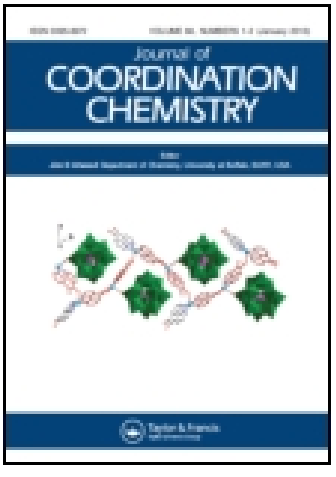


Journal of Coordination Chemistry
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/gcoo20

In silico analysis of metal coordination geometry in arsenic, beryllium, and lead bound structures

R. Jesu Jaya Sudan[a] & C. Sudandiradoss[a]
[a] Bioinformatics Division, School of Bioscience and Technology, VIT University, Vellore, India

Accepted author version posted online: 09 Jun 2014.Published online: 07 Jul 2014.

To cite this article: R. Jesu Jaya Sudan & C. Sudandiradoss (2014) In silico analysis of metal coordination geometry in arsenic, beryllium, and lead bound structures, Journal of Coordination Chemistry, 67:11, 1888-1904, DOI: 10.1080/00958972.2014.931945

To link to this article: http://dx.doi.org/10.1080/00958972.2014.931945

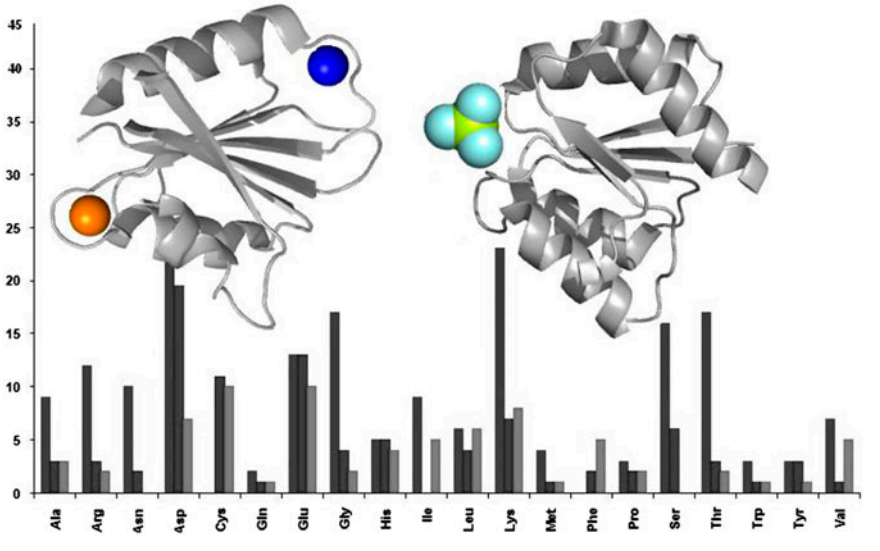
Journal of Coordination Chemistry, 2014
Vol. 67, No. 11, 1888–1904, http://dx.doi.org/10.1080/00958972.2014.931945

In silico analysis of metal coordination geometry in arsenic, beryllium, and lead bound structures

R. JESU JAYA SUDAN and C. SUDANDIRADOSS*

Bioinformatics Division, School of Bioscience and Technology, VIT University, Vellore, India

(*Received 1 August 2013; accepted 23 April 2014*)



Metal toxicity is a potential hazard to health and toxic effects of metals have been implicated in many diseases. Understanding the interaction of toxic metals becomes vital to prevent hazards following its association in living systems. Coordination chemistry helps in predicting the metal environments like coordinating residues, coordination space, metal coordination geometry, etc. Our work aimed at predicting the coordination of toxic metals arsenic, lead, and beryllium. In this work, we analyzed coordination for each metal from a set of arsenic, beryllium and lead bound structures which were retrieved from the Protein Data Bank. The structures were validated using B-factor and occupancy of the coordinating residues towards the metals. Coordination patterns such as chelate residues, chelate length, geometry, coordination number and structural architecture were predicted. Coordination geometry of the metals was exposed beyond the coordination space with their coordination number ranging from 2 to 11. Analysis of metal environment revealed the acidic amino acids aspartic acid, glutamic acid, and the basic amino acids lysine, histidine, and cysteine to be predominant in coordinating with the metals. Chelate patterns like DDVMITAK, DWNVTVK, ESGKNSS for beryllium, CCCSK, DSDWD for lead and FLICVI and LKHHKEE for arsenic were predicted to

*Corresponding author. Email: csudandiradoss@vit.ac.in

be common through extended coordination space. The distinct molecular geometries such as pentagonal bipyramid and square planar were observed only in lead bound structures but not in beryllium and arsenic bound structures. Beryllium had a larger influence than arsenic and lead, based on conformational changes owing to the presence of the metals. Our coordination study puts forth several propositions based on the metal environment that would help in designing chelation strategies.

Keywords: Coordination geometry; Beryllium; Lead; Arsenic; Chelate

1. Introduction

Metal intoxication, particularly carcinogenicity, neurotoxicity and genotoxicity are widely known. Some metals have high affinity for thiol-containing enzymes and proteins, which are responsible for normal cellular defense mechanism. Growth factor receptors, G-proteins, MAP kinases and transcription factors are some of the signaling components affected by toxic metals [1]. Long-term exposure to these metals could lead to apoptosis. The carcinogenicity of toxic metals is well documented, but in a scattered fashion. Trace amount of these elements entering the body via various routes can induce genetic and epigenetic alteration in different cancer-related genes of somatic and stem cells [1, 2]. Arsenic, lead, beryllium, cadmium, mercury, and their compounds are reported as being highly toxic and therefore not essential for nutrition. The toxic effects of these metals may be mediated or enhanced by interactions or deficiencies of nutritionally essential metals such as calcium, iron, zinc, and selenium [3–7].

This study focuses on the binding patterns of the carcinogenic metals, arsenic, beryllium and lead which have not been well discussed. Arsenic exerts its toxicity by inactivating up to 200 enzymes, especially those involved in cellular energy pathways and in DNA synthesis and repair. Inorganic arsenic is considered the most potential human carcinogen and humans are exposed to it from soil, water, air and food. There are no evidence-based treatment regimens to treat chronic arsenic poisoning but antioxidants have been advocated, though benefit is not proven [8, 9]. Lead is a widespread toxic metal which causes potential danger to human health. Recent epidemiological and experimental study on inorganic lead compounds confirms its increased risks of tumorigenesis. Probable mechanisms of lead carcinogenicity include direct DNA damage, clastogenicity, or inhibition of DNA synthesis or repair and altered gene expression [10, 11]. Beryllium has been identified as a human carcinogen on the basis of animal and epidemiological studies. Beryllium induces morphological transformation and enhances viral transformation of mammalian cells and also it decreases fidelity of DNA synthesis [12–14].

The removal of such toxic metals from the biological host involves different methods of chelation that employ synthetic, chemical and peptide chelators. The study of toxic metal association with functionally vital biomolecules will eventually produce significant results on the donor type, geometry and structure of residues that favor such coordination. These results when combined with chelation therapy produce new insights into designing effective peptide chelators. Many studies have employed coordination chemistry in designing effective chelators for toxic metal chelation [15–18]. Coordination geometries of metals have been analyzed to understand the physiochemical properties of metal complexes with ligands and also in solvatochromism studies. However, this paper focuses on studying coordination geometry to gain insight into effective chelate strategies [19, 20]. Our earlier work on cadmium coordination provides useful information in chelate patterns and geometry that can be employed in cadmium chelation [21]. To understand arsenic, beryllium and lead in biological samples and to support chelator design, the present study is aimed at determining the

coordination geometry of the carcinogenic metals. The study was carried out on the basis of the work describing geometry of metal ion-binding sites within proteins by Rulisek and Vondrasek [22] and Harding [23]. This coordination study gives an account of the residue, residue position and donor atom distance that gives insight on the choice of chelators. We also examined bond lengths and coordination numbers, B-factor (displacement parameter sometimes referred as "temperature factor") and relative occupancies of metal ions [24].

