BIOLOGICAL CRYSTALLOGRAPHY
ISSN 1399-0047

Crystallography and chemistry should always go together: a cautionary tale of protein complexes with cisplatin and carboplatin

Ivan Shabalin,[a] Zbigniew Dauter,[b]* Mariusz Jaskolski,[c,d] Wladek Minor[a] and Alexander Wlodawer[e]*

[a]Department of Molecular Physiology and Biological Physics, University of Virginia, 1340 Jefferson Park Avenue, Charlottesville, VA 22908, USA, [b]Synchrotron Radiation Research Section, MCL, National Cancer Institute, Argonne National Laboratory, Argonne, IL 60439, USA, [c]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznan, Poland, [d]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, and [e]Protein Structure Section, MCL, National Cancer Institute, Frederick, MD 21702, USA. *Correspondence e-mail: dauter@anl.gov, wlodawer@nih.gov

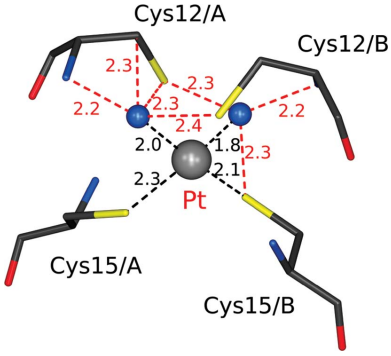
Received 26 February 2015
Accepted 27 March 2015

Edited by T. O. Yeates, University of California, USA

Keywords: cisplatin; carboplatin; crystal structure; error corrections; structural databases; structure re-refinement; validation; data reprocessing; reproducibility.

PDB references: cisplatin bound to a human copper chaperone (monomer), new refinement, 4ydx; cisplatin bound to a human copper chaperone (dimer), new refinement, 4yea; carboplatin binding to HEWL in NaBr crystallization conditions studied at an X-ray wavelength of 0.9163 A, new refinement, 4yem; room-temperature X-ray diffraction studies of cisplatin binding to HEWL in DMSO medium after 14 months of crystal storage, new refinement, 4yen; triclinic HEWL co-crystallized with cisplatin studied at a data-collection temperature of 150 K, new refinement, 4yeo

Supporting information: this article has supporting information at journals.iucr.org/d



The anticancer activity of platinum-containing drugs such as cisplatin and carboplatin is considered to primarily arise from their interactions with nucleic acids; nevertheless, these drugs, or the products of their hydrolysis, also bind to proteins, potentially leading to the known side effects of the treatments. Here, over 40 crystal structures deposited in the Protein Data Bank (PDB) of cisplatin and carboplatin complexes of several proteins were analysed. Significant problems of either a crystallographic or a chemical nature were found in most of the presented atomic models and they could be traced to less or more serious deficiencies in the data-collection and refinement procedures. The re-evaluation of these data and models was possible thanks to their mandatory or voluntary deposition in publicly available databases, emphasizing the point that the availability of such data is critical for making structural science reproducible. Based on this analysis of a selected group of macromolecular structures, the importance of deposition of raw diffraction data is stressed and a procedure for depositing, tracking and using re-refined crystallographic models is suggested.

1. Introduction

Since its birth in 1913, structural X-ray crystallography has been a committed and faithful companion of chemistry, even if certain initial results were at first surprising to some chemists, who had conceptual difficulties in accepting, for example, that there is no such thing as a separate molecule of NaCl (Bragg, 1913) or who questioned the possibility that penicillin contains a four-membered β-lactam ring (Crowfoot *et al.*, 1949). Macromolecular crystallography was, and still is, absolutely indispensable in explaining the atomic details of many biochemical processes of life, such as the transport of oxygen (Perutz, 1970) and various enzymatic reactions, starting from the work of David Phillips on lysozyme (Phillips, 1967) and culminating in the recognition that the ribosome is a ribozyme (Cech, 2000), where protein synthesis is catalyzed by a fragment of ribonucleic acid and not by a protein.

Many crystallographic results were quite unexpected and engendered new avenues of chemistry and biochemistry, but eventually were always in agreement with the evolving chemical principles and chemical common sense. Therefore, the symbiosis between crystallography and chemistry has solid foundations and each of the disciplines has benefitted enormously from the knowledge amassed in the other field.

However, on occasions the results of crystallography are interpreted (or rather overinterpreted) with disregard for the established principles of chemistry. Such results, if proliferated through the various data banks, *e.g.* the CSD (Allen, 2002) or the PDB (Berman *et al.*, 2000), may confuse not only inexperienced but even experienced users of crystallographic data and models, and may upset those who live by those principles.

In this context, it is difficult to overestimate the importance of proper validation of structural results prepared for publication and dissemination to the community of potential 'customers'. Both major repositories, the CSD, which contains over 750 000 small-molecule organic and organometallic crystal structures, and the PDB, which currently holds ∼110 000 models of macromolecular structures established predominantly (>88%) by crystallography, perform a variety of validation procedures using a number of sophisticated software tools, *e.g. PLATON* (Spek, 2009) in the CSD and *SFCHECK* (Vaguine *et al.*, 1999), *PROCHECK* (Laskowski *et al.*, 1993), *WHATCHECK* (Hooft *et al.*, 1996) and *MolProbity* (Chen *et al.*, 2010) in the PDB. However, the latter tools mainly validate the stereochemical plausibility of the protein and nucleic acid portions of macromolecular models. The assessment of small-molecule ligands is much more difficult, and validation tools that check whether ligand placement and conformation are chemically justified are relatively new, for example *ValLigUrl* (Kleywegt & Harris, 2007), *PURY* (Andrejašič *et al.*, 2008), *EDSTATS* (Tickle, 2012), *Twilight* (Weichenberger *et al.*, 2013) and *CheckMyMetal* (Zheng *et al.*, 2014). An excellent recent review has comprehensively addressed the problems of modeling the crystal structures of protein–ligand complexes (Deller & Rupp, 2015). Moreover, many ligand models contain atoms with low or even zero occupancy, which usually means that no electron density was observed in the experimental maps and that the molecule or its fragment was added there only as a placeholder based on the experimenter's expectation (or wishful thinking, as exemplified by Winnie-the-Pooh: 'The more he looked inside the more Piglet wasn't there'; Milne, 1928) of the chemical composition of the molecules. Wishful thinking may not take into account, for example, that crystallizatin cocktail components could lead to decomposition or transformation of the ligand under study. We want to stress that the ultimate responsibility for the proper validity of the final results rests on the shoulders of the depositors, not the annotators, watchdogs or the developers of validation software (Read *et al.*, 2011; Dauter *et al.*, 2014), although the curators of structural data banks and databases should make every effort to ensure that all deposited models faithfully represent the investigated structures.

