Flexibility of Metal Binding Sites in Proteins on a Database Scale

Mariana Babor,[1] Harry M. Greenblatt,[2] Marvin Edelman,[1] and Vladimir Sobolev[1]*
[1]*Department of Plant Sciences, Weizmann Institute of Science, Rehovot, Israel*
[2]*Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel*

ABSTRACT **Protein metal binding sites in the pre-bound (apo) state, and their rearrangements upon metal binding were not analyzed previously at a database scale. Such a study may provide valuable information for metal binding site prediction and design. A high resolution, nonredundant dataset of 210 metal binding sites was created, containing all available representatives of apo–holo pairs for the most populated metals in the PDB. More than 40% of the sites underwent rearrangements upon metal binding. In 30 cases rearrangements involved the backbone. The tendency for side-chain rearrangement inversely correlates with the number of first-shell residues. Analysis of side-chain reorientations as a result of metal binding showed that in 95% of the rigid-backbone binding sites at most one side chain moved. Thus, in general, part of the first coordination shell is already in place in the pre-bound form. The frequencies of side-chain reorientation directly correlated with metal ligand flexibility and solvent accessibility in the apo state. Proteins 2005;59:221–230.** © 2005 Wiley-Liss, Inc.

Key words: metal ligand; pre-bound state; apo–holo pairs; side-chain reorientation; first-shell residues

INTRODUCTION

More than one third of the known proteins require metal ions as cofactors to perform their functions.[1,2] Metal sites in proteins serve a great variety of purposes, including increased structural stability or acting as cofactors in catalytic and regulatory processes. Protein function is the result of dynamic processes,[3,4] which involve conformational changes affecting the protein structure at a variety of levels: motion of a few atoms or amino acid side chains, movements affecting a few secondary structural elements, larger movements within a domain, or movements affecting the orientation of one domain relative to another.[5] Metal binding may involve any of these kind of conformational changes. For example, carboxypeptidase A, a Zn-dependent hydrolase, undergoes small structural rearrangements in the transition from apo to metal-bound form. A salt bridge formed between the metal ligand His 196 and a glutamate is broken and His 196 rotates 110° around the Cβ-Cγ bond to ligate the Zn.[6,7] On the other hand, transferrin, a transport protein with considerable intrinsic flexibility, undergoes major conformational changes involving hinge motions during the uptake and release of iron. These structural changes play a key role in receptor recognition.[8]

Metal-binding sites in proteins are very diverse, varying in their coordination numbers,[9,10] geometries,[11–15] metal preferences,[16,17] and ligands.[18–20] The latter include backbone carbonyl oxygens, side-chain groups of polar residues (mainly aspartate, glutamate, histidine) and cysteine, as well as water molecules. Despite this variety, it has been observed that many metal sites in proteins are centered in a shell of hydrophilic ligands, surrounded by a hydrophobic area.[21] Principal coordination numbers observed in metalloproteins range from three to eight, with six as the most common value. Mg and Mn have a main coordination number equal to six;[16] whereas Zn and Ca show greater coordination flexibility, with values ranging from four to six and six to eight, respectively. Calcium almost always binds oxygens, while the preference of Zn, a softer metal ion, depends upon the ligation geometry. When the coordination number is low, Zn binds nitrogen and sulfur, preferring oxygen for a coordination number equal to six.[10,14,18] The ability of metals to modify the coordination number is also related to the role metals play in biological systems; Mg ions generally play a structural role, while Zn ions are more likely to participate directly in catalytic processes, with possible changes in the coordination number during the reaction.[10,22]

There is considerable interest in modelling and designing metal binding sites.[23] In order to bind a certain metal ion, a designed site must either closely resemble the optimal coordination geometry of the ion, or must be able to reorient and attain the required geometry.[24] Thus, a study of protein pre-bound conformations and the rearrangements they undergo upon metal binding, at a database scale, may provide valuable information for metal binding site prediction and design. Here, we create a nonredundant dataset representing all available apo–holo pairs for the most abundant metals in the PDB[25,26] and

The Supplementary Materials referred to in this article can be found at http://www.interscience.wiley.com/jpages/0887-3585/suppmat

*Correspondence to: Dr. Vladimir Sobolev, Dept. Plant Sciences, Weizmann Institute of Science, Rehovot 76100 Israel. E-mail: vladimir.sobolev@weizmann.ac.il

Received 18 July 2004; Accepted 16 November 2004

Published online 22 February 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20431

analyze in detail the pre-bound conformations and structural effects that result from transition to the bound state.

METHODS

Creation of the a Nonredundant Dataset of Apo–Holo Pairs

A metallo–protein database was built of apo–holo PDB file pairs (resolution higher than 2.5 Å, R factor $\leqslant$ 23%) containing Ca, Mg, Zn, Cu, Fe, Mn, Ni, and Co. Apo–holo pairs involving more weakly binding monovalent Na and K ions were not included at this stage since these are frequently the result of crystallization conditions rather than reported in vivo functions.

Identifying apo–holo pairs

Files with proteins that bind directly to selected metals were extracted from the PDB. Metals complexed to prosthetic groups or with a coordination number smaller than three were excluded. Metal ligands were defined according to the MDB metallodatabase[27] and the domain family according to the SCOP database,[28] version 1.63. In order to search for a particular apo partner, sequence comparisons were performed between each metal-containing domain (holo form) and all other members of the SCOP family. Domain sequences were obtained through the SCOP-linked web tool, Astral.[29] For holo proteins having more than one domain, sequence comparisons were performed only on those domains that contain the metal ligands. Metal sites where the ligands belong to several domains were not analyzed. Only apo–holo pair structures sharing 98% or more sequence identity were retained.

Identifying ligand residues in apo forms

Following sequence alignment, the residues in the putative apo forms corresponding to the metal ligands in the holo form were identified. This proved to be difficult, as it was necessary to take into account modified amino acids, atoms not resolved in the x-ray structures, inconsistencies in the PDB formats and gaps in the alignments, in order to include all possible pairs and to ensure that the assignment of the correspondent ligands was correct. As a further step, contacts established between the ligand residues (first shell) in the putative apo forms and any other residues or heterogroups were sought using the CSU program.[30] If there were no contacts with metals, and neither first-shell residues, nor those in contact with them (second shell), were mutated, the apo protein was accepted as a partner for the holo protein and the pair was retained.

Exclusion of problematic pairs

Nevertheless, pairs with problematic electron densities were excluded where one or more atoms of any amino acid ligand (in apo or holo form) were not resolved or had occupancy smaller than one. In addition, pairs were excluded where ligands interacted with large hetero groups ($>$ five atoms) in only one of the forms. Finally, some interactions are not detected when analysis is restricted to the asymmetric unit of the PDB entry, but appear when neighboring molecules in the crystal generated by symmetry operators are considered (E. Eyal, pers. comm.). Thus, to avoid false apo proteins, we excluded those pairs where the apo and holo forms come from the same PDB file. Finally, a small number of pairs were discarded that showed large "domain root-mean-square deviation (RMSD)" and were documented in the literature as not being due to metal binding.

