Structural insights into protein–uranyl interaction: towards an in silico detection method

O. Pible, P. Guilbaud, J.-L. Pellequer, C. Vidaud, E. Quéméneur

Available online at www.sciencedirect.com
ScienceDirect
Biochimie 88 (2006) 1631–1638
BIOCHIMIE
www.elsevier.com/locate/biochi

Structural insights into protein–uranyl interaction: towards an in silico detection method

O. Pible[a,*], P. Guilbaud[b], J.-L. Pellequer[a], C. Vidaud[a], E. Quéméneur[a]

[a] *CEA VALRHO, DSV-DIEP-SBTN, Service de Biochimie postgénomique et Toxicologie Nucléaire, 30207 Bagnols-sur-Cèze, France*
[b] *CEA VALRHO, DEN-DRCP-SCPS, Service de Chimie des Procédés de séparation, laboratoire de conception des architectures moléculaires, 30207 Bagnols-sur-Cèze, France*

Received 16 January 2006; accepted 17 May 2006
Available online 21 June 2006

Abstract

Documenting the modes of interaction of uranyl (UO_2^{2+}) with large biomolecules, and particularly with proteins, is instrumental for the interpretation of its behavior in vitro and in vivo. The gathering of three-dimensional information concerning uranyl-first shell atoms from two structural databases, the Cambridge Structural Databank and the Protein Data Bank (PDB) allowed a screening of corresponding topologies in proteins of known structure. In the computer-aided procedure, all potentially bound residues from the template structure were granted full flexibility using a rotamer library. The Amber force-field was used to loosen constraints and score each predicted site. Our algorithm was validated as a first stage through the recognition of existing experimental data in the PDB. The coherent localization of missing atoms in the density map of an ambiguous uranium/uranyl–protein complex exemplified the efficiency of our approach, which is currently suggesting the experimental investigation of uranyl–protein binding site.

Keywords: Actinide; Hard Lewis acid; Structural analysis; Metal binding site; Screening; Amber; Uranium

1. Introduction

Despite its natural abundance in the earth crust or in sea water[1], uranium has never been observed within biochemical systems and even appears as a toxic compound at rather low dose. Its toxicity might be exerted through several modes yet to be fully described but all likely resulting from the ability of the uranyl cation (UO_2^{2+}) to bind strongly to biomolecules. As a consequence, information pertaining to interactions between uranyl and especially proteins is highly sought for the elaboration of pertinent toxicological models. Searching for target proteins on a large scale only by biochemical methods has proven to be a very challenging task as exemplified recently for serum binding proteins of uranyl cations [2]. Fortunately, computer-aided approaches have demonstrated their high efficiency in the discovery of natural binding sites for catalytic [3] or structural metal ions [4–6].

There are many possibilities regarding the screening strategy. 1D search in sequence databanks, based on sequence motifs established using 3D coordinates of known metalloproteins were shown to work for the prediction of protein-metal interactions [7]. However, such an approach is probably mostly valid for metal ions involved in a catalytic site and displaying a soft Lewis acid nature associated with polarized and oriented bonding (e.g. Cu(I), some Zn (CxxC) [3], indirect bonding through Fe-S clusters or heme groups being another matter…). Structural cations as Ca^{2+}, Mg^{2+}, or Zn^{2+} but also uranyl are certainly less prone to bind to sequentially constrained amino-acids, especially when bound to "harder" Lewis bases like oxygens [3]. In the case of Ca^{2+} that displays an extreme bonding heterogeneity, several studies [5,8] looked at valence distributions and used this accessible and distributed measurement for screening ion localizations in 3D databases. Since a favorable site for Ca binding might shield it from the solvent and exhibit a low dielectric constant in order to favor the replacement of

* Corresponding author.
E-mail address: olivier.pible@cea.fr (O. Pible).

[1] According to published seawater analysis [1] uranium concentration is about 2.8 ppb by weight, which is in the same range as some biologically important metals, e.g. zinc (1.75 ppb), iron (2.4 ppb), or nickel (0.4 ppb) averaged on the two seawater samples presented.

0300-9084/$ - see front matter © 2006 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2006.05.015

water oxygen atoms by carboxylate charged atoms [9], Yamashita et al. [4] previously examined hydrophobic properties in shells surrounding the putative positions of the cation. The recent Fold-X method for locating metal sites (Ca^{2+}, Mg^{2+}, Mn^{2+}, Zn^{2+}, Cu^{2+}) [6] involved the screening of protein structures for consensus sites that match canonical binding modes, and optimization in an empirical force field. Whatever the strategy used, the large diversity of coordination architectures should be kept in mind [10].

In biological media, uranium predominantly exists under its hexavalent oxidation state, forming a linear di-oxo uranyl cation (UO_2^{2+}) [11]. Uranyl is therefore presenting interesting attributes: a moderate net charge and a mostly electrostatic type of bonding which makes it acceptable for classical force-field (FF) calculations [12], and sterical constraints due to the two axial oxygen atoms which should impose a pentagonal or hexagonal bi-pyramidal planar chelation geometry. Being a hard Lewis acid, uranium is highly oxophilic [13]. Force fields parameters are available for molecular modeling of UO_2^{2+} [14], allowing fast calculations on large systems. Two limitations are well known for such an approach. The first is that short-distance interaction calculations are more questionable than longer ones in a simple FF without polarization nor charge transfer (polarization was not activated in Amber) [15]. More sophisticated force fields for uranyl [16] are not yet parameterized for complete proteins, in addition to high computational costs. The second issue is the fact that using an enthalpy based score—as done after minimization—will imperfectly depict all of the complexation phenomena. On the other hand, the classical Amber FF, using appropriate parameters for UO_2^{2+} [14], is probably better adjusted than most current free energy FF for these calculations.

