Structural biology of the lanthanides—M—mining rare earths in the Protein Data Bank

Kristina Djinovic-Carugo[a, b]

Oliviero Carugo[c, *]

olicar04@unipv.it

[a]Max F. Perutz Laboratories, University of Vienna, 5 Campus Vienna Biocenter, A-1030, Austria

[b]Department of Biochemistry, Faculty of Chemistry and Chemical Technology, University of Ljubljana, Askerceva 5, SI-1000 Ljubljana, Slovenia

[c]Department of Chemistry, University of Pavia, Viale Taramelli 12, I-27100 Pavia, Italy

*Corresponding author.

Abstract

With its about 100,000 three-dimensional structures, the Protein Data Bank is a copious source of information: it contains also some hundreds of structures of macromolecules complexed with lanthanide cations, which are examined here. These cations, which are found in a wide variety of protein types, were introduced to determine the structures, by exploiting their anomalous dispersion (in crystallographic studies, where they are also used as crystallization additives) or the paramagnetic pseudocontact shifts (in NMR analyses). The coordination numbers in the first coordination sphere are very variable, though they tend to be close to those that are observed in small molecules or in water solution. The coordination polyhedra are also quite variable as it can be expected for large cations. Interestingly, lanthanide cations are frequently observed in packing bridges between symmetry equivalent molecules in crystals, where they tend to form polynuclear complexes, with up to seven cations bridged by water/hydroxide ligands.

Keywords: Protein Data Bank; Lanthanides; Coordination number; Coordination polyhedron; Packing bridge; Clusters

1 Introduction

A considerable fraction of the proteome of any organism is made by metalloproteins (up to 50% [1]) and metal cations participate to many (if not most or even all) biological processes [2,3]. Some metals are present in large amounts in living organisms (for example calcium, sodium, potassium and magnesium). Others are present only in very small amounts (for example copper, zinc or iron). Nine transition metals are essential, at low concentration, for any type of organism (bacteria, plants or animals) vanadium, chromium, manganese, iron, cobalt, nickel, copper, zinc, molybdenum [3]. Other trace metals, like tungsten or cadmium, might be essential for some species [3]. The large majority of the other metals are not only not essential for life but also extremely toxic, even at low concentration.

The lanthanides are the 15 element from lanthanum to lutetium. All of them have very similar chemical properties (like for example the unique, stable oxidation state of + 3) and are therefore difficult to separate and purify from each other from the minerals where they tend to be systematically co-present. They find nevertheless several applications and are strategically important for several industries. Complexes of gadolinium are for example routinely used in the hospitals as contrast agents in magnetic resonance imaging [4]. Terbium and europium phosphorescence are exploited in numerous optical devices [5]. Since they are isomorphous with calcium(II), several lanthanides(III) were used as probes of calcium binding to proteins. Holmium based lasers are used in medical and dental applications [6].

Lanthanides are certainly not essential for human life, though they are ubiquitous at relatively low concentrations. Given their modest abundance, their biological activity received modest attention. A general survey of their biochemistry has been published [7]. Much of the biochemical uses of the lanthanides centers is due to their ability to provide information on the interactions of calcium(II) with macromolecules [8]. In particular, the photophysical properties of some lanthanide cations, typically europium(III) and terbium(III), have been widely exploited to examine the calcium(II) surrounding in several biological samples [9–11]. Most of the industrial applications of the lanthanides in biochemistry and medicine have been limited to magnetic resonance contrast agents, where gadolinium(III) compounds are routinely used [12,13].

Some studies are nevertheless recently being published on the physiological roles of the lanthanides, for example about the bioaccumulation in bones [14]. It has been reported that lanthanide cations are necessary for the growth of Methylacidiphilum fumariolicum SolV, an extremely acidophilic methanotropic microbe isolate in volcanic mudpot in southern Italy near Neaples [15]. It has been proposed that this is due to a lanthanide-dependent methanol dehydrogenase [15],

like the recently reported XoxF1 of Methylobacterium extorquens AM1 [16]. Interactions between lanthanide and ribulose-1,5-bisphosphate carboxylase have been proposed to be the reason for the positive effects of fertilizers enriched with lanthanide salts [17]. It must also be remembered that the lanthanides may be calcium antagonists in living systems

Despite their little known biological impact, lanthanides are nevertheless used in structural biology because of their magnetic properties, particularly interesting in Nuclear Magnetic Resonance (NMR) studies, and because of their anomalous scattering of the X-rays, exploited in crystallographic studies to solve the phase problem. Furthermore, they have been used as probes to characterize local structural features, like for example the calcium(II) binding sites in EF-hands and other proteins [18–20]. Since they were not covered in a recent review on the metal cations in protein structures [21], the occurrence on lanthanide cations on the Protein Data Bank [22,23] is summarized and commented in the present paper.

2 Materials and methods

Crystal packing contacts and packing bridges were identified as was previously described [24,25].

The technique used to solve the phase problem in crystal structures was extracted manually from the PDB files [22,23] or from the literature, since it is not always indexed in the Protein Data Bank, especially for old structures.

Molecular graphics and analyses were performed with the UCSF Chimera package, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311) [26].

3 Results and discussion

3.1 Binding to biological macromolecules

Lanthanide cations were found in 194 entries of the Protein Data Bank. This clearly indicates that they are rather uncommon, since there are over 100,000 entries in the Protein Data Bank. Since some entries have more than one chain, a total of 363 chains were found and since some entries have more than one lanthanide cation, a total of 619 lanthanide cations were observed. As it can be seen in Table 1, most of them are in protein crystal structures. All the lanthanide cations are in the oxidation state + 3, with few exceptions: two cerium cations complexed to human lactoferrin are in the oxidation state + 4, stabilized by treatment with H_2O_2 (PDB 1fck) [27], and ytterbium is in the uncommon oxidation state + 2 in a synthetic right-handed coiled-coil tetramer [28].

Table 1 Number of lanthanide cations observed in the Protein Data Bank.

Lanthanide	Total number of cations observed in the PDBf	Number of protein molecules that contain a lanthanide	Number of protein crystal structures that contain a lanthanide
La	19	19	13
Ce	7	7	4
Pr	51	51	51
Nd	0	0	0
Pm	0	0	0
Sm	110	110	110
Eu	49	49	49
Gd	168	168	165
Tb	37	32	28
Dy	0	0	0
Ho	35	35	35
Er	6	6	6
Tm	0	0	0
Yb	143	143	143

Lu	54	53	53

3.2 Uses of the lanthanides in structural biology

In the case of NMR studies, paramagnetic lanthanide cations have been used to exploit the paramagnetic pseudocontact shifts, which provide long-range distance information (up to 40–45 ~~Angstroms~~Å) [29–31]. For example, the positions of the two cerium(III) cations, which replace the physiological calcium(II) cations in the N-terminal domain of calmodulin, were determined by using paramagnetic pseudocontact shifts and T1 relaxation enhancements produced by the lanthanide cations together with more conventional nuclear Overhauser effect (NOE) constraints (PDB file = 1ak8) [32]. This is a typical use of the lanthanide cations: being isomorphous with calcium(II) and owing to their charge being higher than that of calcium(II), they easily replace calcium(II) and can be used to probe the calcium(II) environment.