Calculating domain RMSD

Domain RMSD was calculated by superimposing Cα atoms of apo and holo domains.[31] Large domain RMSD could indicate large structural differences between apo and holo forms caused by metal binding or by other factors (e.g., different crystallization procedures, other hetero groups far from the binding site, large intrinsic flexibility, etc.). Analysis of domain RMSD distribution among apo–holo pairs showed a decreasing distribution that reached a plateau at about 0.8 Å, which was used as a threshold. This value is within the cut-off range of 0.4 to 1.0 Å, obtained in previous studies.[32,33]

Removing redundancy

Because some proteins are overrepresented in the PDB with respect to others, holo protein domains sharing more than 40% sequence identity were grouped and, as is commonly done, the one with the highest resolution was taken as the representative structure. However, analysis of the excluded proteins showed that two of them had metal binding sites located in different regions of the molecule so these were re-included. In cases where there are several apo forms per holo, the one with the highest resolution was retained. Finally, a total of 210 pairs of apo–holo forms were obtained which are listed in Supplementary Material.

Analysis of Backbone Conformational Changes

Backbone conformational changes upon metal binding were analyzed by measuring domain RMSD, binding site Cα RMSD, and maximum Cα displacement. Binding site Cα RMSD was calculated by superimposing first- and second-shell Cα atoms, since it is not possible to superimpose only first-shell ligands for binding sites with fewer than three amino acid residues. RMSD was then calculated only for first-shell ligands. The superimposition was carried out using a program (Potapov, unpublished) that provides an analytical deterministic solution for the problem of finding the minimum RMSD and is based on the singular value decomposition method.[34] Maximum Cα displacement was obtained by calculating the maximum displacement between pairwise distances of any two Cα atoms of first-shell ligands in the apo and holo forms.

Analysis of Side-Chain Conformational Changes

A side-chain conformational change between pair members was considered to have occurred if an absolute difference between two angles larger than 40° exists for at least one dihedral angle of the residue in question. This cutoff value has been widely used.[35–37] Certain amino acids are symmetrical in their terminal dihedral angle:

TABLE I. Number of Apo-holo Pairs and SCOP Families Represented for Each Metal

Metal	Ca	Mg	Mn	Zn	Cu	Ni	Fe	Co
Number pairs	58	37	32	48	8	9	7	11
Number families	37	29	23	33	5	8	7	11

Glu (at χ_3), Asp (at χ_2), Tyr (at χ_2) and Phe (at χ_2). In addition, His (at χ_2), Asn (at χ_2) and Gln (at χ_2) do not exhibit a real symmetry but the position of the C, N, or O atoms in some cases might be wrongly assigned, causing incorrect calculation of the associated dihedral angles.[33] As a consequence, there are two possible values for the differences between the dihedral angles. In all these cases the smaller value was taken as the actual difference. The probability for dihedral angle change was calculated as in Eyal et al:[38]

$$P_i = \frac{N_c^i}{N_t^i} \pm \sqrt{\frac{F_c \times (1 - F_c)}{N_t^i}} \quad (1)$$

$$F_c = \frac{\sum_{i=a\alpha} N_c^i}{\sum_{i=a\alpha} N_t^i} \quad (2)$$

where N_c^i is number of amino acids of type i that undergo conformational change and N_t^i is the total number of amino acids of type i.

Solvent Accessibility

Side-chain solvent accessibility was calculated using CSU software.[30] Relative accessibility of side-chain atoms for any residue was defined as the ratio between absolute solvent accessibility in the experimental structure and solvent accessibility in the extended conformation of Gly-Gly-X-Gly-Gly.[38]

B-factor Normalization

Observed B-factor values may differ considerably between different protein structures. Therefore, for statistical analysis, values were normalized[38,39] as follows: $B' = (B - \langle B \rangle)/\sigma(B)$, where, for a given PDB file, $\langle B \rangle$ and $\sigma(B)$ are the average and standard deviation of the B-factors, respectively.

RESULTS AND DISCUSSION

A Nonredundant Dataset of Apo–Holo pairs

A nonredundant dataset was compiled containing all possible representatives of apo–holo pairs for the eight most populated metals in the PDB. A total of 210 pairs were obtained (see Supplementary Materials). The number of pairs for each metal and the number of SCOP[28] families in which they are represented are shown in Table I.

Backbone Rearrangements Upon Metal Binding

Metal binding can result in different degrees of structural rearrangements. Backbone motions were analyzed by measuring three parameters: domain RMSD, binding-site RMSD and maximum Cα displacement. None of these parameters alone can fully characterize the wide spectrum of possible motions, but a combination of them can approximate it. Domain RMSD indicates global changes encompassing whole domains, binding-site RMSD reflects first-shell motions and to a smaller extent second-shell motions (as detailed in Methods), whereas maximum Cα displacement considers only first-shell residues and is more sensitive than RMSD to changes in which a limited number of points are involved.

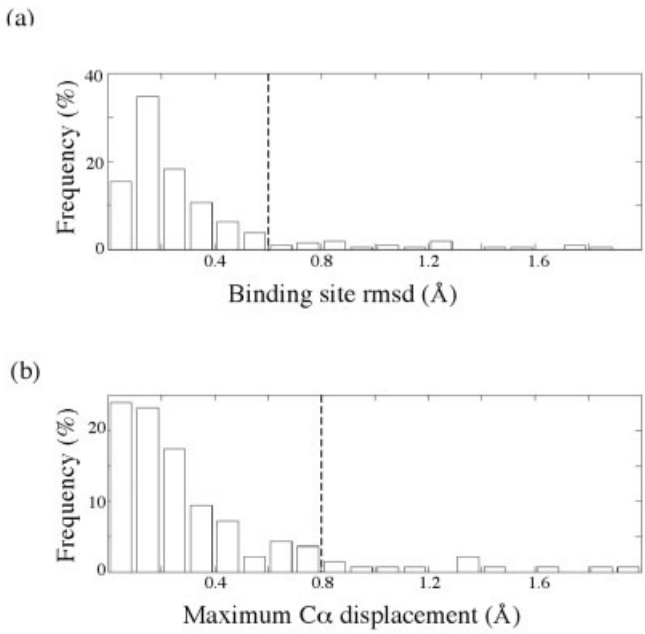

Fig. 1. Distribution of binding site RMSD and maximum Cα displacement. (**a**) Binding site RMSD for all 210 sites. A plateau is reached at 0.6 to 0.7 Å. (**b**) Maximum Cα displacement for the 147 binding sites having two or more amino acid ligands. A plateau is reached at 0.7 to 0.9 Å.