In the field of structure-based drug design, where in silico screening of databases is frequent, flexibility is mostly granted to the ligand whereas the target protein is almost always left rigid [17]. However, incorporating receptor flexibility might be critical for the adaptation of the protein structure to the binding of small ligands. The importance of this parameter could be approached through comparison of structures of apo and holo-proteins [18]. This showed that more than 40% of metal sites undergo rearrangements, with main-chain motion in 14% of the cases, and side-chain reorientations in 35% of rigid-backbone sites. At least one side-chain moves in more than 75% of the cases. In their effort to design novel metal-binding templates by site-directed mutagenesis [19], protein engineers have realized the benefit of using side-chains rotamer libraries [20,21]. To our knowledge this type of method was not applied on more than a few proteins. Our computer approach aimed at implementing this function on a larger scale during the screening.

In this study, we gathered most of the structural information available for protein–uranium interaction from the protein data bank (PDB) [22], which contains more than 30,000 experimental structures of proteins and from the Cambridge Structural Database (CSD). We obtained through this survey significant data on the coordination of a defined ligand as uranyl onto a protein surface, allowing us to implement an original method for the screening of potential uranyl binding sites in proteins of known structure. Our computer approach was developed to perform a large scale screening on the PDB. A major advantage of our approach was to consider side-chain flexibility during the first geometrical screening stage. We calculated a score based on interaction and deformation enthalpy comparisons between identified complexation sites for sorting detected uranyl “sites” and hopefully predicting the "strength" of metal–protein interaction. The screening scheme also proved to be efficient in the re-analysis of the density map of a crystal structure.

2. Materials and methods

2.1. Databases

The April 2001 release of the CSD and the May 2004 release of the PDB database were used for uranyl statistics collection. The same PDB release and the updated version of Dec. 2005 were used for the screening of uranyl binding sites.

2.2. Processing of databases for computer-aided analysis

Only PDB files containing at least one polypeptide chain were kept. Modifications were performed in files where coordinates were missing; i) if one of the main-chain atoms (N, C, CA or O) was missing, the whole residue was removed from processed PDB files; ii) if the main-chain atoms were present but side-chain atoms were missing, all the side-chain atom positions were substituted with the first residue rotamer atoms positions after superimposition of the main-chain atoms; iii) if two cysteines sulfur atoms were found at a distance of less than 2.1 Å, the corresponding residues were written CYX to be treated by Amber as disulfide-bonded.

2.3. Software

Most figures were drawn with WebLab ViewerPro 3.7 for Windows (Molecular Simulations Inc.) and rendered with POV-Ray 3.6 under Linux (Persistence of Vision Pty. Ltd. 2004), except for the electronic density maps which were rendered using xfit [23] and Raster3D [24] under Linux. The screening program was written in Java (PDB files reading and fixing as detailed in the previous paragraph, geometrical screening, PDB writer for Amber, scripts for Amber parameters, Amber output analysis). Minimizations were performed using Amber6 [25] with previously reported parameters for UO_2^{2+} [14]. These parameters were adjusted in order to reproduce experimental hydration first sphere geometry of the cation and the hydration Gibbs free energy difference with other divalent cations. We used the TIP3P water molecule model. TYR had to be treated as deprotonated when involved in uranyl binding. Thus a new residue TYD, based on TYR, was added in Amber. Its atomic charges were calculated using the RESP method [26] applied to an electrostatic potential computed with the Merz–Kollman method in Gaussian 98 [27] using 6-31G*

basis set. Various scripts in csh shell, awk, perl were used for tasks automation including files handling and results processing.

3. Results

3.1. Statistical analysis of uranyl coordination geometries

The first part of this work involved the analysis of experimental 3D structures containing the UO_2^{2+} cation in order to define as accurately as possible the geometric features of the first coordination shell. These statistical analyses were performed on two structural databanks; the Cambridge Structural Database and the Protein DataBank.

The April 2001 release of the CSD contained 233,218 organic molecules and metallic complexes, among which 393 containing UO_2^{2+}; 237 with a coordination number (CN) of 5 and 119 with a CN of 6. UO_2^{2+} appeared bound to at least one oxygen atom in its equatorial plane for 350 complexes, and 209 structures counted only oxygen atoms in the first coordination shell of the cation; 121 with CN = 5 and 88 with CN = 6. Overall, 1578 U-O bonds were counted, compared to only 235 involving U-N bonds.

Average lengths of O–UO_2^{2+} bonds of various types of organic functions with CN = 5 are reported in Table 1. These data allowed us to extract specific U-O bond distances.

Most of these complexes occurred through phenolate, carboxylate or carbonyl functions also present in possible protein–UO_2^{2+} complexation sites.

Even if U-N bonds may be found in some CSD structures, the hard Lewis acid nature of UO_2^{2+} would promote its binding with oxygen atom donor amino-acids. Furthermore, among the 235 structures found in the CSD, none of them involves the nucleophilic "histidine like" imidazol motif, encouraging us to focus only on U-O bonds. As described below, no bonding involving nitrogen atoms were observed in protein complexes, except for one 1,3-diazole group.

The same statistical analysis was performed for CN = 6 uranyl complexes. Most of them implied structurally constrained or specific bidentate anions (oxalates, nitrates, etc) used for nuclear fuel reprocessing. Given these singularities and the fact that coordination number of UO_2^{2+} in proteins should be close to what is observed in water solution, i.e. CN = 5, the result of this analysis is not here detailed.

