and two others are partially exposed at the protein surface, coordinated with one or two direct acidic ligands and other carbonyl attachments. Sonic hedgehog (1vhh) distinguishes a significant {HED} cluster substantially overlapping the zinc environment. Somewhat analogous to thermolysin, there is also an acidic cluster {ED} that does include the zinc metal center. It is possible that, similar to thermolysin, adventitious Ca²⁺ ions might bind in the midst of the {ED} cluster of 1vhh.

Class V. The Zn atom of fructose-1,6-bisphosphatase (1frp) is ligated only by acidic residues plus a fructose 2,6-bisphosphate molecule. The zinc center in each chain is close to the fructose 2,6-bisphosphate interface of the homodimer. Although all residue ligands of enolase (4enl) are acidic and the second shell contains two acidic and two lysine residues, there are no statistically significant charge clusters or residue clusters of any kind. The two lysines K-396 and K-345 salt bridge to the ligands E-295 (or D-246) and D-320, respectively.

Numbers of Hydrophobic (Φ) vs. Polar (Π) Residues of the Second Shell and Numbers of the Different Types of Closest Atoms

Table 4 gives the aggregate residue counts in Φ and Π [glycine (G), proline (P), and water counted separately], and the counts

Table 4. Numbers of residues and closest atoms in the zinc ion environment

Zn	No.			Aggregate residue counts					No. of closest atoms				
	Structures	Ligands	2nd-shell residues	Φ	Π	G	P	HOH	C	N	O	HOH	S
Class I	4	19	55	20	16	3	2	12	11	6	40	12	0
Class II	4	16	36	10	22	0	1	2	9	10	21	2	1
Class III	10	40	80	24	39	2	3	10	30	17	34	10	0
Class IV	4	16	51	11	25	2	1	11	18	2	29	11	0
Class V	2	10	36	7	15	1	2	11	8	15	9	11	0
Class VI	3	13	29	10	12	0	1	6	10	1	16	6	1

of the closest atoms from second-shell residues to ligands [carbon (C), nitrogen (N), oxygen (O), and sulfur (S)] for the six zinc classes. These results demonstrate that for the zinc classes II–V and marginally for class VI the counts unambiguously favor polar over hydrophobic residues in the second shell. Moreover, oxygen is the predominant closest atom contact with zinc ligands. Only in Zn class I does the count of Φ exceed the count of Π .

Functional and Evolutionary Implications

Examination of the extended environment of zinc metal centers reveals the following general features. The second shell generally contains a majority of polar residues, including several charged residues (mostly R but occasionally H) and often one or more acidic residues. Most of the zinc protein structures possess a significant {HED} cluster and/or a {CHY} cluster overlapping the zinc environment.

There are zinc metal centers (class I, a subclass of hydrolase structures) which are coordinated by three histidine residues arranged in the sequence motif $\text{HEX}_2\text{HX}_2\text{G}(\text{M/L/F})\text{XH}$. The geometry of ligation is persistently that of $\text{N}^{\delta 2}$. Generally, Table 1 reveals that zinc ligation by two or more histidine residues having some two histidines proximal in the primary sequence HX_mH , $m < 5$, broadly adopt the $\text{N}^{\delta 2}$ tautomeric conformation (an exception is the metallo- β -lactamase (1bmc) structure). Why is the $\text{N}^{\delta 2}$ geometry advantageous for proximal histidine residues? A $\text{N}^{\delta 1}$ ligation occupies more space (with the backbone closer) in contacting the metal. It is, therefore, difficult for two proximal histidines in the sequence to both ligate a metal ion via $\text{N}^{\delta 1}$ without causing steric conflicts. However, two histidines distant in the sequence can move around to surround a metal ion on different sides. The $\text{N}^{\delta 2}$ approach generally has the histidine backbone pointing away from the metal center and has little surface area of the metal involved in the contact. From this perspective, the following principle applies to histidine ligation configurations: Two close histidine ligands HX_mH , $m < 5$ rarely simultaneously coordinate a single metal ion in the $\text{N}^{\delta 1}$ conformation.

