Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

A complement to the modern crystallographer's toolbox: caged gadolinium complexes with versatile binding modes

Meike Stelter,[a,b,c]* Rafael Molina,[d] Sandra Jeudy,[e] Richard Kahn,[a,b,c]‡ Chantal Abergel[e] and Juan A. Hermoso[f]

[a]University Grenoble Alpes, Institut de Biologie Structurale (IBS), Grenoble, France, [b]Centre National de la Recherche Scientifique, IBS, Grenoble, France, [c]Commissariat à l'Énergie Atomique et aux Énergies Alternatives, IBS, Grenoble, France, [d]Structural Biology and Biocomputing Programme, Spanish Cancer Research Centre (CNIO), Melchor Fdez Almagro, 28029 Madrid, Spain, [e]Laboratoire Information Génomique et Structurale, CNRS – UMR7256, Institut de Microbiologie de la Méditerranée, 163 Avenue de Luminy, 13288 Marseille CEDEX 09, France, and [f]Departamento de Cristalografía y Biología Estructural, Instituto Química Física Rocasolano, CSIC, Serrano 119, 28006 Madrid, Spain

‡ Deceased.

Correspondence e-mail: meike.stelter@ibs.fr

Received 12 February 2014
Accepted 10 March 2014

A set of seven caged gadolinium complexes were used as vectors for introducing the chelated Gd^{3+} ion into protein crystals in order to provide strong anomalous scattering for *de novo* phasing. The complexes contained multidentate ligand molecules with different functional groups to provide a panel of possible interactions with the protein. An exhaustive crystallographic analysis showed them to be nondisruptive to the diffraction quality of the prepared derivative crystals, and as many as 50% of the derivatives allowed the determination of accurate phases, leading to high-quality experimental electron-density maps. At least two successful derivatives were identified for all tested proteins. Structure refinement showed that the complexes bind to the protein surface or solvent-accessible cavities, involving hydrogen bonds, electrostatic and CH–π interactions, explaining their versatile binding modes. Their high phasing power, complementary binding modes and ease of use make them highly suitable as a heavy-atom screen for high-throughput *de novo* structure determination, in combination with the SAD method. They can also provide a reliable tool for the development of new methods such as serial femtosecond crystallography.

1. Introduction

Despite the growing number of available structures allowing the phases of the structure factors to be obtained by molecular replacement, experimental phases remain essential in the absence of suitable homologous structures. Unless the protein contains endogenous anomalous scatterers, introducing a heavy atom (HA) or anomalous scatterer into the inherently fragile protein crystals without perturbing diffraction quality remains one of the bottlenecks of protein structure determination. Phasing methods based on anomalous scattering have the advantage that no isomorphous derivatives are required. Table 1 shows a summary of the compounds used for experimental phasing. Using selenomethionyl protein crystals for MAD phasing is the most used approach (Hendrickson *et al.*, 1990; Dauter & Nagem, 2002), but expression and crystallization is not always straightforward and, although possible (Doublié, 2007), is not yet commonly applied in eukaryotic expression. For a review of derivatization methods using classical HA compounds, see Garman & Murray (2003). Sulfur SAD (Hendrickson & Teeter, 1981; Ramagopal *et al.*, 2003*b*) is limited to relatively well diffracting protein crystals that contain a large proportion of Met or Cys residues, as it provides a comparatively small anomalous signal, and faces technical problems inherent to the use of softer X-rays. Site-specific radiation damage can be exploited as an additional source of phase information (Ravelli *et al.*, 2003; Schiltz *et al.*, 2004). The search for an adequate HA derivative often

Table 1
Examples of heavy-atom compounds or endogenous scatterers used in anomalous phasing methods.

Shown are typical anomalous differences and binding modes. Certain derivatives are also suitable for isomorphous replacement methods. λ_{abs}, absorption edge wavelength; N_H, typical number of anomalous scatterers per 100 amino-acid residues; q_j, typical binding-site occupancy; $\Delta F^{\pm}/F$, theoretical anomalous signal (equation 1).

Type	Compounds	λ_{abs} (Å)	f'' (e^-)	N_H, q_j	$\Delta F^{\pm}/F$ (%)	Binding mode
SelenoMet	Selenium†	0.98	3.8	2, 1	4.1	Selenomethionine
Halide compounds‡	Bromine	0.92	3.8	3, 0.5	2.5	Binding in solvation shell. Arg, Lys. Also as brominated nucleic acids§ or iodination of tyrosines with *N*-iodosuccinimide.
	Iodine triiodide¶	1.54††	6.8	3, 0.5	4.5	
	I_3C‡‡§§ (or B_3C)	1.54††	6.8	6, 0.8	10.2	Hydrogen bonds through functional carboxylate and amine groups
Noble gases	Xe	1.54††	7.3	3, 0.3	2.9	Pressure derivatization. Binding in hydrophobic cavities.
	Kr	0.86	3.8	3, 0.3	1.5	
Weak endogenous scatterers¶¶	S	2.0††	0.90	4, 1	1.4	Endogenous
		1.54††	0.56		0.85	
	P	2.0††	0.70	¶¶	2.7	
		1.54††	0.43		1.7	
	Cl	2.0††	1.1	2, 0.8	0.96	
Endogenous scatterers	Mn	1.89	9.7	1, 1	7.3	
	Fe	1.74	4.5	1, 1	3.4	
Classical heavy-atom salts†††, *e.g.* magic seven	$UO_2(C_2H_3O_2)_2$, $K_3UO_2F_5$, lanthanides	1.54††	13.4	2, 0.3	4.3	Class A metals: charge interactions with electronegative protein residues, *e.g.* carboxylate of Glu/Asp. Lns in Ca sites.
	K_2PtCl_4, K_2HgI_4, $HgCl_2$, PCMBS‡‡‡, $KAu(CN)_2$, Pb, Tl	Pt: 1.1 Hg: 1.0	10 10	1, 0.5 1, 1	3.9 7.7	Class B metals: covalent binding to soft ligands (Cys, His, Met) or interaction with hydroxyl groups.
Caged, cyclen-based Ln complexes	Gd	1.71	28	2, 0.5	21.3	Hydrogen bonds through functional carboxylate and amine groups, water network, CH–π interaction with aromatic residues, ionic interaction of Gd^{3+} ion and carboxylates.
	Gd	1.54††	12	2, 0.5	6.4	
	Yb	1.39	28			
	Eu	1.54††	12			

† Hendrickson *et al.* (1990). ‡ Dauter & Dauter (2007). ¶ Evans & Bricogne (2002). †† Suitable wavelength if the absorption edge is inaccessible. § Ennifar *et al.* (2002). ‡‡ 5-Amino-2,4,6-triiodoisophthalic acid. §§ Beck *et al.* (2008). ¶¶ 10 P atoms in 290 atoms. ††† Blundell & Johnson (1976). ‡‡‡ *p*-Chloromercuriphenylsulfonic acid.

remains a tedious and random stage, as the classical compounds bind only to specific motifs or are deleterious to the crystal quality. Therefore, the development of reliable and straightforward methods for obtaining heavy-atom derivatives remains an important challenge.

The anomalous phasing power of a derivative is proportional to the imaginary part of the atomic scattering factor f'' of the HA and to the occupancy of its binding site. Lanthanide (Ln) atoms provide high anomalous scattering, owing to a white line, with $f'' \simeq 28$ e^-, in their L_{III} absorption edge, which is accessible using synchrotron radiation for certain lanthanides (for Gd the L_{III} wavelength is 1.7 Å). Several Lns also provide high anomalous signal at a home source (for Gd, f'' for Cu $K\alpha$ is 12 e^-). Nevertheless, the use of Ln salts is limited to proteins with calcium-binding sites owing to otherwise insufficient binding. In order to overcome this limited binding spectrum, we propose the use of a set of seven caged gadolinium complexes composed of multidentate ligand molecules, each chelating one Gd^{3+} ion, containing different functional groups to provide a panel of possible interactions with the protein of interest (Fig. 1). Previously, several of these complexes have been used for experimental phasing using a number of test proteins (Girard, Stelter, Anelli *et al.*, 2003; Girard *et al.*, 2002; Girard, Stelter, Vicat *et al.*, 2003). Furthermore, to our knowledge, the use of the complexes has allowed the determination of at least six new structures (Jeudy *et al.*, 2005; Hermoso *et al.*, 2005; Gras *et al.*, 2007; Márquez *et al.*, 2006; Molina, González *et al.*, 2009; Pérez-Dorado *et al.*, 2010). Their compatibility with lipidic mesophases as a crystallization medium has been ascertained (Girard *et al.*, 2004). Very recently, the Gd-HPDO3A complex, also called gadoteridol, was used in the first reported *de novo* phasing by serial femtosecond crystallography (Barends *et al.*, 2014). In the particular cases of lysozyme and *Streptococcus pneumoniae* choline-binding protein F (CbpF), the binding modes of several complexes have been determined (Girard *et al.*, 2002; Molina, Stelter *et al.*, 2009). In these cases, binding mainly occurred *via* hydrophobic CH–π interactions between the cyclen macrocycle of the complex and aromatic tryptophan side chains present in the active site of lysozyme and in the choline-binding motifs (CBMs) of CbpF. On the other hand, the complexes displayed a very different binding behaviour with urate oxidase (Girard, Stelter, Anelli *et al.*, 2003), indicating that binding of the complexes occurs *via* different modes depending on the protein.