Analysis of binding-site RMSD and maximum Cα displacement distributions were performed in order to establish suitable cutoffs to determine whether backbone changes occurred. Binding-site RMSD distribution reached a plateau at 0.6–0.7 Å [Fig. 1(A)]. Maximum Cα displacement showed a similar distribution, reaching a plateau at 0.7–0.9 Å [Fig. 1(B)]. Selecting any cutoff value within these ranges affects only a small number of structures. Thus, for our dataset, 0.6 Å and 0.8 Å were chosen as the cutoff values to be considered as significant backbone rearrangements for binding-site RMSD and maximum Cα displacement, respectively. The selected value for maximum Cα displacement is in agreement with the values obtained by Betts and Sternberg,[33] who analyzed Cα displacements of particular atoms.

Analysis of binding site RMSD and maximum Cα displacement upon metal binding showed that about 14% of the binding sites underwent backbone rearrangements. These are termed mobile-backbone binding sites and the remainder, rigid-backbone sites. Large backbone rearrangement in most cases occurs upon Ca binding when backbone atoms serve as ligands. Figure 2 depicts the relationship between maximum Cα displacement and binding-site RMSD for metal binding sites in the dataset with

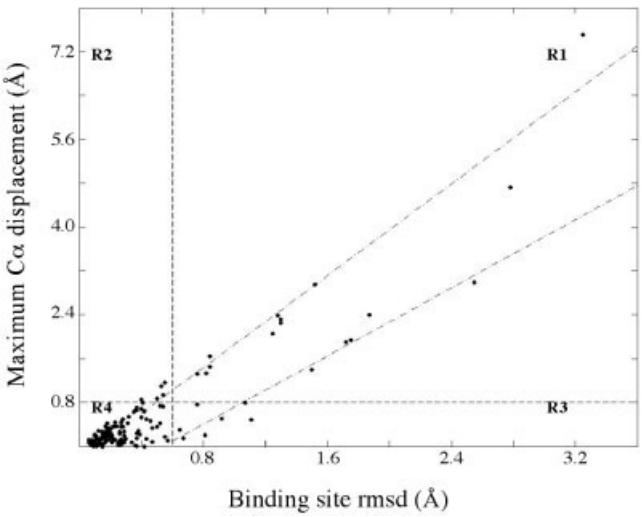

Fig. 2. Backbone rearrangements in the dataset upon metal binding. R1, metal binding sites undergoing backbone motions with significant values of both binding site RMSD and Cα maximum displacement; R2, backbone motions with significant values of only maximum Cα displacement; R3, backbone motions with significant values of only binding site RMSD; R4, binding sites that do not show any significant backbone rearrangements. Points in R1, R2, and R3 appear to fall along two lines (see text).

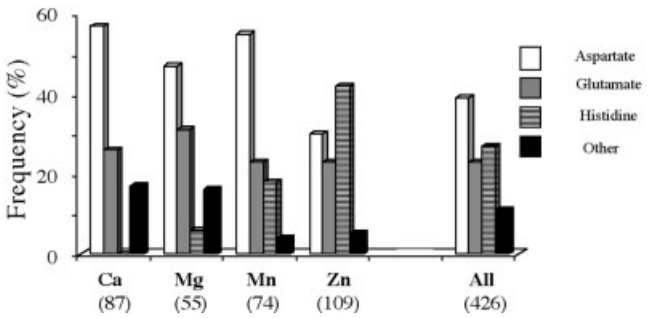

Fig. 3. Distribution frequency of ligands that bind metal by side-chain atoms. Four common metals in our database are shown. The number of cases for each metal is indicated in parenthesis. Metals are arranged from left to right in decreasing hardness. Data for Asp, Glu and His are presented separately. "Other" mainly includes polar and sulfur containing residues.

two or more amino acid ligands. The graph is separated into four regions: mobile-backbone sites with significant binding-site RMSD and maximum Cα displacement (region 1), mobile-backbone sites with significant maximum Cα displacement (region 2), mobile-backbone sites with significant binding-site RMSD (region 3), and a region with rigid-backbone sites (region 4). The 28 pairs with two or more ligand residues in the mobile-backbone site regions could be grouped along two different lines (slope = 2.1, x intercept = 0.1 Å; and slope = 1.6, x intercept = 0.5 Å), each with an $R^2 = 0.95$. Thus, considering that the superimposition used to calculate the binding-site RMSD (shown in Fig. 2) was performed on first- and second-shell amino acid (see Methods), this suggests that the binding-site RMSD values in the pairs located along the lower line are influenced by second-shell residues. Indeed, when the maximum Cα displacement was plotted as a function of first-shell RMSD for apo–holo binding site pairs having at least three amino acid ligands (22 pairs), all the points grouped along a single line (slope = 2.0, $R^2 = 0.91$) that intercepted the x axis close to the origin (x intercept = 0.0) (data not shown).

Figure 2, Region 1 contains sixteen apo–holo pairs. Few showed overall domain rearrangement, possibly because the PDB is underpopulated with apo forms, owing to difficulties in crystallizing the apo form when the metal is important for protein folding. We observed that half of them underwent rearrangements in loops containing metal ligands (and sometimes a neighboring helix as well). Five pairs are in the quadrant of insignificant binding-site RMSD but significant maximum Cα displacement (region 2, Fig. 2). This relation between binding-site RMSD and maximum Cα displacement suggests that a limited number of first-shell ligands underwent backbone motions in these cases. On the other hand, for the seven pairs that show significant binding-site RMSD and small maximum Cα displacement (region 3, Fig. 2), conformational changes involving second-shell residues are indicated. Indeed, the average RMSD calculated for second-shell residues of R3 pairs is 1.3 ± 0.5 Å, whereas that for R2 pairs is 0.45 ± 0.16 Å. All of the five pairs of metal binding sites having significant displacement and small local RMSD (R2) have three or more amino acid ligands, and four of the five have four or more. Conversely, all seven R3 binding sites showing significant binding-site RMSD but small displacement, have only two or three first-shell amino acid residues. Region 4 in Figure 2 contains all the rigid-backbone binding sites. These 119 sites show insignificant local RMSD and insignificant maximum Cα displacement. Finally, there are 63 metal sites in our database having one amino acid ligand. These are not depicted in Figure 2 because it is not possible to calculate maximum Cα displacement with only one ligand. Of the 63, only two (1fwk, 1bc5) showed significant binding-site RMSD.

