Protein Science (1998), 7:1700–1716. Cambridge University Press. Printed in the USA.

Analysis of zinc binding sites in protein crystal structures

IAN L. ALBERTS,[1,2] KATALIN NADASSY,[1,2] AND SHOSHANA J. WODAK[1,3]

[1]EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, England
[2]Department of Biological and Molecular Sciences, University of Stirling, Stirling FK9 4LA, Scotland
[3]Unité de Conformation des Macromolécules Biologiques, Université Libre de Bruxelles, Avenue F. Roosevelt, CP 160/16, B1050, Bruxelles, Belgium

(RECEIVED October 27, 1997; ACCEPTED April 6, 1998)

Abstract

The geometrical properties of zinc binding sites in a dataset of high quality protein crystal structures deposited in the Protein Data Bank have been examined to identify important differences between zinc sites that are directly involved in catalysis and those that play a structural role. Coordination angles in the zinc primary coordination sphere are compared with ideal values for each coordination geometry, and zinc coordination distances are compared with those in small zinc complexes from the Cambridge Structural Database as a guide of expected trends. We find that distances and angles in the primary coordination sphere are in general close to the expected (or ideal) values. Deviations occur primarily for oxygen coordinating atoms and are found to be mainly due to H-bonding of the oxygen coordinating ligand to protein residues, bidentate binding arrangements, and multi-zinc sites. We find that H-bonding of oxygen containing residues (or water) to zinc bound histidines is almost universal in our dataset and defines the elec-His-Zn motif. Analysis of the stereochemistry shows that carboxyl elec-His-Zn motifs are geometrically rigid, while water elec-His-Zn motifs show the most geometrical variation. As catalytic motifs have a higher proportion of carboxyl elec atoms than structural motifs, they provide a more rigid framework for zinc binding. This is understood biologically, as a small distortion in the zinc position in an enzyme can have serious consequences on the enzymatic reaction. We also analyze the sequence pattern of the zinc ligands and residues that provide elecs, and identify conserved hydrophobic residues in the endopeptidases that also appear to contribute to stabilizing the catalytic zinc site. A zinc binding template in protein crystal structures is derived from these observations.

Keywords: bidentate; catalytic; coordination; ligand; sequence pattern; structural; template; zinc

Zinc is an important metal in biological systems as it is a strong Lewis acid, forms a stable Zn^{2+} ion, and can exist in several coordination geometries. In proteins, zinc ions play valuable roles in both enzyme catalysis and maintaining structure.

The catalytic role of zinc involves its electrophilic character. The Zn^{2+} ion can stabilize the negative charges in the reaction intermediate, as in both carboxypeptidase (Christianson & Lipscomb, 1989; Coleman, 1992; Lipscomb & Sträter, 1996) and alcohol dehydrogenase (Ramaswamy et al., 1996; Cho et al., 1997). Catalytic zinc ions can also ionize bound solvent to nucleophilic hydroxyl, as in carbonic anhydrase (Eriksson et al., 1986, 1988; Silverman & Lindskog, 1988; Silverman, 1991; Linskog & Lilsjas, 1993; Lilsjas et al., 1994). In catalytic sites, the zinc ion is usually exposed and bound to a solvent molecule. Its most common ligand is His, followed by Glu, Asp, and Cys (Vallee & Auld, 1990), and the prevailing geometry is tetrahedral.

Reprint requests to: Shoshana J. Wodak, EMBL Outstation, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, England; e-mail: shosh@ebi.ac.uk.

The structural roles of zinc in proteins have been recently reviewed (Coleman, 1992; Berg & Shi, 1996). In "Zn-finger" DNA binding domains, the zinc ion stabilizes the folded conformation of the protein required for interactions with the nucleic acid (Vallee & Auld, 1990; Harrison, 1991; Berg & Godwin, 1997). Zinc ions are also required for the assembly of polymeric species, for example, the hexameric insulins (Derewenda et al., 1989) or can be involved in stabilizing the active site structure in enzymes, such as superoxide dismutase (Lippard et al., 1977). Bound zinc ions playing the latter role do not participate directly in the catalytic reaction and are hence categorized as structural. In general, structural zinc sites adopt a tetrahedral arrangement formed by groups from side chains of Cys, His, and occasionally Asp or Glu.

From analyses of the crystal structures of small molecules in the Cambridge Structural Database (CSD) (Allen et al., 1979) and protein crystal structures in the Protein Data Bank (PDB) (Bernstein et al., 1977), very useful information has been derived about zinc coordination (for reviews, see Christianson, 1991; Glusker, 1991) and the geometry and stereochemistry of zinc ion–ligand interactions (Armstrong, 1988; Christianson & Alexander, 1989, 1990; Chakrabarti, 1990; Vedani & Huhta, 1990; Carrell et al.,

1993). For example, coordination numbers of 4, 5, and 6 were observed for zinc ions in the CSD (Bock et al., 1995; Katz et al., 1996). Higher coordination numbers were also observed, albeit more rarely. Geometric studies of zinc binding to specific ligands include, for example, analyses of zinc-imidazole interactions, in which the metal is usually found to be within 10° of the heterocyclic ring plane in the direction of the coordinating nitrogen lone pair (Chakrabarti, 1990; Vedani & Huhta, 1990; Carrell et al., 1993).

The increase by nearly 10-fold of the number of protein structures in the PDB since the last surveys on metal binding in proteins were performed, and the improved quality of these structures, make it worthwhile to repeat and extend the surveys of zinc binding sites using this expanded sample. In this study, we report such an undertaking. All highly resolved and zinc binding protein crystal structures available in the July 1997 release of the PDB are surveyed to investigate the trends in geometrical and stereochemical parameters of the zinc binding sites.

Although the electron density at the zinc positions in proteins is usually rather well defined, the zinc ligation geometry may still be influenced by the restraints used during the refinement procedure. As the parameters for these restraints are often derived from small molecule data, we also survey the different zinc coordination geometries in zinc complexes in the CSD, comparing, whenever relevant, the observed trends with those of the corresponding coordination types in protein structures.

The trends in these various geometrical and stereochemical properties are surveyed separately for zinc ions involved in catalysis (catalytic zincs), and those playing a structural role (structural zincs), and the observations are related to zinc's biological roles. The reported observations should be helpful in improving the atomic models for zinc-protein complexes obtained from X-ray diffraction data, and provide valuable guidelines for de novo engineering of zinc binding sites in proteins.

Results

1. Zinc containing protein structures in the PDB

Searching the PDB for "HET zinc" gave 387 entries. Of these, 339 are X-ray structures, and the rest are NMR (44) and theoretical (4) structures. Our analysis was restricted to high quality X-ray structures, defined as the subset of well-refined structures with resolution better than 2 Å and R factors better than 0.2. This led to 146 structures. Out of these, a further 35 were removed based on several criteria, leading to 111 structures analyzed in our dataset (Tables 1A and 1B list the proteins containing catalytic and structural zincs, respectively; the 35 excluded proteins are listed in Table 1C). Seventy-eight of the analyzed proteins are enzymes which only contain catalytic zinc sites, 27 are proteins (including enzymes) in which the zinc plays only a structural role, and six are enzymes that contain both a catalytic and structural zinc and so appear in both categories. The dataset consists of native structures, enzyme-inhibitor complexes with zinc bound inhibitors, and mutated species. In cases where the PDB file contains more than one identical chain, only the first chain was used. In 13 cases (7 insulins, 3 thermolysins, pancreatic hormone, drosophila TBP, and myrosinase) where the asymmetric unit of the crystal structure does not include the complete oligomer, the coordinate files generated at the EBI (Henrick, 1997) were used, in which crystallographic symmetry operations have been applied to obtain the oligomer.

The zinc-containing proteins analyzed in this work are divided into families and subfamilies according to sequence homology and function, as shown in Tables 1A and 1B for proteins containing catalytic and structural zincs, respectively. Members from different families are nonhomologous (≤25% sequence identity), while members from different subfamilies show some sequence homology (30–72% sequence identity) and within subfamilies, the sequences are almost identical (>90%). This classification follows closely that of SCOP (Murzin et al., 1995). For further details of the dataset of catalytic and structural zincs, see the legends of Tables 1A and 1B.

2. Properties of ligands in the zinc primary coordination shell

The zinc ligands and coordination number of each zinc binding site are given in Tables 1A and 1B for each family of catalytic and structural zinc-containing proteins, respectively. The frequency of protein residues acting as zinc ligands is given in the histogram in Figure 1. Of the 90 catalytic zinc sites, 39% contain zinc bound water and 53% contain a zinc bound inhibitor, in which the inhibitor has replaced the water in the primary coordination sphere. The remaining 8% are carbonic anhydrases in which the zinc bound inhibitor does not replace the coordinated water (OH), but increases the coordination number by one.

In 22 of the 34 structural zinc sites, the zinc ion is liganded to protein residues only. Of the remaining 12 sites, of which nine are in the insulin family, nine contain zinc bound water, and three contain zinc bound to a negatively charged ion, Cl^- or SCN^-, from the crystallization procedure.

The propensity of zinc bound His, Asp, Glu, and Cys residues to occur in secondary structural elements versus other types of local structure was computed from the representative protein of each family of catalytic and structural zinc-containing proteins in our dataset. This propensity (65%) was found to be very similar to that previously computed by McGregor et al. (1987) (58%) for the same residues, irrespective of their functional or structural roles, in a set of 61 proteins. The individual propensity of His ligands to occur in α-helices (46%), β-strands (21%), and other structural elements (33%) is also similar to that previously computed by McGregor et al. (1987) (40, 21, and 39% for α-helices, β-strands, and other structural elements, respectively). This agreement suggests that zinc ligation does not impose any extra constraints on secondary structure. The number of observations of the other zinc bound residues in the representative proteins is too small to warrant further comment on their secondary structure properties.

The catalytic zincs are exposed to solvent and have a bound water molecule in the native form of the enzyme. The protein residues in the first coordination shell in catalytic sites have buried backbones as they are usually part of secondary structural elements. Side chains may be exposed depending on the shape of the active site. In carbonic anhydrase, for example, the zinc is at the bottom of a narrow channel and only one of the three zinc bound histidines has an exposed side chain. In carboxypeptidase, on the other hand, the active site is more open near the protein surface and zinc bound residues' side chains are more solvent accessible.

Structural zincs are buried, except in insulin where the two zinc ions stabilize the hexameric unit and are solvent bound. The accessibility of zinc bound protein residues varies according to the

Table 1A. *Proteins containing catalytic zinc sites analyzed in this work*[a]

Name of protein family	PDB code for Rep	*R* (Å)	*R*-value	Number of members	Coordinate number	Zinc ligands	PDB codes of other members
Carboxypeptidase	2CTB	1.5	0.151	9	T_5	2H-E-W	2CTC SCPA 1BAV
					T_5 SP	2H-E-Inh	3CPA
					T_5	2H-E-Inh	6CPA 7CPA 8CPA 1CBX
Muramoyl-pentapeptide carboxypeptidase	1LBU	1.8	0.164	1	T_4	2H-D-W	
Bacterial leucyl aminopeptidase	1AMP	1.8	0.161	1	T_5	1H-D-E-W	
						1H-2D-W	
Leucine aminopeptidase	1LAM	1.6	0.172	3	T_5	2D-E-W	
						2D-K-E-W	
					T_5 SP	2D-K-E-Inh	1LAN 1LCP
					T_6	2D-E-Inh	
Thermolysin-like							
Thermolysin	8TLN	1.6	0.176	12	T_4	2H-E-W	3TMN
					T_4	2H-E-Inh	1THL
					T_5	2H-E-Inh	2TMN 5TMN 1HYT 1TMN
Elastase					T_4	2H-E-W	1EZM
Neutral protease					T_5	2H-E-W	1NPC 1LND
					T_6	2H-E-2W	1LNF
					T_6	2H-E-N-Inh	4TMN
Metallo proteases							
Serratia protease	1SAT	1.75	0.188	3	T_4	3H-W	1SRP
Alkaline protease					T_5 SP	3H-Inh	1KAP
Astacin	1AST	1.8	0.158	1	T_5	3H-Y-W	
Atrolysin	1ATL	1.8	0.16	2	T_5	3H-InH	
					T_4	3H-Inh	1DTH
Matrix metalloproteases							
Collagenase	1HFC	1.56	0.174	5	T_5 SP	3H-Inh	1JAP
					T_5	3H-Inh	1CGE
Matrilysin					T_5 SP	3H-Inh	1MMQ
Stromelysin					T_4	3H-C	1SLM
Neutral protease	1KUH	1.6	0.164	1	T_4	2H-D-W	
Carbonic anhydrase II	2CBA	1.54	0.151	37	T_4	3H-W	1CNG 1CNI 1CNJ 1HEA 1HEB 1HEC 1HED 1HUG (CA1) 1YDC 1ZSC
					T_4	3H-Inh	2CBD 3CA2 1AZM (CA I) 1BZM (CA I) 1BCD 1CAI 1CAK 1CAO 1CIL 1CNW 1CNX 1CRA 1CZM 1RAY 1RAZ 1YDB 1ZSB 1UGB
					T_4^*	3H	2CAB 1UGE
					T_5	3H-W-Inh	1CAM 1CAZ 1HCB 2CBC
					T_5 SP	3H-W-Inh	2CA2
					T_5	3H-Inh	1BIC
Alcohol dehydrogenase	3BTO	1.66	0.185	3	T_4	1H-2C-INH	1BTO
					T_4	1H-2C-W	1CDO
Alkaline phosphatase	1ALK	2.0	0.177	1	T_5	2H-D-Inh	
					T_5	1H-2D-S-Inh	
Sonic hedgehog	1VHH	1.7	0.191	1	T_4	2H-D-W	
Enolase	4ENL	1.9	0.149	1	T_5	2D-E-2W	
Phosphomannose isomerase	1PMI	1.7	0.184	1	T_5 SP	2H-Q-E-W	
Phosphate aldolase	1FUA	1.92	0.186	1	T_5 SP	3H-E	
Metallo-beta-lactamase	1ZNB	1.85	0.167	1	T_4	3H-W	
					T_5	1H-C-D-2W	

Note: Tables 1A, 1B, and 1C list the datasets of zinc containing proteins used in our analysis as well as those excluded from the study. Each dataset is divided into protein families whose names are given in column 1. The 4-letter PDB code, resolution and *R*-factor of the representative protein of each family are given in columns 2–4. In columns 5–8, we give the number of members in each family, the coordination number and zinc ligands in the zinc primary coordination sphere and the PDB codes of the family members. Protein residues are represented by their one-letter code, while water is denoted by W and inhibitor by Inh. Coordination numbers 4, 5, and 6 are denoted T_4, T_5, and T_6, respectively. T_4^* represents tetrahedral coordination lacking a zinc-bound water. All T_5 sites are trigonal bipyramidal except the ones marked with SP, which are square-based pyramidal.