In this study, we have carried out, for the first time, a statistically relevant analysis of complex binding and phasing power for a large number of derivatives obtained with various proteins. Using our range of seven test proteins, phasing was successful for more than 50% of the derivatives (25 out of 44) and for at least two of the complexes for each of the tested proteins. We analysed the structures of the different complexes bound to the proteins, which revealed their distinct, complementary binding modes, explaining their unequal binding behaviour with different proteins. Our results highlight the advantage of accurate experimental electron-density

maps for straightforward structure determination and for the direct identification of bound ligands.

The high success rate of this set of complexes is owing to their ease of use, their versatile binding properties and their high phasing power, making them very suitable as a heavy-atom screen, in addition to more traditional heavy-atom compounds, for solving new, challenging structures. They also constitute a powerful tool to develop cutting-edge crystallographic methods, such as the emerging serial femtosecond crystallography on microcrystals. Finally, the complexes are now commercially available.

2. Methods

2.1. Gd complexes and proteins

The complexes were obtained as described in Girard, Stelter, Anelli *et al.* (2003). Gd-HPDO3A, Gd-DO3A, Gd-DOTMA and Gd-DOTA-BOM were kindly provided by Bracco Imaging SpA, Milan, Italy and solutions of Gd-DOTA, Gd-DTPA and Gd-DTPA-BMA by Professor J.-F. Le Bas, CHU-Hôpital Nord, Grenoble, France. The complexes are very stable, nontoxic and have a high solubility in water of above 1 *M*, except for Gd-DOTA-BOM, which has a solubility of 0.6 *M*. They are about 8.5–10 Å in diameter. For thermodynamic characterizations, see Bianchi *et al.* (2000) and Aime *et al.* (1996). Originally used for medical imaging, the Gd complexes generally display a comparatively weak affinity for biological molecules.

Europium and ytterbium complexes of DOTA, DO3A, HPDO3A and DTPA-BMA, as well as Gd-HPDO3A, are now available as part of a Lanthanide Phasing Kit distributed by NatX-ray (http://www.natx-ray.com). Complexes with other Lns show the same binding behaviour with proteins as the corresponding Gd complexes (Girard, Anelli *et al.*, 2003). Europium has very similar anomalous scattering properties as gadolinium. Ytterbium complexes are very well suited for synchrotron experiments, with an L_{III} wavelength of 1.39 Å, but less so for in-house data collection, with an f'' for Cu $K\alpha$ of 5 e^-.

The test proteins were chosen to cover a large range of crystallization conditions based on different salts, PEGs or MPD and a variable pH between pH 4.5 and 9: *Aspergillus flavus* urate oxidase, *Streptomyces rubiginosus* glucose isomerase, *Thaumatococcus daniellii* thaumatin, hen egg-white lysozyme, *Escherichia coli* hypothetical protein YggV and *E. coli* YeaZ. For the purpose of comparison and completeness, we include our previous results with *S. pneumoniae* CbpF (Molina, Stelter *et al.*, 2009).

A. flavus urate oxidase was kindly provided by Sanofi Synthélabo. Hen egg-white lysozyme was purchased from Boehringer, *T. daniellii* thaumatin was purchased from Sigma, *S. rubiginosus* glucose isomerase was purchased from Hampton Research and *E. coli* YeaZ and native crystals of *E. coli* YggV were produced by the IGS laboratory. Production of *S. pneumoniae* CbpF has been described previously (Molina *et al.*, 2007).

2.2. Crystallization

All proteins were used to prepare derivative crystals with seven Gd complexes, except for glucose isomerase and YeaZ, for which six and three complexes were tested, respectively. The proteins were crystallized by the hanging-drop vapour-diffusion technique. Crystallization conditions were optimized based on the known conditions. Urate oxidase crystals were obtained by mixing 4 µl protein solution at a concentration of 11.3 mg ml^{-1} with 2 or 2.5 µl reservoir solution consisting of 75 m*M* Tris pH 8.5, 6–15% PEG 3350 (Bonneté *et al.*, 2001). Crystallization trays were prepared and stored at 281 K. Lysozyme crystals were obtained by mixing the protein at a concentration of 40 mg ml^{-1} with a reservoir solution consisting of 50 m*M* sodium acetate pH 4.5, 0.8–1.2 *M* NaCl at 293 K. Thaumatin crystals were obtained by mixing 6 µl protein solution at a concentration of 50 mg ml^{-1} with 4 µl reservoir solution consisting of 100 m*M* ADA pH 6.6, 0.6–0.9 *M* disodium potassium tartrate, 0–5% ethylene glycol at 293 K (Ko *et al.*, 1994). Glucose isomerase crystals were obtained by mixing the protein at a concentration of 30 mg ml^{-1} with a reservoir solution consisting of 50 m*M* Tris pH 7, 9–14% MPD, 50 m*M* $MgCl_2$ at 293 K (Ramagopal *et al.*, 2003*a*). Native crystals of YggV crystallized in 0.1 *M* Bicine pH 9.0, 33.3–38% ammonium sulfate, 5% glycerol at 293 K. Crystals of YeaZ were obtained by mixing 1 µl 17 mg ml^{-1}

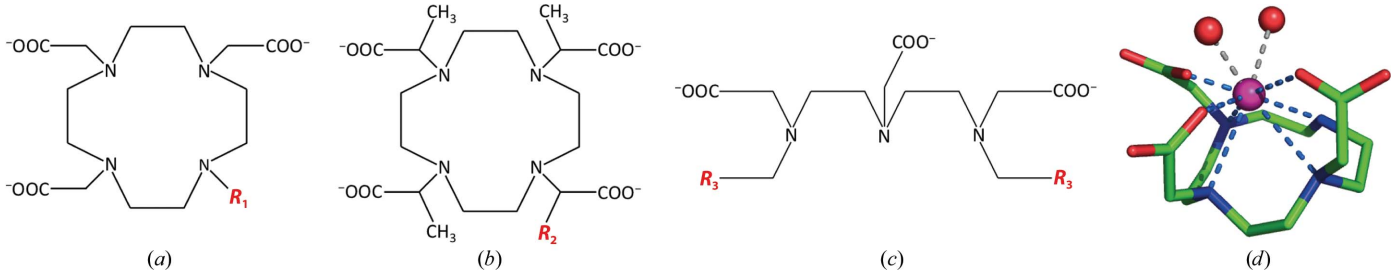


Figure 1
Ligands of the seven Gd complexes used in this study. The ligands are derivatives of the tetraazacyclododecane (cyclen) macrocycle (*a*, *b*) or linear (*c*). (*a*) R_1 = H, DO3A; R_1 = CH_2COO^-, DOTA; R_1 = $CH_2CHCH_3COO^-$, HPDO3A. (*b*) R_2 = CH_3, DOTMA; R_2 = $CH_2OCH_2C_6H_5$, DOTA-BOM. (*c*) R_3 = COO^-, DTPA; R_3 = $CONHCH_3$, DTPA-BMA. The Gd complexes of DO3A, HPDO3A and DTPA-BMA are electrically neutral, those of DOTA, DOTMA and DOTA-BOM bear one negative charge and that of DTPA bears two negative charges. (*d*) The DO3A ligand chelating the Gd^{3+} ion (magenta sphere). In solution, the enneacoordination is completed by two water molecules (red spheres). Blue and grey dashed lines indicate interactions of the ion with the ligand and water molecules, respectively.

protein solution with 0.5 µl reservoir solution consisting of 0.1 *M* sodium acetate buffer pH 4.7–5.5, 4–8% PEG 8000, 0.2 *M* NaCl, 10–20% glycerol (Jeudy *et al.*, 2005). Crystals of CbpF were grown using a protein solution consisting of 140 m*M* choline chloride, 20 m*M* Tris pH 8.0 and 3.9 mg ml^{-1} protein and a reservoir solution consisting of 0.01 *M* $NiCl_2$, 20% PEG MME 2K, 0.1 *M* Tris pH 8.5 (Molina *et al.*, 2007).