Table 1
Average lengths of O-uranyl bonds for different chemical groups within small UO_2^{2+} organic complexes of the CSD displaying five equatorial ligands (CN = 5)

U-bound chemical group	Numbers of CSD entries	Number of bonds	Distance (Å)
Ph-O^-	54	113	2.29 ± 0.12
COO^- monodentate	37	108	2.37 ± 0.05
R_2-C=O	19	33	2.40 ± 0.03
S-O	18	30	2.38 ± 0.04
S=O	18	24	2.38 ± 0.03
N-C=O	16	30	2.40 ± 0.05
P-O	16	49	2.39 ± 0.11
P=O	16	23	2.33 ± 0.04
H_2O	14	17	2.44 ± 0.04
COO^- bidentate	9	9	2.47 ± 0.05
NO_3^- bidentate	4	5	2.51 ± 0.03
NO_3^- monodentate	2	2	2.43 ± 0.06

A search for the items "uranyl", "uranium", "UO2" or "IUM" (regular HETATM residue name for uranyl) was performed in a recent release of the PDB (December 13, 2005) and revealed 64 hits. From these, 31 protein structures were shown to contain uranyl as a IUM heteroatom (see Table 2). The entry 1T9H deposited in May 2004, which displays an interesting small cluster of eight uranyl ions, was not in the database at the time of the primary screening. Therefore it was not included in the statistics. The two files 1AA0 (fibritin) and 4MDH (cytoplasmic malate deshydrogenase) were discarded from our set even though they give coordinates for uranyl in the REMARK section. Information relating to structures in the "Heavy Atom Databank" [28] were as well not satisfactory in terms of site resolution.

Most uranyl molecules are located at the asymmetric unit interface in crystals. This was expected since uranyl is used for phase resolution in multiple isomorphous replacement and is supposed to favor crystal contacts. Thus, we systematically performed crystal reconstruction to complement the full coordination shell before analyzing uranyl-atoms distance statistics on PDB files.

As anticipated, oxygen atoms were found to occur more frequently as ligands of uranyl than nitrogen atoms; i.e. 576 U-O bonds compared to only six U-N bonds in this dataset. Preliminary distance statistics were drawn from this complete data (not shown). A more pertinent reduced set aiming at reducing the crystallographic uncertainty was defined as follows. Among the 31 structures, 21 corresponded to a redundancy of the oligo-peptide binding protein (OppA) associated with different ligands [29]. Concerning OppA structures, Davies et al. [29] already stated that the electron density around the uranyl ions was noisy due to anisotropic thermal motion of the heavy atoms. To ensure that the crystallographic refinement was acceptable, only uranyl molecules including both apical oxygens were selected. This led to a drastic reduction of the dataset since this was the case for only 20 uranyl molecules from six PDB structures among the 180 IUM heteroatoms found originally. Then, only uranyl sites showing more than four first coordination shell ligands and at least three of them originating from protein residues were selected. This gave a small subset of seven well-defined sites from four PDB files (1 FE4 [30], 1EFQ [31], 1NCI [32] and 2OLB [33]), unbiased by questionable crystallographic assignment or refinement. In our view, this subset is representative of a significant uranyl–protein interaction and will constitute the test set. The equatorial U-O distances averaged from these seven sites are presented in Table 3. In the complete set of U-O distances, uranyl bonding to sidechain oxygen atoms of ASP, GLU, GLN, ASN, TYR and SER residues were observed in addition to mainchain carbonyl and water bonding. Some 6 U-N bonds were found and examined. They occurred in areas were uranyl oxygens were unattributed as well as most first shell ligands. The confidence in the electronic density attribution was too low to con-

Table 2
List of PDB files containing uranium as IUM

Protein name	Resolution (Å)	R-value	R-free	PDB identifier(s)
Oligopeptide binding protein (oppa) from *Salmonella typhimurium* complexed with various di/tri-peptides	**1.40 (2OLB)** others from 1.20 (1JET) to 2.30 (1B52)	0.183 (2OLB) others from 0.143 (1OLC) to 0.229 (1JET)	No value for 2OLB others from 0.197 (1OLC) to 0.263 (1B52)	1B1H, 1B32, 1B3F, 1B3G, 1B3L, 1B40, 1B46, 1B4Z, 1B51, 1B52, 1B58, 1B5J, 1B9J, 1JET, 1JEU, 1JEV, 1OLA, 1OLC, 1QKA, 1QKB, **2OLB**, 2RKM
DNA-binding protein from human adenovirus type 5	2.70	0.204	0.310	1ANV
A1 domain of human von Willebrand factor	2.30	0.186	0.248	1AUQ
Cu-Zn superoxide dismutase from *Photobacterium leiognathi*	2.10	0.190	0.260	1BZO
Asparagine synthetase b from *Escherichia coli*	2.00	0.197	0.297	1CT9
Q38d mutant of human Len (kappa-4 immunoglobulin light chain)	**1.60**	0.230	0.289	**1EFQ**
Atox1/Hah1—human copper transport protein, complexed with Hg	**1.75**	0.204	0.218	**1FE4**
Ferredoxin:$NADP^+$ oxidoreductase from spinach	1.70	0.179	–	1FNB
Domain 1 from murine neural cadherin	**1.90**	0.195	0.253	**1NCI**
Two domains fragment from murine N-cadherin	3.40	0.212	0.321	1NCJ
YloQ, an engineered GTPase from *Bacillus subtilis*	1.60	0.146	0.178	1T9H

Table 3
Statistics on average ($O\text{-}U_{uranyl}$) distances for selected PDB protein structures containing uranyl

	Number	Distance (Å)	Min (Å)	Max (Å)
COO^- monodentate	11	2.45 ± 0.10	2.31	2.61
COO^- bidentate	10	2.59 ± 0.13	2.41	2.84
H_2O	10	2.64 ± 0.33	1.94	3.00
Main-chain C=O	2	2.70 ± 0.15	2.59	2.80
Phenolate O^- (Tyr)	1	2.30		

sider this data as relevant, as it is finally the case for most of the 180 reported uranyl localizations.