With regard to crystallographic studies, lanthanide cations have been used in solving the phase problem in about 60% of the cases (~ 40% by multiple wavelength anomalous ~~D~~dispersion (MAD) experiments, ~ 50% by single wavelength anomalous dispersion (SAD) experiments, and the rest by multiple isomorphous replacement (MIR) experiments). The other ~ 40% of the crystal structures that contain lanthanide cations were solved either by molecular replacement (MR) methods (about one half of them), by refinement of a known existing structure (about one half of them) or even, in only one case, by ab initio methods (PDB = 2anv; bacteriophage P22 lysozyme) [33].

The anomalous dispersion signal of several lanthanide cations has been exploited in MAD and SAD experiments. In general the L absorption edges were used, since they have energies accessible at synchrotron beam lines (see Table 2): most of the L absorption edges are in the 1–2 Å range and the L-III absorption edge is one of the main features for anomalous phasing. Moreover, lanthanides have larger anomalous signals than selenium (usually exploited in MAD experiments based on SeMet labelled proteins) and the phasing power of terbium is approximately ~~4~~four times larger than that of selenium. Even the holmium K edge was used in MAD experiments at ultra-high energy, with wavelengths equal to 0.2229, 0.2227 and 0.2200 Å (55.62, 55.68 and 56.34 keV; namely inflection, peak and high remote, respectively) [34].

Table 2 Energy and wavelenght of the X-ray absorption edges of the lanthanide metals. Data taken ~~at~~from http://skuld.bmsc.washington.edu/scatter/AS_periodic.html.

Ln	Energy (keV)							Wavelength (Å)						
	K	L-I	L-II	L-III	M1	M2	M3	K	L-I	L-II	L-III	M1	M2	M3
La	38.9246	6.2663	5.8906	5.4827	1.3613	1.2044	1.1234	0.3185	1.9786	2.1048	2.2614	9.1078	10.2943	11.0365
Ce	40.4430	6.5488	6.1642	5.7234	1.4346	1.2728	1.1854	0.3066	1.8932	2.0114	2.1663	8.6424	9.7411	10.4593
Pr	41.9906	6.8348	6.4404	5.9643	1.5110	1.3374	1.2422	0.2953	1.8140	1.9251	2.0788	8.2054	9.2705	9.9810
Nd	43.5689	7.1260	6.7215	6.2079	1.5753	1.4028	1.2974	0.2846	1.7399	1.8446	1.9972	7.8705	8.8383	9.5564
Pm	45.1840	7.4279	7.0128	6.4593	1.6465	1.4714	1.3569	0.2744	1.6692	1.7680	1.9195	7.5302	8.4263	9.1373
Sm	46.8342	7.7368	7.3118	6.7162	1.7228	1.5407	1.4198	0.2647	1.6025	1.6957	1.8460	7.1967	8.0473	8.7325
Eu	48.5190	8.0520	7.6171	6.9769	1.8000	1.6139	1.4806	0.2555	1.5398	1.6277	1.7771	6.8880	7.6823	8.3739
Gd	50.2391	8.3756	7.9303	7.2428	1.8808	1.6883	1.5440	0.2468	1.4803	1.5634	1.7118	6.5921	7.3437	8.0301
Tb	51.9957	8.7080	8.2516	7.5140	1.9675	1.7677	1.6113	0.2385	1.4238	1.5025	1.6500	6.3016	7.0139	7.6947
Dy	53.7885	9.0458	8.5806	7.7901	2.0468	1.8418	1.6756	0.2305	1.3706	1.4449	1.5916	6.0575	6.7317	7.3994
Ho	55.6177	9.3942	8.9178	8.0711	2.1283	1.9228	1.7412	0.2229	1.3198	1.3903	1.5362	5.8255	6.4481	7.1206
Er	57.4855	9.7513	9.2643	8.3579	2.2065	2.0058	1.8118	0.2157	1.2715	1.3383	1.4834	5.6190	6.1813	6.8432
Tm	59.3896	10.1157	9.6169	8.6480	2.3068	2.0898	1.8845	0.2088	1.2257	1.2892	1.4337	5.3747	5.9328	6.5792
Yb	61.3323	10.4864	9.9782	8.9436	2.3981	2.1730	1.9498	0.2022	1.1823	1.2426	1.3863	5.1701	5.7057	6.3588
Lu	63.3138	10.8704	10.3486	9.2441	2.4912	2.2635	2.0236	0.1958	1.1406	1.1981	1.3412	4.9769	5.4775	6.1269

Lanthanide cations were inserted in protein crystals both by soaking or co-crystallization. Quite often they were used as crystallization additive; for example holmium chloride was employed to obtain well diffracting crystals of *Agaricus bisporus* tyrosinase

(PDB 2y9w) [35] and praseodymium acetate was used for crystallizing the human pox virus and zinc finger (POZ) domain of leukemia/lymphoma related factor (PDB 2if5) [36]; crystallization of the *Mycobacterium tuberculosis* sensory transduction protein regX3 in the presence of lanthanum chloride allowed to improve the crystallographic resolution from 6.0 Å to 2.02 Å (PDB 2oqr) [37]. The influence of lanthanide salts on crystal packing and symmetry is well documented also by the observation that crystals of *Rattus norvegicus* metabotropic glutamate receptor subtype 1 belong to a different space group if they are grown with gadolinium chloride ($P3_22_1$) or without the lanthanide salt ($P4_12_12$) (PDB 1isr) [38]. Similarly, bovine type IV collagen noncollagenous domain-alpha1 crystallized in a smaller unit cell ($P2_1$ space group with a = 80.07 Å, b = 137.96 Å, c = 127.13 Å, β = 90.3° and two hexamers in the asymmetric unit) after soaking with lutetium chloride, while the unit cell of the initial crystals was considerably larger (a = 129.41 Å, b = 143.87 Å, c = 162.92 Å, and β = 91.3° and four hexamers in the asymmetric unit) (PDB 1m3d) [39].