Derivative crystals were obtained by co-crystallization or by soaking the native protein crystals using the complexes at a concentration of 50–100 m*M* (Supplementary Table S1[1]). In most cases derivative crystals could be obtained by either method, and in these cases no difference in the binding behaviour of the complexes was observed. Lysozyme-derivative crystals could only be obtained by co-crystallization because the crystals did not support soaking, which is probably owing to the atypically low solvent content and small solvent channels of lysozyme crystals preventing diffusion of the complexes. For co-crystallization, the Gd-complex solution was added to the drop in order to obtain the final complex concentration. In certain cases, the complexes affected the protein solubility, for which we had to compensate by increasing or decreasing the precipitant concentration of the crystallization buffer (Supplementary Table S1). We found that whenever this was the case the complex actually bound to the protein. Thus, observed modification of protein solubility can be a useful indicator of successful binding, even if in our hands binding did not necessarily modify the crystallization conditions.

For soaking, the native crystals were transferred to a drop of mother liquor containing, in addition, the indicated concentration of the Gd complex and 1–5 mg ml^{-1} protein. Derivative crystals of YggV and YeaZ were exclusively obtained by soaking previously obtained native crystals. Typical soaking times were about 45 min to 1 h (Supplementary Table S1). However, this could vary depending on the binding affinity and the crystal size. Thus, for YeaZ it was necessary to increase the soaking time to 6 h, while for CbpF soaking times were exceptionally short, with 10 s leading to optimum binding, owing to the higher than usual affinity of the complexes for the CBMs of the protein. Similar short soaking times had also been sufficient for Gd-HPDO3A binding to the *S. pneumoniae* phosphorylcholine esterase Pce (Lagartera *et al.*, 2005).

After the desired soaking time or when the co-crystals reached the desired size, the derivative crystals were briefly (∼10 s) soaked in the cryoprotecting solution and cooled in liquid nitrogen or a gaseous nitrogen stream. Solutions for cryoprotection correspond to the crystallization solution without any Gd complex supplemented with 25–30% PEG 400 for urate oxidase and lysozyme crystals, 25% ethylene glycol for thaumatin crystals, 30% MPD for glucose isomerase crystals and 17% glycerol for CbpF crystals. The YggV and YeaZ crystallization solutions were suitable as cryoprotectants.

Complexes that were observed to bind well to lysozyme, urate oxidase and thaumatin were subsequently used at a concentration of 300 m*M* in order to increase the occupancy of the complex binding sites and to refine the structure of the bound complex (Supplementary Table S1).

2.3. Data collection and processing

For all derivatives, redundant diffraction data were collected (Supplementary Table S2) at 100 K. Initial data were collected using Cu $K\alpha$ radiation from a Rigaku RU-200 rotating-anode X-ray generator equipped with a Supper 7600 double-mirror system or a Nonius rotating-anode generator equipped with an Osmic focusing system on a MAR300 or a MAR345 imaging-plate detector, respectively. For derivatives where significant complex binding was detected, additional data were collected using synchrotron radiation on the BM30A beamline at the European Synchrotron Radiation Facility (ESRF), Grenoble, France at the f'' maximum of the Gd L_{III} absorption edge. At the time, owing to the detector size and distance, the resolution was limited to ∼2.7 Å at the wavelength corresponding to the Gd absorption edge. Additional data to the highest possible resolution were collected at a shorter wavelength for derivatives with a high anomalous signal, in anticipation of subsequent structure refinement of the bound complex molecules. These additional data were collected on the ESRF BM30A or ID29 beamlines. Diffraction data were integrated using *XDS* (Kabsch, 2010) and were scaled with *SCALA*, and the solvent content was estimated by *TRUNCATE* (Winn *et al.*, 2011). A summary of the data-collection parameters and processing statistics is given in Supplementary Table S2. The binding of the anomalous scatterers in the derivative crystals was estimated after reducing and scaling complete diffraction data sets using the corresponding statistical indicators of the data integration and scaling program (s_{norm}/s_{ano} in *XDS* and R_{ano} in *SCALA*) and *via* inspection of the anomalous difference Patterson map (an example is shown in Supplementary Fig. S1).

Although concentrations of heavy atoms in the range of those used here in protein crystals are supposed to lead to increased radiation damage (Murray *et al.*, 2004), we did not observe such an effect with our Gd complexes, even for data collected at the Gd absorption edge. We presume that the cage around the heavy atoms formed by the respective ligands might protect the protein from the effects of X-ray absorption by the Ln atom. On the high-intensity beamline ID29, we could observe some radiation damage characterized by a loss of reflections at high resolution, but not more than would be expected for native crystals.

2.4. Phasing and analysis

The initial coordinates of the bound Gd atoms were identified using *SHELXD* (Sheldrick, 2008) and *SnB* (Weeks & Miller, 1999). SAD or MAD phasing was carried out using *SHARP* (Bricogne *et al.*, 2003) and density modification was carried out using *SOLOMON* (Abrahams, 1997) and *DM* (Cowtan & Main, 1996), both of which are accessible *via* the *SHARP* interface. For each derivative, phasing was typically carried out using two data sets: one data set at high resolution

[1] Supporting information has been deposited in the IUCr electronic archive (Reference: TZ5054).

Table 2
Phasing results obtained for 44 different derivatives.

The average solvent content of the crystals is indicated for all proteins. — indicates no significant complex binding; FOM, figure of merit after density modification; $\Delta F^{\pm}/F$, theoretical anomalous signal, as in Table 1, calculated for $f'' = 28\ e^-$ and refined anomalous binding-site occupancies q_j (Table 3).

	Urate oxidase		Glucose isomerase		Thaumatin		YGGV		Lysozyme		CbpF		YeaZ	
Solvent content (%)	52		48		48		49		28		58		48	
	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)	FOM	$\Delta F^{\pm}/F$ (%)
Gd-DO3A	0.77	3.9	0.86	6.4	—	—	0.83	6.1	0.75	14.9	0.84	10.3	0.83	7.0
Gd-HPDO3A	—	—	—	—	—	—	—	—	0.91	16.0	0.86	11.2		
Gd-DTPA-BMA	0.92	10.0	—	—	0.90	6.3	—	—	—	—	0.80	3.4	—	—
Gd-DOTA	—	—	—	—	0.88	5.1	—	—	0.88	11.8	0.78	5.5		
Gd-DOTMA	0.86	8.5	—	—	—	—	0.90	7.1	—	—	0.77	5.0	0.82	5.6
Gd-DTPA	—	—	0.88	3.1	0.91	4.8	—	—	—	—	0.82	7.7		
Gd-DOTA-BOM	—	—			0.91	9.2	0.79	6.8	0.75	8.9	0.81	7.6		

Table 3
Refinement statistics for all derivatives that allowed experimental phasing.

No. of binding sites, number of complex binding sites; No. of refined sites, number of refined complex molecules.