2. Materials and methods

2.1. *Protein data-set*

A search was made across the Protein Data Bank (PDB) to identify all possible protein structures bound with arsenic, lead, and beryllium ions. Only X-ray crystallographic structures of proteins were considered for the study, NMR models and DNA-associated proteins were excluded for the analysis. The crystal structures of the metal bound proteins required for the study were retrieved from the PDB [25]. The data-set was made non-redundant by culling the structures using PISCES [26] such that no two sequences shared >40% sequence identity, to avoid data redundancy due to structurally identical proteins. The other threshold parameters included resolution between 0 and 3 Å, a maximum R value of 0.3, chain length from 40 to 5000 residues. The structures were culled by entries and further culled by chains within entries. Only a single polypeptide chain from each structure was considered for analysis to avoid repetition of data due to homodomain structures. The resultant structures were then subjected to analysis in this study.

2.2. *Analysis of the metal coordination sites*

The metal coordination number (cn) is a key determinant of the structure and properties of metal complexes [27]. The coordination groups were derived on the basis of the work done by Jesu Jaya Sudan and Sudandiradoss [21] and Harding [23]. The metal coordination groups of the three metals i.e. arsenic, beryllium and lead were predicted using ANAMBS [28], a standalone tool that predicts the microenvironment of metals within the protein. The coordination distance cut-off was set to 3 Å to identify residues within the first shell of the coordination sphere [29] as residues within the 3 Å cut-off distance were considered coordinating with the metal [24]. The specificity and frequency of amino acids coordinating with the metals were predicted. The inter-atomic distances of metal and the coordinating atoms were calculated as a factor of their affinity towards the metal. Metal coordination with alternative conformers was ignored. For structures with multiple ions, the first ion according to the PDB residue numbering was taken as the reference. The coordination group also gives an account of the non-residues, namely the number of solvents and ligands within the coordination sphere of the metal. The table also includes the functional folds of the structures predicted from CATH and SCOP databases [30].

2.3. *Correlation of B-factors*

Validation of metal sites requires inspection of occupancy, B values, geometry, bond distances, etc. [21, 22]. Evaluation of the value of B for the metal and the average B-factor for the donors is useful in identifying a metal atom and its environment. B-factors lower than 2.0 were considered flawed and thus were ignored. The B-factor for a metal–ion

environment was calculated as the mean B-factor for all atoms located within 3 Å of the metal ion of interest [24]. Metal sites were also validated using the occupancy assigned to metal and donor. Occupancies falling outside the range of 0.1–1 were ignored.

2.4. *Metal influence on protein structures*

Metals associated with proteins often play a role in stabilizing protein structures or act as cofactors in catalyzing chemical reactions. This is widely observed in essential metals like zinc, magnesium, sodium, etc. Similarly, toxic metals that bind to proteins *in vivo* exert an effect on the protein structure. In order to predict the influence of arsenic, beryllium, and lead on the protein structure, the metal ions were removed and subsequently subjected to an energy minimization for 200 steps of steepest descent by applying GROMOS96 force field included in the SwissPDBviewer [31]. As an estimation of the atomic displacement, the phi and psi angles of the residues within the first shell of the coordination sphere before and after metal removal were predicted and plotted. The average angular deviation of alpha helices and beta sheets of the coordinating residues were analyzed. The extent of deviation reports the degree to which the metal influences structural alterations in proteins.

2.5. *Pattern analysis beyond coordination distance*

Residues beyond the first shell distance have a large structural impact in deciding the geometry of residues that coordinate with the metal. Such residues are conserved among sequences. To explore the sequence conservation among the residues beyond the coordination distance, we predicted all residues within 5 Å from the metal. This was done using ANAMBS [28] by setting the cut-off to a distance of 5 Å. The solvents, ligands, and alternative conformers were ignored for conservation analysis. For any conserved residues, the position with respect to the first shell residue, its frequency, and conservation within the coordination sphere were analyzed.

3. Results and discussion

All the structures required for this study were retrieved from the PDB and were culled at 40% sequence identity. About 166 structures were subjected to culling with PISCES [26]. The culled data-set resulted in a set of 16 arsenic bound structures, 38 beryllium bound structures, and 28 lead bound structures. Although the data-set considered for the analysis is much smaller, the diversity of protein structures sharing a sequence identity of <40% makes the results more specific for an individual protein. However, the structures were validated with the B-factors of metal and metal environment. The results are presented and discussed for the structures that are in agreement with the prescribed range of B values.

3.1. *Validation by B-factor and occupancy*

B-factor which is normally defined as the vibration parameter was predicted for all coordinating atoms and metals [32]. A plot indicating the correlation of metal B-factor with B-factor of coordinating residues is given in figure 1. The B-factors for beryllium and its coordinating residues show a very good correlation of 0.97. This shows that beryllium and the coordinating atoms are in good agreement within the crystal structure. However, the plots of lead and arsenic have a large scatter and show correlations of 0.71 and 0.68,

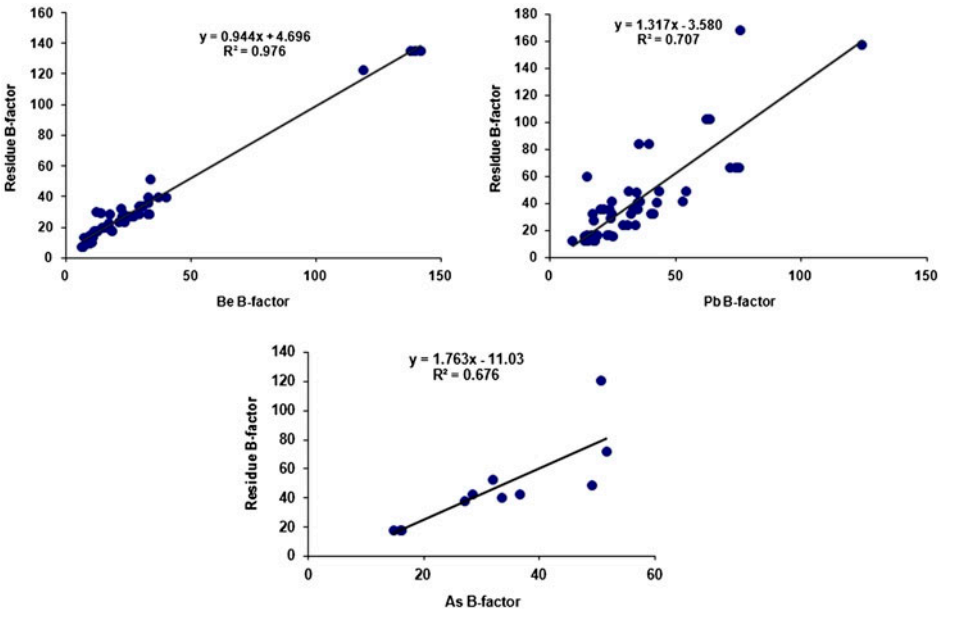


Figure 1. Correlation between B-factors – Beryllium shows a higher correlation between metal and the coordinating atom B-factors than arsenic and lead.

respectively, indicating that the atomic conformation within the crystal structure is less ordered which may arise from weaker bonds. Beryllium, due to its intact structure and the closeness with the nucleus, shows stronger bonding and thus has less scatter [33]. In a similar manner, the occupancy indicating the space occupied by an atom was evaluated. Beryllium and its coordinating atoms had an occupancy of 1 which was in par with the B-factors. For all the metals and the coordinating residue atoms, the B-factor and occupancy were well within the desired range and hence coordination patterns derived from these structures can prove significant.