The relative ease and automated determination of some structures may lead to the temptation to release the results quickly, without proper scrutiny, especially under the pressure of the current paranoia to 'publish or perish'. Analysis of the RSRZ (real-space *R Z*-score) and LLDF (local ligand density fit) parameters recently introduced by the PDB (Sen *et al.*, 2014) indicates that many crystallographers believe so much in automatic programs that they have a tendency to skip the most important validation step, namely visual analysis of the agreement of their model with the experimental electron-density map (Lamb *et al.*, 2015; Deller & Rupp, 2015). Therefore, thorough validation of existing structural models both from the point of view of chemistry and the agreement of the model with the experimental data is not only beneficial, but is quite often necessary.

The present work analyses all of the available structures of protein complexes with two widely used anticancer compounds: cisplatin and carboplatin (for reviews, see Lippard, 1982; Wang & Lippard, 2005). Several laboratories have published structures of such complexes determined at high-to-atomic resolution. Nevertheless, our analysis of the structures deposited in the PDB found quite troubling problems that may require significant reinterpretation from the point of view of acceptable chemistry. We did not aim to simply question the available results and their interpretation, but rather to postulate more robust ways of data analysis, as well as more thorough verification of the macromolecular crystallographic data and results submitted to the PDB. In doing so, we attempted to follow the advice from Hamlet: 'For some must watch, while some must sleep: so runs the world away' (Shakespeare, 1603).

2. Materials and methods

No new experimental data were collected as part of this work, and all analyses involved data collected by others and deposited in the Protein Data Bank. A search conducted on 16 December 2014 for structures of proteins complexed with cisplatin (CPT) identified 25 sets of coordinates (Table 1). In addition, the PDB contained 17 structures of protein complexes with carboplatin (QPT). Not all of these structures are unique, since some represent re-refinement of the same diffraction data processed using different software (Table 2). In addition, in January 2015 the lysozyme–carboplatin data set 4xan superseded entry 4nsf in the PDB; both are analyzed here. We have not analyzed the structures of nucleic acids and protein–nucleic acid complexes in which cisplatin is bound to the nucleic acid component or any Pt-containing structures not related to either carboplatin or cisplatin. When available, electron-density maps were downloaded from the Uppsala EDS (Kleywegt *et al.*, 2004); otherwise, they were calculated using programs from the *CCP*4 suite (Winn *et al.*, 2011). Anomalous difference electron-density maps were calculated if the deposited structure-factor files contained the required information. Selected models were re-refined with *REFMAC*5 v.5.8 (Murshudov *et al.*, 2011) as implemented in *HKL*-3000 (Minor *et al.*, 2006) using the structure-factor data deposited in the PDB or reprocessed (Otwinowski & Minor, 1997) after downloading from the server at http://rawdata.chem.uu.nl. If possible, the reflections selected for $R_{\rm free}$ testing were the same as in the originally deposited data sets. We also discovered that some other small-molecule ligands in the analyzed complexes were modeled incorrectly. Schematic chemical formulae for some of these molecules, including selected bond lengths obtained from the CSD, are shown in Fig. 1.

Table 1
PDB-deposited structures of protein complexes with cisplatin (CPT).

All lysozyme structures are complexes with hen egg-white lysozyme (HEWL).

PDB code	Protein	Resolution (Å)	R, R_{free}	Pt sites	Reference
2aeo	Cu,Zn superoxide dismutase	1.8	0.205, 0.236	2	Calderone *et al.* (2006)
2i6z	Lysozyme, tetragonal	1.9	0.189, 0.243	1	Casini *et al.* (2007)
3iwx	Copper-transport protein ATOX1	2.14	0.188, 0.228	1	Boal & Rosenzweig (2009)
3re0	Cu,Zn superoxide dismutase	2.28	0.216, 0.290	5	Banci *et al.* (2012)
4dd4	Lysozyme, tetragonal	1.7	0.205, 0.259	2	Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.* (2012)
4dd6	Lysozyme, tetragonal	1.7	0.220, 0.268	2	Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.* (2012)
4ddb	Lysozyme, tetragonal	3.0	0.217, 0.280	2	Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.* (2012)
4ddc	Lysozyme, tetragonal	1.8	0.221, 0.253	4	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*b*)
4g49	Lysozyme, tetragonal	2.4	0.177, 0.210	2	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*b*)
4g4a	Lysozyme, tetragonal	2.4	0.165, 0.210	2	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*b*)
4g4b	Lysozyme, tetragonal	2.1	0.162, 0.227	2	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*b*)
4gcb	Lysozyme, tetragonal	1.8	0.176, 0.210	3	Helliwell & Tanley (2013)
4gcc	Lysozyme, tetragonal	2.0	0.189, 0.248	2	Helliwell & Tanley (2013)
4gcd	Lysozyme, tetragonal	2.8	0.178, 0.240	2	Helliwell & Tanley (2013)
4gce	Lysozyme, tetragonal	2.9	0.177, 0.234	2	Helliwell & Tanley (2013)
3txf	Lysozyme, tetragonal	1.69	0.179, 0.249	1	Tanley, Schreurs *et al.* (2013)
3txg	Lysozyme, tetragonal	1.7	0.184, 0.239	2	Tanley, Schreurs *et al.* (2013)
3txk	Lysozyme, tetragonal	3.0	0.214, 0.259	2	Tanley, Schreurs *et al.* (2013)
4ot4	RNase A, *C*2	1.85	0.192, 0.247	3	Messori & Merlino (2014)
4mwm	Lysozyme, triclinic	1.12	0.149, 0.187	4	Tanley & Helliwell (2014*b*)
4mwn	Lysozyme, triclinic	1.42	0.210, 0.235	3	Tanley & Helliwell (2014*b*)
4mwk	Lysozyme, triclinic	0.98	0.119, 0.145	7	Tanley & Helliwell (2014*b*)
4n0z	Glutaredoxin 1 (PfGrx1)	1.7	0.156, 0.171	1	Unpublished
4n10	Glutaredoxin 1 (PfGrx1)	1.87	0.182, 0.223	3	Unpublished
4n11	Glutaredoxin 1 (PfGrx1)	1.97	0.162, 0.208	3	Unpublished

3. Results

3.1. Proteins and platinum complexes analyzed in this work

Structural information about the binding of cisplatin to several model proteins has been reported during the last eight years. Whereas the vast majority of structures represent complexes with the tetragonal or triclinic forms of hen egg-white lysozyme (HEWL; Casini *et al.*, 2007; Tanley, Schreurs, Kroon-Batenburg & Helliwell, 2012*b*; Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012; Tanley, Schreurs *et al.*, 2013; Tanley *et al.*, 2014; Helliwell & Tanley, 2013; Tanley & Helliwell, 2014*a*), data are also available for complexes with bovine Cu,Zn superoxide dismutase (SOD; Calderone *et al.*, 2006; Banci *et al.*, 2012), the human copper-transport protein ATOX1 (Boal & Rosenzweig, 2009), bovine pancreatic RNase A (Messori & Merlino, 2014) and *Plasmodium falciparum* glutaredoxin 1 (PfGrx1; unpublished work). All available structures of carboplatin involve complexes with the tetragonal, monoclinic or trigonal forms of HEWL (Tanley, Schreurs, Kroon-Batenburg & Helliwell, 2012*a*; Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012; Tanley, Schreurs *et al.*, 2013; Tanley *et al.*, 2014; Tanley & Helliwell, 2014*b*).