Although statistical data from the PDB remains poor, the convergence of information gathered from the two databases provided with a simple coordination design suitable for the geometrical screening of protein structures. A Gaussian fit of the complete PDB extracted distances distribution (552 values, excluding water oxygen atoms which could be attributed to unassigned uranyl axial oxygen atoms) yielded a mean distance of 2.51 Å and a standard deviation (S.D.) of 0.20 Å. The CSD set of distances gave a mean distance of 2.38 Å for CN = 5 and 2.51 Å for CN = 6 (S.D. = 0.05 Å). Since the site design is dedicated to PDB structures exploration, our option was to consider primarily PDB values. The model for the first sphere atom positions was thus: UO_2^{2+} with 5 oxygen atoms equally spaced in the equatorial plane, with a U-O distance of 2.51 Å. This geometry is an average construction and is not seen in any available PDB file. Consequently there is no bias from input geometries into the screening procedure or validation process.

The amino-acid atoms considered as favorable to uranyl interaction were, according to atom terminology in the PDB files, OD2 (ASP), OE2 (GLU), OD1 (ASP,ASN), OE1 (GLU,GLN), OXT (CTER, any), OH (TYR), OG (SER), OG1 (THR) and O (main-chain carbonyl, any). These atoms will thereafter be called $O_{\rightarrow U}$. Regarding main chain carbonyls and Asn/Gln sidechains, it is worth recalling the conclusion of Siddons and Hancock [34] that the "amide oxygen is a stronger Lewis base in water than the alcoholic oxygen, or water". The involvement of tyrosine hydroxyl in an uranyl binding site has rather strict requirements due to the generally low flexibility of the side-chain and its high pK_a liable to be lowered by Lys or Arg proximity [35].

Dudev and Lim [9] stated, following free energy calculations, that the maximum number of Mg^{2+} or Ca^{2+} bound carboxylates without compensating effect from the protein matrix is three, and that all Mg sites should display at least one water molecule. We also considered this possibility of leaving 1 equatorial ligand available for a water molecule in our screening procedure by assigning only four of five equatorial positions to protein originated oxygen atoms. The contribution of this additional O atom would easily be discriminated by energetic minimization.

3.2. Design of an algorithm for detecting uranyl binding sites in protein structures

Taking advantage from the data generated by the statistical analyses of known protein–uranyl complexes, we designed a predictive algorithm for locating sites on protein surfaces that exhibit favorable features for a stable interaction with the uranyl cation. Several steps were included in the program to cope with both geometric and energetic criteria.

3.2.1. Geometrical criteria

A set of ranges for U-CA distance was derived from the PDB set and for the different types of amino-acids: Asp (4.7–6.2 Å); Glu (4.3–7.8 Å); Asn (4.7–6.2 Å); Gln (4.5–7.3 Å); Tyr (5.2–9.0 Å); Ser (4.3–5.8 Å); Thr (4.3–5.8); Cter (4.0–5.5). This made a reference table. A three-dimensional grid (1.3 Å mesh) was built for each protein structure to be analyzed. At each node positions, the U-CA distance was compared to the reference Table and only locations fitting with the distance range were kept for the next step of the program.

A library of 109 rotamers [20] was used to take sidechain flexibility into account. For each set of rotamers, $O_{\rightarrow U}$ atoms

were superimposed with the four positions tested from the UO_2^{2+} PDB based first coordination shell considering all possible atoms associations. Only locations rotamers with RMSD below 0.5 Å were accepted. This threshold was heuristically multiplied by 1.4 in case of two carboxylates candidates involved and by 1.6 if they were three. A site for which the U detected position was less than 0.01 Å apart from one of the previous validated grid position is ignored. A rather crude but meaningful estimate of the exposition of residues was then calculated using the OOI(14A) criterion [36], averaged on all the residues involved in the ion coordination. If the position was too exposed (OOI below 29) the IUM position was discarded. A final geometrical validation was performed, mostly to check for sterical clashes and electrostatics constraints: distances to positively charged atoms (the minimal distance is 3 Å to the uranium atom and 2 Å to the apical uranyl oxygen atoms), to backbone atoms (minimum 1.5 Å for U, 2 Å for uranyl oxygen atoms and 1.5 Å for the fifth first shell position) and to already coordinated residues side-chain atoms (minimal distance 1.5 Å to U and 1.2 Å to uranyl oxygen atoms).

3.2.2. Energetical criteria

The geometrical selection was followed by an additional step aiming at evaluating the energy of the protein–uranyl complex through several minimizations in the Amber FF (see Section 2 for Amber settings). A first minimization was performed for reconstructed side chains. Then, for each potential localization of the uranyl cation, three minimization runs were performed. The first one (PS) concerned the original uncomplexed protein structure and side-chain atoms of residues which CA was less than 10 Å away from the detected site. This provided the reference internal energy value. The second one (PU) was performed for the protein–uranyl complex with UO_2^{2+} in its best position (lowest RMSD among the 4 O positions screened) and with coordinated residues rotamer of lowest RMSD. The third one (PUW) was similar to PU, except that one water molecule was added in the fifth position around the U atom and treated as TIP3P in Amber. If a TYR appeared to be linked to UO_2^{2+}, it was considered as deprotonated and written as TYD in the Amber input file (see Section 2 for details).

3.2.3. Score calculation

Depending on the distance between U and O_{TIP3P} in the PUW minimization, the selected score was that of the protein/uranyl minimized structure with either one (d < 3 Å) or no (d > 3 Å) added water molecule. The calculated score took into account both the deformation energy of the protein E_{def} (difference of internal energy between the complexed and uncomplexed forms), the protein-IUM energy of interaction ($E_{P\text{-}IUM}$) plus non bonded terms related to water-protein and water–uranium interactions (E_{PUW}): $Score = E_{def} + E_{P\text{-}IUM} + E_{PUW}$.