3.2. *Coordination geometries of arsenic, beryllium, and lead*

The metal coordination group analyzed for the three metals gives the details of the chelate loop, its length, the coordination number that includes residues, solvents and ligands. Metals associate with proteins through the side chain or the backbone atoms of specific amino acids. Different metals show different level of specificity for the atom types [34]. The type and number of donors contributing to the metal coordination give insight to the coordination geometry of the metal. Table 1 shows the values of coordination groups predicted for arsenic, beryllium and lead within the cut-off distance of 3 Å. Structures that have no observable interactions are not shown in table 1. The metals had their own selectivity for the chelate loops though not much variation was observed among the chelate of a specific metal. The coordination data predicted from the arsenic, beryllium, and lead bound structures are shown in table 1.

Table 1. Coordination group of arsenic, beryllium and lead within the first shell (3 Å).

pid[a]	res[b]	posn[c]	nspan[d]	np[e]	nw[f]	nn[g]	dons[h]	met[i]	sd1[j]	sd2[j]	sd3[j]	sd4[j]	sd5[j]	sd6[j]	sd7[j]	cn[k]	sec.str[l]	Fold
1NQ0	2.4	294	–	1	1	0	C	ARS	−1	–	–	–	–	–	–	2	H	Orthogonal bundle
1WN6	1.8	73	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	S	Cytidine deaminase like fold
2GBM	1.55	24	36	2	2	0	ED	ARS	36	−1	−1	–	–	–	–	4	HT	Roll
2J4Z	2	247	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	H	Orthogonal bundle
2O6 W	2.4	167	–	1	0	0	T	ARS	–	–	–	–	–	–	–	1	T	Nil
2Y0O	1.23	–	–	0	2	0	–	ARS	−1	−1	–	–	–	–	–	2	–	Nil
3CAO	1.6	–	–	0	2	0	–	ARS	−1	−1	–	–	–	–	–	2	–	Alpha beta complex
3FRG	1.7	341	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	H	Nil
3H6 N	2	1592	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	T	Nil
3JRN	2	150	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	S	Nil
3SMP	1.9	397	–	1	0	0	C	ARS	–	–	–	–	–	–	–	1	T	Nil
3SMT	2.04	296	–	1	0	0	C	ARS	–	–	–	–	–	–	–	3	S	Nil
1IW7	2.6	1112	–	1	0	0	C	PB	–	–	–	–	–	–	–	1	T	Alpha beta complex
1N0Y	1.75	58	6	2	0	0	DD	PB	6	–	–	–	–	–	–	2	TT	Orthogonal bundle
1NA0	1.6	99	–	1	0	0	Y	PB	–	–	–	–	–	–	–	1	H	Alpha horseshoe
1QNV	2.5	133	10	3	0	0	CCC	PB	2	8	–	–	–	–	–	3	TTT	Alpha beta barrel
1QR7	2.6	268	58	3	0	0	HED	PB	34	24	–	–	–	–	–	3	TST	Alpha beta barrel
1SN8	2	85	–	1	0	0	H	PB	–	–	–	–	–	–	–	1	T	Beta barrel
1SYY	1.7	270	–	1	0	0	E	PB	–	–	–	–	–	–	–	1	H	Orthogonal bundle
1V0D	2.6	197	–	1	0	0	C	PB	–	–	–	–	–	–	–	1	H	Nil
1XXA	2.2	136	3	2	1	0	AF	PB	3	−1	–	–	–	–	–	3	ST	2–Layer sandwich
1ZHY	1.6	204	32	2	0	0	ER	PB	32	–	–	–	–	–	–	2	ST	Nil
2CH7	2.5	471	–	1	0	0	E	PB	–	–	–	–	–	–	–	1	H	Coiled coil
2EX3	3	169	–	1	0	0	D	PB	–	–	–	–	–	–	–	1	H	2-Layer sandwich
2FJ9	1.6	69	–	1	1	1	D	PB	−1	−1	–	–	–	–	–	3	H	Nil
2FP1	1.55	38	113	3	1	0	QEH	PB	3	110	−1	–	–	–	–	4	THT	Nil
2G0A	2.35	49	193	3	1	0	DDD	PB	189	4	−1	–	–	–	–	4	STH	Nil
2OQ1	1.9	119	–	1	0	0	C	PB	–	–	–	–	–	–	–	1	H	2-Layer sandwich
3EC8	2.6	–	–	0	1	0	–	PB	−1	–	–	–	–	–	–	1	–	Nil
3T8Y	1.9	13	47	3	0	0	DDE	PB	45	2	–	–	–	–	–	3	HTS	Nil
3TWY	1.5	187	61	5	1	0	DDDWD	PB	6	53	1	1	–	–	–	5	TSSST	Nil
1KK8	2.3	182	–	1	0	1	K	BEF	−1	–	–	–	–	–	–	2	H	Orthogonal bundle
1KMI	2.9	57	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	3-Layer sandwich
1L5Y	2.1	55	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	3-Layer sandwich
1W0 J	2.2	162	–	1	2	1	K	BEF	−1	−1	−1	–	–	–	–	4	H	Orthogonal bundle

(*Continued*)

Table 1. (*Continued*).

pid[a]	res[b]	posn[c]	nspan[d]	np[e]	nw[f]	nn[g]	dons[h]	met[i]	sd1[j]	sd2[j]	sd3[j]	sd4[j]	sd5[j]	sd6[j]	sd7[j]	cn[k]	sec.str[l]	Fold
1W7 J	2	169	–	1	1	1	K	BEF	−1	−1	–	–	–	–	–	3	H	Orthogonal bundle
1W9I	1.75	236	–	1	0	1	S	BEF	−1	–	–	–	–	–	–	2	S	Nil
1XHF	2.15	54	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	3-Layer sandwich
1ZES	1.9	53	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	3-Layer sandwich
2BEF	2.3	122	–	1	0	2	H	BEF	−1	−1	–	–	–	–	–	3	S	2-Layer sandwich
2R25	1.7	1144	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Up and down bundle
2V0 N	2.71	53	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
2VB6	2.3	157	–	1	0	1	K	BEF	−1	–	–	–	–	–	–	2	H	Nil
2VUI	2.9	55	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	3-Layer sandwich
2WF8	1.2	8	–	1	0	1	D	BEF	−1	–	–	–	–	–	–	2	S	3-Layer sandwich
2ZBF	2.4	351	–	1	0	1	D	BEF	−1	–	–	–	–	–	–	2	S	3-Layer sandwich
3A10	1.63	52	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
3DHF	1.8	247	–	1	0	0	H	BEF	–	–	–	–	–	–	–	1	H	Nil
3ECC	2.7	92	–	1	0	1	K	BEF	−1	–	–	–	–	–	–	2	H	Nil
3HB0	2.5	274	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
3I61	2.1	158	–	1	1	1	K	BEF	−1	−1	–	–	–	–	–	3	H	Nil
3ICE	2.8	366	–	1	0	0	R	BEF	–	–	–	–	–	–	–	1	T	Nil
3KQN	2.05	210	–	1	1	1	K	BEF	−1	−1	–	–	–	–	–	3	T	3-Layer sandwich
3NNN	2.2	53	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
3NNS	1.9	50	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
3OLV	1.7	57	–	1	0	0	D	BEF	–	–	–	–	–	–	–	1	S	Nil
3PEY	1.4	426	3	2	0	1	RR	BEF	3	−1	–	–	–	–	–	3	TT	Nil
3PUX	2.3	192	–	1	0	1	H	BEF	−1	–	–	–	–	–	–	2	T	Nil
3TMI	2.9	–	–	–	0	3	–	BEF	−1	−1	−1	–	–	–	–	3	–	Nil

[a]Protein ID as given in Protein data bank.
[b]Resolution.
[c]Position of the first residue in the chelate loop.
[d]Length span of the chelate loop.
[e]Number of donor atoms.
[f]Number of water molecules.
[g]Number of non-protein donors.
[h]Donor atoms specified as single letter codes.
[i]Metal ion.
[j]Positioning of the residues from the first residue of the chelate.
[k]Coordination number.
[l]Secondary structural elements of the chelate residues.