3.2. Binding of cisplatin to superoxide dismutase (SOD)

In the complex of cisplatin with SOD (PDB entry 2aeo; Calderone *et al.*, 2006) the Pt atom is coordinated by the $N^{\varepsilon 2}$ atom of His19, with occupancies of 0.7 for molecule *A* and 0.6 for molecule *B*. Two Cl atoms were modeled in each CPT in a *cis* configuration, with occupancies identical to that of the Pt, but, surprisingly, the clear electron density for the fourth ligand completing the square-planar coordination of Pt was not interpreted by the original authors. The residual density in the $mF_o - DF_c$ map reaches 6σ at each of the unmodeled peaks, indicating that an N or O atom could be placed there. The picture is very different in the other SOD structure, PDB entry 3re0 (Banci *et al.*, 2012). In this case, each of the partially occupied (0.4) CPT molecules was modeled as bound alternatively to two closely positioned Cys111 residues of protein molecules *A* and *B*, with another such pair bound to a second dimer, molecules *C* and *D*. The Pt atoms are covalently bound to the cysteine S atoms, and the remaining three atoms were modeled as two *cis* N atoms and a chlorine. Simultaneous occupancy of both sites at the *A*/*B* or *C*/*D* interface is not possible owing to clashes of the N atoms of the two CPTs. The geometry of the CPT complexes is considerably distorted from square planar. Another CPT molecule was modeled in the vicinity of Asp109 of molecule *B*, but since the side chain of this residue sits entirely outside the $2mF_o - DF_c$ electron density contoured at 0.9σ, and the distance of 2.3 Å between N1 of CPT and OD1 of Asp109 is too short for a hydrogen bond, the interpretation of this CPT should be considered as highly doubtful.

3.3. Binding of cisplatin to the copper-transport protein ATOX1

Binding of cisplatin to ATOX1 was modeled differently in the complexes with dimeric and monomeric protein molecules. In the former structure (PDB entry 3iwx; Boal & Rosenzweig, 2009) a CPT molecule was modeled as consisting of a Pt atom and two *cis* ammine groups with an occupancy of 0.4 (Fig. 2*a*). The Pt atom is located between four cysteine S atoms (Cys12

Table 2
PDB-deposited structures of protein complexes with carboplatin (QPT).

All structures represent complexes with hen egg-white lysozyme (HEWL), either tetragonal, triclinic or monoclinic.

PDB code	Resolution (Å)	R, R_{free}	Pt sites	Reference
4dd7, tetragonal	1.6	0.186, 0.223	2	Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.* (2012)
4dd9, tetragonal	1.6	0.206, 0.236	2	Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.* (2012)
4g4c, tetragonal	2.0	0.162, 0.203	2	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*a*)
4g4h, tetragonal	2.0	0.208, 0.268	2	Tanley, Schreurs, Kroon-Batenburg & Helliwell (2012*a*)
3txh, tetragonal	1.69	0.170, 0.233	2	Tanley, Schreurs *et al.* (2013)
3txi, tetragonal	1.6	0.190, 0.233	2	Tanley, Schreurs *et al.* (2013)
4lt0, tetragonal	2.1	0.226, 0.283	1	Tanley *et al.* (2014)
4lt1, tetragonal	2.3	0.189, 0.248	1	Unpublished
4lt2, tetragonal	2.0	0.210, 0.265	2	Unpublished
4lt3, tetragonal	2.0	0.199, 0.257	2	Tanley *et al.* (2014)
4owa, monoclinic	1.8	0.187, 0.241	3	Tanley & Helliwell (2014*a*)
4nsf, tetragonal	1.47	0.180, 0.214	2	Tanley *et al.* (2014)
4xan, tetragonal	1.47	0.129, 0.181	2	Tanley *et al.* (2014)
4nsg, tetragonal	2.0	0.221, 0.271	2	Tanley *et al.* (2014)
4nsh, tetragonal	2.1	0.199, 0.262	2	Tanley *et al.* (2014)
4nsi, tetragonal	2.3	0.224, 0.285	2	Tanley *et al.* (2014)
4nsj, tetragonal	1.7	0.202, 0.258	1	Tanley *et al.* (2014)
4oxe, triclinic	1.13	0.180, 0.221	3	Tanley & Helliwell (2014*b*)