(*Footnotes continue on next page.*)

(*Table 1A footnotes continued*)

[a]The dataset of catalytic zinc containing enzymes. Many of the enzymes are zinc peptidases, which can be subdivided into two main categories, the exo- and endo-peptidases, each containing a number of enzyme families that are detailed below. The exo-peptidases are represented in our dataset by nine examples of *Carboxypeptidase A*, corresponding to the native structure (Rees et al., 1983) and its complexes with different inhibitors, one Muramoyl-pentapeptide carboxypeptidase (Ghuysen et al., 1994) and four Aminopeptidases, comprising three *Leucine aminopeptidases* (Sträter & Lipscomb, 1995a, 1995b) and one *Bacterial leucyl aminopeptidase* (Chevrier et al., 1994). Our dataset includes 10 examples of *Thermolysins* corresponding to the native structure (Holmes & Matthews, 1982) and a series of enzyme-inhibitor complexes and two *Thermolysin-like* structures—the neutral protease (1NPC) (Stark et al., 1992) and elastase (1EZM) (Thayer et al., 1991). We include three *Metalloproteases*, namely alkaline protease (1KAP) (Baumann et al., 1993) and two serratia proteases (1SAT,1SRP) (Baumann, 1994; Hamada et al., 1993), one Astacin (Bode et al., 1992), two *Atrolysins* (Zhang et al., 1994; Botos et al., 1996) and five *Matrix metalloproteases*, consisting of two fibroblast collagenases (1HFC, 1CGE) (Lovejoy et al., 1994; Spurlino et al., 1994; Grams et al., 1995), one neutrophil collagenase (1JAP) (Bode et al., 1994), matrilysin (1MMQ) (Browner et al., 1995), and stromelysin (1SLM) (Gooley, 1994). We also include one *Neutral protease* (1KUH) (Kurisu et al., 1997) in a separate family as it shows only 12.9% sequence identity with the neutral protease (1NPC), which has been included in the thermolysin-like category. Furthermore, our dataset contains 37 *Carbonic Anhydrases*, corresponding to the native structure (Hakansson et al., 1992), complexes with inhibitors, and mutants. The majority of these examples (33) belong to the type II (form B) carbonic anhydrase, while the other four enzymes are examples of type I (form A) carbonic anhydrase. Three *Alcohol Dehydrogenases* are considered in our dataset (Ramaswamy et al., 1996; Cho et al., 1997) and one enzyme from the following families: *Alkaline phosphatase* (Kim & Wyckoff, 1991), *Sonic Hedgehog* (Hall et al., 1995), *Enolase* (Lebioda & Stec, 1989), *Phosphomannose Isomerase* (Cleasby et al., 1996), *Phosphate Aldolase* (Dreyer & Schulz, 1996), and *Metallo-beta-Lactamase* (Concha et al., 1996).

Table 1B. *Proteins containing structural zinc sites analyzed in this work*[a]

Name of protein family	PDB code for representative	*R* (Å)	*R*-value	Number of members	Coordinate number	Zinc ligands	PDB codes of other members
Insulin	4INS	1.5	0.153	7	T_4	3H-W	2TCI 6INS
						3H-Ion	1BEN 3MTH 2TCI 7INS
					T_4^*	3H	1IZB
					T_6	3H-3W	1BEN 1IZB 4INS
Superoxide dismutases	1XSO	1.49	0.104	5	T_4	3H-D	1JCV 1SRD 1SXA 2SOD
Matrix metalloproteases	1HFC	1.56	0.174	3	T_4	3H-D	1MMQ 1JAP
Alpha-lactalbumin	1HML	1.7	0.165	1	T_5	3E-2W	
Rubredoxin	1IRN	1.2	0.107	1	T_4	4C	
Protein kinase C	1PTQ	1.95	0.196	1	T_4	1H-3C	
Pancreatic hormone	1PPT	1.37		1	T_4	1H-N-G	
Enterotoxin	1STE	2.0	0.20	1	T_4	2H-2D	
Plant lectin	1ENR	1.8	0.176	1	T_6	1H-2D-E-2W	
Zif268 zinc finger-DNA	1AAY	1.60	0.195	1	T_4	2H-2C	
Alcohol dehydrogenase	3BTO	1.66	0.185	3	T_4	4C	1BTO 1CDO
Drosophila TBP	1TAF	2.0	0.198	1	T_4	2H-D-W	
					T_4	1D-1E-K	
Myrosinase	2MYR	1.60	0.140	1	T_4	2H-2D	
tRNA-guanine transglycosylase	1PUD	1.85	0.190	1	T_4	1H-3C	
Adenylate kinase	1ZIN	1.60	0.173	1	T_4	4C	

[a]The dataset of structural zinc containing proteins. This dataset consists of 16 families, of which only four have more than one member. There are seven hexameric *Insulins* (Smith et al., 1984; Baker et al., 1988) each of which contain two zinc ions, five *Superoxide Dismutases:* (Tainer et al., 1982; Carugo et al., 1996), three *Matrix Metalloproteases* (Bode et al., 1994; Spurlino et al., 1994; Browner et al., 1995) and three *Alcohol Dehydrogenases* (Ramaswamy et al., 1996; Cho et al., 1997). Only one member is included in our dataset of structural zincs for the following 12 families; *Alpha-Lactalbumin* (Ren et al., 1994), *Rubredoxin* (Dauter et al., 1996), *Protein Kinease C* (Zhang et al., 1995), *Pancreatic Hormone* (Blundell et al., 1981), *Enterotoxin* (Papageorgiou et al., 1995), *Leucine Aminopeptidase* (Sträter & Lipscomb, 1995a, 1995b), *Plant lectin* (Bouckaert, 1995), *Zif268 zinc finger-DNA* (Elrod-Erickson, 1996), *Drosophila TBP* (Xie et al., 1996), *Myrosinase* (Burmeister et al., 1997), *tRNA-Guanine Glycosylase* (Romier et al., 1996), and *Adenylate Kinase* (Berry et al., 1997).

Matrix metalloprotease, alcohol dehydrogenase, and leucine aminopeptidase contain both catalytic and structural zinc sites and so appear in both Tables 1A and 1B. Only three of the matrix metalloproteases and one of the three leucine aminopeptidases were included in the structural zinc section as the structural zincs in the other examples have temperature factors above our cut-off (20.0). Similarly, six of the seven zinc ions in the asymmetric unit of the Drosophila TBP crystal structure are completely excluded from our geometrical analysis due to the high temperature factors (>20.0) associated with the zinc ions.

functional role of the zinc. In insulin, for example, the backbones of the zinc bound histidines are buried, while the side chains are exposed in the region between the monomers. In the DNA binding proteins, the backbone and side chain of zinc bound residues tend to be solvent exposed.

There are two possible tautomeric forms of zinc bound histidine depending on which His nitrogen atom interacts with the metal: NE2 or ND1. Tautomer ε is defined as His NE2 interacting with metal and tautomer δ is a His ND1–metal interaction. Chakrabarti (1990) suggested that NE2 coordinates to the metal ion, giving the ε tautomer. In our analysis with a larger dataset of high quality protein crystal structures, approximately 70% of the zinc bound histidines are ε tautomers and 30% are δ tautomers. These percentages are not dependent on the role of the zinc ion. Our analysis

Table 1C. *PDB codes of structures that were excluded from the dataset*[a]

	$B > 20.0$ on Zn	Identical structures	Mutated species of carbonid anhydrase	Crystallization factor
Catalytic	1AKL 1LAG 1PCA 1BME 1DOS 1FRP	1CA2 2CBB 1YME 6TMN	1MUA 1UGA 1UGC 1UGD 1UGE	
Structural	2EBN 2FBJ 1DVF 4KMB 1CLC 1HXP 1TYM 1TYL 1TRZ	3INS 1JCW 1SXB 1SXC 1ZIO 1YSO		
Not defined				1TON 8RNT 1BRN 1BRK 1BRJ

[a]Lists the 35 protein structures that were excluded from our analysis with the reasons for their exclusion. The temperature factor of the zinc atom in the proteins listed in column 2 is greater than 20.0. The crystal structures of the proteins listed in column 3 involve the same sequence and source as current members of the dataset and differ only in refinement procedure and resolution. Column 4 lists mutated carbonic anhydrases in which the mutated residues are well separated from the zinc binding site. The proteins listed in column 5 contain a zinc ion which is only part of the crystallization process (function of zinc is not defined).

did not identify any significant difference in the coordination distances and angles between the two tautomers and, therefore, we do not comment on them further.

3. Geometrical parameters of the zinc primary coordination sphere

The zinc coordination geometries in the surveyed proteins are classified according to the zinc coordination number in the primary coordination shell (Table 1A,1B). For the zinc coordination numbers 4, 5, and 6, we use the usual notation T_4, T_5, and T_6. These coordination numbers are observed with frequencies of 48, 44, and 6%, respectively, for catalytic sites and 79, 6, and 12%, respectively, for structural sites. In addition we define T_4^* to represent the T_4 coordination case lacking the zinc bound water molecule. This incomplete coordination arrangement is observed in 2 and 3% of catalytic and structural sites, respectively; the tetrahedral coordination geometry dominates for structural zincs. Clearly these frequencies reflect the biases of the available structural data on zinc-containing proteins used in this analysis. However, searching the CSD for all examples of the different zinc coordination numbers, also shows the tetrahedral coordination (T_4) to be the most abundant (50%), but with the T_5 and T_6 coordination numbers occurring with similar frequencies of 26 and 24%, respectively.

Zinc coordination geometries are governed by electrostatics, particularly repulsions between the zinc-coordinating atom bonded electron pairs. Therefore, all T_4 zinc sites are found to have tetrahedral coordination geometries. Among the T_5 sites for catalytic zincs, the majority (74%) have trigonal bipyramid geometries, and 26% have square-based pyramid geometries. There are only two T_5 sites for structural zincs in our dataset and both are trigonal bipyramidal. All the T_6 sites are octahedral. These coordination geometries with their associated ideal coordination angles are depicted schematically in Figure 2.

Applying the simple electron pair repulsion model, expected deviations from the ideal angles in Figure 2 are due to repulsions

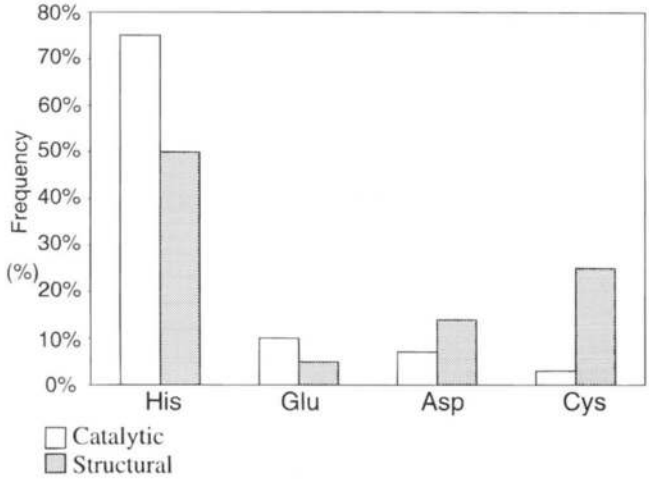


Fig. 1. Frequency (%) of protein residues acting as zinc ligands in catalytic and structural zinc sites in our dataset.