The computational cost of the minimization stage is about 25-fold greater than the geometrical stage which lasts for an average of 2 min per protein.

3.3. Validation of the algorithm on known protein–uranyl complexes

The uranyl detection algorithm was validated on the four pdb structures mentioned previously, i.e. 1FE4, 1EFQ, 1NCI and 2OLB, which form seven complete coordination sites in the reconstructed crystal structures. Among these seven penta-coordinated sites, five involve four O-atoms provided by the protein chain, and one water molecule as screened by our algorithm. In the other two penta-coordinated sites from 2OLB, one has three bound water molecules and the other has two bond water molecules and an acetate ion.

The five site geometries were detected using the standard settings during the complete PDB database screening. Care was taken to add only crystal symmetric protein structures needed to complete the uranyl first shell instead of a complete crystal cell. Predicted uranyl positions were very close to actual positions in the PDB files and positioning errors shown in Table 4 were in the range of the atom position uncertainty, i.e. ~1/5–1/10 of the resolution. They were lowered during the minimization step which illustrates the FF parametering adequacy. Table 4 score rank was the order among the several detected sites for each structure, as ordered using the enthalpy based score.

The case of 1EFQ is exemplary of how efficient the prediction can be when the resolution and the quality of the initial structure is high (Fig. 1A5, B3). This also validated the model used for the unprotonated Tyr residue.

On the other hand, 1NCI challenged the prediction method. Interestingly, two models with two different scores were obtained. The “best” site did not involve the exact experimental ligands, but the Amber minimization restored the binding geometry (Fig. 1B1). The reason for the score ordering has been examined, and, as shown on Fig. 1B2, all detected sites

Table 4
Prediction positioning errors and score ranks

PDB_ID	Resolution	Residues retained during geometrical screening as bond to uranyl	Distance {U(PDB),U} Geometrical screening	Distance {U(PDB),U} Amber minimization	Score rank
1EFQ	1.60	ASP_38;GLN_42;TYR_27	0.28	0.16	1
1FE4	1.75	ASP_32;GLU_5;GLU_68	0.55	0.24	2
1NCI	1.90	GLU_11;ASP_67;GLU_69;GLU_11	0.47	0.30	3
1NCI	1.90	GLU_11;ASP_67;ASP_67;GLU_11	0.12	0.08	8
2OLB	1.40	GLU_53;ASP_362;ASP_410	0.32	0.18	1
2OLB	1.40	ASP_133;HIS_517	0.54	0.66	8