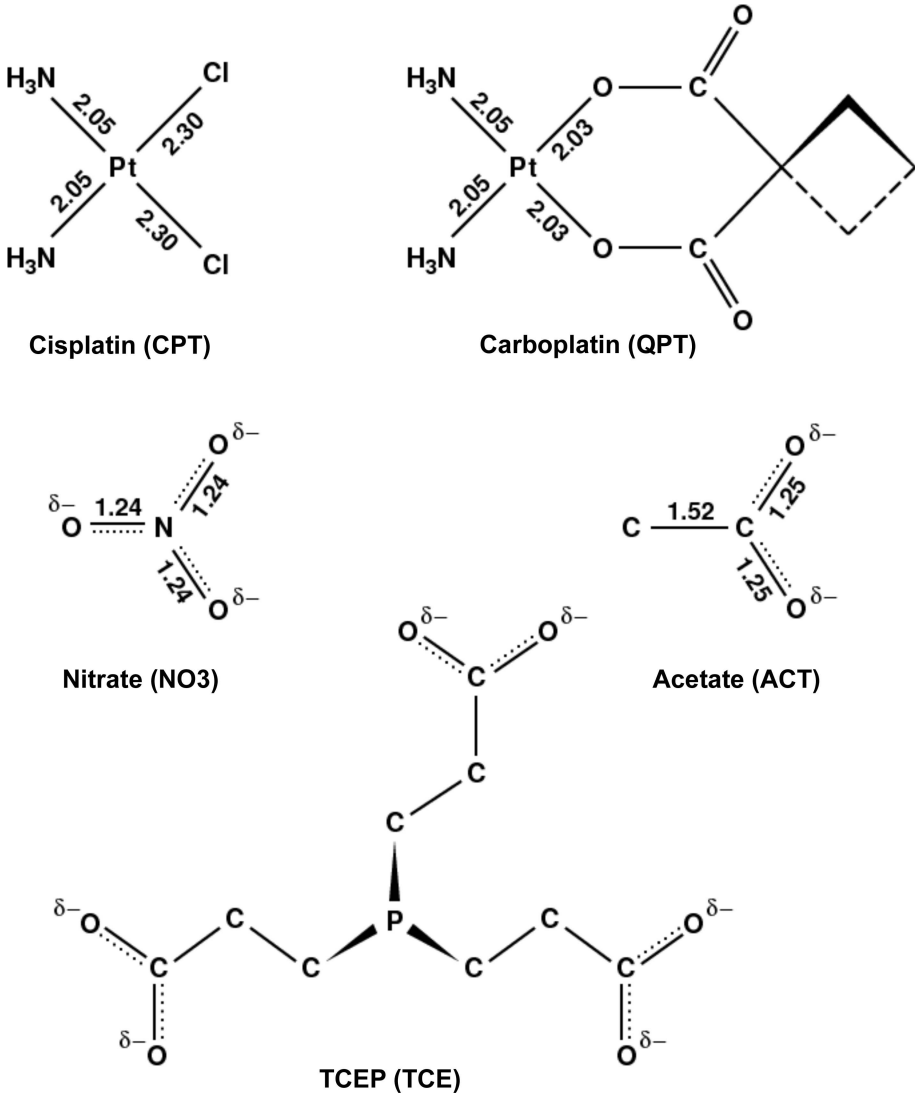


Figure 1
Chemical formulae of cisplatin, carboplatin, nitrate, acetate and tris(2-carboxyethyl)phosphine (TCEP). Selected averaged bond lengths (Å) based on CSD analysis are shown.

and Cys15 from two molecules), with roughly tetrahedral coordination (Pt–S^{γ} distances ranging from 2.10 to 2.48 Å). The two N atoms are placed 1.80–1.99 Å from the Pt at positions extending the Cys15 S^{γ}—Pt bonds, thus providing a hexacoordinated Pt complex that has never been observed in any experimentally determined structures involving platinum(II). Whereas the B factors for the S^{γ} atoms of the four cysteine residues are in the range 19–25 Å^2, the B factor of the partially occupied Pt is 37 Å^2 and those for the ammine N atoms are 63–64 Å^2. Additionally, these ammine groups have unacceptable clashes with the cysteine residues, with each group having three interatomic distances of 2.1–2.3 Å. In the structure of monomeric ATOX1 (PDB entry 3iwl; Boal & Rosenzweig, 2009), only a Pt atom is modeled between the sulfurs of Cys12 and Cys15, also making contacts with the amide N atom of Cys12 (2.27 Å) and the P atom (2.48 Å) of a (2-carboxyethyl)phosphine (TCEP) molecule from the crystallization buffer. Very significant $mF_o - DF_c$ density, both positive and negative, is present in the vicinity of the Pt atom, and the electron density for most of the TCEP molecule is either weak or negative (Fig. 3*a*). In this case, the Pt atom is not designated as part of CPT; thus, this particular structure is not listed in Table 1.

3.4. Binding of cisplatin to ribonuclease A

A complicated arrangement of three partially occupied CPT molecules was modeled in the complex with RNase A (PDB entry 4ot4; Messori & Merlino, 2014). Two molecules were modeled bound to the S^{δ} atom of Met29 of molecule *A*. One, with 0.75 occupancy, has a square-planar configuration (Pt–S^{δ} distance of 2.38 Å) with two *cis* N atoms and a water molecule at 2.32 Å. The second CPT consists of only a Pt atom and one N atom at 0.25 occupancy. Some positive and negative difference electron density is still present in the vicinity of the putative cisplatin molecules. A CPT molecule bound to S^{δ} of Met29 of molecule *B* (occupancy 0.55) also interacts with $N^{\varepsilon 2}$ of Gln28 (Pt–$N^{\varepsilon 2}$ distance of 2.62 Å), but the electron density for the side chain of Gln28 does not fully support its assumed orientation and instead indicates a turn, which would make the Pt–$N^{\varepsilon 2}$ distance, which is highly unlikely to represent a coordination bond, considerably longer.

3.5. Binding of cisplatin to glutaredoxin 1

Three crystallographic models of complexes of cisplatin with PfGrx1 are available in the PDB (PDB entries 4n0z, 4n10 and 4n11; M. Yogavel, T. Tripathi, S. Rahlfs, K. Becker & A. Sharma, unpublished work). Strangely, although the three structures are isomorphous, the protein molecules are placed very differently in the unit cell, making direct comparisons unnecessarily complicated (Kowiel *et al.*, 2014). In the structure 4n0z, a Pt site with 0.26 occupancy is placed 2.64 Å from $N^{\delta 1}$ of His49, with a Cl and an N atom, as well as a water molecule, completing a severely distorted square-planar arrangement. The $N^{\varepsilon 2}$ atom of His49 is not available for CPT binding since it makes an excellent salt bridge with OE1 of Glu51. Very significant residual electron density near the Cl atom indicates that the model is incomplete, although other small molecules present in this structure, such as MES, are in excellent density. Two alternative CPT molecules were found in the corresponding location in the structure 4n10, but very large positive and negative residual electron density indicates that the model is rather unsatisfactory. Additionally, the presence of very significant positive electron density next to $C^{\varepsilon 1}$ of His49 indicates that the interpretation of the ligand binding to this side chain is incomplete. A third CPT moiety, consisting of Pt and Cl atoms at 0.4 occupancy, was found next to S^{δ} of Met59 and within 3.08 Å of the $O^{\varepsilon 1}$ atom of Gln63. Both positive and negative residual electron density cover the CPT molecule. A cisplatin molecule bound to S^{γ} of Cys29 in the structure 4n11 was modeled as containing a single Pt atom at 0.4 occupancy, with two different, superimposed constellations of coordinating groups, including the rare but sometimes observed (Kane & Lippard, 1996) ω-amino group of Lys26. The Pt atom is located in very strong negative difference electron density (*ca.* -30σ), whereas some positive difference density is also present.