Complex	No. of binding sites	No. of refined sites	Binding-site occupancies	R/R_{free}	Resolution (Å)
Urate oxidase					
Gd-DOTMA	3	1	1.00, 0.29, 0.21	0.20/0.21	1.35
Gd-DTPA-BMA	3	3	0.92, 0.48, 0.70	0.19/0.20	1.45
Gd-DO3A	10	0	0.26, 0.21, 0.19, 0.17, 0.14, 0.14, 0.11, 0.08, 0.07, 0.05	0.22/0.23	1.45
Lysozyme					
Gd-HPDO3A	2	2	1.00, 0.87	0.17/0.20	1.53
Gd-DOTA-BOM	1	1	0.74	0.22/0.25	1.53
Gd-DOTA	2	2	0.82, 0.53	0.19/0.19	1.54
Gd-DO3A	4	2	0.96, 0.69, 0.31, 0.18	0.23/0.25	1.54
Thaumatin					
Gd-DOTA-BOM	2	2	0.84, 0.48	0.20/0.20	1.45
Gd-DTPA-BMA	1	1	0.66	0.19/0.20	1.45
Gd-DOTA	1	1	0.54	0.20/0.20	1.45
Gd-DTPA	1	1	0.50	0.19/0.20	1.45
YGGV					
Gd-DOTMA	3	1	0.66, 0.25, 0.17	0.28/0.33	2.70
Gd-DOTA-BOM	3	0	0.47, 0.39, 0.34	0.31/0.38	2.68
Gd-DO3A	4	1	0.51, 0.25, 0.20, 0.17	0.32/0.37	2.68
YeaZ					
Gd-DOTMA	8	1	0.66, 0.42, 0.36, 0.20, 0.17, 0.17	0.22/0.25	2.3
Gd-DO3A	6	2	0.74, 0.67, 0.33, 0.30, 0.29, 0.18	0.23/0.29	2.7
Glucose isomerase					
Gd-DTPA	1	1	0.45	0.19/0.20	1.45
Gd-DO3A	6	3	0.56, 0.56, 0.37, 0.24, 0.15, 0.10	0.21/0.21	1.44
CbpF					
Gd-HPDO3A	5	4	0.87, 0.72, 0.66, 0.60, 0.13	0.21/0.24	2.17
Gd-DO3A	6	3	0.83, 0.66, 0.63, 0.30, 0.29, 0.28	0.22/0.25	2.17
Gd-DOTA-BOM	4	0	0.95, 0.17, 0.10, 0.12	0.23/0.24	1.40
Gd-DTPA	1	1	0.99	0.26/0.30	2.03
Gd-DOTMA	1	1	0.65	0.21/0.25	2.17
Gd-DTPA-BMA	2	0	0.34, 0.28	0.24/0.28	2.98
Gd-DOTA	2	0	0.53, 0.47	0.23/0.24	2.17

with a relatively weak anomalous signal ($f'' \simeq 6\ e^-$) and one data set collected at maximum f'' limited in resolution to 2.7 Å. The density-modification step allowed the extension of the experimental phases to high resolution. A summary of the phasing statistics is given in Supplementary Table S3.

As a first criterion, the outcome of the phasing can be estimated based on phasing statistics, including the figure of merit before and after density modification. Eventually, we only considered as successful the derivatives which led to easily interpretable experimental electron-density maps that permitted straightforward model building. Because map quality is not only influenced by the accuracy of the phases, once the refined model is available a more meaningful measure of the phasing quality is the complex correlation coefficient between the observed and the calculated structure factor as a function of resolution. Phasing results are summarized in Table 2.

2.5. Structure refinement

For all 24 derivatives that allowed determination of the complex binding sites, the structure of the protein and the positions, occupancies and *B* factors of the Gd atoms were refined. Refinement statistics are given in Table 3. All derivative structures were mostly isomorphous with the native structures and did not require major adjustment of the protein model. Refinement of the ligand moieties of the bound complexes was not always possible. Indeed, binding-site occupancies of about 0.2–0.5 can provide sufficient anomalous signal for phasing, but the electron density of the ligand is generally too poor to identify and refine its orientation. The models of the ligand molecules could be refined for 21 derivatives in total (Table 3).

The models of the proteins and bound gadolinium complexes were refined using *CNS* (Brünger *et al.*, 1998) against the data sets with the highest Gd-site occupancies and resolution. For derivatives where phases were obtained by

MAD, experimental phase information was used during refinement in the form of Hendrickson–Lattman coefficients using the MLHL mode. For single-wavelength data F^+ and F^- were used in MLF mode. The PDB codes of the initial atomic protein models used for refinement are 193l for lysozyme (Vaney *et al.*, 1996), 1ws3 for urate oxidase (Retailleau *et al.*, 2005), 1thw for thaumatin (Ko *et al.*, 1994), 1xib for glucose isomerase (Carrell *et al.*, 1994) and 1k7k for YggV (Savchenko *et al.*, 2007), and the models corresponding to the solved structures of YeaZ and CbpF are PDB entries 1okj (Jeudy *et al.*, 2005) and 2v04 (Molina, González *et al.*, 2009), respectively. The progress of refinement was followed using the $R_{\rm free}$ based on an excluded 5% of reflections. All water molecules were removed from the initial models. Initial values for Gd-atom coordinates and occupancies were obtained from *SHARP* phasing. Rigid-body refinement, simulated annealing and refinement of *B* factors and of Gd-site occupancies were followed by automatic introduction of water molecules. The visualization of electron-density maps and manual adjustment of the models was performed in *O* (Jones *et al.*, 1991). Experimental $2F_{\rm o} - F_{\rm c}$ and $F_{\rm o} - F_{\rm c}$ electron-density maps were used to model small molecules and ligands of the Gd complexes. In the cases of Gd-DTPA-BMA (Ehnebom & Pedersen, 1992), Gd-HPDO3A (Kumar *et al.*, 1994), Gd-DOTA (Dubost *et al.*, 1991) and Gd-DO3A (Chang *et al.*, 1993), their known structures were used as initial models. Gd-DOTMA was modelled based on the Gd-DOTA structure, adding methyl groups, and for Gd-DOTA-BOM the Gd-DOTMA model was used, the additional (phenylmethoxy) methyl group of the DOTA-BOM ligand being absent from the electron-density maps. The CbpF Gd-DTPA derivative allowed *de novo* construction of the model of Gd-DTPA. The models of the complexes were adjusted manually using the experimental and the $F_{\rm o} - F_{\rm c}$ electron-density maps and were included in refinement.

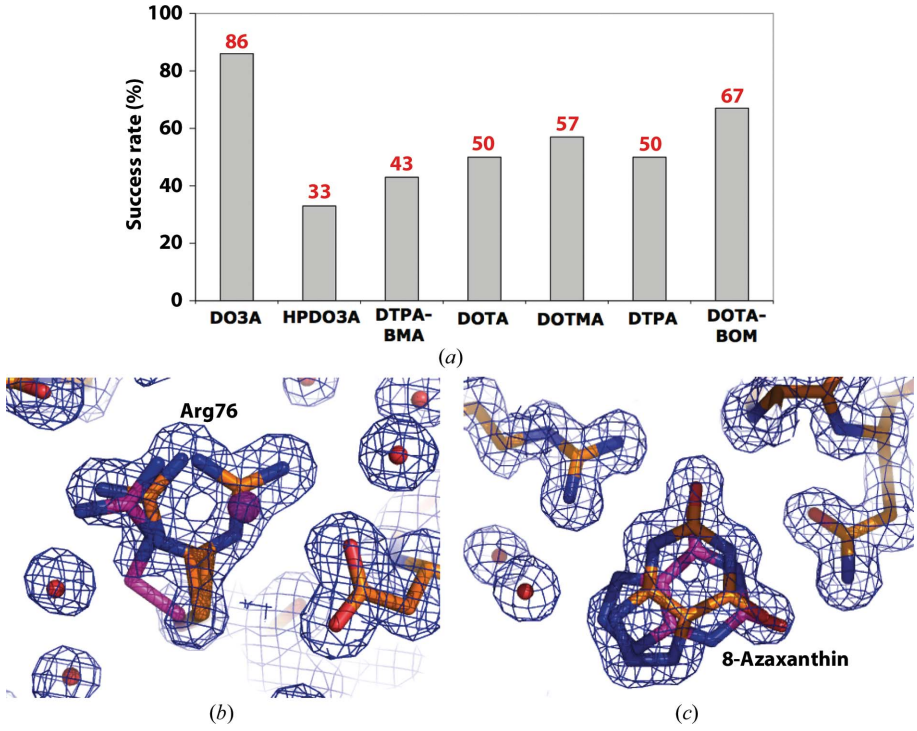


Figure 2
Phasing results. (*a*) Bar graph representing the phasing success rate as a percentage for each complex based on the derivatives described in Table 2. (*b*, *c*) Experimental electron-density map after density modification (contoured at 1σ), our refined model (yellow sticks) and the original incorrect model (magenta sticks). (*b*) Glucose isomerase Gd-DTPA derivative. The map clearly shows the alternate conformations of Arg76. The side chain and water molecule in the original model (PDB entry 1xib) are shown in magenta. (*c*) Urate oxidase Gd-DTPA-BMA derivative. The inhibitor 8-azaxanthine in the refined model and the original orientation (PDB entry 1uox).

3. Results and discussion

3.1. Crystallization, data collection and phasing

Seven proteins, urate oxidase, lysozyme, thaumatin, glucose isomerase, YggV, YeaZ and CbpF, have been used to prepare derivative crystals with the seven different Gd complexes which were then used for experimental phasing. Among these, the structures of YeaZ and CbpF were effectively solved using the complexes (Jeudy *et al.*, 2005; Molina, González *et al.*, 2009). Generally, the derivatives were highly isomorphous to the native crystals and the initial diffraction quality was preserved.