Side-Chain Reorientations Upon Metal Binding

Side-chain reorientations may also occur during the transition from pre-bound to bound form. Among the 180 pairs with rigid-backbone binding sites, 170 contain side-chain ligand atoms. Figure 3 depicts the amino acids most frequently found as side-chain ligands for the most common metals in our dataset. As expected, there is a clear correlation between the hardness of the metal and its propensity to interact with hard ligands.[16,18,20] Hard metals tend to bind hard ligands by ionic forces, while soft metals tend to bind soft ligands by forming partial covalent bonds. Here, we observed that the hard metals, Ca and Mg, are coordinated mainly by Asp and Glu side chains and very infrequently by His. The softer character of Mn allows it to accept His ligands more often than Mg, while Zn, which is considerably softer (more polarizable) than the other metals shown here, has the highest tendency to interact with His (i.e., nitrogen-containing) ligands. It should be mentioned that Cys, a common amino acid in Zn structural binding sites (e.g., zinc finger proteins), is under-represented in our dataset. Presumably,

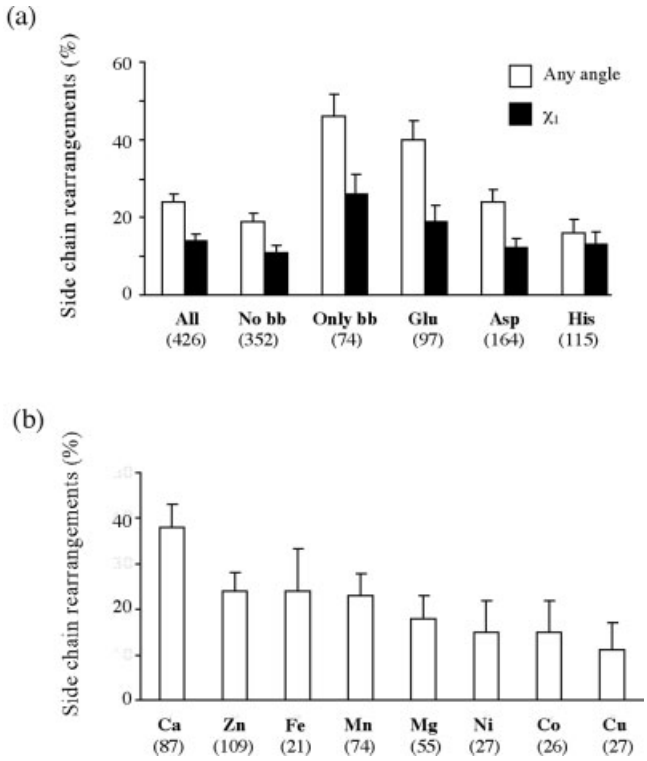

Fig. 4. Frequency of side-chain reorientations in the first shell upon metal binding. (**a**) All, all side chains; No bb, sites without backbone motion; Only bb, sites with backbone motion; Glu, Asp, and His are self explanatory. White bars, any dihedral angle changes. Black bars, only χ_1 motion. (**b**) Different metals, as indicated. The number of side chains for each case is given in parenthesis.

many proteins require Zn for proper folding and, thus, in general, are more difficult to crystallize in the apo form. Otherwise, the trends seen in Figure 3 are in general agreement with those found in previous studies[14,20] or derivable from the MSD site (http://www.ebi.ac.uk/msd-srv/msdsite). Deviations that exist among the analyses probably arise from the different criteria used in data selection, such as the degree of homology allowed, whole-residue versus side-chain analysis and the availability of crystallized apo forms.

The tendency of side chains to undergo reorientation upon metal binding is summarized in Figure 4(a). We observed side chain motions in 14% of χ_1 angles and 24% of cases where any dihedral angle changed. Considering only the cases with rigid-backbone sites, the levels of side-chain reorientation are 11% and 19%, respectively. These percentages for mobile-backbone sites increase dramatically to 26% and 46%, respectively. Thus, the tendency of a side chain to reorient upon metal binding is greatest when it is presented with a new structural milieu, as often occurs when there are backbone rearrangements. However, because there are only 74 side chains in this category versus 352 from rigid-backbone sites, the majority of side-chain changes in our database occur in sites that do not show backbone rearrangements.

As shown in Figure 4(a), the percent of side-chain reorientation (side-chain flexibility) of Glu (40%) is significantly larger than that of Asp (24%) or His (16%). The percentage increases to 24% if putative labeling discrepancies due to ambiguities in the PDB files[40] are not corrected. This observation agrees with previous flexibility studies that compared pairs of uncomplexed proteins[37] and proteins with and without large hetero groups.[41] The amount of side-chain reorientation decreases to 37%, 17%, and 11% for Glu, Asp, and His, respectively, if only rigid-backbone sites are considered.

TABLE II. Side–Chain Reorientation

	Number of ligand residues[a]			
Side chains	One	Two	Three	Four
Total number	53	86	102	101
Number reoriented	21	25	18	3
Frequency of reorientation	0.4	0.29	0.18	0.03

[a]Only side chains of rigid-backbone sites were considered.

Analysis of side-chain rearrangements for each metal is shown separately in Figure 4(b). Calcium, with the highest atomic radius among the group, showed the largest percentage of side-chain motions. The remaining metals showed statistically similar levels of side-chain reorientations, with Zn and Mn having a slightly higher tendency to reorient ligands. For most metals, the percentage of side-chain reorientations tended to decrease slightly if only rigid-backbone sites are considered. The exception is Fe, which showed a larger difference, possibly related to the small total number of Fe binding sites in our dataset and the fact that about half of them underwent backbone motions. A more exact picture of the behavior of each metal will require more data to become available in the PDB.

Side-Chain Rearrangements within the Binding Site

As described above, metal binding sites undergoing backbone conformational changes have a higher tendency to undergo side-chain reorientations. Indeed, as a consequence of backbone motion, side chains, perforce, may be embedded into a different environment. In order to focus directly on the effect of the metal in triggering side-chain reorientations, further analysis is restricted to rigid-backbone sites. About 35% of these sites undergo side-chain reorientations. Table II shows that the frequency of a side chain to reorient decreases with increasing number of ligands. The frequency decreases linearly going from one to three ligand residues but more precipitously from three to four. Probably, as the first-shell ligands increase in number, their interaction with the protein matrix (second shell) becomes stronger and movement of their side chains is restricted. However, this does not explain the sharp decrease between three and four ligand residues. One possibility is the small number of side chains (only three) that reoriented in binding sites with four ligand residues (Table II).

Table III summarizes the frequency of side-chain reorientations at the level of the binding site. For the 170 sites analyzed, 35% had at least one side chain that reoriented.

Only eight sites (or 14% of those reorienting) experienced multiple side-chain movements. This suggests that, often, part of the binding site is already in place in the pre-bound state. An analysis of changes in putative hydrogen bonds (CSU software;[30] distance restriction of 3.5 Å) between first and second shells of the metal binding sites in our data set supports this suggestion. Hydrogen bonds formed by ligand backbones, or ligand side chains, were conserved in $>$ 90% and 80% of cases, respectively. Thus, it should be possible to restrict side-chain movement to a single ligand in docking predictions for metal binding sites. An analogous conclusion was reached in studies of conformational changes around point mutations[38] and for large hetero groups.[41,42] Therefore, this behavior seems to be inherent to protein binding sites independent of the bound partner.