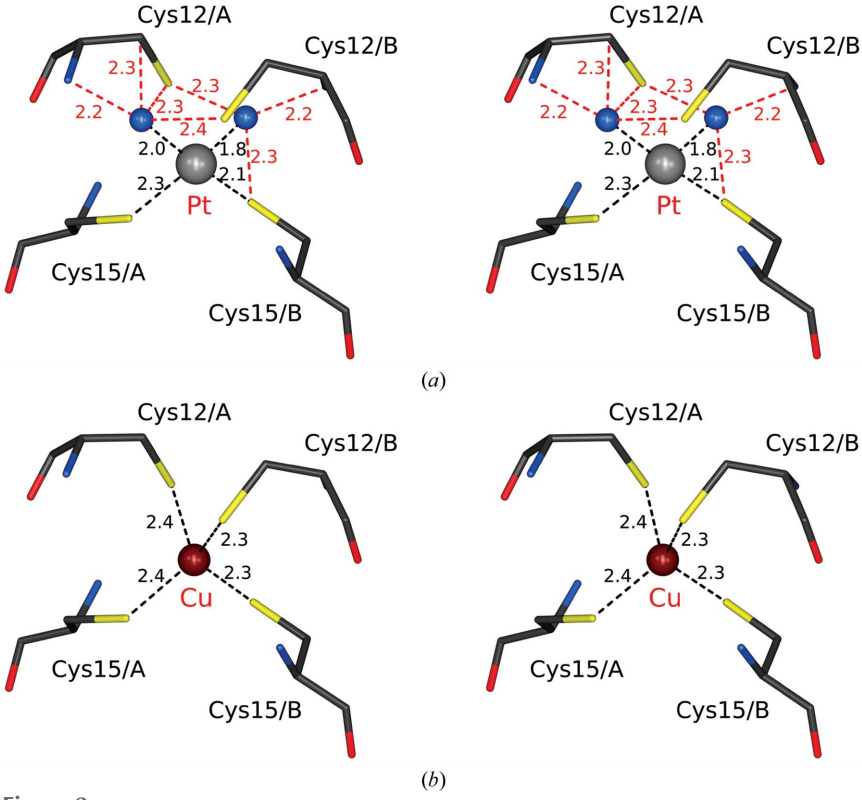


Figure 2
Stereoview of metal binding at the dimer interface of the ATOX1 dimer. (*a*) The original PDB deposition 3iwx. The ammine groups of cisplatin are shown as blue spheres and the S^{γ} atoms of cysteine residues are shown as yellow cylinders. Platinum–ligand interactions within the originally proposed cisplatin adduct with ATOX1 are depicted as black dashed lines with distances marked in Å. Steric clashes between the ammine groups and protein atoms are highlighted as red dashed lines with distances in Å. (*b*) After re-refinement. A copper ion and the S^{γ} atoms of the cysteine residues form a typical tetrahedral complex with close to ideal distances.

3.6. Binding of cisplatin and carboplatin to hen egg-white lysozyme

All other available structures of protein–cisplatin complexes, as well as of carboplatin complexes, involve different crystal forms of hen egg-white lysozyme. Four structures of triclinic HEWL are at atomic or near-atomic resolution (0.98–1.42 Å), whereas a much larger number of tetragonal structures are in the resolution range 1.6–3.0 Å. A single structure of monoclinic lysozyme with bound carboplatin was determined at 1.8 Å resolution.

Although the structure with PDB code 4mwk, determined at 150 K, was refined with data extending to better than 1 Å resolution and the resulting *R* factors (R = 0.119, $R_{\rm free}$ = 0.145) are quite acceptable, the electron density for the CPT ligands is far from clear, and its interpretation leaves a number of unanswered questions. In addition, the r.m.s. deviation from accepted bond

distances of 0.039 Å is too high even for this resolution, although it may mostly reflect the deviations from ideality of the CPT groups and not just the protein. Another structure, PDB entry 4mwm, with data collected at 200 K, exhibits much smaller distortions of bond lengths, but the electron density in the vicinity of His15 is equally ambiguous. Unlike in PDB entry 4mwk, a DMSO molecule was modeled in the immediate vicinity of the Pt207 atom (with an occupancy of 0.44 and a completely unrealistic B value of 169.32 Å^2) located close to N$^{\varepsilon2}$ of His15. A single Pt atom coordinated by two Cl atoms (occupancy of 0.29) was modeled next to N$^{\delta1}$ of His15, but very strong positive residual density indicates that the models are, at best, incomplete. Finally, a slightly lower resolution (1.42 Å) room-temperature crystallographic model (PDB entry 4mwn) contains a Pt atom with 0.29 occupancy close to N$^{\delta1}$ of His15, with a single Cl atom 2.28 Å away but with an occupancy of 0.75. Two Pt atoms, with occupancies of 0.30 and 0.36, share a patch of electron density disconnected from N$^{\varepsilon2}$ of His15 and are not accompanied by any other atoms in their coordination spheres. Significant negative residual electron density covers both Pt atoms.

The structure with PDB code 4oxe represents a complex with carboplatin determined at a resolution of 1.13 Å but with refinement parameters (R = 0.180, $R_{\rm free}$ = 0.221) that are rather high compared with other structures at similar resolution. The Pt atom closest to N$^{\varepsilon2}$ of His15 is 2.58 Å away and at an angle that significantly differs from the configuration expected for platinum(II) ions. Another Pt atom is located further away from N$^{\varepsilon2}$ (3.12 Å). A single Pt (occupancy 0.2) is located 2.45 Å from N$^{\delta1}$ of His15. Although some positive density is present nearby, this Pt atom is not accompanied by other ligands. Some parts of the model distal to the putative carboplatin binding site may also be suboptimal, for example Na$^+$ ions coordinated in a manner more appropriate for water molecules, DMSO molecules outside of electron density and a (gaseous in normal conditions!) MEB (2-methylpropene) molecule not covered by electron density.

As shown by Gust & Schnurr (1999), carboplatin is not stable in solution, especially in the presence of Cl$^-$ ions. Since the standard crystallization conditions for tetragonal lysozyme include as much as 1 M NaCl, it could be expected that a significant fraction of carboplatin would be converted to cisplatin or aqua complexes during the crystallization experiments. However, proving such a conversion solely on the basis of electron-density maps corresponding to partial ligand occupancy may not be very convincing. To address this question, Tanley *et al.* (2014) determined a number of structures of tetragonal lysozyme crystallized in the presence of the much heavier NaBr or under conditions from which all halides were excluded. One data set (PDB entry 4nsf) was collected at a wavelength of 0.9163 Å in order to maximize the anomalous signal of Br, and an equivalent data set (PDB entry 4nsg) was also collected at a wavelength of 1.5418 Å. Five additional data sets were collected at the latter wavelength using crystals grown under conditions that excluded halides from the crystallization medium.