Experimental phasing was carried out for all derivatives that allowed determiniation of the complex-binding sites (Supplementary Table S3). For about 50% of the tested derivatives, binding-site occupancies of 0.3–1.0 allowed the straightforward calculation of accurate experimental phases.

Table 2 summarizes the phasing results obtained with seven systematically tested proteins, represented by the figures of merit after density modification. In order to compare the different derivatives among them and with other derivatization methods (Table 1), we calculated the theoretical anomalous signal, $\Delta F^{\pm}/F$, provided by each derivative (Table 2), where

$$\frac{\Delta F^{\pm}}{F}\ (\%) = \frac{100}{Z_{\rm eff}}\left(\frac{2\sum q_2^j f_j''^2}{N_{\rm p}}\right)^{1/2} = 0.76 N_{\rm H}^{1/2} q_j f_j'' \qquad (1)$$

(Hendrickson & Ogata, 1997) assuming equal anomalous binding-site occupancies q_j; the average atomic scattering factor of the native protein $Z_{\rm eff}$ is 6.6 e$^-$ for $\theta = 0$ and $N_{\rm p}$ is the number of protein atoms. Derivatives that allowed phasing have minimum and maximum anomalous signals of 3.1 and

16% with data collected at the f'' maximum, corresponding to 1.3 and 6.9% using Cu $K\alpha$ radiation. By comparison, a selenomethionine and a mercury derivative containing three methionines and one fully occupied Hg site per 150 residues, using data collected at the absorption edges, provide anomalous signals of 4.1 and 6.2%, respectively.

Overall, more than 50% (25 out of 44) of the tested derivatives allowed successful phasing, as indicated by well defined experimental electron-density maps, and for all of the proteins at least two complexes provided suitable derivatives.

Fig. 2(*a*) represents the success rate for each complex, which approximates to 50% for most complexes. Gd-DO3A proved to be the most versatile, by binding to all of the proteins except thaumatin, while Gd-HPDO3A only bound to lysozyme and CbpF but provided the best derivatives in these cases.

The accurate experimental electron-density maps enable straightforward bias-free structure determination and model building. Notably, the alternate conformation of Arg76 and two MPD molecules, which are very well defined in the experimental map of our glucose isomerase derivative, are absent in the original model, where they were modelled as water molecules (Fig. 2*b*). Urate oxidase was crystallized in the presence of its inhibitor 8-azaxanthin. Our experimental electron-density map directly pointed out the wrong orientation of the inhibitor in the initial model (Fig. 2*c*). These examples demonstrate how accurate experimental electron-density maps allow direct, bias-free identification and modelling of bound ligands, which can be crucial in drug development.

3.2. Structure refinement and binding-mode analysis

Generally, the extent to which each complex bound was different from one protein to the next. The numbers and occupancies of Gd-complex binding sites varied, with multiple (four to ten) binding sites of varying occupancies in the case of Gd-DO3A, three to eight sites for Gd-DOTMA and between one and three sites in the cases of the other complexes (Table 3). Binding sites are situated on the protein surface or in solvent-accessible cavities. They can either vary from one complex to the next, or complexes can share the same binding sites, as is the case for the major sites in certain derivatives of lysozyme, thaumatin or CbpF. The refined structures of the bound complexes in 21 different derivatives gave insight into their distinct binding modes (Table 4), which will be detailed below, and explained the unequal binding behaviour with different proteins.

In all of the complexes, the ligand arms carrying the functional groups bend towards the Gd^{3+} ion, forming a cage (Figs. 1*d*, 3, 4, 5, 6 and 7). The cyclen macrocycle forms the hydrophobic back of the circular complexes, and the linear ligands form a similar structure by folding onto themselves (Fig. 6*b*). The circular back side can establish a CH–π hydrophobic interaction with aromatic residues of the protein surface. Aromatic residues may also coordinate the sides of certain complexes, forming an approximate plane (Figs. 5 and 7*b*). The N atoms of the back side and the O atoms of the ligand arms, which are mainly negatively charged carboxylate O atoms, coordinate the ion (Fig. 1). Thus, eight (or seven in the case of DO3A) ligand atoms occupy the first coordination sphere. In solution, the enneacoordination of the Gd^{3+} ion is completed by one water molecule (two in the case of Gd-DO3A; Fig. 1*b*). Contributing to complex binding, the Gd^{3+} ion can interact with the electronegative carboxylate or carboxamide O atom of Asn, Gln, Asp or Glu, or with two carboxylic O atoms in the case of the Gd-DO3A complex, replacing the water molecules (GC; gadolinium coordination). The Gd^{3+} ion can also coordinate a water molecule, which in turn is coordinated by protein residues (iGC; indirect gadolinium coordination). The outward-facing O atoms of the ligand arms can accept hydrogen bonds (HB) or interact electrostatically with positively charged protein residues, and

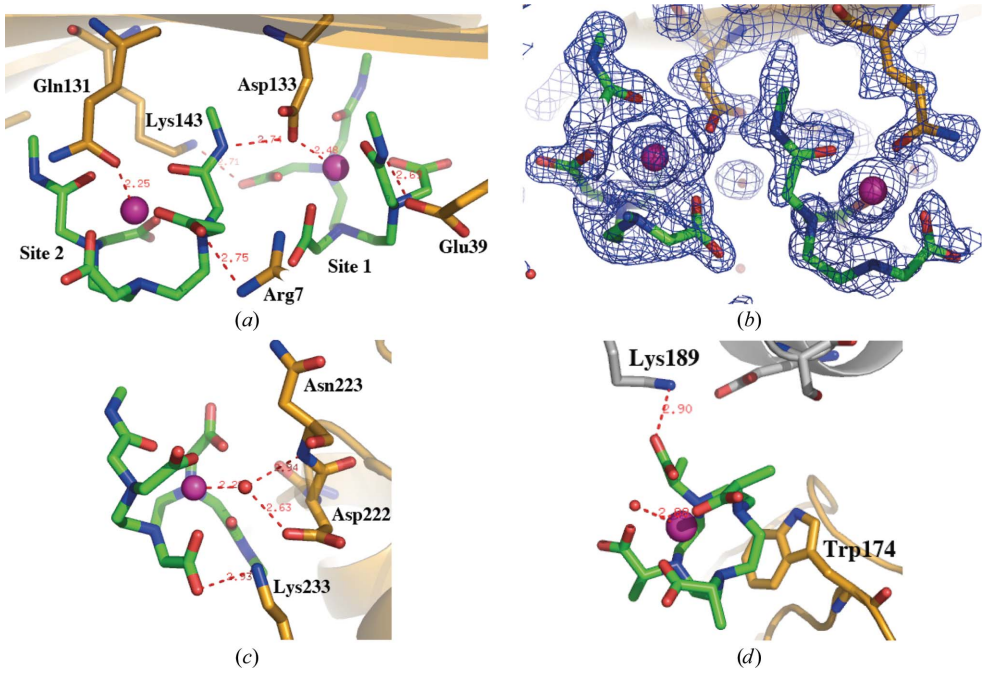


Figure 3
Refined models of Gd complexes bound to urate oxidase. The cartoon representation of proteins is shown in gold, or in grey for symmetry-related molecules. Only side chains of protein residues involved in complex binding are shown. The ligands are shown as green sticks and the Gd^{3+} ion is shown as a magenta sphere. (*a*, *b*, *c*) Binding of Gd-DTPA-BMA to urate oxidase. (*a*, *b*) Sites 1 and 2. (*b*) Refined $2F_o - F_c$ electron density of protein and complex molecules at 1.45 Å resolution. (*c*) Binding site 3. (*d*) Major binding site of Gd-DOTMA.

Table 4
Observed types of interaction leading to complex binding.

Numbers in parentheses indicate the overall charge of the complex. Abbreviations indicate the kind of observed interaction between the complex molecule and the protein. GC, gadolinium coordination: interaction of the Gd^{3+} ion with the carboxylate or carboxamide O atom of Asn, Gln, Asp or Glu. iGC, indirect gadolinium coordination: the Gd^{3+} ion coordinates a water molecule, which in turn is coordinated by protein residues. HB, hydrogen bonds or electrostatic interaction between the ligand and protein residues. W, hydrogen-bond network between the ligand and protein surface involving water molecules of the hydration shell of the protein. CO, cooperative binding of two neighbouring complex molecules.