Is there substantial divergent or convergent evolution among metal centers, or neither? How do similarities and differences in the metal, in the ligand set, and in the second-shell constituents reflect evolutionary and functional processes? Recall (see Legend to Table 1) that the protein 3D structures analyzed were collected as a representative set (443 nonredundant structures) having pairwise sequence identity less than 25%. There are numerous instances of structural conservation without sequence conservation. For example, the two manganese protein structures 2chr and 2mnr discussed in ref. 1 show <10% sequence identity but possess rather similar structural architecture. However, 2chr possesses two mixed-charge residue clusters on the protein surface, whereas 2mnr shows no residue cluster of any kind. A possible case of convergent evolution: Almost all mononuclear type I Cu ion proteins are of the all- β structural class.

There are abundant differences. In fact, although the type I Cu metal centers share the same ligand set, they adopt at least three different secondary structural motifs of the ligand set (1). As a similarity, the second shell about each type I Cu ion contains on

average three exposed residues, mostly located in coil elements. These solvent-accessible coil residues suggest an intrinsic relationship between the type I Cu metal environment and the protein surface. From this perspective, the second shell provides flexibility in the metal environment that may facilitate electron transfer activity. Also, the histidine tautomeric conformation for all type I Cu ligands invariably uses the nitrogen $\text{N}^{\delta 1}$ ligand–metal contact. This bonding preference contrasts sharply with type II Cu ion and multinuclear Cu ligation patterns, which ligate predominantly via $\text{N}^{\delta 2}$. An opposite pattern occurs with respect to iron metal centers. Mononuclear iron coordination universally use $\text{N}^{\delta 2}$ tautomeric contacts, whereas diiron centers predominantly make $\text{N}^{\delta 1}$ contacts.

Mononuclear zinc environments are rather diverse. Are the members within each zinc class (I–V) evolutionarily related? All examples of class I follow the ligand motif $\text{HX}_3\text{HX}_2\text{H}$ and always have the first two histidine ligands part of the same α -helix facing the metal. By contrast, the examples of class II having histidine ligands (HXH) divide into at least two local structural motifs. Class IV members, which share an identical ligand array, mostly diverge in their metal bonding geometry. The charge environment about class V zinc centers consists of ligands that are mostly acidic balanced by positively charged residues in the second shell. A paramount second-shell influence is stabilizing the metal–ligand organization and establishing catalytic function. A histidine–acidic cluster intermeshed with hydrogen-bonding networks, which commonly traverses the metal environment of active zinc sites, may be conducive to these objectives.

Further questions and issues include the following: How does the second shell contribute in function, in structural stability, in protein–protein interactions, and in substrate channeling? Is there a correlation between protein tertiary structure (e.g., structural class, quaternary structure) and the extended metal environment? Are there situations where it is clear that second-shell residues are “preorganizing” direct ligands via hydrogen-bonding and other specific interactions, thereby controlling metal-binding affinity as well as situations in which the second shell appears to have little directing function on the direct ligands? The answers may be helpful for predicting metal sites, for purposes of protein engineering, and for suggesting designs for testing function/structure characteristics. The second-shell residues, in particular, offer attractive sites for single or cassette mutagenesis.

S.K. was supported in part by National Institutes of Health Grants 5R01GM10452-33 and 5R01HG00335-09 and National Science Foundation Grant DMS9403553-002.

1. Karlin, S., Zhu, Z.-Y. & Karlin, K. D. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 14225–14230.
2. Hobohm, U. & Sander, C. (1994) *Protein Sci.* **3**, 522–524.
3. Christianson, D. W. (1991) *Adv. Protein Chem.* **42**, 281–350.
4. Coleman, J. E. (1992) *Annu. Rev. Biochem.* **61**, 897–946.
5. Lipscomb, W. N. & Sträter, N. (1996) *Chem. Rev.* **96**, 2375–2433.
6. Karlin, S. & Zhu, Z.-Y. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8344–8349.
7. Zhu, Z.-Y. & Karlin, S. (1996) *Proc. Natl. Acad. Sci. USA* **93**, 8350–8355.
8. Concha, N. O., Rasmussen, B. A., Bush, K. & Herzberg, O. (1996) *Structure* **4**, 823–836.
9. Howard, J. B. & Rees, D. C. (1991) *Adv. Protein Chem.* **41**, 199–273.
10. Creighton, T. E. (1993) *Proteins: Structures and Molecular Properties* (Freeman, New York).