Anomalous diffraction data were collected for all seven crystals, with the resulting maps shown in Figs. 2 and 4 of Tanley *et al.* (2014). Unfortunately, the structure factors deposited together with PDB models 4lt0 and 4nsi do not contain anomalous data, making independent recalculation of the anomalous difference maps impossible. As expected, anomalous signal for Pt is present in the maps calculated for the remaining five data sets (although very weak for PDB entry 4lt3), but only that for PDB entry 4nsf has additional peaks, presumably originating from the scattering of Br. Two peaks (at 12 and 16σ) are located within ∼2.5 Å of the Pt bound to N$^{\delta1}$ of His15, whereas a 10σ peak was interpreted as a Br$^-$ ion bound to a putative Pt located 2.63 Å from the N$^{\varepsilon2}$ atom. The occupancy of this Pt atom was refined to only 0.13 (in a 7σ anomalous difference density peak), whereas the density for Br is at 10σ, with an occupancy refined as 0.65. By comparison, the

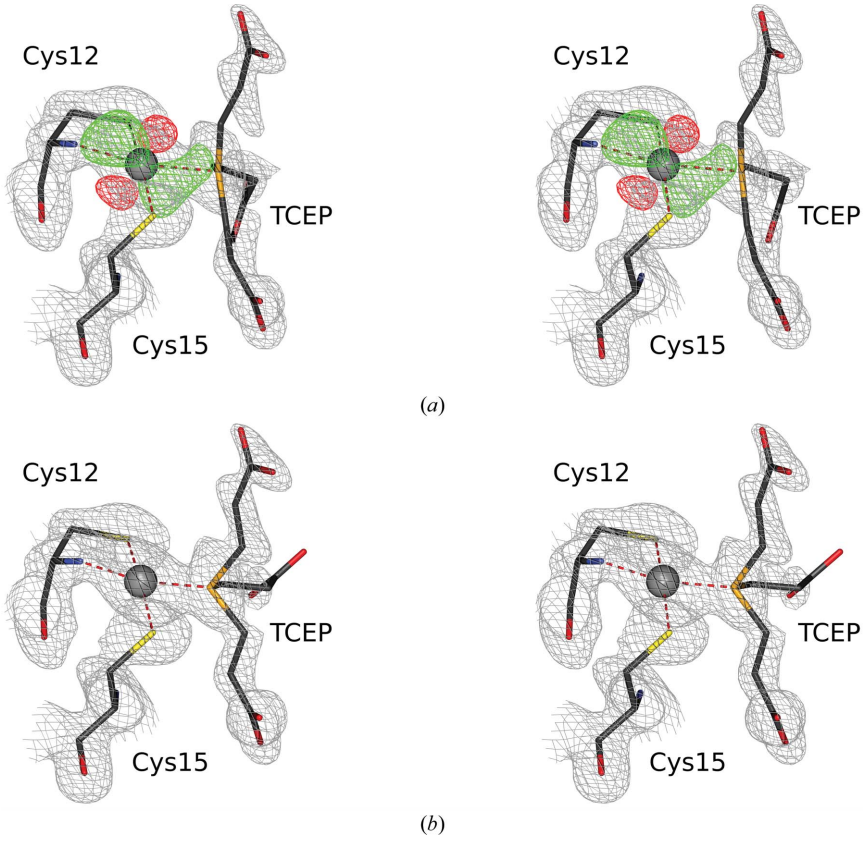


Figure 3
Stereoview showing binding of platinum(II) and TCEP to the ATOX1 monomer. $2mF_o - DF_c$ maps are displayed in gray contoured at a level of 1.2σ. $mF_o - DF_c$ difference maps are contoured at the 4σ level in green (positive) and red (negative). (*a*) The original PDB deposition 3iwl. (*b*) After re-refinement.

anomalous difference density of some peaks interpreted as Br^- ions not bound to Pt is as high as 21σ. Analysis of the $2mF_o - DF_c$ map indicates clear electron density at a position *trans* to $N^{\delta1}$ of His15 that is not interpreted by any atom included in the PDB file. Incidentally, the originally deposited file 4nsf appears to be an earlier version than the file described in Tanley *et al.* (2014), since it is lacking the anisotropic ANISOU records and the *R* factors are much higher than those reported in the publication. The subsequently deposited data set 4xan is much closer (but not identical) to the description in the publication, but does contain several additional atoms *trans* to $N^{\delta1}$ of His15 representing a putative fragment of carboplatin. However, some of these atoms are involved in serious clashes (for example, O1 is not in the square-planar configuration and is located only 1.9 Å from Br202). The $2mF_o - DF_c$ electron density for the data set 4nsg is very similar to that for 4nsf; in particular, electron density for the atom(s) *trans* to $N^{\varepsilon2}$ of His15 is present although not modeled, and the peak for the putative Pt site close to $N^{\delta1}$ is much weaker than for the adjacent Br^- ion.

Some of the structures of carboplatin complexes determined earlier by the same team (Tanley, Schreurs, Kroon-Batenburg & Helliwell, 2012*a*; Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012; Tanley, Schreurs *et al.*, 2013) provide a different interpretation of the His15 ligand. For example, in the structure with PDB code 4dd7 (Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012) the vicinity of $N^{\delta1}$ is modeled as a fully occupied Pt coordinated by two *cis* N atoms and a single water molecule. However, the N atom *cis* to $N^{\delta1}$ is still covered by positive $mF_o - DF_c$ electron density at 5.5σ, whereas the Pt atom is in a negative (-10σ) peak. The Pt atom bound to $N^{\varepsilon2}$ and its surrounding ligands are also fully occupied, although it coincides with a strong negative electron-density peak of -14σ. This carboplatin fragment is modeled with two *cis* N atoms, as well as with another N atom, $N^{\eta1}$ of Arg14 (2.35 Å from Pt). However, a somewhat different model is shown in the structure with PDB code 3txh (Tanley, Diederichs *et al.*, 2013) based on the same diffraction images but processed with different software. The positions of the N atom and water *cis* to $N^{\delta1}$ have been swapped, much more positive difference density is seen in the vicinity of the Pt ligands and the Pt is located in a -12σ $mF_o - DF_c$ electron-density peak. The environment of the Pt atom next to $N^{\varepsilon2}$ is similar to its counterpart in 4dd7, with the exception of an additional DMSO molecule, the O atom of which forms a Pt—O bond of 2.63 Å that is perpendicular to the plane of the other ligands. However, there is no electron density to support the presence of a DMSO molecule in this location.

The model with PDB code 4dd9 (Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012) includes a water molecule and two N atoms around the Pt site adjacent to $N^{\delta1}$, whereas two ammines and the $N^{\eta1}$ of Arg14 surround the Pt site next to $N^{\varepsilon2}$. Both Pt atoms are covered by large negative difference electron density, and positive density is seen close to the ligands. In PDB entry 3txi, which was based on the same diffraction data processed with a different program, the model is similar, with the exception of an additional N atom modeled along a perpendicular direction to the square-planar ligand plane adjacent to $N^{\varepsilon2}$. This N atom, however, is not supported by electron density.

The problems found in the examples listed above are representative of the other reported crystallographic models of the complexes of carboplatin and cisplatin.