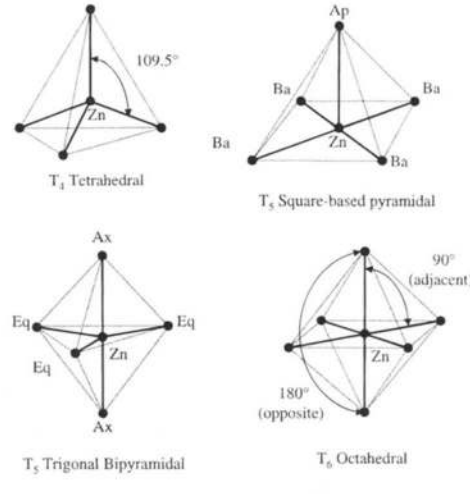


Fig. 2. Ideal angles in the different zinc coordination geometries. The ideal angles for T_5 square-based pyramidal geometry are the following: Ap-Zn-Ba, 90°; Ba-Zn-Ba (adjacent), 90°; Ba-Zn-Ba (opposite), 180°; ideal angles for T_5 trigonal bipyramidal geometry are: Ax-Zn-Ax, 180°; Ax-Zn-Eq, 90°; Eq-Zn-Eq, 120°.

involving lone pairs on the coordinating atoms. Further deviations that are not described by this simple model are caused by the following:

1. H-bonding involving zinc ligands.
2. Bidentate binding arrangements. Thirty-two cases are observed of which 23 involve carboxylate groups and two involve phosphate groups in which two oxygens are coordinated to the zinc giving rise to four-membered chelate rings. Five-membered chelate rings are formed by mainchain $NC_{\alpha}CO$ (3 cases) and NCPO (1 case) groups through the nitrogen and oxygen atoms and by ONCO (3 cases) through the two oxygens binding bidentately to zinc.
3. Multi-zinc catalytic sites in which the zinc ions are directly bridged by one or more oxygen atoms. Asp is the most common ligand in multi-zinc sites followed by His and Glu (Vallee & Auld, 1993), while in mono-zinc sites, His predominates.

These deviations will be discussed in detail later in this section. Meaningful comparisons with the coordination angles in the CSD could not be made due to the lack of resemblance between the zinc binding environments found in small zinc complexes and proteins. Such comparisons are made for coordination distances, however, as they are much less dependent on coordination geometry than the angles. The CSD distances provide a guide for expected trends. For the zinc coordination distances and angles considered in this section, only cases with more than five observations warrant comment.

3.1. Distances between the zinc and coordinating atom

In Tables 2A and 2B, we report the distances between the zinc ion and the coordinating atoms for the analyzed catalytic and structural zinc sites, respectively. Geometrical parameters are only included if the atoms involved have temperature factors less than 20.0. In the protein structure data, Zn-N distances have been subdivided into two categories, one labeled Zn-N(h) when histidine is the coordinating nitrogen ligand which is the majority (92%) and the other labeled Zn-N(x) when the coordinating nitrogen ligand is anything other than histidine (8%). The Zn-O distances have been subdivided into three categories depending on the type of oxygen coordinating ligand; water, protein residue or inhibitor, labeled henceforth as Zn-O(w) (25%), Zn-O(p) (56%), and Zn-O(i) (19%), respectively.

In compiling distances from the CSD, we chose structures that resembled the zinc binding sites in our dataset as closely as possible. Zn-N(h) distances surveyed in the CSD represent zinc bound to imidazole nitrogen, and Zn-N(x) is compiled from structures with zinc bound to other nitrogen-coordinating ligands. Zn-O(w) distances are compiled from structures with zinc bound to water; however, Zn-O(p) and Zn-O(i) could not be distinguished in the CSD and so we give the same values for these distances. Further details on the compilation of CSD distances are given in the legend of Table 2.

In general, we find that in small molecule crystal structures, Zn-N(h) and Zn-O distances increase with coordination number, as expected, due to the effect of electrostatic repulsions between the coordinating atoms and steric hindrance between other atoms belonging to the zinc binding residues. For all coordination types, the Zn-O distances are slightly smaller than Zn-N(h) in agreement with the size of the atomic radii. The standard deviations for Zn-O(p) are larger than for Zn-N(h) and Zn-O(w) due to the greater variation in the type of oxygen-coordinating ligands used in compiling the distances. In the analyzed proteins, departures from these trends are mainly due to H-bonding, bidentate arrangements, and multi-zinc sites as described in the following.

3.1A. Catalytic zinc sites. From the values in Table 2A, the main observation is that, in general, zinc-coordinating atom distances and their standard deviations are larger than those obtained in the CSD, particularly for the various Zn-O distances. The larger standard deviations are probably due to the lower accuracy of atomic coordinates for protein structures. Furthermore, in contrast to the result from the CSD, average Zn-O distances are in general larger than Zn-N(h). For Zn-O(w) this may be because, in 85% of the cases analyzed, the coordinated water forms at least one H-bond to a protein group that is not in the zinc primary coordination sphere. This hydrogen bonding may pull the zinc bound water away from the ion, resulting in longer than expected Zn-O(w) distances, particularly in T_4, as illustrated in Figure 3 for the catalytic site in thermolysin. In T_5, the effect of the small O-Zn-O angle in bidentate carboxylate arrangements (see section 4 below) is to provide more room around the zinc for the other ligands. This allows the water oxygen to get closer to the zinc ion, leading to a slightly smaller Zn-O(w) in T_5 than in T_4.

Average Zn-O(p) distances are compiled mainly from structures containing zinc bound carboxylate groups from Asp and Glu residues. In T_4, the carboxylate groups bind to the zinc in a monodentate arrangement, leading to a relatively small Zn-O(p) distance, while in T_5, bidentate connections lead to a relatively large Zn-O(p) distance.

Interestingly, when an inhibitor replaces the zinc bound water, the coordinated atom occupies a similar position to the water oxygen. The average T_4 Zn-O(i) and Zn-N(x) distances are, therefore, close to Zn-O(w) (although there are only five observations of Zn-O(i) distances in our dataset). In T_5, bidentate arrangements and multi-zinc sites lead to a larger Zn-O(p) distance. As a result of these features, standard deviations for the various Zn-O distances are larger than for Zn-N(h). Finally, the average Zn-S distance in T_4 is similar to that obtained in the CSD.

3.1B. Structural zinc sites. The three distances in Table 2B that satisfy our criterion of more than five observations are Zn-N(h), Zn-O(p), and Zn-S. The average Zn-N(h) distance and its standard deviation are larger than that obtained from the CSD, but close to that found in catalytic zinc sites. The other two distances are close to those obtained in the CSD.

3.2. Angles between zinc and the coordinating atoms

The angles between the zinc ion and the coordinating atoms in catalytic and structural sites are reported in Tables 3A and 3B, respectively. In the simple electron pair repulsion model, expected deviations are due to repulsions involving lone pairs. Further deviations, not described by this simple model, are due to H-bonding, bidentate binding arrangements, and multi-zinc sites. These effects lead to large standard deviations in most of the angles listed. The most salient observations are discussed below.

3.2A. Catalytic zinc sites. **T_4^* and T_4 Coordination:** From the values in Table 3A, we see that the four-coordinate zinc sites are all distorted tetrahedra. On the basis of the simple electron pair repulsion model, the ordering of angles involving N and O atoms is expected to be; O-Zn-O > N-Zn-O > N-Zn-N. This trend is

Table 2A. *Distances between zinc and the coordinating atoms in catalytic zinc sites according to coordination number*[a]

Coordination number	Zn-N(h)	Zn-O(w)	Zn-O(p)	Zn-O(i)	Zn-N(x)	Zn-S
T_4^*	**1.97(0.06), 6**					
	2.01(0.03), 27					
T_4	**2.07(0.09), 116**	**2.12(0.15), 17**	**2.04(0.17), 6**	**2.15(0.11), 5**	**2.08(0.13), 12**	**2.21(0.13), 7**
	2.01(0.03), 27	*2.01(0.03), 24*	*1.97(0.06), 61*	*1.97(0.06), 61*	*2.02(0.04), 186*	*2.25(0.08), 10*
T_5	**2.11(0.10), 80**	**2.09(0.21), 16**	**2.16(0.16), 53**	**2.13(0.20), 21**	**2.09(0.13), 4**	**2.30, 1**
	2.07(0.08), 30	*2.05(0.07), 12*	*2.06(0.12), 106*	*2.06(0.12), 106*	*2.11(0.09), 131*	*2.31(0.01), 2*
T_6	**2.04(0.07), 4**		**2.08(0.06), 12**	**2.24(0.15), 5**	**2.07(0.24), 3**	
	2.16(0.08), 12		*2.12(0.11), 23*	*2.12(0.11), 23*	*2.16(0.07), 254*	

Note: Distances are given in Å, standard deviations are given in parentheses followed by the number of observations of the geometrical parameters in our dataset. All distances and standard deviations are listed with two decimal places to allow the computation of statistical significance tests. The Student *t*-test performed on all the listed values with more than 10 observations, showed that with very few exceptions these values correspond to non-overlapping distributions with significance level well below 5%.

Values in bold type are for catalytic zinc sites in protein crystal structures from the PDB and values in italics are from small molecule crystal structures in the CSD. In the protein structure data, the Zn-N distances have been subdivided into two categories, one labeled Zn-N(h) when histidine is the coordinating nitrogen ligand, and the other denoted Zn-N(x) when the coordinating nitrogen ligand is anything other than histidine. The Zn-O distances have been subdivided into three categories depending on the type of oxygen coordinating ligand; water, protein residue or inhibitor, labeled as Zn-O(w), Zn-O(p), and Zn-O(i), respectively. The number of observations of these geometrical parameters that are included in the analysis has been reduced due to our atomic *B* factor cut-off of 20.0. For this reason no T_6 Zn-O(w) distances are reported in Table 2A.

In compiling the CSD distances, we chose structures that resemble the zinc binding sites in our dataset of proteins as closely as possible. Zn-N(h) and Zn-O(w) distances were compiled from structures that contain zinc bound to imidazole nitrogen and water oxygen, respectively, in the appropriate coordination number. Zn-O(p) and Zn-O(i) could not be distinguished in the CSD and we give the same values for these distances. They were compiled by considering structures involving the same type of oxygen-coordinating groups (carbonyl, monodentate and bidentate carboxylate, and zinc-bridging oxygens) that are found in our dataset of protein structures, in the appropriate coordination number. We restricted the compilation so that bidentate carboxylate arrangements and zinc-bridging oxygens were included in T_5 and T_6, but not in T_4, as was observed in the zinc-binding sites in our dataset of protein structures.

Nitrogen and sulfur-coordinating zinc ligands in the CSD (except for imidazole) showed little similarity with those in the zinc binding sites in our protein dataset. Zn-N(x) and Zn-S distances in the CSD were, therefore, compiled by consideration of structures with the same atom-coordination types (which describe the number and type of atoms coordinated to the zinc ion) that are found in the zinc sites in our protein dataset. All structures that contain the correct atom coordination type were included in the compilation. For example, in compiling the Zn-N(x) distance in T_4, we considered structures with the atom-coordination type 4N (meaning four nitrogen atoms coordinated to zinc), in T_5, we considered 4N1O, 3N2O, and 1N4O structures and in T_6, 2N4O and 3N3O.

[a]Distances between the zinc and its coordinating atoms in catalytic zinc sites. Average Zn-N(h) distances compiled for the δ and ϵ His-Zn tautomers in the catalytic zinc sites in our dataset are 2.09 ± 0.10 Å for ϵ and 2.06 ± 0.10 Å for δ. In compiling the Zn-S distance in the CSD, we considered 3N1S and 1N1O2S structures in T_4 and 1N3O1S in T_5.

Table 2B. *Distances between zinc and the coordinating atoms in structural zinc sites according to coordination number*[a]

Coordination number	Zn-N(h)	Zn-O(w)	Zn-O(p)	Zn-N(x)	Zn-S	Zn-Cl
T_4^*	**2.29, 1**					
	2.01(0.03), 27					
T_4	**2.09(0.12), 42**	**2.15(0.30), 4**	**1.95(0.08), 15**	**1.94(0.36), 2**	**2.35(0.09), 30**	**2.20(0.09), 2**
	2.02(0.03), 27	*2.01(0.03), 24*	*1.97(0.06), 61*	*2.02(0.04), 186*	*2.30(0.05), 147*	*2.18(0.04), 6*
T_5		**2.23, 1**	**2.28(0.33), 3**			
		2.05(0.07), 12	*2.06(0.12), 106*			
T_6	**2.09(0.03), 4**	**2.26(0.05), 3**	**2.18(0.06), 3**			
	2.16(0.08), 12	*2.11(0.07), 59*	*2.12(0.11), 23*			

[a]Distances between the zinc and its coordinating atoms in structural zinc sites. Average Zn-N(h) distances compiled for the δ and ϵ His-Zn tautomers in the structural zinc sites in our dataset are 2.05 ± 0.12 Å for ϵ and 2.14 ± 0.09 Å for δ. In compiling the Zn-S distance in the CSD, we considered the atom-coordination types 2N2S, 1N3S, and 4S. The CSD Zn-Cl distance in T_4 was compiled from structures with three imidazole nitrogens and one chlorine atom coordinated to the zinc.

observed in Table 3A, although the differences between the average values of these angles are small compared to the standard deviations.