Gd-HPDO3A (0)	
CbpF	CH–π, iGC, HB, W
Lysozyme	CH–π, W, CO
Gd-DO3A (0)	
CbpF	GC, CH–π ,W
YeaZ	GC, CH–π, HB
Lysozyme	GC, CH–π, W, CO
Glucose isomerase	GC, HB, W
YGGV	GC
Gd-DOTA (−1)	
Lysozyme	CH–π, GC, HB, CO
Thaumatin	CH–π, HB
CbpF	CH–π
Gd-DTPA-BMA (0)	
Thaumatin	CH–π
Urate oxidase	GC, HB, W
Gd-DOTA-BOM (−1)	
Lysozyme	CH–π, HB, W
Thaumatin	CH–π
Gd-DTPA (−2)	
CbpF	GC, HB, W
Glucose isomerase	iGC, HB, W
Thaumatin	CH–π, HB, W
Gd-DOTMA (−1)	
CbpF	W
Urate oxidase	CH–π, HB, W
YeaZ	GC, CH–π, HB, W
YGGV	W

the DTPA-BMA amide N atoms can participate in hydrogen bonds. The hydrogen-bond network between the ligand functional groups and protein surface generally involves water molecules of the hydration shell of the protein (W). In several cases, interaction between two neighbouring complex molecules contributes to their cooperative binding (CO; Figs. 3 and 7). Table 4 summarizes the types of interaction involved in complex binding for the different derivatives. In order to illustrate the observed binding modes, examples are described for all of the complexes in more detail.

Of the three well phased urate oxidase derivatives (Table 2), the Gd-DTPA-BMA and Gd-DOTMA derivatives allowed refinement of the complex binding modes (Table 3). Gd-DO3A binds at multiple sites with relatively low occupancies, providing only poor electron density for the ligand molecules. In this case, all of the complexes bound to different binding sites. Binding sites 1 and 2 of the electrically neutral Gd-DTPA-BMA are adjacent and are situated in the tunnel formed by the tetramer of the protein. The molecules bind *via* the coordination of the Gd ion by O atoms of the protein surface (GC; Fig. 3*a*) and *via* hydrogen bonds between ligand and protein atoms (HB) and water molecules, with six and four refined water molecules bound to each ligand molecule respectively (not shown) (W). The $2F_o - F_c$ electron density of the refined complex models is very well defined (Fig. 3*b*). The third binding site is located in a cavity formed by two symmetry-related protein molecules, and the ionic side of the complex is turned towards the protein surface (Fig. 3*c*). In contrast to sites 1 and 2, the Gd^{3+} ion is coordinated by a water molecule (interaction type iGC), which forms two hydrogen bonds to the protein surface. Binding also involves interaction

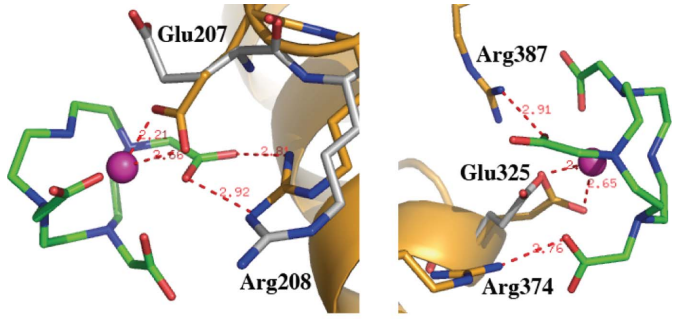
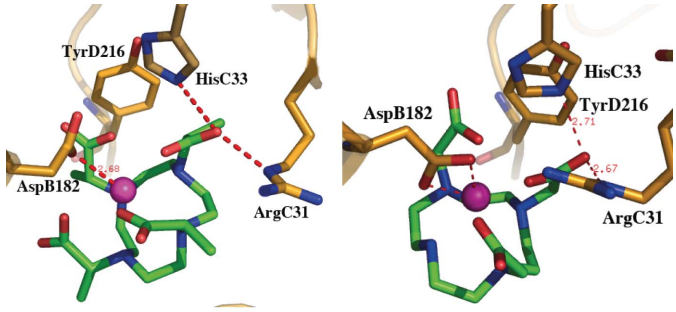
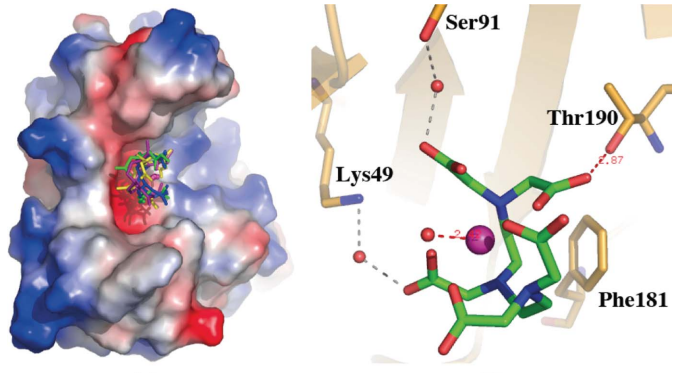


Figure 4 (*a*) (*b*)
Refined models of Gd-DO3A bound to glucose isomerase at site 1 (*a*) and site 2 (*b*). Alternate conformations of protein residues involved in complex binding were refined, corresponding to their conformations in occupied (yellow) and empty (grey) complex-binding sites.



(*a*) (*b*)

Figure 5
Major shared binding site of Gd-DOTMA (*a*) and of Gd-DO3A (*b*) in YeaZ. Residues from three protein chains line the binding cavity.



(*a*) (*b*)

Figure 6
Binding of four Gd complexes to thaumatin at the common binding site. (*a*) Electrostatic surface representation of the protein monomer showing the binding cavity, with the refined models of the complexes shown as sticks. Red, Gd-DTPA-BMA; blue, Gd-DTPA; green, Gd-DOTA; yellow, Gd-DOTA-BOM. The (phenylmethoxy)methyl group of Gd-DOTA-BOM is not defined in the electron-density maps and therefore is not included in the model. (*b*) Detailed binding mode of Gd-DTPA.

of the ligand with a surface lysine and an extensive network of ordered water molecules. The almost fully occupied major Gd-DOTMA binding site is situated in a shallow cavity formed at the interface of two symmetry-related protein molecules (Fig. 3*d*). The complex binds *via* a CH–π interaction between the complex macrocycle and an aromatic tryptophan side chain. Interaction with a lysine and several water molecules that are coordinated by the O atoms of the ligand and the protein surface stabilize the binding (not shown). The Gd^{3+} ion is turned towards the solvent.

In glucose isomerase, Gd-DTPA binds to a unique half-occupied site *via* the coordination of a water molecule between the Gd^{3+} ion and the protein surface and further hydrogen bonds between the ligand and water molecules (not shown). Gd-DO3A binds to six sites. In all sites a glutamate or an aspartate is close to the ion as well as one or several arginine residues, except at site 5. The refined structure of the bound complex at three sites showed that the binding mode at these sites is very similar. The complex binds through the coordination of the Gd^{3+} ion by both carboxylate O atoms of a glutamate residue situated on the protein surface (Fig. 4). Hydrogen bonds between the carboxylate O atoms of the ligand arms and the guanidinium N atoms of neighbouring arginine residues (site 1 and 2) or tryptophan NE1 (site 3; not shown) stabilize complex binding. The similarity of the protein residues constituting binding sites 4 and 6 suggest a similar binding mode as in the three major sites.

In YggV, Gd-DOTMA binds *via* hydrogen bonding involving a water network, while Gd-DO3A binds to multiple, less occupied binding sites through direct coordination of the Gd ion. In the YggV Gd-DOTA-BOM derivative all of the binding sites have solvent-exposed aromatic residues in proximity. Interaction between these residues and the phenyl ring of the ligand possibly contributes to the binding of the complex, although no electron density corresponding to the phenyl group of the ligand is visible in any of the derivatives.

We previously solved the structure of *E. coli* YeaZ using a Gd-DOTMA derivative (Jeudy *et al.*, 2005). The protein crystallizes with four molecules of 251 residues each per asymmetric unit. The Gd-DOTMA and Gd-DO3A complexes have their two major binding sites in common. The main site is located in a cavity lined by residues from three protein chains (Fig. 5). Binding of the complexes involves coordination of the Gd ion by one and two carboxylic O atoms of an aspartate for Gd-DOTMA and DO3A, respectively. Furthermore, in both complexes one ligand carboxylate O atom interacts with histidine NE2 and arginine NE, and the side of the complex possibly interacts through CH–π interaction with the adjacent parallel tyrosine side chain.