3.2.1. Residue coordination with arsenic. Arsenic was analyzed in its +3 oxidation state. In most of the structures, arsenic association was observed to be merely due to the influence of buffer. However, the major concern was to predict the coordination of arsenic irrespective of its role in the proteins to which it is bound. In most structures, arsenic was located at the surface of the protein. However, in the zinc-binding region of D-lyxose isomerase (PDB ID 2Y0O) arsenic coordinates with zinc through the oxygen of water forming a Zn–O–As bond. No other non-protein ligands apart from Zn were noted among the arsenic bound structures. The Zn–O–As coordination illustrates the chelation ability of Zn when supplemented with arsenic [17]. For arsenic, the chelate loops were largely single residue when coordination number was 1. The increase in coordination number was largely associated with solvents as a factor of space filling or charge balancing around the metal ion. The highest coordination number (cn) observed with arsenic bound structures was 4, constituting two donors and two solvent molecules within the coordination sphere. No non-protein ligand atoms coordinate with the ion. The analysis of the structures showed arsenic to coordinate more frequently with sulfur of cysteine within the first shell through a single bond and with oxygens of two waters giving the appearance of a bent molecular geometry. Arsenic interacts with cysteine thiolate with an average distance of 2.62 Å. An important mechanism of arsenic-induced disorders is its ability to bind with sulfydryl (–SH) containing molecules [17]. However, the trigonal geometry of arsenic via thiolates of three cysteines reported by Shi *et al.* [35] was not observed among the arsenic bound structures as there was only a single cysteine residue within a distance of 3 Å from the metal. Even if present in a sequence order, the residue was not close enough to coordinate with the metal. In the absence of cysteine residues, we observed the metal to coordinate with the charged and polar residues, aspartate, glutamate, and threonine, respectively, as in case of 2O6 W and 2GBM given in table 1. The coordination number, position of arsenic, and the donor types from the observed structures suggest that arsenic may not have any structural influence on the protein. Also, the coordination residues followed no specific secondary structural patterns and were noted to coordinate with residues of sheets, helices, and coils.

3.2.2. Residue coordination with beryllium. Since the configuration of beryllium does not allow it to remain a single ion [33], the coordination of beryllium was studied in monoanionic species as beryllium trifluoride (BeF_3^-). Other forms of beryllium, namely beryllium difluoride and beryllium tetrafluoride, were excluded for the study due to a limited data-set. The coordination study of beryllium bound structures reported the metal to be largely associated with single donor atoms; however, the coordination number varied between structures and ranged up to 4. Structures having a coordination number >1 were noted to be due to solvent and non-protein ligand atoms at the vicinity of the metal ion. The beryllium structures also showed similar chelate results as observed with arsenic. The length of the chelate for all beryllium coordination was observed to be 1 due to a single donor within the cut-off distance. Residues like aspartic acid, lysine, and arginine were predominant in coordinating with the metal. Beryllium trifluoride forms coordination with both the delta oxygens (δ1 and δ2) of aspartic acid. It was noted that beryllium also uses nitrogens as electron donors to form complexes. The delta nitrogens of histidine and arginine and the zeta nitrogen of lysine coordinate with the metal ion. The average metal–atom distances for the coordinating residues were predicted to be 2.263 Å for aspartic acid, 2.883 Å for nitrogen of arginine, 2.677 Å for nitrogen of lysine, and 2.551 Å for nitrogen of histidine. The average value of the individual metal–atom distances of aspartic acid was

observed to be 2.59 Å for gamma carbon, 1.84 Å for δ1, and 2.41 Å for δ2 oxygens, respectively. The closeness of delta oxygen of aspartic acid is probably due to electron sharing by the oxygen of the carboxylate. On analyzing the secondary structural patterns of the chelate residues, we observed that aspartic acids prefer sheet and lysines prefer helix structures to coordinate with the metal. Residues at the coils were less frequent in the coordination group.

3.2.3. Residue coordination with lead. Lead was analyzed in its +2 oxidation state. Unlike arsenic and beryllium, lead showed preference for multiple donors with solvents as part of the coordination sphere. However, solvents were rare and apart from 2FJ9 no other structures showed lead coordination with non-protein ligand atoms. The highest coordination number was 6 with five donor atoms within the first shell. Lead(II) forms stable complexes with both soft and hard donors. Under similar coordination environments, the affinity of Pb(II) towards sulfur-based ligands tends to be higher compared to harder oxygen-donor and nitrogen-donor ligands [36]. However, we noted only a few structures with Pb–S coordination through cysteine residues and more complexes with nitrogen and oxygen electron-donating compounds through aspartic acid, glutamic acid, and histidine. In cysteine-based coordination, as much as three residues coordinated with the metal through the gamma sulfur as observed in 1QNV. Though histidine and arginine possess strong electron donors, only a few structures show coordination with these residues. The distance between the metal and coordinating atoms were predicted to be 2.62 Å for delta oxygen of aspartic acid, 2.69 Å for epsilon oxygen of glutamic acid, 2.79 Å for gamma sulfur of cysteine and 2.57 Å for epsilon amino groups of histidine. In spite of its preference for acidic aminoacids, the coordination distance was noted to be unbiased with other residues. The closer distance of Pb with gamma sulfur of cysteine reveals its affinity for the soft ligand over the other donors [36]. The secondary structural features of the coordinating patterns indicated that all the single-residue chelates occur in the helical region of the protein, whereas multiple donors were located at the sheet or coil regions. This indicates that lead coordination with multiple donors prefers a linear arrangement of the residues. There was no symmetry observed between the residue distances even between identical chelates like the DDD patterns. The chelates were noted to be specific in their arrangement and positions. However, the most common and preferred chelate patterns indicated aspartates to co-occur with aspartates and cysteines with other cysteines. Patterns involving three residues namely histidine, aspartate, and glutamate were also preferred.

3.3. *Influence of structural folds in arsenic, beryllium and lead bound proteins*

The folds or architectures of the structures were identified from CATH and SCOP databases [30] and used to predict any influence of structural folds in metal coordination. Structures belonging to different functional folds like orthogonal bundle, alpha–beta barrel, 2-layer sandwich, 3-layer sandwich and many other folds were observed. Most of the arsenic bound proteins were uncharacterized, though structural folds like orthogonal bundle, roll, cytidine deaminase and alpha–beta complex were noted to coordinate with the metal. However, these folds did not show any specificity for the chelate residues. In lead bound structures, the alpha–beta barrel and orthogonal bundle were predominant. Similar to arsenic, there was no fold specificity observed among the chelate residues. Most of the beryllium bound structures belonged to the orthogonal bundle and 3-layer sandwich motif. All beryllium coordination shows single residue chelate except ATP-dependent RNA helicase (PDB ID

3PEY), which shows two donor residues within the coordination sphere. The 3-layer sandwich motif favored coordination with the metal through aspartic acid, preferably located in the sheet. Likewise, orthogonal bundles show preference over aspartic acid and lysine with the residues mostly located in a helix. Secondary structural preferences of the chelate residues were observed to be helices and sheets. Residues of turns are observed when the chelate length is >1.