In the large majority of the crystal structures of the PDB that contain lanthanide cations, these were introduced either by soaking or co-crystallization as simple hydrated ions. In few cases, on the contrary, the lanthanide cations were contained in larger coordination compounds. For example terbium and europium dipicolinate (dpa) complexes of generic formula $M_3[Ln(dpa)_3]$ (Fig. 1; Ln = Tb^{3+} or Eu^{3+}; M = Li^+, Na^+, Cs^+ or NH^{4+}) were co-crystallized with hen egg-white lysozyme. The [Ln(dpa) 3]3 + complex cations were found at the interface between symmetry equivalent protein molecules, interacting preferentially with arginine side chains (PDB 2pc2) [40]. This finding was confirmed by the co-crystallization of lanthanide complexes of dipicolinate with thaumatin from *Thaumatococcus daniellii*, urate oxidase from *Aspergillus flavus*, and xylanase from *Trichoderma reesei* (PDB 2pe7, 2pes, and 3lgr) [41]. Hydroxymethyltriazole and hydroxyethyltriazole dipicolinate (4-(4-(hydroxymethyl)-1 h-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylate and 4-(4-(2-hydroxyethyl)-1H-1,2,3-triazol-1-yl)pyridine-2,6-dicarboxylic acid, respectively) europium complexes were used to determine the crystal structure of *Gallus gallus* lysozyme C and of ~~Thaumatococcus~~ *T. daniellii* thaumatin-1 [42]. Also a hydroxyethylcholinetriazoledipicolinate (N-((1-(2,6-dicarboxypyridin-4-yl)-1 h-1,2,3-triazol-4-yl)methyl)-2-hydroxy-N,N-dimethylethanaminium) europium complex was used to solve the crystal structure of ~~Gallus~~ *G. gallus* lysozyme C [42]. Also a tetraaza macrocycle derivative complexed with gadolinium (Fig. 2), more hydrophobic than the dipicolinate complexes, was co-crystallized with hen egg-white lysozyme and two gadolinium complexes were found in the enzyme active site (PDB 1-h87) [43]. A gadolinium complex with a tetraaza macrocycle w~~ere~~as used also to determine the crystal structures of mouse monoclonal antibodies [44].

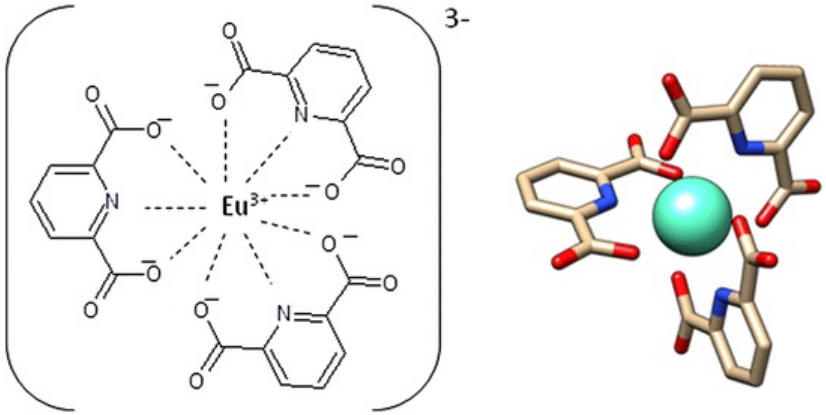


Fig. 1 Scheme and structure of the tris(pyridine-2,5-dicarboxylate)europate(III) or tris(dipicolinate)europate(III) (data taken from a crystal structure of the hen egg-white lysozyme; PDB 2pc2).

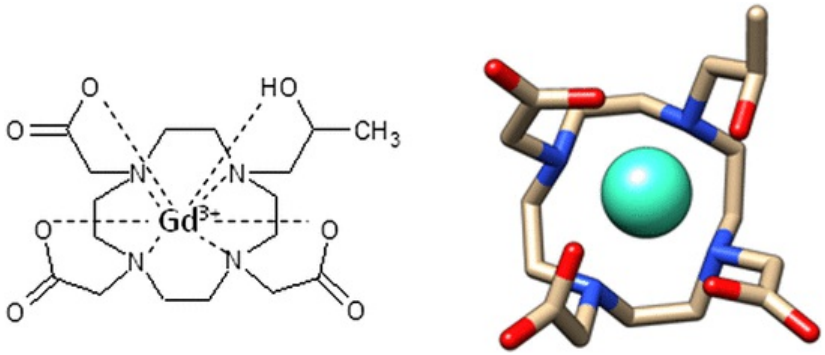


Fig. 2 Scheme and structure of (10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetato)gadolinium(III) (data taken from a crystal structure of the hen egg-white lysozyme; PDB 1-h87).

Also lanthanide-binding tags were designed and used to introduce lanthanide cations into proteins. A 38 residue long polypeptide, able to coordinate two terbium cations (or other lanthanides), was N-linked to human ubiquitin and used to solve the

phase problem for the structure determination (Fig. 3) (PDB 2ojr) [45]. Lanthanide containing tags were used also in NMR structure determinations. A 19 residue long polypeptide, able to coordinate lanthanum (or other lanthanides), was N-tagged to Streptococcus immunoglobulin G-binding protein G, and used to solve the solution structure by using the paramagnetic pseudo-contact shifts (PDB 2rpv) (Fig. 4) [46]. Interestingly, the first residues of this N-terminal tag is a cysteine that forms a disulfide bond with Cys 38 of the protein, making the overall construct more rigid. Several other lanthanide tags have been studied and used as structural probes [47–49].

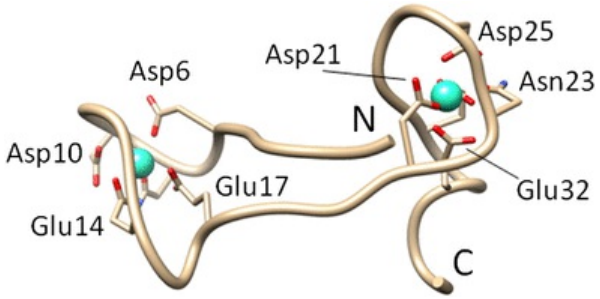


Fig. 3 N-terminal bilanthanide-binding tag (GPGYIDTNNDGWIEGDELYIDTNNDGWIEGDELLA) used to determine the crystal structure of human ubiquitin (PDB 2ojr).

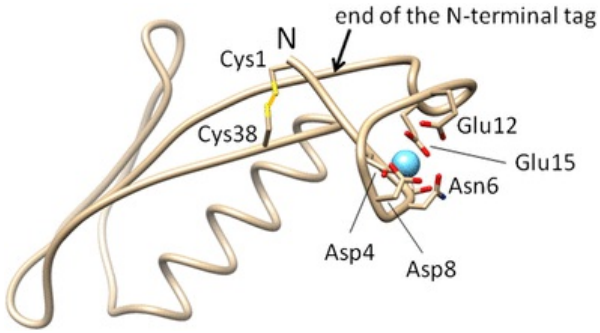


Fig. 4 N-terminal anchoring lanthanide-binding tag (CYVDTNNDGAYEGDELSGT) used to determine the solution structure of Streptococcus immunoglobulin G-binding protein G (PDB 2rpv).

3.3 Lanthanides as packing bridges

As it is shown above, lanthanide cations have been used to solve the crystallographic phase problem, by exploiting their anomalous diffraction and their atomic mass. In several cases, however, it seems that they have been used only as crystallization additives. The use of these compounds for improving the quality of macromolecular crystals is likely related to the fact that these metal cations resemble closely to the much more common alkaline and alkaline-earth metals, in the sense that they lack any redox chemistry, though they have a larger electric charge. It is thus interesting to examine the data of the Protein Data Bank to see if it is possible to extract some information about the role of the lanthanide cations in the crystals of the macromolecules: indeed, several examples of lanthanide packing bridges have been observed.