A majority of flexible binding sites (13 of 18) with three or more ligand residues are catalytic sites, while only a small minority (4 of 36) with two or one ligand residues are so. The vast majority of the rest are seen by visual inspection to be on the protein surface. The acidic residues (Glu and Asp), represent 76% of these cases. We checked also if binding sites with at least two ligands of the same charge show a tendency for more frequent reorientation upon metal binding. They did not.

In many cases, exposed metal binding sites contribute to protein conformational stability,[43] crystal packing,[44] metal sinks,[45] or protein oligomerization.[46] In this regard, Mg and Ca in surface-located sites were identified that mediate protein–protein[47,48] and lipid–protein[49] interactions. Therefore, surface-located binding sites can be crucial for protein function.

Ligand Flexibility and Solvent Accessibility

B-factor is a measure of atomic disorder in crystal structures and can be used to analyze side-chain flexibility. Analysis of normalized B-factors distribution for apo and holo forms in our database showed mean values of 0.19 and −0.36, respectively. These differences are statistically significant (t-test) and indicate, as expected, that ligand residues generally become more ordered upon metal binding. This is in agreement with the observation that catalytic residues in active sites have smaller B-factors in the presence of a cofactor.[50]

A comparison was made of ligand flexibility in the pre-bound state between ligand residues with or without side-chain reorientations upon metal binding. The results in Figure 5 show that the normalized, mean B-factor for ligand side chains that do not reorient is −0.10, whereas the mean for those that do is 0.96. This indicates that, on average, reoriented residues were already more disordered in the pre-bound state than those that did not move; i.e., they are, in general more mobile, perhaps establishing less interactions with the protein scaffold.

First-shell residues whose side chains moved upon metal binding, also tended to be more accessible to solvent in the pre-bound state [Fig. 6(a)], the degree of solvent accessibility being inversely correlated with the number of ligand residues in the metal binding site [Fig. 6(b)]. This can be explained by the fact that binding sites with more ligand residues have stronger second-shell interactions and therefore, might be pre-organized in a more rigid framework.

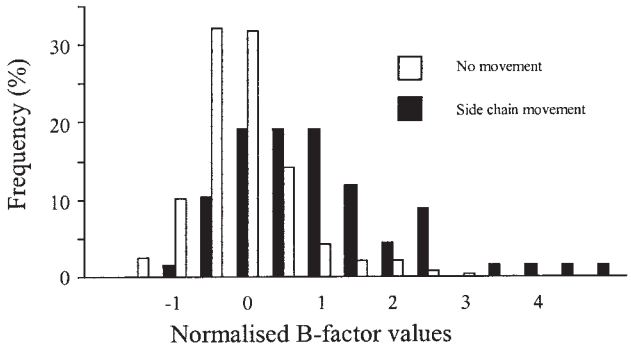

Fig. 5. Normalized B-factor distribution of first-shell side chains in apo forms. Black bars, side chains undergoing conformational changes; white bars, side chains with no conformational changes.

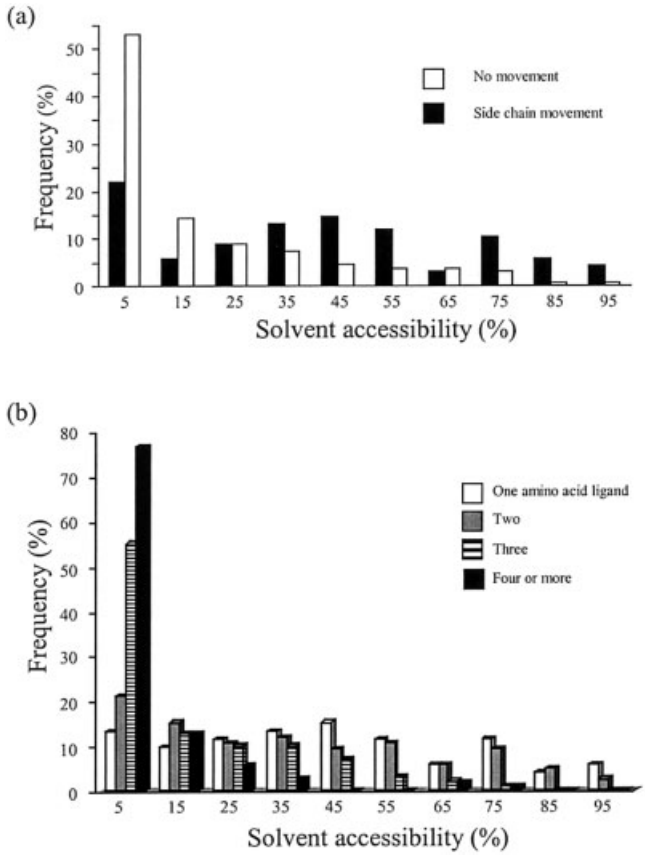

Fig. 6. Frequency of side-chain reorientations in the first shell as a function of solvent accessibility in apo forms. (**a**) Black bars, side chains undergoing conformational changes; white bars, side chains with no conformational changes. (**b**) Number of amino acids in the first shell as indicated in key.

Metal Substitutions

The flexibility of a binding site is reflected not only in the reorientations that occur upon metal binding but also in the geometry it adopts upon binding different metals.

TABLE III. Binding Site Mobility

	Number of ligand residues per binding site[a]				
Binding sites	One	Two	Three	Four or more	Totals
Total number	53	46	39	32	170
Number with reoriented side chains	21	20	14	4	59
Number with one side chain reoriented	21	15	11	4	51
Number with two side chains reoriented	-	5	2	0	7
Number with three side chains reoriented	-	-	1	0	1
Frequency of mobility	0.4	0.43	0.36	0.13	0.35 ave.

[a]Only rigid-backbone sites were considered.

TABLE IV. Listing of Substituted Metal Binding Sites

Protein	Apo form[a]	Holo form[b]
Concavalin A (lectin), (site 1)	d1dq0a_	d1nls_(Mn), d1scs_(Co), d1scr_(Ni), d1enr_(Zn)
Sarcoma virus integrase	d1a5x_	d1vsi_(Ca), d1vsd_(Mg), d1a5v_(Mn), d1vsh_(Zn)
Azurin (electron transport)	d1e65a_	d1jzga_(Cu), d1vlxd_(Co), d1nzrb_(Ni), d1e67c_(Zn)
Carbonic anhydrase	d2cbe_	d2cba_(Zn), 1cah_(Co), 1rzc_(Cu), d1rze_(Ni), d1rzd_(Mn)
D-Xylose isomerase (site 1)	d1xlaa_	d1xlbb_(Mg), d1xlkb_(Mn), d1xlla_(Zn)
D-Xylose isomerase (site 1)	d7xima_	d4ximc_(Co)
D-Xylose isomerase (site 2)	d1xlaa_	d1xlkb_(Mn), (d1xlla_(Zn)
Astacin (endopeptidase)	d1iad_	d1ast_(Zn), d1iab_(Co), d1iaa_(Cu), d1iae_(Ni)
Chey (signal transduction)	d3chy_	d1chn_(Mg), d1ffga_(Mn), d5chy_(Ca)
Pyrophosphatase (site 1)	d1jfda_	d1obwc_(Mg), d1ino_(Mn)
2-phospho-glycerate-hydrolase	d3enl_1	d1ebhb1 (Mg), d4enl_1 (Zn)
2-phospho-glycerate-hydrolase	d1pdy_1	d1pdz_1 (Mn)
Fructose-1,6-biphosphatase (site 1)	d5fbpa_	d1eyja_(Mg), d1cnqa_(Zn)
Staphylococcal nuclease	d1ey0a_	d1sty_(Ca), d1stg_(Co)
Protein kinase DAPK	d1jkla_	d1jkka_(Mg), d1igla_(Mn)
Ribonuclease T1	d1det	d1i0va_(Ca), d8rnt_(Zn)

[a]SCOP identifier of metal binding domain.
[b]SCOP identifier; in parenthesis, a metal is indicated.