T_5 Coordination: Trigonal bipyramid geometry: The main observed feature is that the average eq-Zn-ax angle for O-Zn-O is significantly lower than the ideal value of 90° with a large associated standard deviation (19°). This is due to the contribution of the bidentate coordination of carboxylate groups, which includes values as low as 55°, as illustrated in Figure 4A. In contrast, the corresponding distribution of the eq-Zn-ax angle for N-Zn-O

Table 3A. *Angles between zinc and the coordinating atoms in catalytic zinc sites according to coordination geometry*

	Ideal angles	N-Zn-N	N-Zn-O	O-Zn-O	N-Zn-S	S-Zn-S	O-ZN-S
T_4^* Tetrahedral-water		108 (6), 6					
T_4 Tetrahedral	109.5	108 (6), 140	110 (9), 67	111 (9), 5	112 (4), 8	126 (4), 3	105 (7), 6
T_5 Trigonal bipyrimidal							
Eq-Zn-Ax	90	94 (20), 31	101 (12), 57	80 (19), 46	107, 1	99 (3), 2	
Eq-Zn-Eq	120	109 (9), 10	117 (12), 40	121 (13), 14			115, 1
Ax-Zn-Ax	180		154 (8), 13	157 (18), 6			
T_5 Square-base pyrimid							
Ap-Zn-Ba	90	103 (5), 13	100 (9), 18				
Ba-Zn-Ba (opposite)	180		155 (8), 11	155 (8), 2			
Ba-Zn-Ba (adjacent)	90	100 (5), 5	92 (6), 14	81 (19), 9			
T_6 Octahedral							
Adjacent	90	84 (38), 5	97 (13), 10	90 (10), 26			
Opposite	180		152 (19), 4	161 (12), 6			

Note: Angles between the zinc ion and its coordinating atoms for tetrahedral, trigonal bipyramidal, square-based pyramidal, and octahedral geometries observed in our dataset of protein crystal structures. In column 1 we label the coordination angles for each coordination number and geometry, in accordance with Figure 1. The ideal value for each angle is given in column 2, and in columns 3–8 we give the values for these angles.

Angles are given in degrees, standard deviations are given in parentheses followed by the number of observations of the geometrical parameters in our dataset. T_4^* represents T_4 coordination lacking the zinc-bound water molecule.

Table 3B. *Angles between zinc and the coordinating atoms in structural zinc sites according to coordination geometry*

	Ideal angles	N-Zn-N	N-Zn-O	O-Zn-O	N-Zn-Cl	N-Zn-S	S-Zn-S
T_4^* Tetrahedral-water		107, 1					
T_4 Tetrehedral	109.5	112 (7), 35	107 (12), 43	101 (8), 7	114 (1), 2	107 (7), 7	110 (6), 36
T_5 Trigonal bipyrimidal							
Eq-Zn-Ax	90			87 (23), 4			
Eq-Zn-Eq	120			104, 1			
Ax-Zn-Ax	180			148, 1			
T_6 Octahedral							
Adjacent	90	99 (2), 3	93 (3), 6	86 (10), 9			
Opposite	180		165 (16), 2	171 (8), 2			

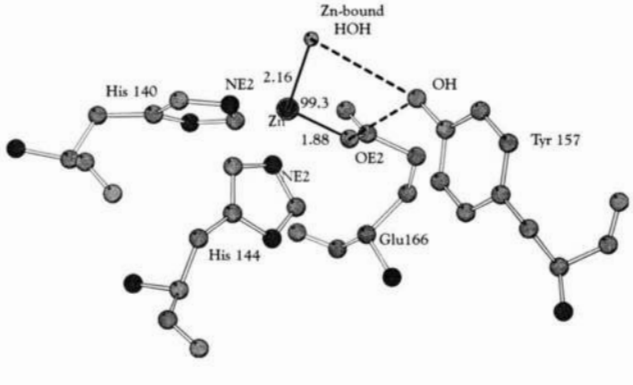


Fig. 3. Zinc binding site of thermolysin taken from PDB file 8TLN (Holland et al., 1992). This figure shows hydrogen bonding interactions that pull the zinc bound water away from the ion, resulting in a relatively large Zn-O(w) distance (2.10 Å).

(Fig. 4B) clearly shows that, although there is considerable variation of the values around the mean, there is no contribution from the low values associated with bidentate carboxylate arrangements.

Square based pyramid geometry: Here, the main observations are that the ap-Zn-ba angles are larger than the ideal value of 90° and the ba-Zn-ba (opposite) angles are about 25° below the ideal value of 180° (Fig. 2). This is a common feature that results from skewing of the basal atoms away from the apical atom due to electrostatic repulsions and, therefore, the zinc is not coplanar with the four basal atoms (Cotton & Wilkinson, 1988).

T_6 Coordination: The average angles for T_6 are reasonably close to ideal. The standard deviation for the N-Zn-N (adjacent) angle is, however, particularly large, 37.5°, due to the contribution from a T_6 site in a dimeric thermolysin (Holden et al., 1987), which has an N-Zn-N (adjacent) angle of 18°, as a result of close van der Waals contact between coordinating nitrogens from different monomers, possibly indicating an inaccuracy in the atomic model.

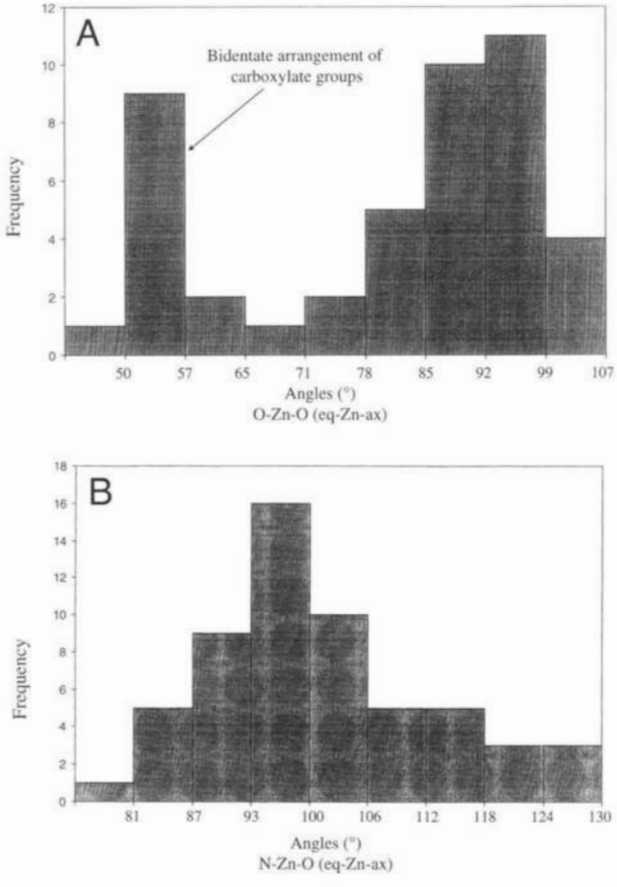


Fig. 4. Distribution of coordination angles in the zinc primary coordination sphere. **A:** Distribution of O-Zn-O (eq-Zn-ax) angles in trigonal bipyramidal catalytic zinc sites in the analyzed proteins. **B:** Distribution of N-Zn-O (eq-Zn-ax) angles in trigonal bipyramidal catalytic zinc sites in the analyzed proteins.

3.2B. Structural zinc sites. Structural zinc sites exhibit tetrahedral, trigonal bipyramid, square-based pyramid, and octahedral coordination geometries with trends in angles similar to those for catalytic sites.

T_4^* and T_4 Coordination: T_4^* and T_4 structural sites are close to tetrahedral. In the T_4 sites, the observed ordering of angles involving nitrogen and oxygen atoms is N-Zn-N > N-Zn-O > O-Zn-O, which is the reverse of the expected ordering. In particular, the average O-Zn-O angle is very low (101 ± 8°), probably because, in each of the four sites that include this angle, the oxygen coordinated atoms are part of H-bonded networks involving protein residues and water (Papageorgiou et al., 1995; Burmeister et al., 1997). The S-N-S angle is close to the ideal tetrahedral angle of 109.5°, in contrast to that in catalytic sites, for which we only have three observations (Table 3A), probably because five of the eight structural sites in which this angle is observed contain zinc liganded to four cysteines which give, on average, a tetrahedral angle that is close to the ideal value.

T_5 and T_6 Coordination: Trigonal bipyramidal T_5 involves a relatively low eq-Zn-ax angle for O-Zn-O with a large standard deviation (87 ± 23°) due to bidentate carboxylate coordination, whereas the observed angles for T_6 coordination are reasonably close to the ideal values.

4. The bidentate arrangement

Many of the deviations from the expected coordination distances and angles are due to the bidentate arrangement of ligands. Here we give the specific geometrical details of this arrangement. The distances and angles associated with bidentate carboxylate binding are depicted in Figure 5A. One of the connections in the bidentate case is necessarily longer than the other. The two distances for bidentate carboxylate are 2.16 ± 0.10 Å and 2.40 ± 0.14 Å, which are significantly longer than the Zn-O distance of 2.01 ± 0.11 Å for monodentate carboxylate. The O-Zn-O angle for the bidentate carboxylate connection is 55.9 ± 2.6°.

Bidentate angles, as well as distances, are not too dependent on coordination geometry and, therefore, meaningful comparisons can be made with bidentate arrangements in small zinc complexes in the CSD. As we can see from the geometrical parameters in Figure 5, bidentate distances and angles in zinc binding sites in proteins and small zinc complexes are rather similar. The standard deviations of the corresponding distances and angles are, however, consistently larger in the protein sample than in the CSD.

5. Multi-zinc sites

Figure 6 schematically summarizes the Zn–Zn and Zn–O (bridging) distances in multi-zinc sites involving closely separated zinc ions and bridging oxygen atoms in alkaline phosphatase, leucine and

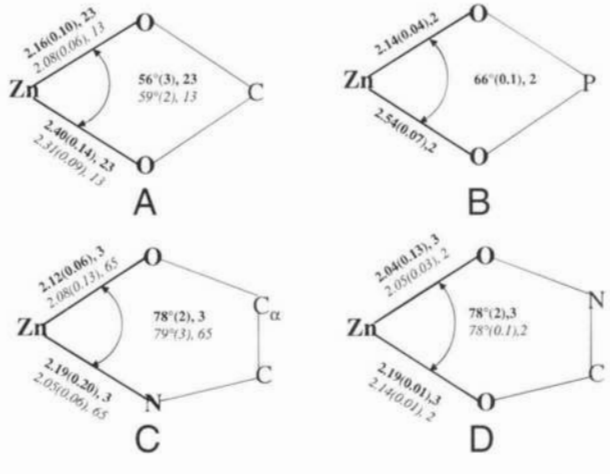


Fig. 5. Geometry of bidentate arrangements in zinc binding sites. Geometrical parameters associated with the bidentate arrangement in zinc binding sites in protein crystal structures from the PDB and in small molecule crystal structures from the CSD are summarized. Mean Zn-O distances (Å) and O-Zn-O angles (°) are given. Standard deviations are given in parentheses and the number of observations of the parameters is given after the comma. The values compiled from proteins are given in bold. Values compiled from the CSD are given below those from proteins in italics. **A:** Bidentate arrangement of carboxylate groups giving rise to a four-membered chelate ring. They are from T_5 and T_6 zinc sites and involve carboxylate oxygens from Asp and Glu residues and inhibitors such as benzyl succinate (BZS), 1-carboxy-3-phenylpropyl (CLT), bicarbonate (BCT), and the tetrapeptide inhibitor, G-S-N-S. **B:** Bidentate arrangement of phosphate oxygens from phosphonaminated and phosphonated peptide inhibitors giving a 4-membered chelate ring. **C:** Bidentate arrangement of mainchain $NC_\alpha CO$ groups from leucine and glycine residues giving a 5-membered chelate ring. **D:** Bidentate arrangement of an ONCO group from hydroxyaminated peptide inhibitors giving a 5-membered chelate ring.

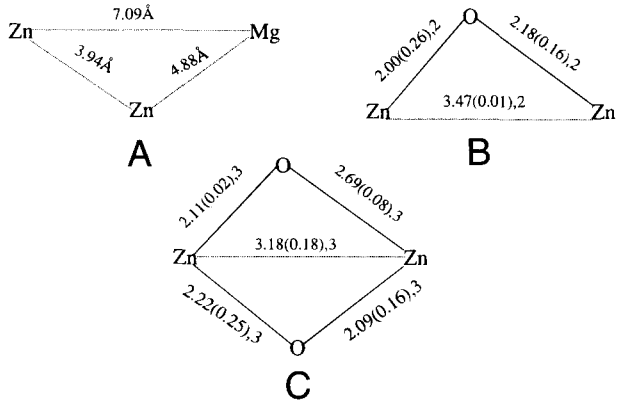


Fig. 6. Zinc–zinc and zinc–oxygen distances in multi-zinc binding sites. This figure summarizes the Zn–Zn and Zn–O distances in multi-zinc sites. These sites have T_5 and T_6 coordination and involve bridging oxygens from water, Asp, and the bound inhibitors l-leucinal and l-leucine phosphonic acid. In all cases, intermetal distances are given, together with bridging Zn–O distances. Values from the CSD are given below those from proteins in italics. **A:** Multi-zinc site in alkaline phosphatase, which has no zinc-bridging oxygen atoms. **B:** Multi-zinc site in leucyl aminopeptidase and β-lactamase showing one zinc-bridging oxygen atom. **C:** Multi-zinc site in leucine aminopeptidase showing two zinc-bridging oxygen atoms.

leucyl bacterial aminopeptidase, and metallo-β-lactamase. We see that, in general, the Zn–Zn distance decreases as the number of bridging oxygen atoms increases. This is due to attractive Zn–O (bridging) electrostatic interactions that screen the Zn–Zn repulsions.