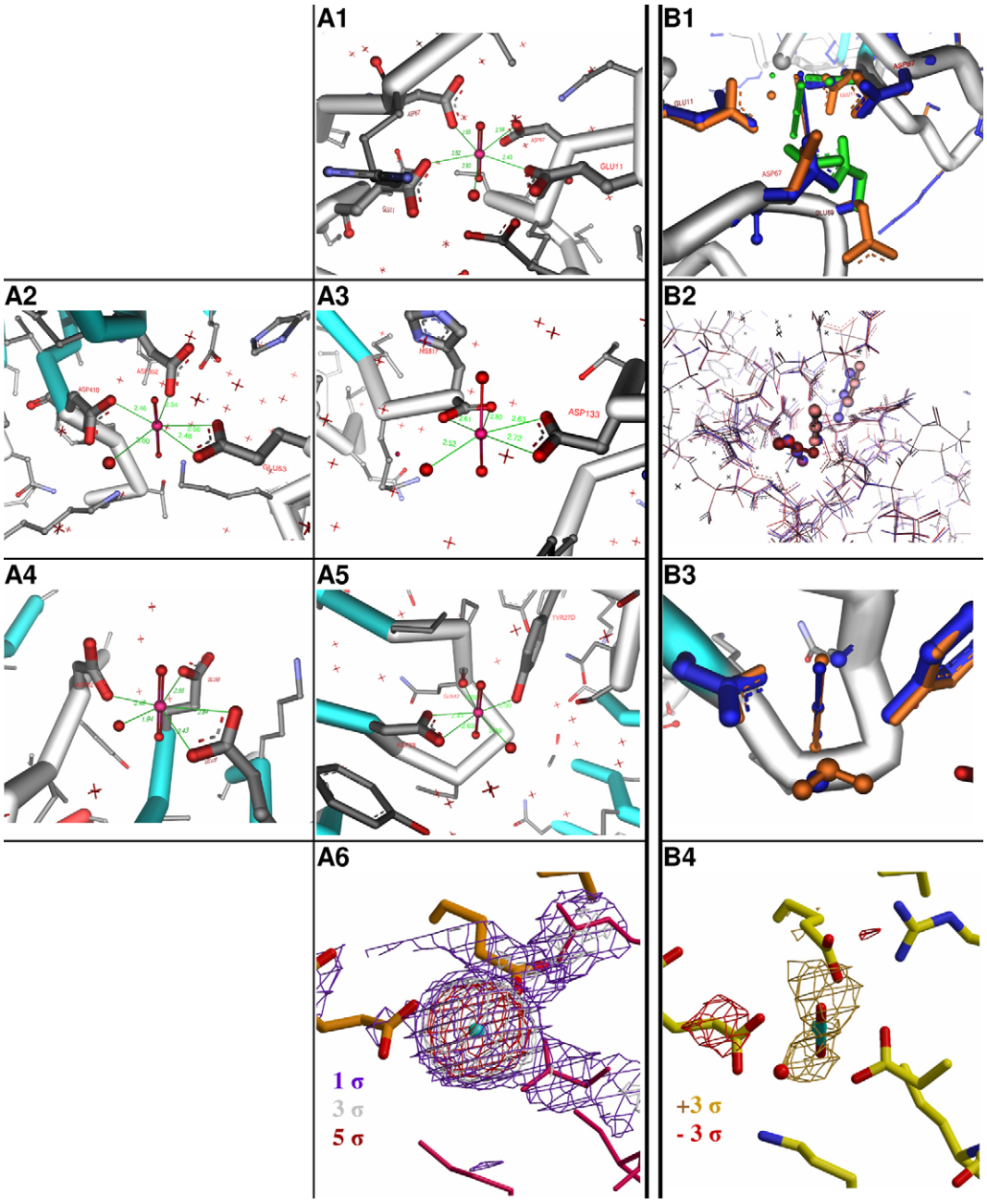

Fig. 1. Experimental (A) and predicted (B) uranyl binding sites.
A1: 1NCI 2 monodentate ASP and 2 monodentate GLU
A2: 2OLB 2 monodentate ASP and 1 bidentate GLU
A3: 2OLB 1 bidentate ASP and 1HIS Cter
A4: 1FE4 1 monodentate ASP, 1 monodentate GLU and 1 bidentate GLU
A5: 1EFQ 1 bidentate ASP, 1TYR and 1 main-chain GLN carbonyl
A6: original 1UYJ structure, symmetric atoms in magenta and 2mFo-DFc electronic density map
B1: 1NCI *Blue*: data from 1nci.pdb with crystal neighbor added, *Green*: rotamers as placed for best RMSD detected, *Orange*: minimized site
B2: 1NCI All of the 8 detected sites are shown here, with the crystal data in black (uranyl in the middle position), and colors from dark red for the best scoring site to dark blue for the worst scoring site.
B3: 1EFQ *Blue*: data from 1efq.pdb with crystal neighbor added, *Orange*: minimized site
B4: original 1UYJ uranium position, predicted uranyl orientation, minimized side-chains and water, and mFo-DFc electronic density map

were rather close to each other and at the interface between molecules. This suggests that the affinity for U of such site is partly entropy driven.

As a conclusion, our algorithm proved successful in detecting uranyl sites since all experimental sites could be detected and since all discrepancies between predicted and actual sites

could be resolved. The scoring function should not be regarded as a quantitative result because solvent interaction was overlooked, and did not prevent from the careful visual examination of putative sites.

3.4. Electronic density maps for 1UYJ

The structure of Clostridium perfringens epsilon toxin was recently collected at a resolution of 2.6 Å (PDB file 1UYJ) [37]. The crystal was found to include several uranium atoms (not identified as uranyl) among which one is bound to 3 Glu residues with an occupancy factor of 1. However, this site failed to be detected as an uranyl binding site by our screening algorithm. O-U distances were measured at 2.56, 2.59, 2.71, 3.07 and 3.57 Å. A water molecule possibly missing in the structure might explain the last 2 higher than expected distances. If uranium was considered as bound under its ionic uranyl form -which is the case in most normal conditions except for highly reducing solvent- the two uranyl oxygen atoms might also be missing (see Fig. 1A6). Furthermore, one of the GLU side-chain experimental geometry is not represented correctly by any of the rotamers in the limited set used. So a larger rotamer database might improve the detection quality in some conditions. The actual site is however found when allowing a greater number of sites to be detected (5–50), and is ranked third among detected sites using the standard scoring scheme.

The examination of the electronic density maps (obtained using the EDS service [38] and the PDB 1UYJ phase file) showed a good agreement between unattributed electronic density and the predicted position of the uranyl oxygen apical atoms (Fig. 1B4), confirming the supposition that uranium might need to be replaced by uranyl in this structure. The predicted water position is also inside the mFo-DFc electronic density map with a 2 sigma threshold.

4. Discussion

The approach described here proved successful in detecting the significant uranyl positions in the few known protein structures including uranyl as an heteroatom. It was then used for an exhaustive PDB screening, and to our knowledge this has never been performed previously for any cation using a full protein flexibility method. The computational cost was acceptable allowing multiple and frequent screenings of large structural databases. Among the many predicted uranyl localizations detected, some will be explored in more depth in further studies involving experimental confirmation of plausible targets all of which are much more time and resources demanding than computer-aided approaches such as this one.

An unsolved issue at this time remains the limited representativity of the PDB compared to large protein sequences databanks. The current UniProtKB/Swiss-Prot Release 48.8 database holds 205780 entries, whereas the January 10th, 2006 PDB database holds 32918 proteins or protein–nucleic acids complexes. Many structures represent only part of large proteins. For instance, the only available structure for human transferrin corresponds to its N-terminal lobe. Some are mutated, many loops are missing among other well known protein structures flaws. Molecular modeling is however important for the rationalization of biochemical hypotheses [2]. This might be more problematic then envisioned, especially because of the insufficient quality of most homology models. Even for existing structures, the many higher than 2.5 Å resolution structures are seldom well defined enough for adequate uranyl binding prediction.

A striking conclusion of our PDB analysis is that crystallographic data for uranyl containing areas is frequently questionable. This could even be misleading should raw statistics be performed. The reasons are multiple. They might relate to imperfectly parameterized refinement force-fields for such cations, high thermal motion associated to low affinity “sites” or high solvent exposure of most uranyl sites. A possible use of our program is to help crystallographic assignment for uranyl derivatives.

Protein main-chain flexibility is a source of false negatives and false positives in such a rigid backbone approach. A mere 14% of metal sites involving backbone motion greater than 0.8 Å for the maximum CA displacement and 0.6 Å ligand atoms RMSD was reported [18]. This issue might then be less problematic than expected.

