Journal of Inorganic Biochemistry 143 (2015) 69–76

Contents lists available at ScienceDirect

Journal of Inorganic Biochemistry

journal homepage: www.elsevier.com/locate/jinorgbio

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

Kristina Djinovic-Carugo [a,b], Oliviero Carugo [c,*]

[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

ARTICLE INFO

Article history:
Received 10 September 2014
Received in revised form 2 December 2014
Accepted 2 December 2014
Available online 11 December 2014

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

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.

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

* Corresponding author.
E-mail address: olicar04@unipv.it (O. Carugo).

http://dx.doi.org/10.1016/j.jinorgbio.2014.12.005
0162-0134/

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 H2O2 (PDB 1fck) [27], and ytterbium is in the uncommon oxidation state +2 in a synthetic right-handed coiled-coil tetramer [28].

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 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.

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

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 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 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].

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