A packing bridge is an atom that is in contact with two protein molecules that are adjacent in the crystal because of a symmetry operation [24]. The analysis of 1346 carefully selected protein crystal structures showed that about 45% of the protein crystal structures deposited into the Protein Data Bank contain at least one packing bridge and about 10% of the crystal packing contacts occur in the presence of a packing bridge [24]. These data suggested that there are small molecules or monoatomic ions that can bridge two protein molecules in solution, initiating the nucleation of the protein crystals. This also reflects the importance of the crystallization additives that are commonly used to improve the quality of protein crystals: an additive may favor the reciprocal orientation of the protein molecules that progressively allow crystals to grow.

~~50%~~Fifty percent of the protein crystal structures examined here have at least one packing bridge and 9% of the crystal packing contacts occur in the presence of a packing bridge. About 20% of the lanthanide cations observed in the Protein Data Bank form packing bridges and about 50% of the crystal structures examined here have at least one packing bridge that involves a lanthanide cation. These numbers compare favorably with those observed in previous studies [24], suggesting the lanthanide cations act as typical packing bridges. Their high electric charge and their modest dimension, comparable to calcium(II), make these cations particularly able to mediate inter-molecular interactions between surface patches. A visual inspection of a randomly selected sub-set of the data indicates that lanthanide cations are often bridging anionic side-chains at the protein surface of adjacent proteins. An example is shown in Fig. 5: in the crystal structure of human DNA repair endonuclease HAP1 (space group C2; PDB 1bix) [50], the molecule present in the asymmetric unit forms a crystal packing contact with a molecule that is related to the first one by a crystallographic two-fold axis parallel to the axis b of the unit cell; a samarium cation is sandwiched in between the two protein molecules

and bridges them by interacting with three glutamate and one glutamine side-chains.

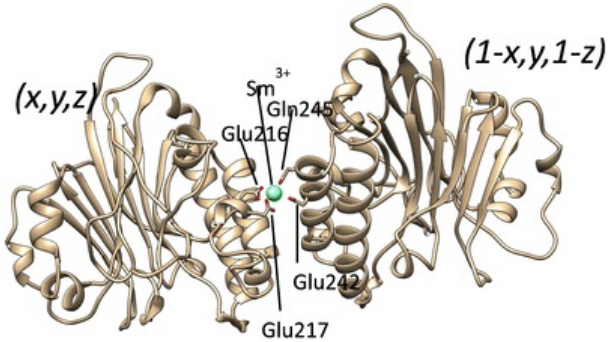


Fig. 5 Samarium cation forming a packing bridge in the crystal structure of human DNA repair endonuclease HAP1 (monoclinic space group C2; PDB 1bix).

This observation is supported by the presence of picolinate lanthanide coordination compounds at the interface between symmetry equivalent protein molecules, interacting preferentially with arginine side chains (PDB 2pc2) [40]. Interestingly, some of these lanthanide coordination compounds are commercially available as heavy atom anomalous scatterers for MAD/SAD experiments (http://www.jenabioscience.com/cms/en/1/catalog/2181_lanthanide_phasing_kit.html).

3.4 Coordination numbers and polyhedra

The first coordination sphere around a lanthanide cation is described by two variables, the coordination number and the symmetry of the coordination polyhedron. Both of them were monitored with the utility “Metal geometry” of the software suite Chimera [26] and validated by visual inspection. Only the structures determined by crystallographic methods and refined at a resolution of at least 2 Å were considered.

It is important here to make clear that the coordination numbers that can be extracted from the Protein Data Bank are not necessarily the genuine coordination numbers, since some atoms or groups of atoms might have been undetected experimentally, like for example some water molecules that are conformationally too disordered. For this reason, we use here the expression “observed coordination number” to indicate the coordination number extracted from the Protein Data Bank.

Both the observed coordination number and polyhedron can be computed by considering the raw data as they are deposited in the Protein Data Bank, and by considering dummy donor atoms, located at a reasonable distance from the metal center, in coordination positions that are empty in the crystal structure (this second option was applied only if the rough coordination number was at least 3) [26]. An example of this procedure is shown in Fig. 6: an ytterbium cation was found at the surface of human leukotriene A-4 hydrolase (PDB: 3fh8); it is coordinated by the side-chain of an aspartate and by three water molecules in a geometry that might resemble an extremely distorted tetrahedron; by using the utility “Metal geometry” of the software suite Chimera, two additional water molecules (two dummy atoms) can find their place close to the metal cation, the coordination polyhedron of which becomes a nearly regular octahedron.

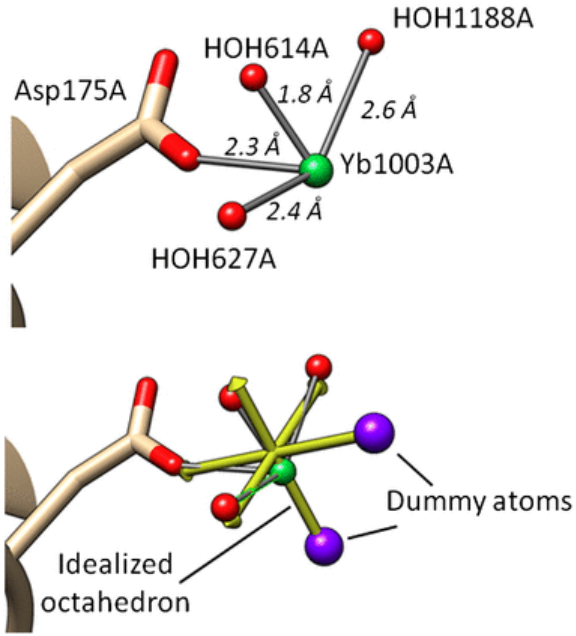


Fig. 6 Example of dummy atom identification.

The distribution of the observed coordination numbers is shown in Table 3. Obviously, if necessary, dummy donor atoms are considered (right column of the table), the observed coordination numbers are larger than if only the rough data are considered (central column of the table). More than 70% of the metal centers are surrounded by at least eight donor atoms if dummy donor atoms are taken into account. On the contrary, by examining only the rough data of the Protein Data Bank and by ignoring these additional, dummy donor atoms, more than 60% of the metal centers have a observed coordination number lower that eight; about 5% of the metal centers are completely isolated, in the sense that they are not in proximity (a threshold value of 3 Å was arbitrarily fixed) of any other atom (either a protein atom or an atom of a solvent molecule or an atom of any other small molecule present in the crystal structure); 20% of the lanthanide cations are observed to be coordinated by less than ~~3~~three atoms.

Table 3 Distribution of the observed coordination numbers of the lanthanide cations in the Protein Data Bank. There are two cases: on the one hand, only the donor atoms deposited in the Protein Data Bank are considered; on the other hand, additional atoms are located around the metal center, if there is enough room for them.