Known questions raised include post-translational modifications. They are indeed under-represented in the PDB databank. Oxygen atoms of phosphate groups being among the best ligands for uranyl, phosphorylated residues might mediate uranyl complexation. This property is overlooked in this study. Glycosylation of asparagine, serine or threonine would adversely modify local binding conditions.

An important question is also the cell control over the internal uranyl concentration. Competition between Mg and Ca or Zn have been reported for low affinity binding sites depending relative ion concentrations [19,9,39]. Our data do not exclude the possibility that uranyl occupies other hard Lewis metals sites at high concentration. This should also be kept in mind in the case of bone hydroxyapatite which is a final target of uranyl in the human body.

The study of protein interaction networks is one of the current “hot” topics in today’s biology. Disturbance of such networks was invoked as a major mode of toxicity of uranium in kidney and lung cells [40,41]. Structural data as obtained through this work will provide convergent perspectives with the toxico-genomic and proteomic approaches.

Acknowledgements

We wish to thank Gilles Imbert for various perl and shell scripts and Isabelle Vergely for help on CIF data extraction help.

This study has been supported by a grant of the French National program “Toxicologie Nucléaire et Environnementale” to the project “Cibles Moléculaires des Actinides”.

References

[1] A.P. Mykytiuk, D.S. Russell, R.E. Sturgeon, Simultaneous determination of iron, cadmium, zinc, copper, nickel, lead, and uranium in seawater by stable isotope-dilution spark source-mass spectrometry, Anal. Chem. 52 (1980) 1281–1283.

[2] C. Vidaud, A. Dedieu, C. Basset, S. Plantevin, I. Dany, O. Pible, E. Quemeneur, Screening of human serum proteins for uranium binding, Chem. Res. Toxicol. 18 (2005) 946–953.

[3] I.L. Alberts, K. Nadassy, S.J. Wodak, Analysis of zinc binding sites in protein crystal structures, Protein Sci. 7 (1998) 1700–1716.

[4] M.M. Yamashita, L. Wesson, G. Eisenman, D. Eisenberg, Where metal ions bind in proteins, Proc. Natl. Acad. Sci. USA 87 (1990) 5648–5652.

[5] M. Nayal, E. Di Cera, Predicting Ca(2+)-binding sites in proteins, Proc. Natl. Acad. Sci. USA 91 (1994) 817–821.

[6] J.W. Schymkowitz, F. Rousseau, I.C. Martins, J. Ferkinghoff-Borg, F. Stricher, L. Serrano, Prediction of water and metal binding sites and their affinities by using the Fold-X force field, Proc. Natl. Acad. Sci. USA 102 (2005) 10147–10152.

[7] C. Andreini, I. Bertini, A. Rosato, A hint to search for metalloproteins in gene banks, Bioinformatics 20 (2004) 1373–1380.

[8] P. Mueller, S. Koepke, G.M. Sheldrick, Is the bond-valence method able to identify metal atoms in protein structures, Acta Crystallogr. D Biol. Crystallogr. D59 (2003) 32–37.

[9] T. Dudev, C. Lim, Principles governing Mg, Ca, and Zn binding and selectivity in proteins, Chem. Rev. 103 (2003) 773–788.

[10] M.M. Harding, The architecture of metal coordination groups in proteins, Acta Crystallogr. D Biol. Crystallogr. 60 (2004) 849–859.

[11] C. Voetglin, H.C. Hodge, Pharmacology and Toxycology of Uranium Compounds, McGraw-Hill Book Company, New York, 1953.

[12] L. David, P. Amara, M.J. Field, F. Major, Parametrization of a force field for metals complexed to biomacromolecules: applications to Fe(II), Cu (II) and Pb(II), J. Comput. Aided Mol. Des. 16 (2002) 635–651.

[13] A.E. Gorden, J. Xu, K.N. Raymond, P. Durbin, Rational design of sequestering agents for plutonium and other actinides, Chem. Rev. 103 (2003) 4207–4282.

[14] P. Guilbaud, G. Wipff, Force field representation of the UO22+ cation from free energy MD simulations in water. Tests on its 18-crown-6 and NO3- adducts, and on its calix [6]arene(6-) and CMPO complexes, J. Mol. Struct. Theochem 366 (1996) 55–63.

[15] L. Hemmingsen, P. Amara, E. Ansoborlo, M.J. Field, Importance of Charge Transfer and Polarization Effects for the Modeling of Uranyl-Cation Complexes, J. Phys. Chem. A 104 (2000) 4095–4101.

[16] C. Clavaguera-Sarrio, V. Brenner, S. Hoyau, C.J. Marsden, P. Millie, J.P. Dognon, Modeling of Uranyl Cation-Water Clusters, J. Phys. Chem. B 107 (2003) 3051–3060.

[17] M.L. Teodoro, L.E. Kavraki, Conformational flexibility models for the receptor in structure based drug design, Curr. Pharm. Des. 9 (2003) 1635–1648.

[18] M. Babor, H.M. Greenblatt, M. Edelman, V. Sobolev, Flexibility of metal binding sites in proteins on a database scale, Proteins 59 (2005) 221–230.

[19] W. Yang, H.W. Lee, H. Hellinga, J.J. Yang, Structural analysis, identification, and design of calcium-binding sites in proteins, Proteins 47 (2002) 344–356.

[20] P. Tuffery, C. Etchebest, S. Hazout, R. Lavery, A new approach to the rapid determination of protein side chain conformations, J. Biomol. Struct. Dyn. 8 (1991) 1267–1289.

[21] H.W. Hellinga, F.M. Richards, Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry, J. Mol. Biol. 222 (1991) 763–785.

[22] H.M. Berman, T.N. Bhat, P.E. Bourne, Z. Feng, G. Gilliland, H. Weissig, J. Westbrook, The Protein Data Bank and the challenge of structural genomics, Nat. Struct. Biol. 7 Suppl (2000) 957–959.