In thaumatin, the Gd-DOTA-BOM, Gd-DTPA-BMA, Gd-DOTA and Gd-DTPA complexes share their major binding site, consisting of a cavity of about 14 Å in diameter and 11 Å in depth (Fig. 6*a*). All complexes bind through CH–π interactions between their cyclic back side and the parallel aromatic side chain of Phe181. Their binding involves a network of ordered water molecules between the ligand and the protein surface, and additional hydrogen bonds for Gd-DTPA (Fig. 6*b*) and Gd-DTPA-BMA. In the absence of distinct interactions through the ligand arms, the electron density of the arms is poorly defined, despite relatively high binding-site occupancies, which is probably owing to a rotational variation of the bound complexes.

In lysozyme, three of the complexes bind to two adjacent sites (Table 4) situated in a cavity formed by three protein molecules that is destined for binding the oligosaccharide substrate (Fig. 7). Complex binding involves CH–π interactions of the cyclic back side, or the side formed by two ligand arms, with side chains of tryptophan residues lining the cavity. The larger Gd-DOTA-BOM complex only binds to one site. The small size of the protein and the low solvent content of the crystals (Table 2) possibly limit the number of alternative binding sites. In all cases, binding of the adjacent molecules is cooperative, involving the coordination of one Gd ion by a carboxylate O atom of the neighbouring molecule or a network of symmetric direct and water-mediated interactions between the complex molecules (Figs. 7*a* and 7*b*). In addition, aspartate residues coordinate the Gd ions of the Gd-DO3A molecules (Fig. 7*a*), and ND2 of Asn103 hydrogen bonds to a ligand carboxylate O atom of Gd-DOTA and Gd-DOTA-BOM. Our accurate electron density of the Gd-HPDO3A derivative allows the three carboxylated ligand arms to be distinguished from that carrying the hydroxypropyl group,

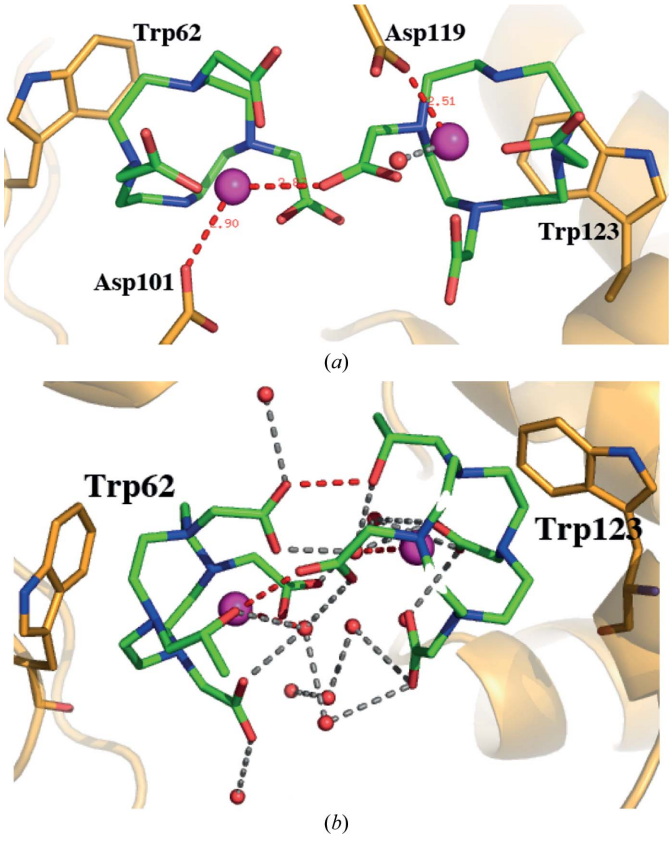


Figure 7
Binding of Gd-DO3A (*a*) and Gd-HPDO3A (*b*) to lysozyme at binding site 1 (left) and binding site 2 (right). Dashed red and grey lines indicate the interactions of the complexes with the protein or the adjacent complex molecule and with water molecules, respectively.

Table 5
Derivatives where complexes were observed to bind *via* CH–π interaction (indicated with a +).

	Thaumatin	Lysozyme	CbpF	Urate oxidase	YeaZ
Gd-HPDO3A		+	+		
Gd-DO3A		+	+		+
Gd-DTPA	+				
Gd-DOTMA				+	+
Gd-DOTA	+	+	+		
Gd-DTPA-BMA	+		+		
Gd-DOTA-BOM	+	+	+		

which allowed us to correct the previously described orientation (Girard *et al.*, 2002).

We have previously solved the structure of CbpF using a Gd-HPDO3A derivative (Molina, González *et al.*, 2009) and analysed the binding of five Gd complexes (Molina, Stelter *et al.*, 2009). Briefly, the C-terminal domain of CbpF is composed of repeated 20-amino-acid CBMs, forming cavities lined by the perpendicularly arranged aromatic rings of two or more tryptophan residues and one tyrosine residue. Three of the complexes bind through hydrophobic CH–π interactions between the complex macrocycle and the side chain of one of the tryptophan residues shaping the choline-binding cavity.

In summary, several complexes were found to bind preferably to specific motifs. Thus, Gd-HPDO3A, Gd-DOTA-BOM and Gd-DOTA were exclusively observed to bind through CH–π interactions with aromatic protein residues. The use of the most versatile Gd-DO3A often provides derivatives with multiple, partly occupied binding sites. Its binding generally relies on the coordination of the Gd^{3+} ion by both carboxyl O atoms of an aspartate or a glutamate, and is stabilized by further optional interactions. Alternatively, binding of Gd-DO3A *via* CH–π interactions between the ligand macrocycle and the aromatic cycle of tryptophan residues has been observed. The complexes Gd-DTPA-BMA, Gd-DTPA and Gd-DOTMA bind to proteins *via* varying modes.

4. Conclusions

The binding of seven Gd complexes in crystals of seven proteins has been investigated. The preparation of the derivatives by soaking or co-crystallization is rapid and is nondisruptive to diffraction quality. The obtained derivative crystals can be used for diffraction experiments using synchrotron or in-house Cu $K\alpha$ radiation. The in-house data may serve to screen for complex binding or, in the case of sufficiently well diffracting crystals, to determine the phases and solve the structure, as performed for *S. pneumoniae* LytC (Pérez-Dorado *et al.*, 2010).

The complexes proved to be highly suitable for derivatization by binding sufficiently for successful phasing in about 50% of the overall cases, and for each tested protein at least two of the complexes allowed *de novo* phasing. The derivatives are mostly isomorphous with the native structures, apart from occasional side-chain movements owing to complex binding. The complexes bind to the protein surface or accessible cavities within the protein or formed by crystal stacking.

Different complexes bind *via* different binding modes (Table 4) resulting from a combination of polar, electrostatic and/or hydrophobic CH–π interactions between the Gd ion, ligand atoms and residues of the protein surface or the hydration shell of the protein. Shape and charge complementarity contribute to binding, as the complexes are observed to match their orientation to the surface shape and negatively charged complexes bind to positively charged surface regions more often than electrically neutral molecules. Thus, binding depends in a complex manner on the chemical structure of the complex, including its functional groups, charge and size, and on the surface charge, topology and exposed residues of the protein.

The observed binding modes explain the occasional influence of the complexes on the protein solubility. Their binding to hydrophobic residues *via* CH–π interaction may have a solubilizing effect. Conversely, binding of the ionic side may cover charged surface residues and expose the hydrophobic complex back side, thus decreasing the protein solubility.

Depending on the protein, different complexes were observed to bind to different sites or, in other cases, to share binding sites and major types of interaction but with varying complex orientations and binding-site occupancies owing to differing additional interactions.

Generally, predicting complex-binding behaviour is not straightforward. Complexes that are different in overall charge, size and functional ligand groups may nevertheless share common binding sites, as in the case of Gd-DO3A and Gd-DOTMA binding to YeaZ. On the other hand, stereochemically similar complexes can show entirely different binding behaviour, as is the case for Gd-DOTMA and Gd-DOTA-BOM, which differ only by the additional phenylmethoxy group, and Gd-HPDO3A and Gd-DOTA, which differ by one methyl group. The intricate nature of complex binding is particular striking for sites of CH–π interaction, where, depending on the protein, entirely different sets of complexes bind (Table 5). This suggests that the type of aromatic residue, its environment and possibly the chemical conditions differently affect the tendency of each complex to bind *via* CH–π interactions.