Coordination number	Without additional atoms	With additional atoms
0	4.7%	
1	3.9%	
2	11.4%	
3	5.7%	
4	7.6%	
5	7.5%	0.1%
6	9.9%	12.6%
7	10.1%	15.6%
8	24.6%	22.2%
9	14.6%	43.8%

10		5.7%

All these metal centers exhibiting low observed coordination numbers in the rough data set must be considered with some caution, since they are impossible from a chemical–physical perspective: each atom in a material in the solid state must be in contact with other atoms. In principle, it is possible that these isolated atoms are interpretation errors [51], for example metal cations erroneously interpreted as lanthanide cations. However, their detection by the crystallographers is likely to be, in most cases at least, a genuine observation and not an artifact, since atoms as heavy as the lanthanide cations can hardly be confused with other chemical species. It is possible, on the contrary, that some lanthanide cations are missed when their occupancy is low.

The most common observed coordination numbers are 8, if the rough data deposited in the Protein Data Bank are considered, and 9, if additional, dummy donor atoms are taken into account. The prevalence of the latter values is extremely marked: more than 40% of the lanthanides seem to be surrounded by nine potential donor atoms. These data agree with the observation that the coordination number of the lanthanides is ca. 8–9 in aqueous solution [52,53]. Actually, the coordination number tends to decrease along the lanthanide series, according to the contraction of the ionic radius of the cations. However, such a periodic trend cannot be detected on the basis of the data available in the Protein Data Bank, the resolution of which is likely to be insufficient to allow one to appreciate subtle differences between individual lanthanide elements.

The distribution of the coordination polyhedra is shown in Table 4 and some examples are given in Fig. 7. The coordination polyhedra are in general quite distorted, in particular those associated with low coordination numbers. Surprisingly, when there are six donor atoms coordinated to the central cation, the more common geometry is not octahedral but trigonal prismatic. The octahedron is in fact much more common in the coordination chemistry of the transition metals [54,55].

Table 4 Frequency of observation of the coordination polyhedra around the lanthanide cations observed in the Protein Data Bank. There are two cases: on the one hand, only the donor atoms deposited in the Protein Data Bank are considered; on the other hand, additional atoms are located around the metal center, if there is enough room for them.

Observed coordination number	Coordination polyhedron	Without additional atoms	With additional atoms
4	Irregular	32%	
	Square plane	37%	
	Tetrahedron	31%	
5	Square pyramid	25%	91%
	Trigonal bipyramid	75%	9%
6	Octahedron	37%	39%
	Trigonal prism	63%	61%
7	Irregular	22%	21%
	Pentagonal bipyramid		10%
	Trigonal prism, square face monocapped	11%	20%
	Monocapped octahedron	67%	49%
8	Irregural	25%	29%
	Trigonal prism, square face bicapped	75%	71%
9	Irregular	24%	13%
	Trigonal prism, square face tricapped	45%	69%
	Monocapped square antiprism	31%	18%
10	Bicapped square antiprism		100%

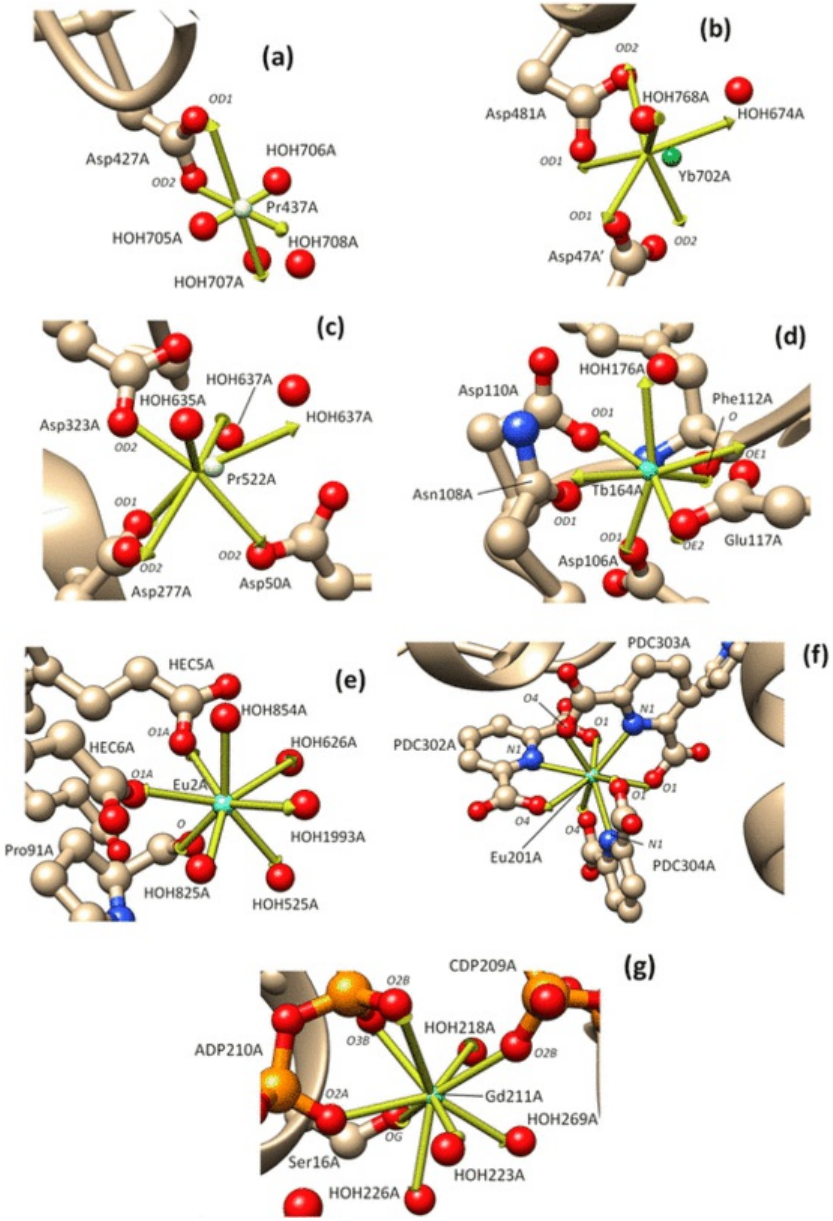


Fig. 7 Examples of lanthanide cations with various coordination numbers and polyhedra (the idealized geometry is indicated by green arrows): (a) octahedron (PDB 2znr); (b) trigonal prism (PDB 3fuk): a triangular face is define by the atoms OD2 of Asp481A, HOH768A, and HOH674A; the other by OD1 of Asp481A and OD1 and OD2 of Asp47A'; (c) trigonal prism, square face monocapped (PDB 2isn): the square face defined by OD2 of Asp277A, HOH635A, HOH637A and OD2 of Asp50A is capped by HOH637A; (d) octahedron monocapped (PDB 1ncz): the triangular face defined by the atoms OE2 of Glu117A, O of Phe112A, and HOH167A is capped by atom OE1 of Glu117a; (e) Trigonal prism, square face bicapped (PDB 3tor): the two, opposite triangular faces of the trigonal prism are defined by the atoms O1A of HEC5A, HOH854A, and HOH626A (upper face) and by the atoms O of Pro91A, HOH825A, and HOH525A (lower face); the atom O1A of HEC6A caps the square face defined by the atoms O1A of HEC5A, O of Pro91A, HOH854A, and HOH825A; the atom HOH1993A caps the square face defined by the atoms HOH854A, HOH626A, HOH828A, and HOH525A (HEC is Heme C); (f) trigonal prism, square face tricapped (PDB 2pc2): the upper triangular face is defined by one atom of PDC302A (O1) and two atoms of PDC303A (O4 and N1); the lower triangular face is defined by one atom of PDC302A (O4) and two atoms of PDC304A (N1 and O4); the other three atoms (N1 of PDC302A, O1 of PDC304A, and O1 of PDC303A) cap the three square faces of the trigonal prism (PDC is pyridine-2,6-dicarboxylic acid); (g) monocapped square antiprism (PDB 3akc): the atom HOH226A caps the square face defined by the atoms OG of

Ser16A, O2A of ADP210A, HOH223A, and HOH269A; the other square face is defined by the atoms O2B and O3B of ADP210A, HOH218A, and O2B of CDP209A (ADP is adenosine-5′-diphosphate and CDP is cytidine-5′-diphosphate).