Figure 6 summarizes the distances for multi-zinc sites with 0, 1, and 2 bridging oxygen atoms, respectively. In alkaline phosphatase, which is a dimer where each monomer carries three closely spaced metal ions, two Zn^{2+} ions and one Mg^{2+} (Fig. 6A), there are no bridging oxygen atoms between the zinc ions, and the Zn–Zn distance is 3.94 Å. In the dizinc catalytic sites of leucyl bacterial aminopeptidase and metallo-β-lactamase, the zinc ions are bridged by a single water oxygen, and the Zn–Zn distance is reduced to 3.46 Å (Fig. 6B). The three leucine aminopeptidases in our dataset contain a dizinc site with two bridging oxygens, yielding the ZnOOZn dibridging unit (Fig. 6C), and an average Zn–Zn distance, which is reduced further to 3.18 ± 0.18 Å.

6. *The elec-His-zinc motif*

The existence of H-bonds between zinc ligands and protein residues that are not part of the zinc primary coordination shell has been studied by a number of workers (Christianson, 1991; Kiefer et al., 1995; Lesburg & Christianson, 1995). For example, a carboxylate-His H-bond is believed to increase the basicity and ligand strength of the His and arrange it correctly for interaction with the metal (Argos et al., 1978; Carver & Bradbury, 1984; Perutz et al., 1985).

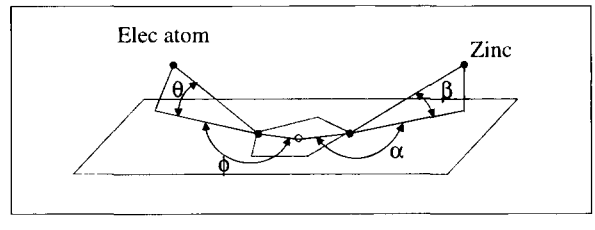
Here, we searched for the existence of a general motif in which zinc bound histidine makes an H-bond to an oxygen atom from a protein residue or water, denoted as an elec atom, to give the elec-His-Zn motif (Wallace, 1996). We find the presence of this motif to be nearly universal in both structural and catalytic zinc sites. In this section, we consider in detail the geometrical parameters associated with this motif.

6.1. *General trends*

We define three types of oxygen elec atom in the motif: carboxyl, carbonyl, and water oxygen. For each coordination number and elec atom type, average elec-N(h) and Zn-N(h) distances and two angles to describe the orientation of the elec atom and zinc ion relative to the histidine ring are defined (Fig. 7). The average values, standard deviations, and the number of observations of these parameters are given in Tables 4A and 4B for catalytic and structural zinc sites, respectively. Of the 206 zinc bound histidines in the catalytic sites analyzed, 97% (200) have an elec H-bonded to give the catalytic elec-His-Zn motif. In the structural zinc sites, 75% (35) of zinc bound histidine residues have an elec atom associated with them to give structural elec-His-Zn motifs.

6.2. *Specific trends in geometrical parameters*

6.2A. Elec-N(h) and Zn-N(h) distances.. From the values in Tables 4A and 4B we see that in both catalytic and structural elec-His-Zn motifs, the ordering of elec-N(h) distances is carboxyl elec-N(h) < carbonyl elec-N(h) < water elec-N(h). The major contribution to elec-N(h) H-bonding is electrostatic interactions. Not surprisingly, therefore, carboxyl oxygen gives the shortest elec-N(h) distance, since the carboxylate group has a formal negative charge. Water oxygen gives the longest distance, since sp^3 hybridization of the water oxygen leads to a lower partial charge than the sp^2 hybridized carbonyl oxygen. The Zn-N(h) distances in these motifs show little variation with elec atom type.



	ϕ	θ
Catalytic		
COO-	180 (20), 63	8 (6), 63
CO	185 (14), 121	7 (7), 121
HOH	182 (16), 16	12 (14), 16
Structural		
COO-	183 (8), 5	8 (4), 5
CO	181 (16), 11	12 (7), 11
HOH	184 (12), 19	11 (13), 19

	α	β
Catalytic	179 (8), 210	7 (6), 210
Structural	182 (6), 48	8 (6), 48

Fig. 7. Geometric parameters of the zinc-His-elec triad. This figure summarizes the observed trends in the geometric parameters that define the position of the elec atom and zinc relative to the imidazole ring of the histidine residue in the elec-His-Zn motif. Φ is the angle between the elec-N(h) vector projected into the ring plane and the line joining the centroid of the imidazole ring to the hydrogen bonded nitrogen atom, termed the in-plane angle for the elec atom. θ is the angle between the elec-N(h) vector and the imidazole ring plane, termed the out-of-plane angle for the elec atom. α is the angle between the Zn-N(h) vector projected into the ring plane and the line joining the centroid of the imidazole ring to the zinc-coordinated nitrogen atom, termed the in-plane angle for the zinc ion. β is the angle between the Zn-N(h) vector and the imidazole ring plane, termed the out-of-plane angle for the zinc ion. Below the figure, values for each of the angles (α, β, θ, and ϕ) in degrees for both catalytic and structural motifs, are reported. Standard deviations are given in parentheses, followed by the number of observations of the parameter. The parameters are given according to elec atom type in the elec-His-zinc motif. Three different elec atom types are defined: carboxyl, carbonyl and water, denoted COO^-, CO, and HOH, respectively.

Table 4A. *Elec-N(h) and Zn-N(h) distances according to coordination number and elec atom type in Elec-His-Zn moifs in catalytic zinc sites*

	Oxygen-N(h)			Zn-N(h)		
Coordinate number	Carboxyl	Carbonyl	Water	Carboxyl	Carbonyl	Water
T_4^*	2.64 (0.04), 2	2.79 (0.08), 4		1.96 (0.08), 2	1.97 (0.05), 4	
T_4	2.67 (0.12), 36	2.73 (0.10), 71	2.89 (0.13), 4	2.06 (0.10), 35	2.09 (0.10), 69	2.02 (0.06), 4
T_5	2.67 (0.10), 23	2.78 (0.11), 44	2.77 (0.11), 12	2.09 (0.09), 23	2.10 (0.08), 44	2.12 (0.14), 12
T_6	2.67 (0.00), 2	2.79 (0.07), 2		2.04 (0.08), 2	2.05 (0.08), 2	
Total	2.67 (0.11), 63	2.75 (0.10), 121	2.80 (0.12), 16	2.07 (0.09), 62	2.09 (0.09), 119	2.10 (0.13), 16

Note: Distances are given in Å; standard deviations are given in parentheses followed by the number of observations of the geometrical parameters in our dataset. Elec-N(h) is the distance between the elec and the H-bonded nitrogen of the His ligand. These parameters are given according to coordination number and elec atom type in the elec-His-zinc motif. The values labeled Total give overall averages of these parameters for each elec atom type. Three different elec atom types are defined: carboxyl, carbonyl, and water. Zn-N(h) indicates the distance from the zinc to the coordinating nitrogen of the His ligand. Zn-N(h) distances are given for the carboxyl, carbonyl, and water elecs in the carboxyl-His-Zn, carbonyl-His-Zn, and water-His-Zn motifs, respectively. T_4^* represents T_4 coordination lacking the zinc-bound water molecule.

Table 4B. *Elec-N(h) and Zn-N(h) distances according to coordination number and elec atom type in Elec-His-Zn motifs in structural zinc sites*

	Oxygen-N(h)			Zn-H(h)		
Coordinate number	Carboxyl	Carbonyl	Water	Carboxyl	Carbonyl	Water
T_4	2.65 (0.17), 5	2.84 (0.11), 9	2.86 (0.16), 17	2.14 (0.07), 5	2.12 (0.15), 9	2.04 (0.09), 16
T_6		2.90 (0.14), 2	3.04 (0.07), 2		2.10 (0.01), 2	2.08 (0.04), 2
Total	2.65 (0.17), 5	2.85 (0.11), 11	2.88 (0.16), 19	2.14 (0.07), 5	2.12 (0.13), 11	2.05 (0.09), 18

6.2B. Out-of-plane and in-plane angles of elec-N(h) and Zn-N(h) vectors relative to the histidine ring. From the values of the out-of-plane angles in Figure 7, carboxyl and carbonyl oxygen elec atoms in both catalytic and structural elec-His-Zn motifs are closer to planarity with the imidazole rings of the histidine residues than the water elec atoms. The planarity of carboxyl elec atoms with the ring is probably due to π electron delocalization extending over the ring and the carboxylate group. Water and carbonyl groups do not have such π systems. However, for carbonyl elec atoms, which are mostly from protein backbones, steric interactions between the backbone atoms and the imidazole ring tend to force the elec atom back to an almost coplanar arrangement with the ring. Water elec atoms show the largest out-of-plane angles with the largest standard deviations, which may be due to high temperature factors generally associated with their positions in the analyzed crystal structures. Similarly, as can be seen in Figure 7, in-plane angle distortions for all elec atom types are also small (Φ close to 180°), although standard deviations are large (up to 20°).

The distribution of oxygen elec atom types for all catalytic and structural sites in our dataset is presented in Figure 8. The results indicate that catalytic sites almost always include an elec-His-Zn motif that involves carbonyl or carboxyl oxygen atoms due to the electrostatic stabilizing role played by these groups. Structural sites still have a high proportion of elec-His-Zn motifs, but, in contrast to the catalytic zinc sites, the elec atoms involve a higher proportion of water oxygens, and a lower proportion of carbonyl and carboxyl oxygens. This suggests that the orientation of the elec atom with respect to the imidazole ring shows more variation in structural than catalytic zinc sites. This is confirmed by the superimposed elec-His-Zn motifs for 3N1O (meaning three nitrogen atoms and one oxygen atom coordinated to the zinc ion) catalytic (Fig. 9) and structural (Fig. 10) zinc sites, which give total RMS deviations (RMSDs) of 1.1 and 2.8 Å, respectively, for the elec

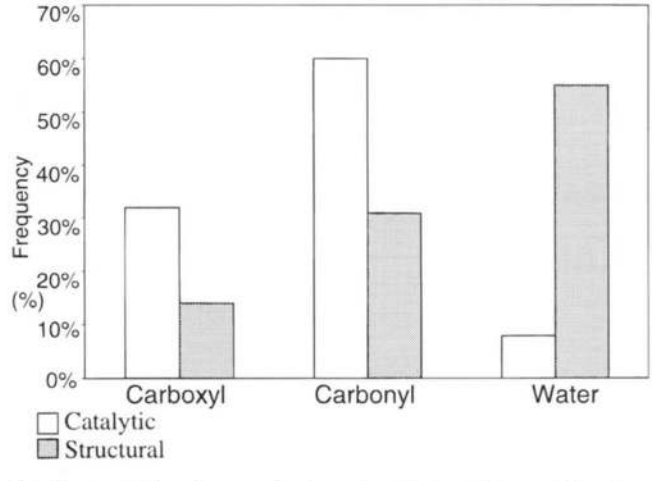


Fig. 8. Distribution of oxygen elec types in catalytic and structural zinc sites.

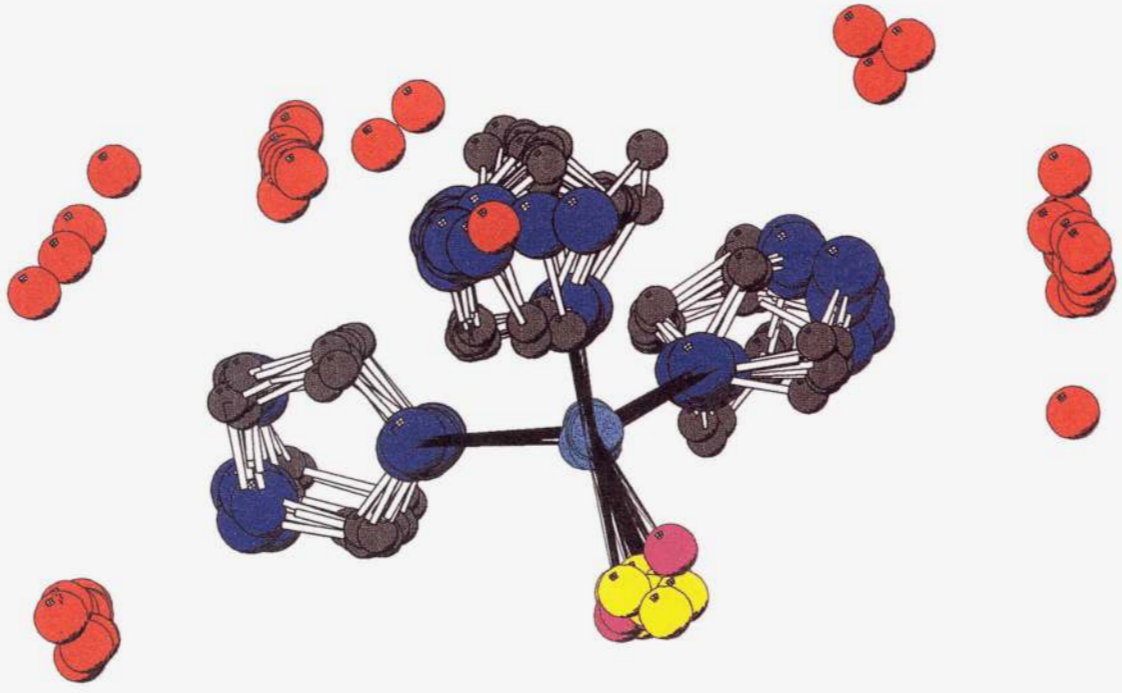


Fig. 9. Superposition of elec-His-Zn motifs for the 3N1O atom-coordination type in catalytic zinc sites in the analyzed proteins. The superposition was carried out using the program superpose, which is part of the CCP4 suite. The transformation matrix for the superposition was determined by superimposing the zinc ion and the coordinated histidine nitrogens. The following color scheme is used in the figure: Cyan, zinc; blue, histidine; yellow, zinc-coordinated solvent oxygens; red, elec atoms hydrogen-bonded to His; purple, zinc-coordinated oxygens from inhibitors; green, zinc-coordinated oxygens from protein residues.

atoms. These superpositions also indicate that the imidazole ring orientation is determined by the elec atom and serves to bring the histidine into the correct orientation for zinc binding (Argos et al., 1978; Christianson, 1991; Kiefer et al., 1995; Lesburg & Christianson, 1995).