"Firm" binding sites bind different metals using the same geometry and through the same set of ligands, whereas "adaptable" binding sites accommodate different metals in different ways. Table IV lists the binding sites in our dataset with metal substitutions. In twelve of fourteen cases water molecules or other hetero groups participate in metal ligation. This suggests that at least part of the coordination shell has a heightened degree of flexibility to accommodate different metals. Metal binding site S1 of concavalin A is an example of a "firm" site that binds different metals of similar hardness using the same set of ligands and octahedral geometry [Fig. 7(a)].[51,52] No backbone or side-chain conformational changes were observed for any substituted metal when apo and holo forms were compared.

"Adaptable" sites exploit different strategies for binding different metals, leading to an alteration of geometry and/or the use of a different set of ligands. For example, the buried binding site of the electron transfer protein azurin, surrounded by a network of hydrogen bonds, is able to accommodate different metals by altering the number of ligand residues in the first coordination shell. As observed in Figure 7(b), Met 121 participates in the trigonal bipyramidal coordination sphere of Cu, but is absent in the distorted tetrahedral coordination sphere of Zn.[53,54] Apo azurin presents two different molecular forms arranged as hetero-dimers.[55] Comparison of one of the conformers with the holo forms showed maximum Cα displacements ranging from 0.68 to 0.87 Å, depending on the metal, whereas no conformational changes were observed when the comparison was performed with the second conformer that contains a water molecule at the position of the metal.

Exposed binding sites can accommodate different metals by using a different number of water molecules and/or

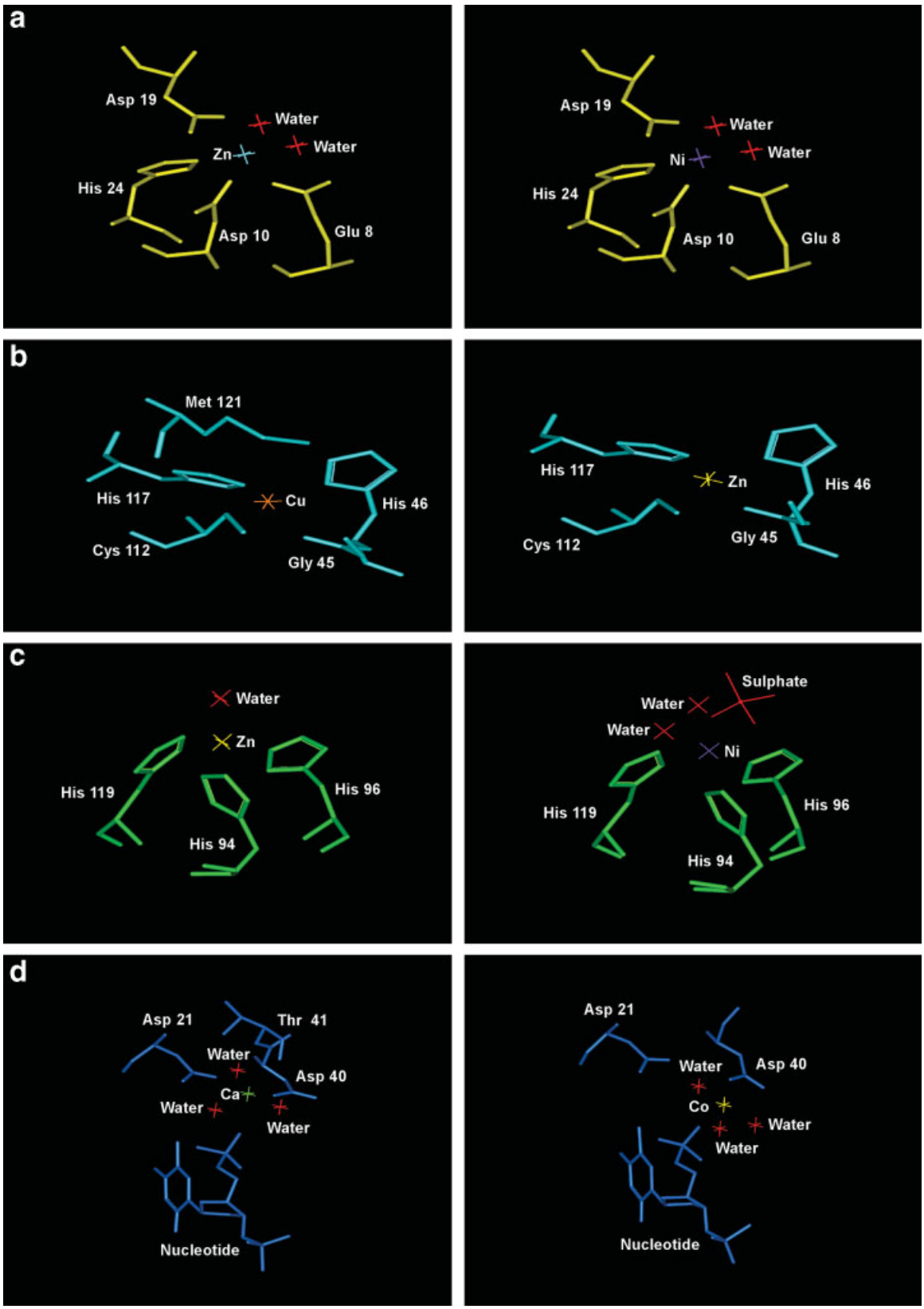

Fig. 7. Metal substitutions in different metal binding sites. First-shell coordination sphere for: (**a**) Concavalin A bound to Zn (**right** panel) and Ni (**left** panel); (**b**) Azurin bound to Cu and Zn; (**c**) Carbonic anhydrase bound to Zn and Ni; (**d**) *Staphylococcal* nuclease bound to Ca and Co.

ions as ligands. For example, carbonic anhydrase binds Zn and Co with a tetrahedral geometry including a water molecule. However upon binding Cu, Ni, or Mg, the coordination number increases to either five or six by the incorporation of water molecules and/or ions. In particular, Ni-bound carbonic anhydrase has one additional water molecule and one sulfate ion with respect to Zn in the first coordination shell [Fig. 7(c)].[56,57] No conformational changes were observed in the ligand residues of carbonic anhydrase, when the apo form was compared with different holo forms. Other sites bind metals with different radii by exchanging direct contact with ligand residues for water-mediated interactions. For instance, staphylococcal nuclease binds both Ca and Co through the octahedral geometry, however Asp 21 interacts directly with Ca, but through a water molecule with Co [Fig. 7(d)].[58,59] In both the Co- and Ca-bound cases, side-chain conformational changes occur at residue 40 when comparing the apo and holo states.