3.4. *Structural deviation due to metal coordination*

The structural distortions in proteins due to binding of arsenic, beryllium and lead were predicted as a factor of phi and psi angle displacements of the residues within the coordination sphere. The empirical distribution of the phi–psi angles across five regions of the Ramachandran plot for the coordinating residues is plotted in figure 2. The five regions of Ramachandran plot were defined based on the values given by Deane and Blundell [37]. The regions A, B, and E refer to the alpha helix ($\varphi = -180$ to 0, $\psi = -120$ to 60), beta sheet ($\varphi = -180$ to 0, $\psi = 60$ to -240), and left-handed alpha helix ($\varphi = 90$ to 100, $\psi = -20$–80), respectively. Regions C and D refer to the partially allowed region ranging between $\varphi = -180$ and -40, $\psi = 0$ to -40. In all the metal bound structures, the alpha helical residues

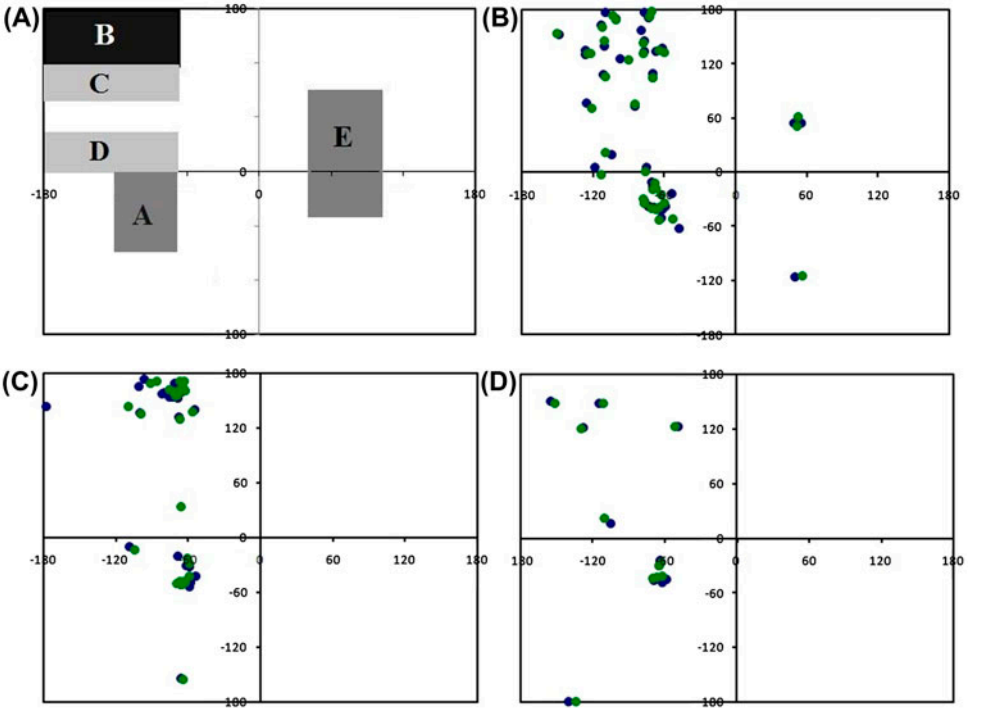


Figure 2. Ramachandran Plot displaying the phi–psi angle displacement.
Notes: (A) Five regions of Ramachandran Plot indicating A – alpha helices, B – beta sheets, C&D – additional allowed regions E- left handed alpha helices. (B) Plot showing phi–psi angle displacement of lead bound structures before and after metal removal. Plot of beta sheet residues mostly occurs in the additionally allowed regions. (C) Plot showing phi–psi angle displacement of beryllium bound structures before and after metal removal. The residues are confined within the core regions of alpha helices and beta sheets. The differences in the angles pre and post metal removal is evident. (D) Plot showing phi–psi angle displacement of arsenic bound structures before and after metal removal.

were confined with the region in the third quadrant of the plot. In lead and arsenic bound structures, the beta sheet residues were more scattered. The beryllium bound structures indicated a larger deviation than lead and arsenic. The predicted average deviation of phi–psi angles for the coordinating residues in the presence and absence of the metals were $1.48^{o}\varphi-2.10^{o}\psi$ for lead, $1.48^{o}\varphi/1.04^{o}\psi$ for arsenic, and $3.32^{o}\varphi/2.23^{o}\psi$ for beryllium. This higher phi–psi angle deviation in beryllium bound structures indicates the influence of the metal on the geometry of the residues in comparison with the other metals of this study. No major structural distortions were observed among the proteins.

3.5. *Coordination patterns beyond the first shell*

Since the coordination groups within 3 Å of the metal show a limited number of donor atoms, we increased the coordination distance to 5 Å in order to predict the nature of the metal environment. The coordination table describing the donor patterns within the extended coordination sphere for arsenic, lead, and beryllium is provided in supplementary table S1, see online at http://dx.doi.org/10.1080/00958972.2014.931945. Comparatively, a higher number of donor atoms for all the metals were observed beyond the first shell. For arsenic, the donor atoms ranged from two to seven compared to one within 3 Å, for lead it ranged from one to seven compared to five and for beryllium between one and eleven compared to one within the first shell. The residue specificities for the individual metals were analyzed and the predominant residues within 5 Å for all the metals are plotted as a histogram in figure 3. From figure, it is evident that the extended space is dominated by charged residues, aspartic acid, glutamic acid, and lysine. The predominance of these residues within the extended space creates an electrostatic potential that favors metal coordination. Nevertheless, in arsenic structures, this region was rich in cysteine. The different residue patterns of the donors within 5 Å for each metal are presented in table 2. Many dominating sequence patterns were observed in lead and beryllium structures. In spite of

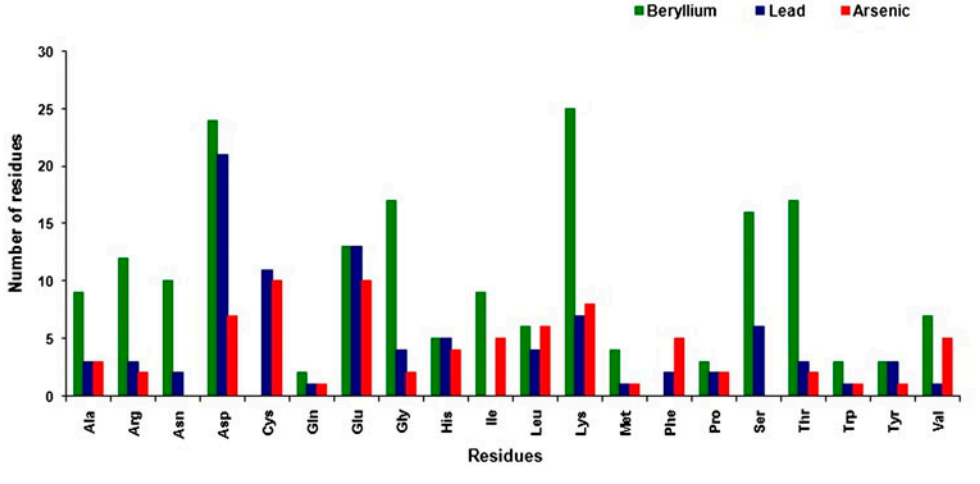


Figure 3. Predominant residues within the extended coordination space.
Notes: The extended coordination space is dominated by aspartic acid, lysine and glutamic acids in beryllium and lead bound structures. Arsenic is dominated by cysteine and glutamate residues.

Table 2. Dominating coordination patterns within 5 Å from the metal.