The values in Figure 7 also show that the orientation of the zinc ion with respect to the ring is even more stringent in terms of out-of-plane and in-plane deviations, since electrostatic interactions are stronger for the cation, in agreement with previous results (Chakrabarti, 1990; Carrell et al., 1993). This is confirmed in Figures 9 and 10, which show that there is a small deviation in the position of the zinc-coordinated His nitrogens (total RMSDs of 0.10 and 0.15 Å in catalytic and structural 3N1O elec-His-Zn motifs, respectively). On the other hand, the zinc-coordinated oxygen atoms show larger deviations (total RMSDs of 0.50 and 0.35 Å for coordinated O atoms in catalytic and structural 3N1O, respectively) in agreement with the larger standard deviations for Zn–O distances.

From the above results concerning the orientation and properties of elec atoms in the elec-His-Zn motifs, it is clear that structural zinc binding sites are less rigid geometrically than catalytic zinc sites. Biologically, this is clearly important, as a small deviation in position of the catalytic zinc could have significant effects on catalysis. In structural sites, carboxyl elec-His-Zn motifs only occur in the superoxide dismutase family. In these enzymes, the zinc is only about 6.3 Å away from the catalytic copper ion and is believed to help orient the Cu site for the catalytic reaction; thus, a relatively rigid zinc binding framework is required.

Superpositions of carbonic anhydrase zinc binding sites in Figure 11 show that, within enzyme families, the zinc binding environment is geometrically fixed, including the orientation of elec-His-Zn motifs and the position of atoms from residues not directly coordinated to the zinc but involved in the catalysis, e.g., ND1 of His64 and OE1 of Glu106. Thus, differences within this family, probably caused by different refinement techniques/programs or resolution, are small. The resulting total RMSD is 0.33 Å between family members for all the atoms shown in Figure 11.

6.3. Sequence patterns associated with zinc binding sites

Sequence patterns associated with the zinc ligands and residues that provide elecs in the catalytic and structural zinc sites in our dataset were derived and compared with the sequence patterns from PROSITE (Bairoch et al., 1997), which are used to characterize the various protein families from the sequence. We find a number of interesting trends associated with the proteins that contain catalytic zincs, but much less for structural zincs. The observed trends are described in this section.

The sequence pattern for the zinc ligands and elecs for each family of catalytic-zinc containing enzymes is given in Table 5, which shows that consistent sequence patterns for zinc ligands and residues providing elecs can be observed for the thermolysin and other endo-peptidases. These proteins are all members of the same superfamily, denoted Metzincins in SCOP (Murzin et al., 1995) and a divergent evolutionary relationship has been established between them (Stocker et al., 1995). The thermolysin and thermolysin-like enzymes contain two domains, one is mainly alpha-beta (roll), the other is alpha (nonbundle). The zinc binding domain in the other endopeptidases belongs to the α-β class (aba sandwich).

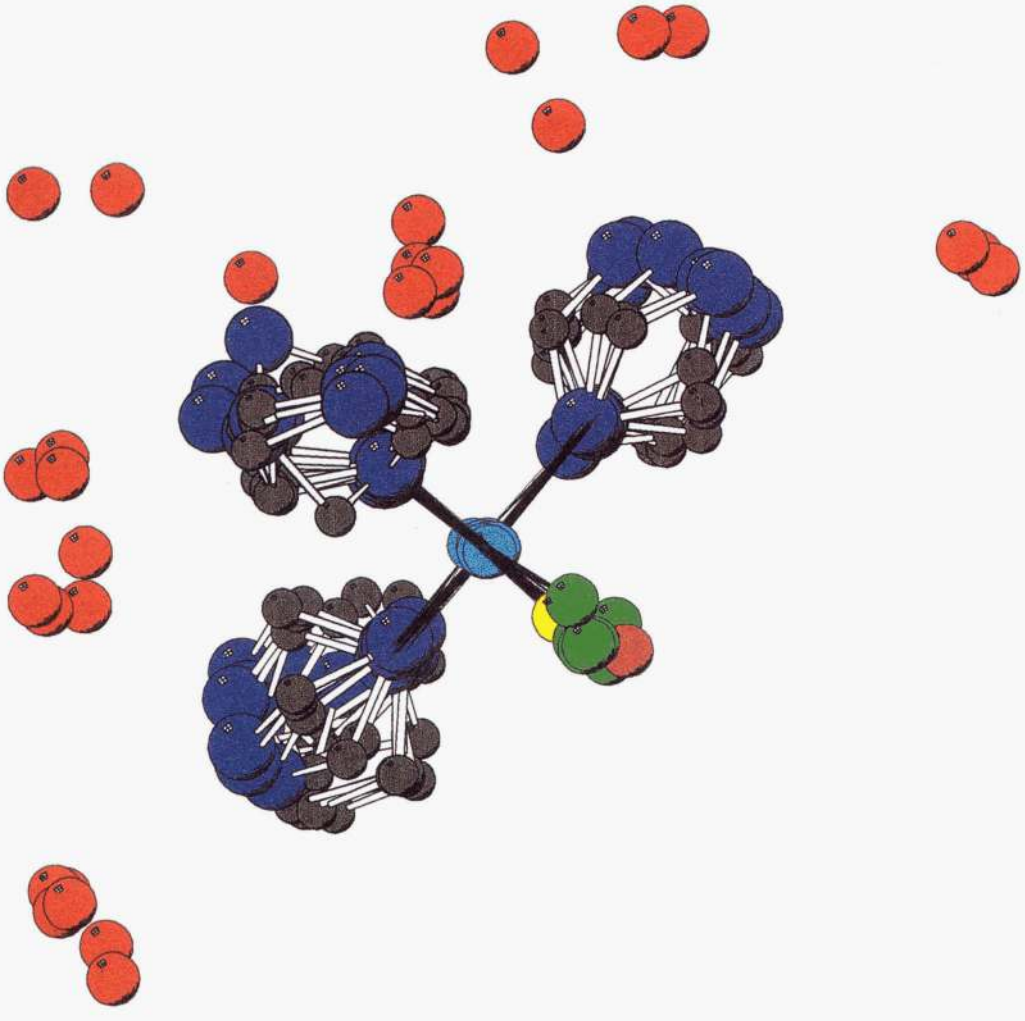


Fig. 10. Superposition of elec-His-Zn motifs for the 3N1O atom-coordination type in structural zinc sites in the analyzed proteins. The superposition was carried out using the program superpose, which is part of the CCP4 suite. The transformation matrix for the superposition was determined by superimposing the zinc ion and the coordinated histidine nitrogens. The following color scheme is used in the figure: cyan, zinc; blue, histidine; yellow, zinc-coordinated solvent oxygens; red, elec atoms hydrogen-bonded to His; purple, zinc-coordinated oxygens from inhibitors; green, zinc-coordinated oxygens from protein residues.

The consistent sequence pattern for the zinc binding site in the endo-peptidases includes two elec-His H-bonded pairs. The first His is H-bonded to an elec that is far removed in sequence (outer pair). This interaction brackets the second His-elec pair (inner pair) in sequence as shown in Table 5. The His and elec of the inner pair are separated by three residues in the endo-peptidases, except in the thermolysin and thermolysin-like enzymes in which the gap is larger (18 residues) because the His and its elec belong to different domains. This bracketed arrangement of His-elec H-bonded pairs is also observed in carbonic anhydrase, alkaline phosphatase, sonic hedgehog, and phosphomannose isomerase.

In the thermolysin and thermolysin-like enzymes, the two elec atoms belong to polar residues, which are in α-helices, as are the two histidines. In the other endo-peptidases, the His ligands are in α-helices, while both elecs are backbone carbonyl oxygens belonging to bulky hydrophobic residues (Stocker et al., 1995) in β-strands, which are highly conserved in each family as highlighted in Table 5. Conserved hydrophobic residues are thought to stabilize the interior of globular proteins. Our analysis suggests that the backbones of these conserved residues also provide elec atoms to stabilize the catalytic site in the correct orientation.

In the PROSITE database (Bairoch et al., 1997) one can find a general sequence pattern, H-E-X-X-H, that describes the zinc proteases, including all the endopeptidases. This pattern includes two zinc bound histidine ligands, but does not include any of the residues that provide elecs as indicated in Table 5. Since the endo-peptidases all display the bracketed arrangement of His-elec pairs, including the residues that provide elecs may lead to a more specific pattern. However, due to the variation in gap sizes between the histidine and the more widely spaced elec in the different subfamilies, one cannot derive a consensus sequence pattern for all the endo-peptidases that includes the residues providing elecs.

The sequence patterns for the other families in Table 5 are family specific. The PROSITE pattern for carbonic anhydrase includes only one of the zinc bound histidines and its elec. Our sequence pattern in Table 5 includes the three zinc bound histidines and their elecs. If we exclude the elec belonging to Asn as it is far

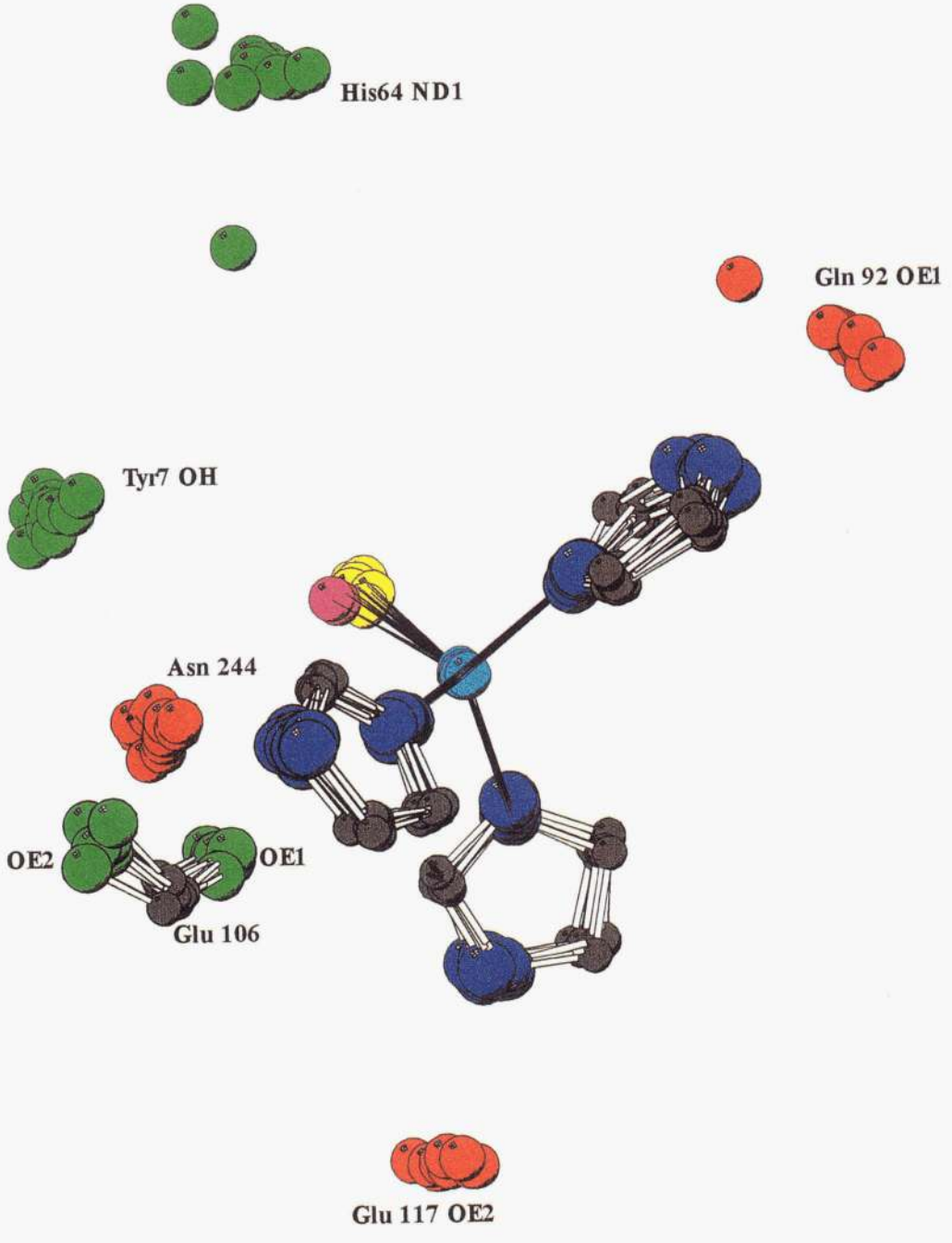


Fig. 11. Superposition of elec-His-Zn motifs and other catalytically important residues in the family of carbonic anhydrases in our dataset. The superposition was carried out using the program superpose, which is part of the CCP4 suite. The transformation matrix for the superposition was determined by superimposing the zinc ion and the coordinated histidine nitrogens. The His64 ND1 and Gln92 OE1 atoms that deviate significantly from the rest are from carbonic anhydrase I, whereas all others are from carbonic anhydrase II. The following color scheme is used in the figure: cyan, zinc; blue, histidine; yellow, zinc-coordinated solvent oxygens; red, elec atoms hydrogen-bonded to His; purple, zinc-coordinated oxygens from inhibitors; green, atoms from other catalytically important residues (His64, Tyr7, Glu106) that are not zinc bound.

away from its His (148 residues) and search SWISS-PROT with the remaining pattern, we can identify 25 of the 30 hits that are picked up with the PROSITE pattern. The remaining five sequences, which are carbonic anhydrase precursors, have a 16 residue gap between a His ligand and a Glu, rather the 20 residue gap in our pattern and, therefore, could not be picked up.