3.7. The presence of mystery molecules in the deposited coordinate sets

Several of the coordinate data sets analyzed in this work also contain small molecules other than cisplatin or carboplatin. Some of those molecules are components of the crystallization buffers, such as acetate, nitrate or dimethyl sulfoxide (DMSO). Whereas their presence might in principle be justified, their identification is not always based on unambiguous fit to the electron density. Some other molecules were also modeled, although the reasons for their inclusion were not clearly stated. In particular, the molecule with the PDB designation MEB (2-methylpropene; isobutylene) was included in seven depositions, some of which are listed in Tables 1 and 2, and others related to these entries (PDB entries 4oxe, 4owb, 4mwk, 4mwm, 3txe, 4dd1 and 4dd3). This extremely hydrophobic gaseous molecule consisting of only four C atoms and eight H atoms was apparently derived from Paratone used for cryopreservation (Tanley, Schreurs, Kroon-Batenburg, Meredith *et al.*, 2012), although this polymer has a molecular weight of ~20 000 Da. In any case, it is unlikely that MEB would be soluble in water under any protein crystallization conditions. In addition, the molecule should be planar and have C—C bonds of different lengths (single and double), but, for example in the data set 3tx3, it is modeled with tetrahedral geometry and with all C—C bonds equal. None of the MEB molecules were placed in any significant electron density; thus, their presence is highly dubious.

Some other ligands found in the crystallographic models discussed here

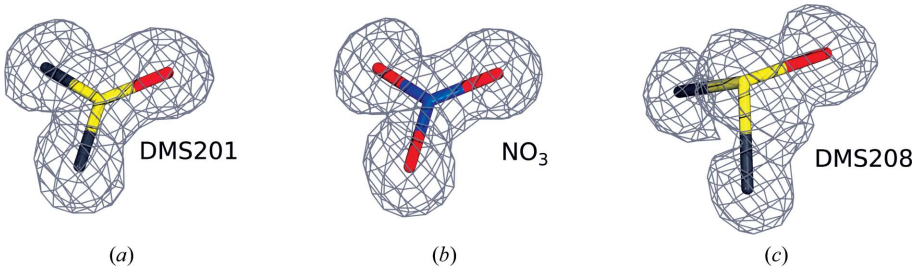


Figure 4
Misinterpretation of a nitrate anion as a DMSO molecule in PDB deposition 4mwk. $2mF_o - DF_c$ maps are displayed in gray contoured at 1.0σ. S, N, O and C atoms are shown in yellow, blue, red and black, respectively. (*a*) A DMSO molecule in the original deposition. (*b*) Reinterpretation as a nitrate anion (this work). (*c*) An example of a properly modeled DMSO molecule in the original deposition 4mwk.

appear to be identified incorrectly, even if they are within visible electron density. For example, DMSO molecules, in which the S atom is expected to be pyramidal, are often planar, violating the stereochemistry and only then fitting the electron density. A good example is provided by DMS207 in the atomic resolution structure with PDB code 4mwk. This molecule, with an occupancy assigned as 0.8, is located in a completely planar patch of electron density. The atomic displacement parameter for S is 14.98 Å^2, whereas the *B* factors for the other three atoms in this molecule vary between 5 and 6 Å^2. An almost identical situation is found in the case of DMS201 (Fig. 4). It is very likely that DMSO was misidentified in these cases, and either a nitrate or acetate ion should occupy each of these positions. On the other hand, the *B* factor for the N atom in the putative nitrate molecule NO3/210 is 8.55 Å^2, whereas the *B* factors of the O atoms range between 28 and 40 Å^2, indicating that this is most likely to be a water molecule. A more difficult case is exemplified by NO3/211, in which the *B* factors of all atoms are similar (11–13 Å^2) but one of the O atoms is located 2.79 Å from the CD atom of Arg114.

Table 3
Selected parameters of the originally refined (first column) and re-refined (second column) ATOX1 structures.

Values in parentheses are for the last resolution shell.

	ATOX1 monomer		ATOX1 dimer	
PDB code	3iwl	4ydx	3iwx	4yea
Resolution (Å)	46.83–1.60 (1.64–1.60)	27.03–1.60 (1.64–1.60)	28.76–2.14 (2.20–2.14)	25.63–2.14 (2.20–2.14)
No. of reflections (refinement/R_{free})	12060/621	12057/621	10032/504	10030/503
Completeness (%)	99.60 (97.30)	99.54 (97.51)	99.90 (100)	99.82 (99.74)
R/R_{free}†	0.179/0.210	0.136/0.156	0.186/0.228	0.160/0.197
No. of atoms				
Protein	496	517	1011	1017
Heterogens	22	22	18	16
Water	91	96	105	112
Ramachandran plot statistics (%)				
Favored	98.44	96.83	99.23	100.0
Outliers	0	0	0	0
R.m.s. deviations from target values				
Bond lengths (Å)	0.010	0.013	0.009	0.012
Bond angles (°)	1.25	1.52	1.09	1.45
LLDF > 2 [> 10]	1 [0]	1 [0]	1 [0]	0 [0]

† $R = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} was defined by Brünger (1992).

3.8. Erroneous description of ligands in the refinement libraries and in the PDB

It has been known for more than a hundred years (Werner, 1893) that the ligands in cisplatin are arranged in a square-planar configuration, with the two chlorines and the two ammines *cis* to each other. However, the restraints for the cisplatin (CPT) group found in the CCP4 ligand library treat this molecule as tetrahedral. Although in this case there would be no distinction between the *cis* and *trans* configurations, this is no justification for showing (in the PDB) and using the chemical formula of **cis**platin with the wrong stereochemistry!

The standard group found in the PDB for the acetate ion (ACT; Fig. 1) sets all valence angles to 120°, whereas the proper O—C—O angle should be close to 128°. The nitrate ion (NO3; Fig. 1) is defined with two single bonds and one double bond, while all three bonds should be the same owing to resonance. The standard description of the TCEP buffer molecule forces the P atom to be coplanar with its substituents (C1, C2 and C3) instead of being pyramidal (TCEP; Fig. 1). These examples emphasize the fact, noted previously (Jaskolski, 2013), that the definitions and nomenclature of standard ligand groups found in the restraint libraries are often nonstandard and even erroneous. Therefore, stereochemical restraints derived from such data should be treated with extreme caution, if not with suspicion. Some servers, such as *PRODRG* (http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg), take electron resonance into account and give acceptable description of many ligands, although they still produce incorrect results in some situations.

3.9. Problems with the calculation of electron-density maps

The Uppsala Electron Density Server (EDS; Kleywegt *et al.*, 2004), accessed directly from the graphics display program *Coot* (Emsley & Cowtan, 2004), is an indispensable tool that allows even nonspecialists to visualize electron-density maps. However, we detected serious problems with some of the EDS maps. For example, some of the atomic resolution maps, such as that for data set 4mwk, had completely spurious residual difference density covering well determined amino-acid residues, including both the main chain and side chains, that was clearly an artifact. The download of some other maps from the EDS failed or was associated with nonreliability warnings. Such warnings may arise from various errors in the deposited structure factors. An example is data set 4nsi, for which one finds a serious discrepancy between the claimed data resolution of 2.3 Å and the resolution limit of 1.209 Å noted in the cif file, with most reflections marked as not measured.