S. No.	Beryllium	Lead	Arsenic
1	DDIMLTAK	AE	DEEDGH
2	DDIRVTGHK	C	ECKIA
3	**DDVMITAK**	**CACGKCC**	EE
4	**DDVMITAK**	CC	EKDD
5	DIELTAYK	**CCCSK**	FCQKMK
6	DLDGASAKE	CRKHED	FCVP
7	DLDVTTKE	DDDEMKS	**FLICVI**
8	DQRITGYK	DDPE	**FPGDC**
9	**DWNVTAK**	**DDTKDSD**	IFLCY
10	**DWNVTVK**	**DSDDWD**	ILCVRD
11	EDINLTGK	EEY	KC
12	EDWMLTAK	EF	**LHE**
13	EHD	GTL	**LKHHKEE**
14	**ESGKNSS**	H	TD
15	**ESGKNSS**	H	VAECK
16	**ESGKNSS**	HEERGY	VALCRTW
17	GDKTITGKDND	NDEKD	
18	H	PANGFT	
19	KDE	QLEH	
20	KYRIHG	RLE	
21	PGKTDNR	SD	
22	PKR	SED	
23	PTGKEANRR	VKD	
24	**SGKNSS**	YLDD	
25	**SGKQEH**		
26	SGRR		
27	SR		
28	TGKEGRR		

Notes: The table shows the coordinating patterns of residues predicted within a distance of 5 Å from the metals. Most similar and dominating patterns are indicated as bold letters.

similar number of donors and residue composition, the residue patterns of arsenic bound structures varied widely due to the fact that it does not represent any functional role within the protein. However, the metal environment provides useful information for chelation studies. The geometry of metal coordination from the most common patterns observed in the extended space was analyzed using PyMOL visualization software [38]. We noted that the coordination geometry for the three metals was well defined only within the extended region of the metal coordination space. Figure 4 illustrates the different types of coordination patterns by lead. Along with aspartates, lead also formed stable associations with cysteine which was the next most important among the residues of the extended coordination space. Different molecular geometries representing tetrahedral in 1QNV [figure 4(A)], linear as in 1IW7 [figure 4(B)], tetrahedral in 2G0A [figure 4(C)], and pentagonal bipyramidal in 3TWY [figure 4(D)] coordinations of lead are observed [36], however, not in the first shell. No specific non-protein ligand coordination was noted within the first shell except in 2FJ9, where Pb coordinates with chlorine at a distance of 1.83 Å. In the extended coordination space, only two structures showed association with non-protein atoms. The extended space in beryllium was equally dominated by hydroxyl amino acids along with first shell residues, lysine and aspartic acid. The beryllium associations with aspartic acids are given in figure 5. The monodentate coordination extended through the carbonyl oxygen of aspartate is illustrated using the structures 3NNN [figure 5(A)], 2WF8 [figure 5(B)] and 1KMI [figure 5(C)]. The extended coordination space in beryllium was equally populated with

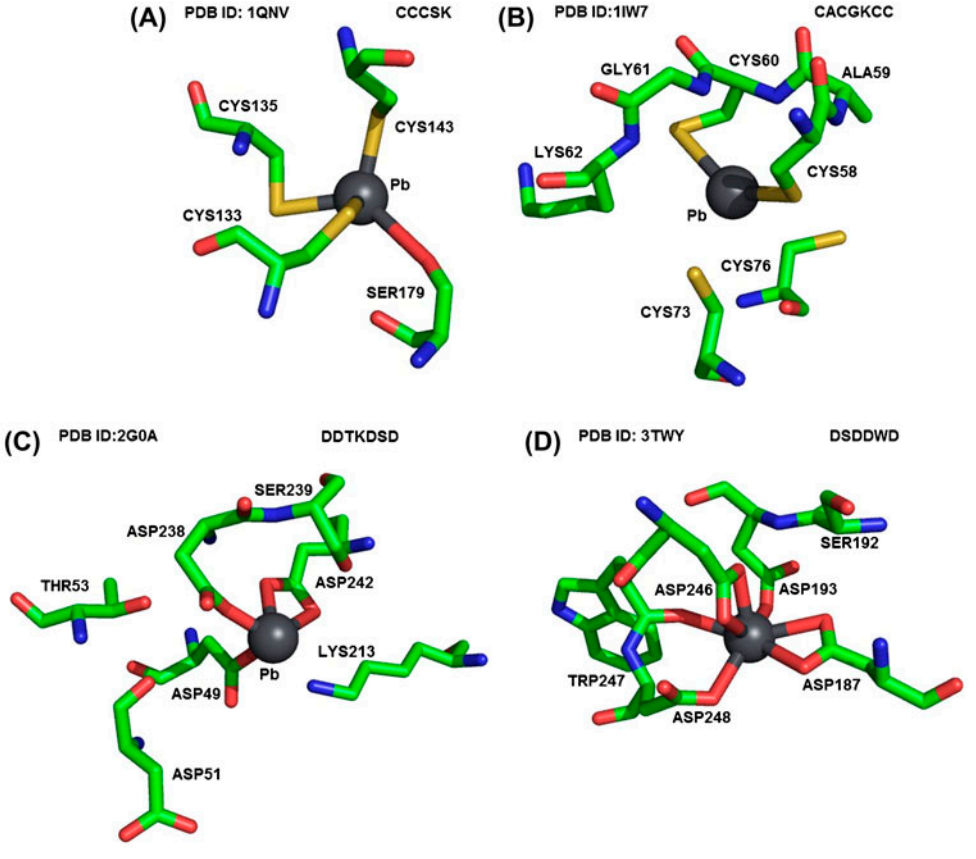


Figure 4. Different coordination patterns of lead.
Notes: (A) represents the tetrahedral, (B) shows the linear geometry, (C) represents tetrahedral coordination and (D) represents the pentagonal bipyramidal geometries of the metal. The coordination of the oxygen atoms and thiolates are evident. (C) and (D) shows the bidentate coordination of the delta oxygens of aspartic acid.

solvents and non-protein ligands, most of which were Mg^{2+} ions and ADP. BeF_3^- that mimics phosphate is usually used to inhibit the uptake of ATP by the enzyme and hence its association with ADP in most structures is noted. The complex formed by beryllium fluoride with the phosphate ligands, ADP, and pyrophosphate (PPi), individually forms a bidentate complex that possesses a chelating ability [39]. BeF_3^- coordinates with ADP in a tetrahedral geometry through the beta phosphate [40]. Though the BeF_3^-–ADP complex was noted among the structures, the bidentate coordination of BeF_3^-–ADP was observed only in 3TMI [39]. The beryllium bound structures were found to be associated with magnesium ions; however, no bonding between the metals was noted. About 50% of the beryllium bound structures showed magnesium association only in the extended coordination space and no specific residue patterns were noted [38]. Mg^{2+} displaces BeF_3^- by forming Mg–ADP complex and is generally used in crystallographic studies. These structures therefore do not represent the chelating property of beryllium, but the use of BeF_3^- and Mg^{2+} in crystallographic studies. The extended space also shows more solvent to as much as six water molecules, indicating preference for solvent reaction [41]. Unlike the first shell which