Finally, there are exposed sites in which the clathrate effect dominates over metal coordination. This is the case for ribonuclease T1, binding Ca or Zn. Binding of one amino acid and six water molecules is similar even though the Zn radius is much smaller than that of Ca.[43,60] In addition, in both cases, the protein undergoes side-chain conformational changes upon metal binding.

Conclusions

The transition from the pre-bound to the metal-bound state is accompanied by changes in the binding site in more than 40% of the pairs present in our nonredundant dataset containing all available representatives of apo–holo pairs for the most populated metals in the PDB. This suggests that almost half of the metal binding sites have a capacity for flexibility.

We observed that in a small number of cases metal binding triggers entire domain conformational changes, however, in general the rearrangements are restricted to the binding site region. In some cases (14% of the total) backbone motions, principally in loop regions, occur and very often are accompanied by side-chain motions. Side-chain reorientations are observed in 35% of rigid-backbone sites. In at least 75% of the cases only one side chain moved, regardless of the number of amino acid ligands per binding site (Table III). This indicates that in these cases, part of the first coordination shell is already in place in the pre-bound state and side-chain flexibility can be restricted to one first-shell ligand in docking prediction algorithms for metal binding sites. For example, in procedures based on searching for a constellation of ligand–backbone positions, consideration of side-chain flexibility would be restricted to one residue at a time.

Analysis of side-chain reorientations triggered upon metal binding for the most frequent ligands showed that Glu, the one with the largest number of degrees of freedom, is the most flexible. When the study was performed for different metals, ligand residues that bind Ca showed the highest tendency to move. In addition, the frequency of first-shell residues to undergo side-chain reorientations was shown to be inversely correlated with the number of amino acid ligands in the binding site. On the other hand, the tendency of a binding site to undergo side-chain reorientations upon metal binding was similar for those sites having one to three amino acid ligands but decreases considerably for those having four or more first-shell residues.

In general, in apo form, those ligands that moved tended to be more disordered (higher B factor) and more accessible to the solvent than those that did not. These direct correlations between ligand disorder and accessibility to the solvent in the apo form and side-chain rearrangements may be useful for prediction of ligand motion based on known pre-bound state conformations.

REFERENCES

1. Ibers JA, Holm RH. Modeling coordination sites in metallobiomolecules. Science 1980;209:223–235.
2. Tainer JA, Roberts VA, Getzoff ED. Protein metal-binding sites. Curr Opin Biotechnol 1992;3:378–387.
3. Volkenstein MV. Electronic-conformational interactions in biopolymers. Pure Appl Chem 1979;51:801–829.
4. Karplus M, McCammon JA. Dynamics of proteins: elements and function. Annu Rev Biochem 1983;52:263–300.
5. Janin J, Wodak SJ. Structural domains in proteins and their role in the dynamics of protein function. Prog Biophys Mol Biol 1983;42:21–78.
6. Feinberg H, Greenblatt HM, Shoham G. Structural studies of the role of the active-site metal in metalloenzymes. J Chem Inform Comp Sci 1993;33:501–516.
7. Greenblatt HM, Feinberg H, Tucker PA, Shoham G. Carboxypeptidase A: native, zinc-removed and mercury-replaced forms. Acta Crystallogr D Biol Crystallogr 1998;54:289–305.
8. Sun HZ, Li HY, Sadler PJ. Transferrin as a metal ion mediator. Chem Rev 1999;99:2817–2842.
9. Vallee BL, Auld DS. Zinc coordination, function, and structure of zinc enzymes and other proteins. Biochemistry 1990;29:5647–5659.
10. Katz AK, Glusker JP, Beebe SA, Bock CW. Calcium ion coordination: a comparison with that of beryllium, magnesium, and zinc. J Am Chem Soc 1996;118:5752–5763.
11. Chakrabarti P. Geometry of interaction of metal-ions with sulfur-containing ligands in protein structures. Biochemistry 1989;28:6081–6085.
12. Chakrabarti P. Geometry of interaction of metal-ions with histidine-residues in protein structures. Protein Eng 1990;4:57–63.
13. Chakrabarti P. Interaction of metal-ions with carboxylic and carboxamide groups in protein structures. Protein Eng 1990;4:49–56.
14. Aberts IL, Nadassy K, Wodak SJ. Analysis of zinc binding sites in protein crystal structures. Protein Sci 1998;7:1700–1716.
15. Harding MM. Geometry of metal-ligand interactions in proteins. Acta Crystallogr D Biol Crystallogr 2001;57:401–411.
16. Bock CW, Katz AK, Markham GD, Glusker JP. Manganese as a replacement for magnesium and zinc: functional comparison of the divalent ions. J Am Chem Soc 1999;121:7360–7372.
17. Dudev T, Lim C. Metal selectivity in metalloproteins: Zn^{2+} vs Mg^{2+}. J Phys Chem B 2001;105:4446–4452.
18. Glusker JP. Structural aspects of metal liganding to functional-groups in proteins. Adv Protein Chem 1991;42:1–76.
19. Karlin S, Zhu ZY, Karlin KD. The extended environment of mononuclear metal centers in protein structures. Proc Natl Acad Sci USA 1997;94:14225–14230.
20. Dudev T, Lin YL, Dudev M, Lim C. First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations. J Am Chem Soc 2003;125:3168–3180.
21. Yamashita MM, Wesson L, Eisenman G, Eisenberg D. Where metal-ions bind in proteins. Proc Natl Acad Sci USA 1990;87:5648–5652.
22. Glusker JP. Directional aspects of intermolecular interactions. Top Curr Chem 1998;198:1–56.
23. Hellinga HW, Richards FM. Construction of new ligand-binding