It is also quite surprising that the monocapped octahedron is more common than the monocapped trigonal prism or the pentagonal bipyramid when the coordination number is equal to seven. In fact, there are no large potential energy barriers between these polyhedra and several types of geometry are expected when the seven donor atoms are equal and independent [54]. This unexpected finding might be related to stereochemical constraints provided by protein atoms and by water molecules at the protein surface. However, it might also depend on the paucity of the data or on their local inaccuracy.

Geometries related to the trigonal prism are the most common also for higher coordination numbers. The bicapped trigonal prism is the most common geometry for coordination number equal to eight and the tricapped trigonal prism is the most frequent geometry for coordination number equal to nine.

This contrasts to the observation that the square antiprism is the principal structure for coordination number eight when the eight ligands are not involved in chelation or bridging to other atom [56]. However, in protein structures, other structural constraints are certainly and systematically present. Moreover, one must also remember that the several geometries, which are observed for high coordination numbers, are energetically close and can be inter- converted by small angular changes [56]. The prevalence of the tricapped trigonal prism over the monocapped square antiprism, on the contrary, agrees with the fact that the tricapped trigonal prism is known to be slightly more stable [54].

3.5 Clusters

The lanthanide cations are observed quite frequently to form polynuclear complexes with proteins, which are usually named clusters in protein biochemistry [57]. This was first observed at high resolution in the crystal structure of the adhesive domain of the yeast flocculin Flo5 soaked with gadolinium (PDB 4ahw): a cluster of seven gadolinium cations was observed on a bowl-shaped surface patch framed by four aspartates (Fig. 8); several water molecules bridge the gadolinium cations [58]. The kinetics of the cluster formation were studied by determining the structure of a series of crystals that were soaked with gadolinium for variable times, and it was proposed that this tendency of the lanthanides to form clusters at the protein surface is due to the fact that the rate of hydrolysis and of protein-template cluster formation is faster than the oligomerization in solution at pH 7.0–7.5 [58].

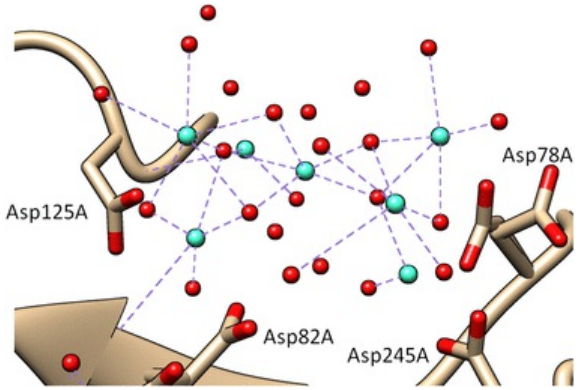


Fig. 8 Cluster of seven gadolinium cations observed in the crystal structure of the adhesive domain of the yeast flocculin Flo5 soaked with gadolinium (PDB 4ahw): gadolinium cations are shown in light blue and water oxygen atoms in red.

Several other lanthanide clusters have been observed in lower resolution structures, like for example a trinuclear gadolinium cluster in the structure of mannosyl-3-phosphoglycerte phosphatase from Thermus thermophilus (PDB 3zty) [59] and a tetrahedral ytterbium cluster at the dimer interface of human blood coagulation factor XIII (PDV 1ggy) [60].

Frequently these clusters behave as packing bridges. About 60% of the lanthanide clusters in the structures determined at a crystallographic resolution better than (or equal to) 2.0 Å are found at the interface between symmetry equivalent protein molecules. For example, in the orthorhombic (space group P21212) crystals of human leukotriene a-4 hydrolase (Fig. 9) (PDB 1hs6) a dinuclear ytterbium cluster is at the interface between two symmetry equivalent protein molecules [61]; in the crystal structure of the POZ domain of human LRF (Fig. 10) (tetragonal space group I4122; PDB 2if5), a binuclear praseodymium site is observed at the interface between two symmetry equivalent molecules. If these polynuclear packing bridges are formed by crystal soaking with lanthanide salts, this would indicate that polynuclear lanthanide complexes are preferentially formed at the interface between adjacent molecules because these are the zones where the templating power of the protein surface is doubled. Alternatively, if they are formed during protein crystallization, this would reflect the stronger power of a multicharged cluster to act as a template for the formation of nascent intermolecular contacts. Therefore one can conclude that these cations have a remarkable influence in crystallogenesis and in ensuring the quality of the protein crystals [36].

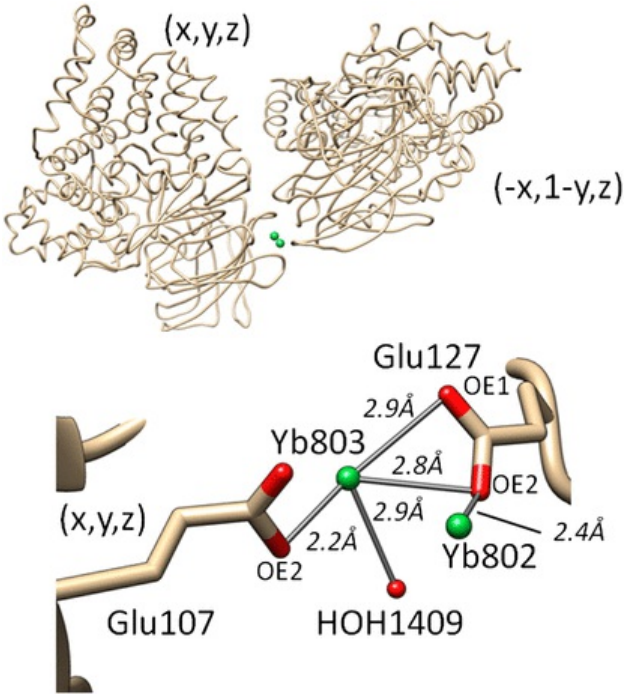


Fig. 9 Binuclear packing bridge observed in the crystal structure of human leukotriene A4 hydrolase (orthorhombic space group $P2_12_12$; PDB 1hs6).