Discussion and conclusions

This study has extended earlier works (Armstrong, 1988; Chakrabarti, 1990; Christianson, 1991; Glusker, 1991), which analyzed the ligation geometries of zinc in protein crystal structures. It analyzed trends in geometrical and stereochemical parameters of zinc bind-

Table 5. *Sequence patterns for each family of catalytic zinc-containing enzymes*

Name of protein family	Sequence pattern	
Carboxypeptidase	H X X E X_{69} D X_{54} H	
Myramoyl-pentapeptide carboxypeptidase	G X_{17} H X_6 D X_{21} E X_{13} H	
Bacterial leucyl aminopeptidase	D X_{34} E X_{103} H	(zinc501)
	H X D X_{17} D X_{61} D	(zinc502)
Leucine aminopeptidase	D X_{76} D X E	(zinc488)
	K X_4 D X_{17} D X_{60} E	(zinc489)
Thermolysin and thermolysin like	H X X X H X_{18} E X X X	
Metalloproteases	H X X X H X X X X H X_{26}	
Astacin	H X X X H X X X X H X_{43} X X Y	
Atrolysin	H X X X H X X X X H X_{12}	
Matrix metalloproteases	H X X X H X X X X H X_6	
Neutral protease	H X X X H X X X X D X_8	
Carbnic anhydrase	Q X H X H X_{20} E X H X_{124} N	
Alcohol dehydrogenase	C X X D X_{17} H X_{106} C	
Alkaline phosphatase	T X_{226} D X X X H X_{78} Q X H	(zinc450)
	D X_{50} S X_{266} D H A	(zinc451)
Sonic hedgehog	E X_{73} G X_{12} H X_6 D X_{34} H	
Enolase	D X_{48} E X_{24} D	
Phosphomannose isomerase	Q X H X_{24} E X_{144} D X H X_{15} N	
Phosphate aodolase	E X_{18} H X H X_{60} H X L	
Metallo β-lactamase	N H X H X_{60} H X_{33} D	(zinc 1)
	P X_{49} D X_{77} C X_{41} H	(zinc 2)

Note: This table shows the sequence pattern for the zinc ligands and elecs for each family of enzymes containing catalytic zinc sites. The name of each enzyme family is given in column 1 and the sequence pattern in column 2. The residues are represented by their one-letter code and X denotes any amino acid residue. In the four families that contain enzymes with a multi-zinc site, a sequence pattern is given for each zinc. In these cases, the residue number of the appropriate zinc ion is given in parentheses after the pattern. The numbers in subscripts indicate the gap size in residues between zinc ligands and/or elecs. The symbol ⌣ denotes the elec-His hydrogen bonded pairs. The residues present in the PROSITE pattern are indicated by the symbol ^. Hydrophobic residues that provide the elecs and are conserved in sequence in the endo-peptidases are boxed.

ing sites using the increased number of proteins determined to high resolution. Our main conclusions can be summarized as follows.

1. From analysis of the geometry of the zinc primary coordination shell, we see: (a) Average Zn-O distances and their standard deviations are larger than those for Zn-N (h). (b) The coordination geometries are those expected for T_4, T_5, and T_6. (c) Deviations from ideal geometries are probably due to imprecisions in atomic coordinates and also to bidentate arrangements of, particularly, carboxylate groups, H-bonding between O coordinating ligands and protein residues and multi-zinc sites.
2. We find a very general elec-His-Zn motif that is almost universally present in our dataset with the following properties: (a) There are three types of oxygen elec atom: carboxyl, carbonyl, and water. (b) Catalytic zinc motifs include a high proportion of carboxyl and carbonyl elec atoms and are geometrically fixed motifs, thus providing a rigid zinc binding framework. (c) Structural zinc motifs include a much higher proportion of water elec atoms that show more geometrical variation; therefore, structural zincs are not held as rigidly. (d) In the endopeptidases, the sequence pattern in the zinc binding site includes two His-elec pairs, in which one pair brackets the other in sequence. Except for the thermolysin and thermolysin-like enzymes, the elecs belong to the backbones of bulky hydrophobic residues that are highly conserved. Our study, therefore, suggests that these residues play a role in stabilizing the zinc binding site and maintaining its geometry.

Our analysis clearly shows some bias due to the large number of carbonic anhydrases included which have very rigid zinc binding sites of 3 histidines + water and/or bound inhibitor. Nevertheless, sufficient proteins from other families included in our dataset display similar trends to justify our conclusions.

The results presented here are valuable for the following two purposes. First, they can be used as an aid in the validation of zinc binding site structures during model building. The geometrical parameters in the primary coordination sphere and elec-His-Zn motifs can be used as a guide to validate the data derived from crystallographic experiments and to specify improved parameters for structure refinement procedures.

The second purpose concerns the de novo design of zinc binding sites in proteins. In general, the design of a metal binding site is an important step toward the engineering of proteins with specific properties (Regan, 1995). To design a metal binding site in a protein, a geometric template of the site is derived from the analysis of sites in known protein structures. The protein structure at hand is then searched with the template to find a suitable anchoring site for this template within the structure (Hellinga & Richards, 1991; Gregory et al., 1993). In designing a zinc binding site, one should use a template that includes a set of elec-His-Zn motifs to stabilize the site. Catalytic zinc motifs require carboxyl oxygen elecs, which are coplanar with the rings and point along the nitrogen lone-pair direction, for greater rigidity. Water elec atoms are suitable for structural motifs. The coordination geometry of the zinc should be tetrahedral, including three zinc bound histidines, with N-Zn-N angles close to the ideal tetrahedral angle and Zn-N(h) distances, of about 2.0 Å, close to those in small zinc complexes. The open fourth site in the tetrahedral coordination geometry is usually occupied by water (or hydroxide) in catalytic zinc sites and Asp or Cys in structural zinc sites. In catalytic sites, the zinc bound water forms at least one H-bond with a protein residue, which is part of an extensive H-bonding network, essential for catalysis. Structural sites, in contrast, do not involve such extensive H-bonded residue clusters. The template should be placed in regions of the protein such that the zinc ligands and residues that provide elec groups are part of appropriate secondary structural elements with suitable solvent accessibilities.

Acknowledgments

The authors would like to thank Kim Henrick for useful discussions and help with software used in the analysis, and Andrew Wallace for providing us with his thesis chapter. We would also like to thank the University of Stirling for providing a studentship for one of us (K.N.) and the European Bioinformatics Institute for their system support.

References

Allen FH, Bellard S, Brice MD, Cartwright BA, Doubleday A, Higgs H, Hummelink T, Hummelink-Peters GG, Kennard O, Motherwell WDS, Rodgers JR, Watson DG. 1979. The Cambridge Crystallographic Data Centre: Computer-based search, retrieval, analysis and display of information. *Acta Crystallogr Sect B 35*:2331–2339.

Argos P, Garavito RM, Eventoff W, Rossmann MG, Brändén C. 1978. Similarities in active centre geometries of zinc-containing enzymes, proteases and dehydrogenases. *J Mol Biol 126*:141–158.

Armstrong WH. 1988. Metalloprotein crystallography. Survey of recent results and relationships to model studied. *ACS Symp Ser 372*:1–27.

Bairoch A, Bucher P, Hofmann K. 1997. The PROSITE database, its status in 1997. *Nucl Acid Res 25*:217–221.

Baker EN, Blundell TL, Cutfield JF, Cutfield SM, Dodson EJ, Dodson GG, Hodgkin DMC, Hubbard RE, Isaacs NW, Reynolds CD, Sakabe K, Sakabe N, Vijayan NM. 1988. The structure of 2ZN pig insulin crystals at 1.5 Å resolution. *Philos Trans R Soc Lond B Biol Sci 319*:369.

Baumann U. 1994. Crystal structure of the 50 kDa metallo protease from *Serratia marcescens*. *J Mol Biol 23*:244–251.

Baumann U, Wu S, Flaherty KM, McKay DB. 1993. Three-dimensional structure of the alkaline protease of *Pseudomonas aeruginosa*: A two-domain protein with calcium binding parallel beta roll motif. *EMBO J 12*:3357–3364.

Berg JM, Godwin HA. 1997. Lessons from zinc-binding peptides. *Annu Rev Biophys Struct 26*:357–371.

Berg JM, Shi YG. 1996. The galvanization of biology: A growing appreciation for the roles of zinc. *Science 271*:1081–1085.

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Simanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival system for macromolecular structures. *J Mol Biol 112*:535–542.

Berry MB, Phillips GN. 1997. Crystal structures of *Bacillus stearothermophilus* adenylate kinase with bound Ap 5 A reveals an intermediate lid position and a fully coordinated Mg-2. *Biophys J 72*:WAMB7.

Blundell TL, Pitts JE, Tickle IJ, Wood SP, Wu CW. 1981. X-ray analysis 1.4 Å resolution of avian pancreatic polypeptide small globular protein hormone. *Proc Natl Acad Sci USA 78*:4175–4179.

Bock CW, Katz AK, Glusker JP. 1995. Hydration of zinc ions: A comparison with magnesium and beryllium ions. *J Am Chem Soc 117*:3754–3765.

Bode W, Gomis-Ruth FX, Huber R, Zwilling R, Stocker W. 1992. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature 358*:164–167.

Bode W, Reinemer P, Huber R, Kleine T, Schnierer S, Tschesche H. 1994. The X-ray crystal structure of the catalytic domain of human neutrophil collagenase inhibited by a substrate analogue reveals the essential for catalysis and specificity. *EMBO J 13*:1263–1269.

Botos I, Scapozza L, Zhang DC, Liotta LA, Meyer EF. 1996. Batimasta, a potent matrix metalloproteinase inhibitor, exhibits an unexpected mode of binding. *Proc Natl Acad Sci USA 93*:2749–2754.

Bouckaert J, Poortmans F, Wyns L, Loris R. 1996. Sequential structural-change upon zinc and calcium-binding to metal-free concanavalin A. *J Biol Chem 271*:16144–16150.

Browner MF, Smith WW, Castelhano AL. 1995. Matrilysin-inhibitor complexes: Common themes among metalloproteases. *Biochemistry 34*:6602–6610.

Burmeister WP, Cottaz S, Driguez H, Iori R, Palmieri S, Henrissat B. 1997. The crystal structure of sinapis alba myrosinase and a covalent glycosyl-enzyme intermediate provide insights into the substrate recognition and active-side machinery of an S-glycosidase. *Structure 5*:663–675.

Carrell AB, Shimoni L, Carrell CJ, Bock CW, Murray-Rust P, Glusker JP. 1993. The stereochemistry of the recognition of nitrogen-containing heterocycles by hydrogen-bonding and by metal-ions. *Receptor 3*:57–76.

Carver JA, Bradbury JH. 1984. Assignment of H-1-NMR resonances of histidine and other aromatic residues in met-myoglobin, cyano-myoglobin, oxy-myoglobin, and carbon monoxy myoglobin. *Biochemistry 23*:4890–4905.

Carugo KD, Battistoni A, Carri MT, Polticelly F, Desidery A, Rotilio G, Coda A, Wilson KS, Bolognesi M. 1996. 3-Dimensional structure of Xenopus-Laevis Cu,Zn superoxide-dismutase-B determined by X-ray crystallography at 1.5 angstrom resolution. *Acta Crystallogr D 52*:176–188.

Chakrabarti P. 1990. Geometry of interaction of metal ions with histidine residues in protein structures. *Protein Eng 4*:57–63.

Chevrier B, Schalk C, D'Orchymont H, Rondeau J-M, Moras D, Tarnus C. 1994. Crystal structure of aeromonal-proteolytica aminopeptidase: A prototypical member of the co-catalytic zinc enzyme family. *Structure 2*:283–291.

Cho H, Ramaswamy S, Plapp BV. 1997. Flexibility of liver alcohol dehyrogenase in stereoselective binding of 3-buylthiolane 1 oxides. *Biochemistry 36*:382–389.