sites in proteins of known structure. 1. Computer-aided modeling of sites with predefined geometry. J Mol Biol 1991;222:763–785.
24. Wray JW, Baase WA, Ostheimer GJ, Zhang XJ, Matthews BW. Use of a non-rigid region in T4 lysozyme to design an adaptable metal-binding site. Protein Eng 2000;13:313–321.
25. Bernstein FC, Koetzle T.F, Williams GJ, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. Nucleic Acids Res 2000;28:235–242.
27. Castagnetto JM, Hennessy SW, Roberts VA, Getzoff ED, Tainer JA, Pique ME. MDB: the Metalloprotein Database and Browser at the Scripps Research Institute. Nucleic Acids Res 2002;30:379–382.
28. Murzin AG, Brenner SE, Hubbard T, Chothia, C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J Mol Biol 1995;247:536–540.
29. Brenner SE, Koehl P, Levitt, M. The ASTRAL compendium for protein structure and sequence analysis. Nucleic Acids Res 2000; 28:254–256.
30. Sobolev V, Sorokine A, Prilusky J, Abola EE, and Edelman M. Automated analysis of interatomic contacts in proteins. Bioinformatics 1999;15:327–332.
31. Dror O, Benyamini H, Nussinov R, Wolfson H. MASS: multiple structural alignment by secondary structures. Bioinformatics 2003;19 Suppl 1:I95–I104.
32. Flores TP, Orengo CA, Moss DS, Thornton JM. Comparison of conformational characteristics in structurally similar protein pairs. Protein Sci 1993;2:1811–1826.
33. Betts MJ, Sternberg MJ. An analysis of conformational changes on protein-protein association: implications for predictive docking. Protein Eng 1999;12:271–283.
34. Arun KS, Huang TS, Blostein SD. Least-squares fitting of 2 3-D point sets. IEEE Transact Pattern Anal Machine Intelligence 1987;9:699–700.
35. Schrauber H, Eisenhaber F, Argos P. Rotamers: to be or not to be? An analysis of amino acid side-chain conformations in globular proteins. J Mol Biol 1993;230:592–612.
36. Bower MJ, Cohen FE, Dunbrack RL, Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. J Mol Biol 1997;267:1268–1282.
37. Zhao S, Goodsell DS, Olson AJ. Analysis of a data set of paired uncomplexed protein structures: new metrics for side-chain flexibility and model evaluation. Proteins 2001;43:271–279.
38. Eyal E, Najmanovich R, Edelman M, Sobolev V. Protein side-chain rearrangement in regions of point mutations. Proteins 2003;50:272–282.
39. Parthasarathy S, Murthy MR. Analysis of temperature factor distribution in high-resolution protein structures. Protein Sci 1997;6:2561–2567.
40. McDonald IK, Thornton JM. The application of hydrogen bonding analysis in X-ray crystallography to help orientate asparagine, glutamine and histidine side chains. Protein Eng 1995;8:217–224.
41. Najmanovich R, Kuttner J, Sobolev V, Edelman M. Side-chain flexibility in proteins upon ligand binding. Proteins 2000;39:261–268.
42. Fradera X, De La Cruz X, Silva CH, Gelpi JL, Luque FJ, Orozco M. Ligand-induced changes in the binding sites of proteins. Bioinformatics 2002;18:939–948.
43. Deswarte J, De Vos S, Langhorst U, Steyaert J, Loris R. The contribution of metal ions to the conformational stability of ribonuclease T1: crystal versus solution. Eur J Biochem 2001;268: 3993–4000.
44. Fisher AJ, Thompson TB, Thoden JB, Baldwin TO, Rayment I. The 1.5 Å resolution crystal structure of bacterial luciferase in low salt conditions. J Biol Chem 1996;271:21956–21968.
45. Ellis MJ, Prudencio M, Dodd FE, Strange RW, Sawers G, Eady RR, Hasnain SS. Biochemical and crystallographic studies of the Met144Ala, Asp92Asn and His254Phe mutants of the nitrite reductase from Alcaligenes xylosoxidans provide insight into the enzyme mechanism. J Mol Biol 2002;316:51–64.
46. Hemmens B, Goessler W, Schmidt K, and Mayer B. Role of bound zinc in dimer stabilization but not enzyme activity of neuronal nitric-oxide synthase. J Biol Chem 2000;275:35786–35791.
47. McLaughlin PJ, Gooch JT, Mannherz HG, Weeds AG. Structure of gelsolin segment 1-actin complex and the mechanism of filament severing. Nature 1993;364:685–692.
48. Lee JO, Rieu P, Arnaout MA, Liddington R. Crystal structure of the A domain from the alpha subunit of integrin CR3 (CD11b/CD18). Cell 1995;80:631–638.
49. Zhang X, Rizo J, Sudhof TC. Mechanism of phospholipid binding by the C2A-domain of synaptotagmin I. Biochemistry 1998;37: 12395–12403.
50. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM. Analysis of catalytic residues in enzyme active sites. J Mol Biol 2002;324:105–121.
51. Emmerich C, Helliwell JR, Redshaw M, Naismith JH, Harrop SJ, Raftery J, Kalb AJ, Yariv J, Dauter Z, Wilson KS. High-resolution structures of single-metal-substituted Concanavalin-A—the Co,Ca-protein at 1.6 Å and the Ni,Ca-Protein at 2.0 Å. Acta Crystallogr D Biol Crystallogr 1994;50:749–756.
52. Bouckaert J, Poortmans F, Wyns L, Loris R. Sequential structural changes upon zinc and calcium binding to metal-free concanavalin A. J Biol Chem 1996;271:16144–16150.
53. Nar H, Huber R, Messerschmidt A, Filippou AC, Barth M, Jaquinod M, van de Kamp M, Canters GW. Characterization and crystal structure of zinc azurin, a by-product of heterologous expression in Escherichia coli of Pseudomonas aeruginosa copper azurin. Eur J Biochem 1992;205:1123–1129.
54. Crane BR, Di Bilio AJ, Winkler JR, Gray HB. Electron tunneling in single crystals of Pseudomonas aeruginosa azurins. J Am Chem Soc 2001;123:11623–11631.
55. Nar H, Messerschmidt A, Huber R, van de Kamp M, Canters GW. Crystal structure of Pseudomonas aeruginosa apo-azurin at 1.85 Å resolution. FEBS Lett 1992;306:119–124.
56. Hakansson K, Carlsson M, Svensson LA, Liljas A. Structure of native and apo carbonic anhydrase II and structure of some of its anion-ligand complexes. J Mol Biol 1992;227:1192–1204.
57. Hakansson K, Wehnert A, Liljas A. X-ray analysis of metal-substituted human carbonic anhydrase II derivatives. Acta Crystallogr D Biol Crystallogr 1994;50:93–100.
58. Keefe LJ, Sondek J, Shortle D, Lattman EE. The alpha aneurism: a structural motif revealed in an insertion mutant of staphylococcal nuclease. Proc Natl Acad Sci USA 1993;90:3275–3279.
59. Loll PJ, Quirk S, Lattman EE, Garavito RM. X-ray crystal structures of staphylococcal nuclease complexed with the competitive inhibitor cobalt(II) and nucleotide. Biochemistry 1995;34: 4316–4324.
60. Ding J, Choe HW, Granzin J, Saenger W. Structure of ribonuclease T1 complexed with zinc(II) at 1.8 Å resolution: a Zn^{2+}.6H2O.carboxylate clathrate. Acta Crystallogr B 1992;48:185–191.