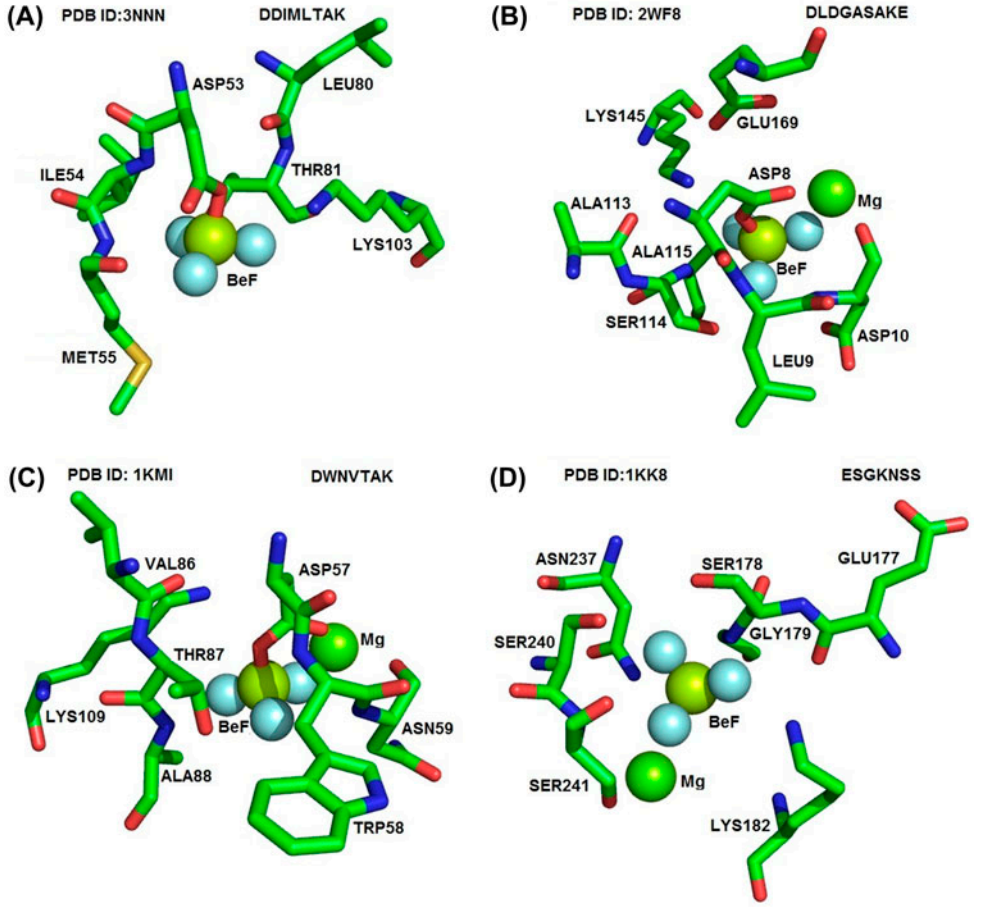


Figure 5. Different coordination patterns of beryllium.
Notes: The coordination of beryllium is shown in the form of beryllium trifluoride. In (A), (B) and (C) the coordination of the metal with aspartate oxygen is observed. In (D) the amino and oxygen atoms directed towards the metal is seen.

was rich in cysteine residues, the extended space in arsenic coordination is predominant in glutamic acids due to its association with zinc ion within the active site of the protein as seen in 1Y0R [figure 6(A)] and 2Y0O [figure 6(B)]. The coordination geometry of arsenic with zinc and the gamma sulfur of cysteines is given in figure 6(C) and (D). Though many common patterns were noted among the extended space residues, no noticeable sequence conservation was observed. However, in beryllium, most of the lysine and arginine residues were associated with glycine. These glycines reduce the steric hindrance around the metal coordination caused by the large side chains of the neighboring residues. These wide differences in coordination numbers and patterns between metals are justified by the differences in their valence shell electrons. However, variations within the same metal ions are largely due to the structural diversity of the protein folds.

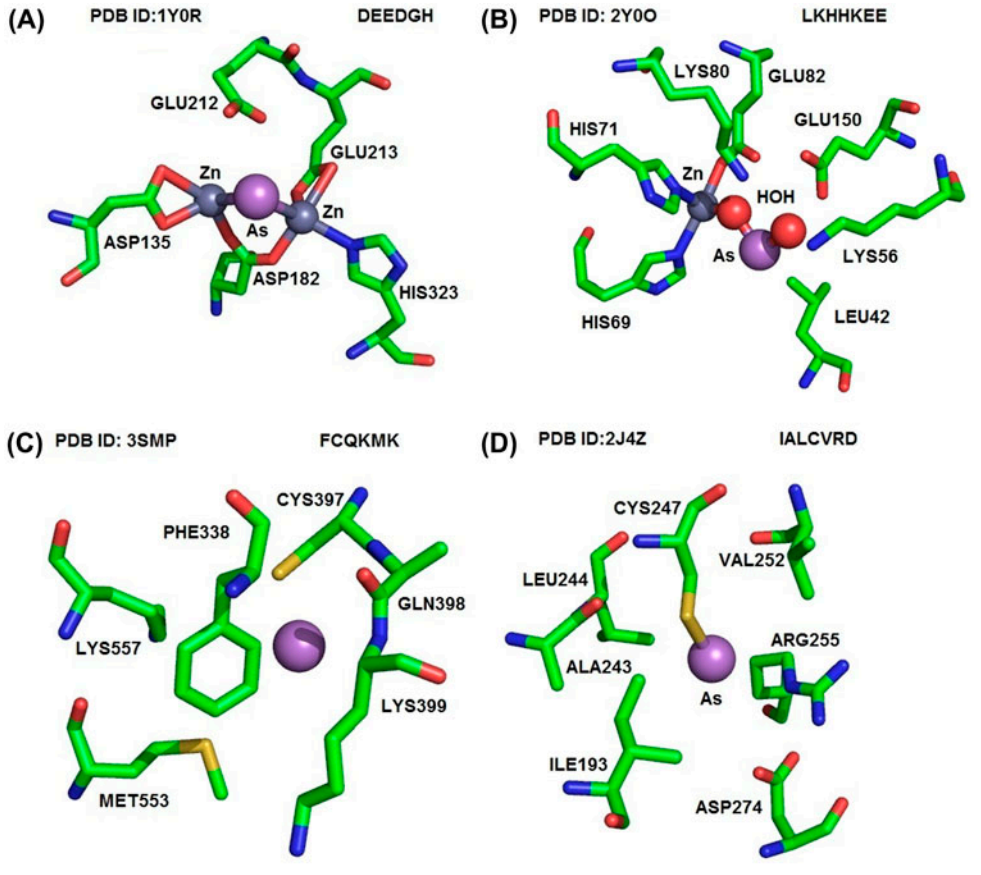


Figure 6. Different coordination patterns of arsenic.
Notes: The arsenic coordination for some of the chelates predicted within 5 Å of the metal is displayed. (A) and (B) show the side chains of the acidic and basic aminoacids oriented towards arsenic. In (C) and (D) the cysteine thiolate coordinates with arsenic.

3.6. *Propositions derived from the coordination study*

The results derived from the coordination analysis of the three metals provide useful insights into chelation. Analysis of the coordinating residues shows the preference for cysteines, glutamates, aspartates, and lysines for arsenic. The coordination number of one within the 3 Å metal environment further states the preference of a single donor in arsenic chelation. A maximum number of seven residues as observed with lead and arsenic in the extended coordination space indicates an ideal length for defining effective chelators. Similarly, beryllium coordination suggests the use of aspartates as the major donor. The presence of hydroxyl amino acids like serine and therionine along with the charged residues substantiates the presence of solvent around the metal center. However, bridging with glutamates, lysines, and arginines in a chelator length of 11 residues will eventually aid in chelation. The lead coordination gives clear insight into coordination due to the well-defined coordination geometries. The specific positioning of cysteine in the order of n, $n + 2$ and aspartates in the order of n, $n + 1/2$ in a chelate length of seven would prove significant in chelation.