We suggest that a promising approach for identifying a suitable heavy-atom derivative for a new protein is to screen three or more of the Ln complexes with observed complementary binding modes; for example, in a first round, Ln-DO3A, Ln-HPDO3A and Ln-DTPA-BMA. In our case, using this trio only would have provided good derivatives for all seven test proteins (Table 2). In practice, screening implies preparing derivative crystals *via* soaking or co-crystallization, ideally in parallel with several complexes, the collection of at least complete diffraction data using Cu $K\alpha$ or synchrotron radiation, data integration, inspection of anomalous Patterson maps and possibly the detection of heavy-atom binding sites. For proteins known to contain surface-accessible tryptophan residues, Ln-HPDO3A should preferably be tested. For proteins that bind choline, and other proteins that bind their ligands through CH–π interactions, Ln-HPDO3A and Ln-DO3A are likely to bind.

Table 1 shows a comparison of the anomalous scattering properties and binding modes of conventional scatterers and the described Ln complexes. It illustrates that by binding through versatile binding modes and providing high anomalous scattering, the complexes constitute a useful complement to the classical heavy-atom screening compounds in the crystallographer's toolbox.

These complexes can be used for the develoment of new methods such as serial femtosecond crystallography and, when included in a screen, in combination with SAD phasing, for high-throughput protein structure determination.

We thank the European Synchrotron Radiation Facility beamline staff for their support. We thank M. Martínez-Ripoll for discussion and helpful comments. We thank Gordon Leonard for advice. This work was supported by the Spanish Ministry of Economy and Competitiveness (JCI-2011-09308 to RM and BFU2011-25326 to JAH) and the Government of the Community of Madrid (S2010/BMD-2457 to JAH).

References

Abrahams, J. P. (1997). *Acta Cryst.* D**53**, 371–376.
Aime, S., Botta, M., Fasano, M., Crich, S. G. & Terreno, E. (1996). *J. Biol. Inorg. Chem.* **1**, 312–319.
Barends, T. R., Foucar, L., Botha, S., Doak, R. B., Shoeman, R. L., Nass, K., Koglin, J. E., Williams, G. J., Boutet, S., Messerschmidt, M. & Schlichting, I. (2014). *Nature (London)*, **505**, 244–247.
Beck, T., Krasauskas, A., Gruene, T. & Sheldrick, G. M. (2008). *Acta Cryst.* D**64**, 1179–1182.
Bianchi, A., Calabi, L., Corana, F., Fontana, S., Losi, P., Maiocchi, A., Paleari, L. & Valtancoli, B. (2000). *Coord. Chem. Rev.* **204**, 309–393.
Blundell, T. & Johnson, L. (1976). *Protein Crystallography*. New York: Academic Press.
Bonneté, F., Vivarès, D., Robert, C. & Colloc'h, N. (2001). *J. Cryst. Growth*, **232**, 330–339.
Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* D**59**, 2023–2030.
Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Carrell, H. L., Hoier, H. & Glusker, J. P. (1994). *Acta Cryst.* D**50**, 113–123.
Chang, C. A., Francesconi, L. C., Malley, M. F., Kumar, K., Gougoutas, J. Z., Tweedle, M. F., Lee, D. W. & Wilson, L. J. (1993). *Inorg. Chem.* **32**, 3501–3508.
Cowtan, K. D. & Main, P. (1996). *Acta Cryst.* D**52**, 43–48.
Dauter, M. & Dauter, Z. (2007). *Methods Mol. Biol.* **364**, 149–158.
Dauter, Z. & Nagem, R. A. P. (2002). *Z. Kristallogr.* **217**, 694–702.
Doublié, S. (2007). *Methods Mol. Biol.* **363**, 91–108.
Dubost, J.-P., Leger, J.-M., Langlois, M.-H., Meyer, D. & Schaefer, M. (1991). *C. R. Acad. Sci. Paris*, **312**, 349–354.
Ehnebom, L. & Pedersen, B. (1992). *Acta Chem. Scand.* **46**, 126–130.
Ennifar, E., Carpentier, P., Ferrer, J.-L., Walter, P. & Dumas, P. (2002). *Acta Cryst.* D**58**, 1262–1268.
Evans, G. & Bricogne, G. (2002). *Acta Cryst.* D**58**, 976–991.
Garman, E. & Murray, J. W. (2003). *Acta Cryst.* D**59**, 1903–1913.
Girard, É., Anelli, P. L., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* D**59**, 1877–1880.
Girard, É., Chantalat, L., Vicat, J. & Kahn, R. (2002). *Acta Cryst.* D**58**, 1–9.
Girard, É., Pebay-Peyroula, E., Vicat, J. & Kahn, R. (2004). *Acta Cryst.* D**60**, 1506–1508.
Girard, É., Stelter, M., Anelli, P. L., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* D**59**, 118–126.
Girard, É., Stelter, M., Vicat, J. & Kahn, R. (2003). *Acta Cryst.* D**59**, 1914–1922.
Gras, S., Chaumont, V., Fernandez, B., Carpentier, P., Charrier-Savournin, F., Schmitt, S., Pineau, C., Flament, D., Hecker, A., Forterre, P., Armengaud, J. & Housset, D. (2007). *EMBO Rep.* **8**, 569–575.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hendrickson, W. & Ogata, C. (1997). *Methods Enzymol.* **276**, 494–523.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Hermoso, J. A., Lagartera, L., González, A., Stelter, M., García, P., Martínez-Ripoll, M., García, J. L. & Menéndez, M. (2005). *Nature Struct. Mol. Biol.* **12**, 533–538.
Jeudy, S., Stelter, M., Coutard, B., Kahn, R. & Abergel, C. (2005). *Acta Cryst.* F**61**, 848–851.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Ko, T.-P., Day, J., Greenwood, A. & McPherson, A. (1994). *Acta Cryst.* D**50**, 813–825.
Kumar, K., Chang, C. A., Francesconi, L. C., Dischino, D. D., Malley, M. F., Gougoutas, J. Z. & Tweedle, M. F. (1994). *Inorg. Chem.* **33**, 3567–3575.
Lagartera, L., González, A., Stelter, M., García, P., Kahn, R., Menéndez, M. & Hermoso, J. A. (2005). *Acta Cryst.* F**61**, 221–224.
Márquez, J., Reinelt, S., Koch, B., Engelmann, R., Hengstenberg, W. & Scheffzek, K. (2006). *J. Biol. Chem.* **281**, 32508–32515.
Molina, R., González, A., Moscoso, M., García, P., Stelter, M., Kahn, R. & Hermoso, J. A. (2007). *Acta Cryst.* F**63**, 742–745.
Molina, R., González, A., Stelter, M., Pérez-Dorado, I., Kahn, R., Morales, M., Moscoso, M., Campuzano, S., Campillo, N. E., Mobashery, S., García, J. L., García, P. & Hermoso, J. A. (2009). *EMBO Rep.* **10**, 246–251.
Molina, R., Stelter, M., Kahn, R. & Hermoso, J. A. (2009). *Acta Cryst.* D**65**, 823–831.
Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.
Pérez-Dorado, I., González, A., Morales, M., Sanles, R., Striker, W., Vollmer, W., Mobashery, S., García, J. L., Martínez-Ripoll, M., García, P. & Hermoso, J. A. (2010). *Nature Struct. Mol. Biol.* **17**, 576–581.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003*a*). *Acta Cryst.* D**59**, 868–875.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003*b*). *Acta Cryst.* D**59**, 1020–1027.
Ravelli, R. B. G., Leiros, H.-K. S., Pan, B., Caffrey, M. & McSweeney, S. (2003). *Structure*, **11**, 217–224.
Retailleau, P., Colloc'h, N., Vivarès, D., Bonneté, F., Castro, B., El Hajji, M. & Prangé, T. (2005). *Acta Cryst.* D**61**, 218–229.
Savchenko, A., Proudfoot, M., Skarina, T., Singer, A., Litvinova, O., Sanishvili, R., Brown, G., Chirgadze, N. & Yakunin, A. F. (2007). *J. Mol. Biol.* **374**, 1091–1103.
Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst.* D**60**, 1024–1031.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Vaney, M. C., Maignan, S., Riès-Kautt, M. & Ducruix, A. (1996). *Acta Cryst.* D**52**, 505–517.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.