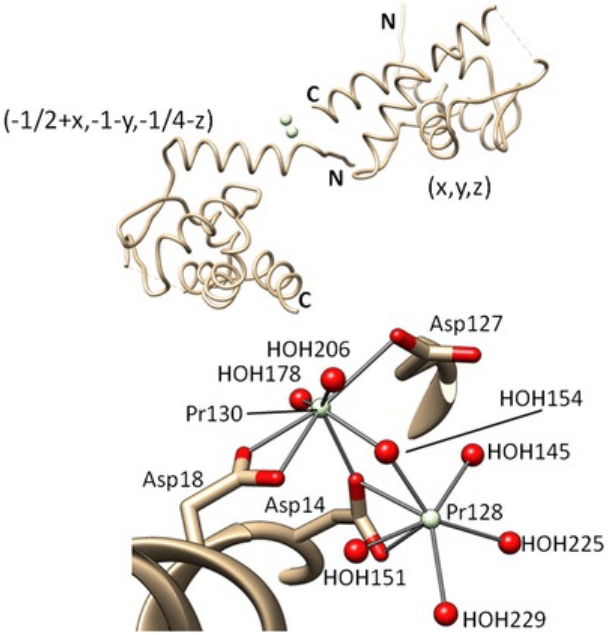


Fig. 10 Binuclear packing bridge observed in the crystal structure of the POZ domain of human LRF (tetragonal space group $I4_122$; PDB 2if5).

4 Conclusions

Lanthanide salts and complexes are used in macromolecular crystallography not only because of the anomalous scattering that may allow the solution of the phase problem but also because they often allow one to improve the diffraction quality of the crystals. This is due, at least in part, to their ability to form mononuclear or polynuclear packing bridges that stabilize the crystal lattice. Coordination numbers and polyhedra are quite variable, though they tend to be similar to those observed in small molecules, with eight to nine donor atoms that surround the metal center in a capped trigonal prismatic geometry. The results of these analyses should prove useful to examine and validate new protein crystal structures that contain lanthanide cations.

References

[1]

Y. Lu, N. Yeung, N. Sieracki and N.M. Marshall, *Nature* **460**, 2009, 855–862.

[2]

R.R. Crichton, Biological Inorganic Chemistry, 2012, Elsevier.

[3]

J.J.R. Frausto da Silva and R.J.P. Williams, The Biological Chemistry of the Elements, 2001, Oxford University Press.

[4]

L.M. Mitsumori, P. Bhargava, M. Essig and J.H. Maki, *Top. Magn. Reson. Imaging* **23**, 2014, 51–69.

[5]

R.C. Evans, P. Douglas and C.J. Winscomb, *Coord. Chem. Rev.* **250**, 2006, 2093–2126.

[6]

N. Krishnamurthy and C.K. Gupta, Extractive Metallurgy of Rare Earths, 2004, CRC Press.

[7]

C. Evans, Biochemistry of the Lanthanides, 1990, Springer.

[8]

P.M. Colman, L.H. Weaver and B.W. Matthews, *Biochem. Biophys. Res. Commun.* **46**, 1972, 1999–2005.

[9]

J.C. Buenzli, *Interface Focus.* **3**, 2013, 2013003.

[10]

J.C. Buenzli, *Chem. Rev.* **110**, 2010, 2729–2755.

[11]

Y.Y. Song, Y.Z. Xu, S.F. Weng, L.B. Wang, X.F. Li, T.F. Zhang and J.G. Wu, *Biospectroscopy* **5**, 1999, 371–377.

[12]

Y. Chang, G.H. Lee, T.J. Kim and K.S. Chae, *Curr. Top. Med. Chem.* **13**, 2013, 434–445.

[13]

P.L. de Sousa, J.B. Livramento, L. Helm, A.E. Merbach, W. Même, B.T. Doan, J.C. Beloeil, M.I. Prata, A.C. Santos, C.F. Geraldes and E. Tóth, Contrast Media Mol, *Imaging* **3**, 2008, 78–85.

[14]

C. Vidaud, D. Bourgeois and D. Meyer, *Chem. Res. Toxicol.* **25**, 2012, 1161–1175.

[15]

A. Pol, T.R.M. Barends, A. Dietl, A.F. Khadem, J. Eygensteyn, M.S.M. Jetten and H.J.M. Op den Camp, *Environ. Microbiol.* **16**, 2014, 255–264.

[16]

T. Nagakawa, R. Mitsui, A. Tani, K. Sasa, S. Tashiro, T. Iwama, T. Hayakawa and K. Kawai, *PLoS ONE* **7**, 2012, e50480.

[17]

C. Lui, F.S. Hong, Y. Tao, T. Liu, Y.N. Xie, J.H. Xu and Z.R. Li, *Biol. Trace Elem. Res.* **143**, 2011, 1110–1120.

[18]

R.H. Kretsinger and C.D. Barry, *Biochim. Biophys. Acta* **405**, 1975, 40–52.

[19]

T.D. Veenstra, M.D. Gross, W. Hunziker and R. Kumar, *J. Biol. Chem.* **270**, 1995, 30353–30358.

[20]

N. Coruh and J.P. Riehl, *Biochemistry* **31**, 1992, 7970–7976.

[21]

M.M. Harding, M.W. Nowicki and M.D. Walkinshaw, *Crystallogr. Rev.* **16**, 2010, 247–302.

[22]

H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov and P.E. Bourne, *Nucleic Acids Res.* **28**, 2000, 235–242.

[23]

F.C. Bernstein, T.F. Koetzle, G.J. Williams, E.F. Meyer, Jr., M.D. Brice, J.R. Rodgers, O. Kennard, T. Shimanouchi and M. Tasumi, *J. Mol. Biol.* **112**, 1977, 535–542.

[24]

O. Carugo and K. Djinović-Carugo, *J. Appl. Crystallogr.* **47**, 2014, 458–461.

[25]

O. Carugo and K. Djinović-Carugo, *J. Struct. Biol.* **180**, 2012, 96–100.

[26]

E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng and T.E. Ferrin, *J. Comput. Chem.* **25**, 2004, 1605–1612.

[27]

H.M. Baker, C.J. Baker, C.A. Smith and E.N. Baker, *J. Biol. Inorg. Chem.* **5**, 2000, 692–698.

[28]

M. Sales, J.J. Plecs, J.M. Holton and T. Alber, *Prot. Sci.* **16**, 2007, 2224–2232.

[29]

M. Gochin and H. Roder, *Protein Sci.* **4**, 1995, 296–302.

[30]

R.D. Guiles, S. Sarma, R.J. DiGate, D. Banville, V.J. Basus, I.D. Kuntz and L. Waskell, *Nat. Struct. Biol.* **3**, 1996, 333–339.

[31]

L. Banci, I. Bertini, K.L. Bren, M.A. Cremonini, H.B. Gray, C. Luchinat and P. Turano, *J. Biol. Inorg. Chem.* **1**, 1996, 117–126.

[32]

D. Bentrop, I. Bertini, M.A. Cremonini, S. Forsen, C. Luchinat and A. Malmendal, *Biochemistry* **36**, 1997, 11605–11618.