[23] D.E. McRee, XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density, J. Struct. Biol. 125 (1999) 156–165.

[24] E.A. Merritt, D.J. Bacon, Raster3D: photorealistic molecular graphics, Methods Enzymol. 277 (1997) 505–524.

[25] D.A. Case, D.A. Pearlman, J.W. Caldwell, I.T.E. Cheatham, W.S. Ross, C.L. Simmerling, T.A. Darden, K.M. Merz, R.V. Stanton, A.L. Cheng, J.J. Vincent, M. Crowley, V. Tsui, V. Radmer, Y. Duan, J. Pitera, I. Massova, G.L. Seibel, U.C. Singh, P.K. Weiner, P.A. Kollman, AMBER 6, San Francisco, 1999.

[26] C.I. Bayly, P. Cieplak, W.D. Cornell, P.A. Kollman, A Well-Behaved Electrostatic Potential Based Method Using for Deriving Atomic Charges: The RESP Model, J. Phys. Chem. 97 (1993) 10269–10280.

[27] M.J.T. Frisch, G.W. Schlegel, H.B. Scuseria, G.E.M.A.C. Robb, J.R. Zakrzewski, V.G. Montgomery, J.A.R.E.B. Stratmann, J.C. Dapprich, S. Millam, J.M. Daniels, A.K.N.S.D. Kudin, M.C. Farkas, O. Tomasi, J. Barone, V. Cossi, R.M.M. Cammi, B. Pomeli, C. Adamo, C. Clifford, S. J.P. Ochterski, G.A. Ayala, P.Y. Cui, Q. Mokoruma, K. Malick, A.D.R.D.K. Rabuck, K. Foresman, J.B. Cioslowski, J.J.V.S. Ortiz, B.B. Liu, G. Liashenko, A. Piskorz, P. Komaromi, R.M.L. Gomperts, R.L. Fox, D.J. Keith, T. Al-Laham, M.A. C.Y.N. Peng, A. Gonzalez, C. Challacombe, M. Gill, P.M.B.G.C.W. Johnson, W. Wong, M.W. Andres, J.L. Head-Gordon, E.S.P.M. Replogle, J.A. Gaussian 98, Gaussian Inc, Pittsburgh, PA, 1998.

[28] S.A. Islam, D. Carvin, M.J. Sternberg, T.L. Blundell, HAD, a data bank of heavy-atom binding sites in protein crystals: a resource for use in multiple isomorphous replacement and anomalous scattering, Acta Crystallogr. D Biol. Crystallogr. 54 (1998) 1199–1206.

[29] T.G. Davies, R.E. Hubbard, J.R. Tame, Relating structure to thermodynamics: the crystal structures and binding affinity of eight OppA-peptide complexes, Protein Sci. 8 (1999) 1432–1444.

[30] A.K. Wernimont, D.L. Huffman, A.L. Lamb, T.V. O'Halloran, A.C. Rosenzweig, Structural basis for copper transfer by the metallochaperone for the Menkes/Wilson disease proteins, Nat. Struct. Biol. 7 (2000) 766–771.

[31] P.R. Pokkuluri, M. Gu, X. Cai, R. Raffen, F.J. Stevens, M. Schiffer, Factors contributing to decreased protein stability when aspartic acid residues are in beta-sheet regions, Protein Sci. 11 (2002) 1687–1694.

[32] L. Shapiro, A.M. Fannon, P.D. Kwong, A. Thompson, M.S. Lehmann, G. Grubel, J.F. Legrand, J. Als-Nielsen, D.R. Colman, W.A. Hendrickson, Structural basis of cell-cell adhesion by cadherins, Nature 374 (1995) 327–337.

[33] J.R. Tame, E.J. Dodson, G. Murshudov, C.F. Higgins, A.J. Wilkinson, The crystal structures of the oligopeptide-binding protein OppA complexed with tripeptide and tetrapeptide ligands, Structure 3 (1995) 1395–1406.

[34] C.J. Siddons, R.D. Hancock, Possible insights into metal ion recognition in calcium-binding proteins provided by complexing properties of ligands containing amide oxygen donors, Chem. Commun. (Camb.) (2004) 1632–1633.

[35] P. Chakrabarti, B.T. Hsu, Cation binding by the phenolate group in small molecules and proteins, inorg, Chem 33 (1994) 1165–1170.

[36] K. Nishikawa, T. Ooi, Radial locations of amino acid residues in a globular protein: correlation with the sequence, J. Biochem. (Tokyo) 100 (1986) 1043–1047.

[37] A.R. Cole, M. Gibert, M. Popoff, D.S. Moss, R.W. Titball, A.K. Basak, Clostridium perfringens epsilon-toxin shows structural similarity to the pore-forming toxin aerolysin, Nat. Struct. Mol. Biol. 11 (2004) 797–798.

[38] G. Kleywegt, M. Harris, J. Zou, T. Taylor, A. Wählby, T. Jones, The uppsala electron density server, Acta Cryst., D/CCP4 Proceedings, 2004.

[39] J.J. Falke, E.E. Snyder, K.C. Thatcher, C.S. Voertler, Quantitating and engineering the ion specificity of an EF-hand-like Ca^{2+} binding, Biochemistry 30 (1991) 8690–8697.

[40] O. Prat, F. Berenguer, V. Malard, E. Tavan, N. Sage, G. Steinmetz, E. Quemeneur, Transcriptomic and proteomic responses of human renal HEK293 cells to uranium toxicity, Proteomics 5 (2005) 297–306.

[41] V. Malard, O. Prat, E. Darrouzet, F. Berenguer, N. Sage, E. Quemeneur, Proteomic analysis of the response of human lung cells to uranium, Proteomics 5 (2005) 4568–4580.