4. Conclusion

Arsenic, lead and beryllium are implicated in carcinogenesis and mutagenesis. The association of these metals in protein structures is not much elaborated. The coordination analysis of these metals reported in the study provides better insight into the metal environment. The coordination patterns derived for arsenic and beryllium indicated single donor coordination, whereas, coordination number of 3 was observed with lead within the first shell. Similar to most metals, arsenic, lead, and beryllium also prefer coordination with aspartates and glutamates, imidazole nitrogens of histidine and sulfur of cysteine. However, according to analysis made from the coordination study, not only the residues but also the distances between the metal and residue atom play an important role. This is evident from analysis of the extended coordination sphere which revealed significant coordinating patterns demonstrating the dependency of the metal with the residues beyond the first shell. Also, the structural integrity and spatial positioning of the residues around the metal centers were largely derived from extended space. In lead bound structures, the clear geometry of metal coordination is observed with aspartate and cysteine clusters in a spatial positioning of 2 residue distance. A chelate length of seven residues was preferred as indicated by lead coordination. Similarly for beryllium, aspartates serve as major donors along with small and hydroxyl amino acids. Arsenic coordinates with cysteine as single donor within the first shell. No specific patterns within the extended space were noted. However, basic and non-polar amino acids were dominant. The coordination data derived from the study also provides the atomic and molecular details on metal atom interaction like donors, metal–donor distances, coordination number, residue preferences, and chelate patterns as predicted from the first shell and extended coordination space. These findings should be pivotal in understanding the metal environment and aid in designing chelation strategies based on the coordination pattern.

Acknowledgements

The authors thank the management of the VIT University for providing the facilities and encouragement to carry out this work. The authors also thank the editor and anonymous reviewers for their valuable comments and suggestions that helped to strengthen the quality of the work.

References

[1] S.J.S. Flora, M. Mittal, M.A. Mehta. *Indian J. Med. Res.*, **128**, 501 (2008).
[2] S. Mishra, S.P. Dwivedi, R.B. Singh. *Open Nutraceuticals J.*, **3**, 188 (2010).
[3] R.A. Goyer. *Am. J. Clin. Nutr.*, **61**, 646S (1995).
[4] O.A. Levander. *Environ. Health Perspect.*, **25**, 77 (1978).
[5] S. Soghoian, R.H. Sinert. eMedicine from WebMD (2009). Available online at: http://emedicine.medscape.com/article/814960-overview (accessed 4 May 2012).
[6] B. Thivierge, R. Frey. *Gale Encyclopedia of Medicine*, 3rd Edn (2006). Available online at: http://www.encyclopedia.com/topic/Heavy_metal_poisoning.aspx (accessed 10 June 2013).
[7] E.J. Tokar, L. Benbrahim-Tallaa, M.P. Waalkes. *Life Sci.*, **8**, 375 (2011).
[8] J.W. Hamilton, R.C. Kaltreider, O.V. Bajenova, M.A. Ihnat, J. McCaffrey, B.W. Turpie, E.E. Rowell, J. Oh, M.J. Nemeth, C.A. Pesce, J.P. Lariviere. *Environ. Health Perspect.*, **106**, 1005 (1998).
[9] R.N. Ratnaike. *Postgrad. Med. J.*, **79**, 391 (2003).
[10] T.A. El-Masry, A.M. Emara, N.A. El-Shitany. *J. Evol. Biol. Res.*, **3**, 4 (2011).

[11] E.K. Silbergeld, M. Waalkes, J.M. Rice. *Am. J. Ind. Med.*, **38**, 316 (2000).
[12] M. Kuschner. *Environ. Health Perspect.*, **40**, 101 (1981).
[13] M.K. Schubauer-Berigan, J.A. Deddens, J.R. Couch, M.R. Petersen. *Occup. Environ. Med.*, **68**, 354 (2011).
[14] C. Steinmaus, J.R. Balmes. *Environ. Health Perspect.*, **108**, 1003 (2000).
[15] S.J. Archibald. *J. Inorg. Chem.*, **107**, 274 (2011).
[16] S.R. Banerjee, K.P. Maresca, L. Francesconi, J. Valliant, J. Babich, W Zubieta. *J. Nucl. Med. Biol.*, **32**, 1 (2005).
[17] S.J.S. Flora, V. Pachauri. *Int. J. Environ. Res. Public Health*, **7**, 2745 (2010).
[18] M.M. Jones. In *Coordination Chemistry, A Century of Progress*, G.B. Kauffman (Ed.), pp. 427–438, American Chemical Society, Washington, DC (1994).
[19] A. Messadi, A. Mohamadou, I. Déchamp-Olivier, L. Dupont. *J. Coord. Chem.*, **65**, 2442 (2012).
[20] H. Golchoubian, E. Rezaee, G. Bruno, H.A. Rudbari. *J. Coord. Chem.*, **66**, 2250 (2013).
[21] R. Jesu Jaya Sudan, C. Sudandiradoss. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, **68**, 1346 (2012).
[22] L. Rulisek, J. Vondrasek. *J. Inorg. Biochem*, **71**, 115 (1998).
[23] M.M. Harding. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, **60**, 849 (2004).
[24] H. Zheng, M. Chruszcz, P. Lasota, L. Lebioda, W. Minor. *J. Inorg. Biochem.*, **102**, 1765 (2008).
[25] H.M. Berman, T. Battistuz, T.N. Bhat, W.F. Bluhm, P.E. Bourne, K. Burkhardt, Z. Feng, G.L. Gilliland, L. Iype, S. Jain, P. Fagan, J. Marvin, D. Padilla, V. Ravichandran, B. Schneider, N. Thanki, H. Weissig, J.D. Westbrook, C. Zardecki. *Nucleic Acids Res.*, **28**, 235 (2000).
[26] G. Wang Jr., R.L. Dunbrack. *Bioinformatics*, **19**, 1589 (2003).
[27] M. Dudev, J. Wang, T. Dudev, C. Lim. *J. Phys. Chem. B*, **110**, 1889 (2006).
[28] B.K. Kuntal, P. Aparoy, P. Reddanna. *Protein Pept. Lett.*, **17**, 765 (2010).
[29] I.N. Kasampalidis, I. Pitas, K. Lyroudia. *Protein*, **68**, 123 (2007).
[30] G. Csaba, F. Birzele, R. Zimmer. *BMC Struct. Biol.*, **9**, 23 (2009).
[31] N. Guex, M.C. Peitsch. *Electrophoresis*, **18**, 2714 (1997).
[32] P.G. Jones. *Chem. Soc. Rev.*, **13**, 157 (1984).
[33] R. Adair. *Beryllium*, Rosen publishing group, New York (2007).
[34] I. Bertini, P. Turano. In *Biological Inorganic Chemistry: Structure and Reactivity*, I. Bertini (Ed.), pp. 31–41, University Science Books, New Jersey, NJ (2007).
[35] W. Shi, J. Don, R.A. Scott, M.Y. Ksenzenko, B.P. Rosen. *J. Biol. Chem.*, **27**, 9291 (1996).
[36] R.L. Davidovich, V. Stavila, K.H. Whitmire. *Coord. Chem. Rev.*, **254**, 2193 (2010).
[37] C.M. Deane, T.L. Blundell. In *Perspectives in Structural Biology*, M. Vijayan, N. Yathindra, A.S. Kolaskar (Eds.), pp. 196–208, Indian Academy of Sciences, Hyderabad (1999).
[38] W.L. DeLano. *The PyMOL Molecular Graphics System*, DeLano Scientific, San Carlos, CA (2002). Available online at: http://www.pymol.org (accessed 19 May 2013).
[39] J.P. Issartel, A. Dupuis, C. Morat, J.L. Girardet. *Eur. Biophys. J.*, **20**, 115 (1991).
[40] Y.W. Xu, S. Morera, J. Janin, J. Cherfils. *Biochemistry*, **94**, 3579 (1997).
[41] M. Walther, A. Budimir, R. Puchta. *J. Coord. Chem.*, **65**, 4359 (2012).