[33]

B.H. Mooers and B.W. Matthews, *Acta Crystallogr.* **D62**, 2006, 165–176.

[34]

J. Jakoncic, M. Di Michiel, Z. Zhong, V. Honkimaki, Y. Jouanneau and V. Stojanoff, *J. Appl. Crystallogr.* **39**, 2006, 831–841.

[35]

W.T. Ismaya, H.J. Rozeboom, A. Weijn, J.J. Mes, F. Fusetti, H.J. Wichers and B.W. Dijkstra, *Biochemistry* **50**, 2011, 5477–5486.

[36]

F.D. Schubot, J.E. Tropea and D.S. Waugh, *Biophys. Res. Commun.* **351**, 2006, 1–6.

[37]

J. King-Scott, E. Nowak, E. Mylonas, S. Panjikar, M. Roessle, D.I. Svergun and P.A. Tucker, *J. Biol. Chem.* **282**, 2007, 37717–37729.

[38]

D. Tsuchiya, N. Kunishima, N. Kamiya, H. Jingami and K. Morikawa, *Proc. Natl. Acad. Sci. U. S. A.* **99**, 2002, 2660–2665.

[39]

M. Sundaramoorthy, M. Meiyappan, P. Todd and B.G. Hudson, *J. Biol. Chem.* **277**, 2002, 31142–31153.

[40]

G. Pompidor, A. D'Aléo, J. Vicat, L. Toupet, N. Giraud, R. Kahn and O. Maury, *Angew. Int. Ed.* **47**, 2008, 3388–3391.

[41]

G. Pompidor, O. Maury, J. Vicat and R. Kahn, *Acta Crystallogr.* **D66**, 2010, 762–769.

[42]

R. Talon, L. Nauton, J.-L. Canet, R. Kahn, E. Girard and A. Gautier, *Cem. Commun.* **48**, 2012, 11886–11888.

[43]

E. Girard, L. Chantalat, J. Vicat and R. Kahn, *Acta Crystallogr.* **D58**, 2002, 1–9.

[44]

T.M. Corneillie, A.J. Fisher and C.F. Meares, *J. Am. Chem. Soc.* **125**, 2003, 15038–15048.

[45]

N.R. Silvaggi, L.J. Martin, H. Schwalbe, B. Imperiali and K.N. Allen, *J. Am. Chem. Soc.* **129**, 2007, 7114–7120.

[46]

T. Saio, K. Ogura, M. Yokochi, Y. Kobashigawa and F. Inagaki, *J. Biomol. NMR* **44**, 2009, 157–166.

[47]

T. Saio, M. Yokochi, H. Kumeta and F. Inagaki, *J. Biomol. NMR* **46**, 2010, 271–280.

[48]

Y. Kobashigawa, T. Saio, M. Ushio, M. Sekiguchi, M. Yokochi, K. Ogura and F. Inagaki, *J. Biomol. NMR* **53**, 2012, 53–63.

[49]

K.N.I. Allen, *B Curr. Opin. Chem. Biol.* **14**, 2010, 247–254.

[50]

M.A. Gorman, S. Morera, D.G. Rothwell, E. de La Fortelle, C.D. Mol, J.A. Tainer, I.D. Hickson and P.S. Freemont, *EMBO J.* **16**, 1997, 6548–6558.

[51]

H. Zheng, M.D. Chordia, D.R. Cooper, M. Chruszcz, P. Müller, G.M. Sheldrick and W. Minor, *Nat. Protoc.* **9**, 2014, 156–170.

[52]

G.V. Ionova, C. Madic and R. Guillamont, *Russ. J. Coord. Chem.* **27**, 2001, 439–442.

[53]

T. Lowall, F. Foglia, L. Helm and A.E. Merbach, *J. Am. Chem. Soc.* **117**, 1995, 3790–3799.

[54]

B.W. Clare and D.L. Kepert, Encyclopedia of Inorganic Chemistry Online, 2006, Wiley & Sons.

[55]

F.A. Cotton, G. Wilkinson, C.A. Murillo and M. Bochmann, Advanced Inorganic Chemistry, 1999, Wiley & Sons.

[56]

S. Adam, A. Ellern and K. Seppelt, *Chem. Eur. J.* **2**, 1996, 398–402.

[57]

H. Matsubara, K. Wada and R. Masaki, *Adv. Exp. Med. Biol.* **74**, 1976, 1–15.

[58]

M. Veelders and E. L.-O., *ChemBioChem* **13**, 2012, 2187–2190.

[59]

B.A. Fox, V.C. Yee, L.C. Pedersen, I. Le Trong, P.D. Bishop, R.E. Stenkamp and D.C. Teller, *J. Biol. Chem.* **274**, 1999, 4917–4923.

[60]

S. Goncalves, A.M. Esteves, H. Santos, N. Borges and P.M. Matias, *Biochemistry* **50**, 2011, 9551–9555.

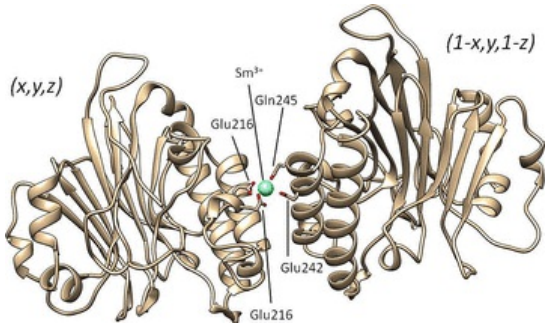
[61]

M.M. Thunnissen, P. Nordlund and J.Z. Haeggstrom, *Nat. Struct. Biol.* **8**, 2001, 131–135.

elsevier_JIB_9628

Graphical abstract

The protein crystal structures of the Protein Data Bank that contains lanthanide cations are inspected to determine the importance of lanthanides in crystallization and structure determination. Moreover, the lanthanide coordination numbers and geometries are scrutinized together with the packing bridges determined by the lanthanides.



Highlights

- We examine the coordination chemistry of the lanthanides in protein structures.
- The coordination numbers in the first coordination sphere are very variable.
- The coordination numbers resemble those that are observed in small molecules or in water solution.
- The coordination polyhedra are very variable.
- Lanthanides are frequently observed in packing bridges between symmetry equivalent molecules.

Queries and Answers

Query:

Please confirm that given names and surnames have been identified correctly.

Answer: Names and surnames have been identified correctly.

Query:

Your article is registered as a regular item and is being processed for inclusion in a regular issue of the journal. If this is NOT correct and your article belongs to a Special Issue/Collection please contact r.balan@elsevier.com immediately prior to returning your corrections.

Answer: This article is a regular item that must being processed for inclusion in a regular issue of the journa.

Query:

"There are 2 graphical abstract captions provided. The Synopsis has been followed. Please check if appropriate and amend if necessary."

Answer: It is fine.

Query:

Please confirm that surnames have been identified correctly and are presented in the desired order.

Answer: A modification is needed; the authors of this article are: M. Veelders and L.-O. Essen