Christianson DW. 1991. Structural biology of zinc. *Adv Protein Chem 42*:281–355.

Christianson DW, Alexander RS. 1989. Carboxylate histidine zinc interactions in protein-structure and function. *J Am Chem Soc 111*:6412–6419.

Christianson DW, Alexander RS. 1990. Another catalytic triad. *Nature 346*:225.

Christianson DW, Lipscomb WN. 1989. Carboxypeptidase-A. *Acc Chem Res 22*:62–69.

Cleasby A, Wonacott A, Skarzynski T, Hubbard RE, Davies GJ, Proudfoot AEI, Bernard AR, Payton MA, Wells TNC. 1996. The X-ray crystal structure of phosphomannose isomerase from candida-albicans at 1.7 angstrom resolution. *Nat Struct Biol 3*:470–479.

Coleman JE. 1992. Zinc proteins: Enzymes, storage proteins, transcription factors, and replication proteins. *Annu Rev Biochem 61*:897–946.

Concha NO, Rasmussen BA, Bush K, Herzberg O. 1996. Crystal structure of the wide-spectrum binuclear zinc beta-lactamase from bacterorides-fragilis. *Structure 4*:823–836.

Cotton FA, Wilkinson G. 1988. *Advanced inorganic chemistry*. New York: Wiley.

Dauter Z, Wilson KS, Sieker LC, Moulis JM, Meyer J. 1996. Zinc- and iron-rubredoxins from *Clostridium pasteurianum* at atomic resolution: A high-precision model of a ZnS4 coordination unit in a protein. *Proc Natl Acad Sci USA 93*:8836–8840.

Derewenda U, Derewenda Z, Dodson EJ, Dodson GG, Reynolds CD, Smith GD, Sparks C, Swenson D. 1989. Phenol stabilizes more helix in a new symmetrical zinc insulin hexamer. *Nature 338*:594–596.

Dreyer MK, Shulz GE. 1996. Refined high-resolution structure of the metal-ion dependent L-fuculose-1-phosphate aldolase Class-II from *Escherichia coli*. *Acta Crystallogr D 52*:1082–1091.

Elrod-Erickson M, Rould MA, Nekludova L, Pabo CO. 1996. Zif268 protein-DNA complex refined at 1.6 Å: A model system for understanding zinc-finger DNA interactions. *Structure 4*:1171–1180.

Eriksson AE, Jones TA, Lilsjas A. 1986. Zinc enzymes. In: Bertini I, Luchinat C, Maret W, Zeppezauer M, eds. *Progress in inorganic biochemistry and biophysics. Vol. 1*. Cambridge, MA: Burkhauser. pp 317–328.

Eriksson AE, Jones TA, Lilsjas A. 1988. Refined structure of human carbonic anhydrase-II at 2.0 Å resolution. *Proteins 4*:274–282.

Ghuysen JM, Lamottebrasseur J, Joris B, Shockman GD. 1994. Binding site-shaped repeated sequences of bacterial wall peptidoglycan hydrolases. *FEBS Lett 342*:23–28.

Glusker JP. 1991. Structural aspects of metal liganding to functional groups in proteins. *Adv Protein Chem 42*:1–76.

Gooley PR, O'Connell JF, Marcy AI, Cuca GC, Salowe SP, Bush BL, Hermes JD, Esser CK, Hagmann WK, Springer JP. 1994. The NMR structure of the inhibited catalytic domain of human stromelysin-1. *Nat Struct Biol 1*:111–118.

Grams F, Reinemer P, Powers JC, Kleine T, Peiper M, Tschesche H, Huber R, Bode W. 1995. X-ray structures of human neutrophil collagenase complexed with peptide hydroxamate and peptide thiol inhibitors: Implications for substrate-binding and rational drug design. *Eur J Biochem 228*:830–841.

Gregory DS, Martin ACR, Cheetham JC, Rees AR. 1993. The prediction and characterization of metal binding sites in proteins. *Protein Eng 6*:29–35.

Hakansson K, Wehnert A. 1992. Structure of cobal carbonic anhydrase complexed with bicarbonate. *J Mol Biol 228*:1212–1218.

Hall TM, Porter JA, Beachy PA, Leahy DJ. 1995. A potential catalytic site revealed by the 1.7 Å crystal structure of the amino-terminal signalling domain of Sonic Hedgehog. *Nature 378*:212–216.

Hamada K, Hiramatsu H, Katsuya Y, Hata Y, Matsuura Y, Katsube Y. 1993. Structural analysis of serratia protease. *Acta Crystallogr A 49*:153.

Harrison SC. 1991. A structural taxonomy of dna-binding domains. *Nature 353*:715–719.

Hellinga HW, Richards FM. 1991. Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modelling of sites with pre-defined geometry. *J Mol Biol 222*:763–785.

Henrick K. 1997. Crystal macromolecule files-description [on-line]. http://www2.ebi.ac.uk/pdb/ebi/html/macmol_doc.shtml.

Holden HM, Tronrud DE, Monzingo AF, Weaver LH, Matthews BW. 1987. Slow-binding and fast-binding inhibitors of thermolysin display different modes of binding-crystallographic analysis of extended phosphonamidate transitions state analogs. *Biochemistry 26*:8542–8553.

Holland DR, Tronrud DE, Pleyk HW, Flaherty KM, Start W, Jansonius JN, McKay DB, Matthews BW. 1992. Structural comparison suggests that ther-

molysin and related proteases undergo hinge-bending motion during catalysis. *Biochemistry 31*:11310–11316.

Holmes MA, Matthews BW. 1982. Structure of thermolysin refined at 1.6 Å resolution. *J Mol Biol 160*:623–639.

Hurley TD, Bosron WF, Stone CL. 1994. Structures of 3 human β alcohol dehydrogenase variants; correlations with their functional differences. *J Mol Biol 239*:415–429.

Katz AK, Glusker JP, Beebe SA, Bock CW. 1996. Calcium ion coordination: A comparison with that of beryllium, magnesium, and zinc. *J Am Chem Soc 118*:5752–5763.

Kiefer LL, Paterno SA, Fierke CA. 1995. Hydrogen-bond network in the metal-binding site of carbonic anhydrase enhances zinc affinity and catalytic efficiency. *J Am Chem Soc 117*:6831–6837.

Kim EE, Wyckoff HW. 1991. Reaction mechanism of alkaline phosphatase based on crystal structures. Two-metal ion catalysis. *J Mol Biol 218*:449–464.

Kurisu G, Kinoshita T, Sugimoto A, Nagara A, Kai Y, Kasai N, Harada S. 1997. Structure of the zinc endoprotease from streptomyces caespitosus. *J Biochem 121*:304–308.

Lebioda L, Stec B. 1989. Crystal structure of holoenolase refined at 1.9 Å resolution-trigonal bipyramidal geometry of the cation binding-site. *J Am Chem Soc 111*:8511–8513.

Lesburg CA, Christianson DW. 1995. X-ray crystallographic studies of engineered hydrogen-bond networks in a protein-zinc binding-site. *J Am Chem Soc 117*:6838–6844.

Lilsjas A, Håkansson K, Jonsson BH, Xue Y. 1994. Inhibition and catalysis of carbonic-anhydrase-recent crystallographic analyses. *Eur J Biochem 219*: 1–10.

Linskog S, Lilsjas A. 1993. Carbonic anhydrase and the role of orientation in catalysis. *Curr Opin Struct Biol 3*:915–920.

Lippard SJ, Burger AR, Ugurbil K, Pantoliano MW, Valentine JS. 1977. Nuclear magnetic resonance and chemical modification studies of bovine erythrocyte superoxide dismutase: Evidence for zinc-promoted organization of the active site structure. *Biochemistry 16*:1136–1141.

Lipscomb WN, Sträter N. 1996. Recent advances in zinc enzymology. *Chem Rev 96*:2375–2433.

Lovejoy B, Hassell AM, Luther MA, Weigl D, Jordan SR. 1994. Crystal structures of recombinant 19-kDa human fibroblast collagenase complexed to itself. *Biochemistry 33*:8207–8217.

McGregor MJ, Islam SA, Sternberg MJE. 1987. Analysis of the relationship between side-chain conformation and secondary structure in globular proteins. *J Mol Biol 198*:295–310.

Murzin AG, Brenner SE, Hubbard T, Chothia C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol 247*:536–540.

Papageorgiou AC, Acharya KR, Shapiro R, Passalacqua EF, Brehm RD, Tranter HS. 1995. Crystal structure of the superantigen enterotoxin C2 from *Staphylococcus aureusn* reveals a zinc-binding site. *Structure 3*:769–779.

Perutz MF, Gronenborn AM, Clore GM, Fogg JH, Shih DTB. 1985. The pK_a values of 2 histidine residues in human haemoglobin, the Bohr effect and the dipole-moments of alpha-helices. *J Mol Biol 183*:491–498.

Ramaswamy S, Eklund H, Plapp BV. 1994. Structures of horse liver alcohol-dehydrogenase complexed with NAD+ and substituted benzyl alcohols. *Biochemistry 33*:5230–5237.

Ramaswamy S, El Ahmad M, Danelsson O, Jornwall H, Eklund H. 1996. Crystal structure of cod liver class I alcohol dehyrogenase: Substrate pocket and structurally variable segments. *Protein Sci 5*:663–671.

Rees DC, Lewis M, Lipscomb WN. 1983. Refined crystal structure of carboxypeptidase-A at 1.54 Å resolution. *J Mol Biol 168*:367–387.

Regan L. 1995. Protein design: Novel metal-binding sites. *TIBS 20*:280–285.

Ren J, Stuart DI, Acharya KR. 1993. Alpha-lactalbumin possesses a distinct zinc binding site. *J Biol Chem 268*:19292–19298.

Romier C, Reuter K, Suck D, Ficner R. 1996. Crystal structure of transfer RNA-guanine transglycosylase-RNA modification by base-exchange. *EMBO J 15*:2850–2857.

Silverman DN. 1991. The catalytic mechanism of carbonic anhydrase. *Can J Bot 69*:1070–1078.

Silverman DN, Lindskog S. 1988. The catalytic mechanism of carbonic anhydrase: Implications of a rate-limiting proteolysis of water. *Acc Chem Res 21*:30–36.

Smith GD, Swenson DC, Dodson EJ, Dodson GG, Reynolds CD. 1984. Structural stability in the 4-zinc human insulin hexamer. *Proc Natl Acad SciUSA 81*:7093–7097.

Spurlino JC, Smallwood AM, Carlton DD, Bank TM, Vavra KJ, Johnson JS, Cook ER, Falvo J, Wahl RC, Pulvino TA. 1994. 1.56 Å Structure of mature truncated human fibroblast collagenase. *Proteins 19*:98–109.

Stark N, Pauptit TA, Wilson KS, Jansonius JN. 1992. The structure of neutral protease from *Bacillus cereus* at 0.2-NM resolution. *Eur J Biochem 207*:781–791.

Stocker W, Grams F, Baumann U, Reinemer P, Gomisruth FX, McKay DB, Bode W. 1995. The metzincins—Topological and sequential relations between the astacins, adamalysins, serralysins and matrixins (collagenases) define a superfamily of zinc-peptidases. *Protein Sci 4*:823–840.

Sträter N, Lipscomb WN. 1995a. 2-Metal ion mechanism of bovine lens aminopeptidase: Active-site solvent structure and binding mode of L-leucinal, a gem-diolate transition-state analog, by X-ray crystallography. *Biochemistry 34*:14792–14800.

Sträter N, Lipscomb WN. 1995b. Transition-state analog L-leucinephosphonic acid bound to bovine lens leucine aminopeptidase: X-ray structure at 1.65 Angstrom resolution in a new crystal form. *Biochemistry 34*:9200–9210.

Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC. 1982. Determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J Mol Biol 160*:181–217.

Thayer MM, Flaherty KM, McKay DB. 1991. 3-Dimensional structure of the elastase of pseudomonas-aeruginosa at 1.5 Å resolution. *J Biol Chem 266*:2864–2871.

Vallee BL, Auld DS. 1990. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry 29*:5647–5659.

Vallee BL, Auld DS. 1993. New perspective on zinc biochemistry: Cocatalytic sites in multi-zinc enzymes. *Biochemistry 32*:6493–6500.

Vedani A, Huhta DW. 1990. A new force-field for modelling metalloproteins. *J Am Chem Soc 112*:4759–4767.

Wallace A. 1996. The role of His in metal binding sites. Chapter 7 [PhD thesis]. London, UK: University College London.

Xie X, Kokubo T, Cohen SL, Mirza UA, Hoffmann A, Chait BT, Roeder RG, Nakatani Y, Burley SK. 1996. Structural similarity between TAFs and the heterotetrameric core of the histone octamer. *Nature 380*:316–322.

Zhang D, Botos I, Gomis-Rueth F-X, Doll R, Blood C, Njoroge FG, Fox JW, Bode W, Meyer EF. 1994. Structural interaction of natural and synthetic inhibitors with the venom melalloproteinase, Atrolysin C Form D. *Proc Nat Acad Sci USA 91*:8447–8451.

Zhang GG, Kazanietz MG, Blumberg PM, Hurley JH. 1995. Crystal structure of the cys 2 activator-binding domain of protein kinase C-delta in complex with phorbol ester. *Cell 81*:917–924.