3.10. Re-refinement of the structures of the copper-transport protein ATOX1

The structure 3iwx, with cisplatin bound by an ATOX1 dimer, as well as 3iwl, in which the monomeric form of ATOX1 was interpreted as binding just a single Pt atom with a TCEP molecule serving as another ligand, have no immediately obvious technical problems other than some ligand bonds that significantly differ from the ideal values (3iwl) and a number of protein–ligand clashes (3iwl and 3iwx). However, we were troubled by some puzzling aspects of their interpretation in terms of chemistry, and thus these structures have been re-refined and reinterpreted. Structure amplitudes and

starting models were taken from the relevant PDB depositions and the resolution limits were the same as in the original files.

3.10.1. Structure of the ATOX1 monomer in complex with cisplatin (PDB entry 3iwl). Our re-refinement of the structure 3iwl resulted in a decrease in R_{free} of over 5%, partly owing to the employment of TLS parametrization, which was not used in the original refinement (Table 3). Notably, ∼1% of the improvement in R_{free} came from anisotropic refinement of the platinum ion and the P atom of the TCEP molecule. Even if the resolution of the data (1.6 Å) is too low for full anisotropic refinement, such a procedure can still be applied to atoms with well defined electron density, especially to heavy scatterers such as Pt. The necessity of anisotropic refinement was revealed by a characteristic pattern of positive and negative difference peaks around Pt (Fig. 3*a*) which disappeared after anisotropic refinement (Fig. 3*b*). In addition, anisotropic refinement of these atoms allowed the assignment of almost full occupancy to the Pt atom (0.9 instead of the original 0.75), bringing it into better agreement with the 1:1 protein:Pt ratio indicated by ICP-AES and ESI-MS data (Boal & Rosenzweig, 2009). Another important correction was applied to the geometry of the TCEP ligand. The restraints for the TCEP C—P—C bond angles in the *REFMAC* dictionary are set to 120°, forcing the P atom to be coplanar with its substituents (C1, C2 and C3) instead of being pyramidal (Fig. 1). As a result, the Pt–P distance (2.5 Å) was longer than expected (2.3 Å) and the C atoms C1, C2 and C3 of the TCEP ligand were forced out of electron density. We corrected the library entry for TCEP by assigning pyramidal coordination to the P atom. As a consequence, the new model of TCEP fits the electron density much better (Fig. 3) and the Pt–P distance is 2.3 Å. In addition, we set the occupancies of Pt and TCEP to the same value of 0.9, because it is unlikely that TCEP could be bound without Pt present. It is worth emphasizing that in this structure, which is meant to represent a protein complex of cisplatin, all of the original ligands of platinum in cisplatin have been replaced by protein ligands and TCEP from the buffer solution. Therefore, it was not cisplatin itself, but rather its derivatives, that interacted with the protein and then crystallized. It is by far not the first example of interactions between cisplatin and the buffer components used in the preparatory protocol. For example, it was shown by Appleton *et al.* (1984) that when acetate or phosphate buffers are used, the corresponding acetate/phosphate complexes of platinum are present in significant quantities in any solution. In conclusion, despite some flaws in the structure refinement, the overall interpretation of Pt binding in the structure 3iwl was not substantially changed by our re-refinement.

3.10.2. Structure of the ATOX1 dimer in complex with cisplatin (PDB entry 3iwx). Despite quite good overall quality indicators and acceptable agreement between the crystallographic model and the experimental data (Table 3), the presence of a Pt atom at the dimer interface with tetrahedral coordination by four S atoms (in a rather copper-characteristic binding site) in the 3iwx structure is rather suspicious. In addition, the originally deposited model had two ammine groups squeezed between cysteine residues and the platinum ion (Fig. 2*a*). Boal & Rosenzweig (2009) mention two persistent positive peaks in their difference Fourier map, together with ESI-MS data, as a justification for modeling these ammine groups. However, these peaks almost disappeared in our re-refinement after anisotropic modeling of the metal (the highest residual $mF_o - DF_c$ peak is at 4.8σ). Without the ammine groups, the metal atom has tetrahedral coordination provided by four S atoms (Fig. 2*b*). Our search of the CSD found no structures with Pt in such a tetrahedral coordination. We must therefore conclude that this option is highly unlikely. In addition, the occupancy of the platinum ion, which at best can be raised to 0.5, means that in the actual crystal the dimer molecules without platinum will have to accommodate some other metal, because such a close arrangement of four cysteine side chains without a coordinated metal ion would be rather unstable. On the other hand, the ICP-AES data of Boal & Rosenzweig (2009) suggest that there are 0.63 ± 0.09 Pt ions per ATOX1 monomer, meaning that there should be more than one Pt ion per protein dimer. This observation is also supported by ESI-MS data, which show an ATOX1 dimer with two Pt ions. Therefore, it is entirely possible that Pt was bound not in the copper-binding site but by some other residues. A review of cisplatin binding to human albumin (Ivanov *et al.*, 1998) suggests that cisplatin can be bound to methionine, histidine or cysteine residues or by the N-terminus. The ATOX1 dimer has Cys41, His46 and Met48 exposed on the surface of the protein that could potentially bind cisplatin. In conclusion, the postulated presence of platinum in the copper-binding site is very unlikely from the crystallographic/chemical point of view, and the experiments presented in the paper do not necessarily support this interpretation.

In our re-refined crystallographic model we interpreted the anomalous peak at the dimer interface simply as a copper(II) ion. In view of the above analysis it is the most straightforward interpretation of the crystallographic data, and is supported by the following observations: (i) copper could be refined with 100% occupancy and has *B* factors similar to the surrounding atoms; (ii) the Cu–S^{γ} distances are within the reasonable range of 2.3–2.4 Å; (iii) the coordination of the copper ion is very similar to the situation in the structure with PDB code 1fee, identified as a copper complex of ATOX1; (iv) ATOX1 is a copper-binding protein, thus it could retain copper during purification and dialysis. However, we must stress that our interpretation is only hypothetical as it is based solely on structural evidence and chemical common sense. In our opinion, if the authors of the original paper were absolutely certain about placing a Pt ion at this site, they should have buttressed their interpretation with a deeper discussion of this very unusual proposition.

3.11. Re-refinement of selected HEWL structures containing cisplatin or carboplatin

We have also re-refined several representative structures of HEWL listed in Tables 1 and 2. Unless noted below, structure amplitudes and starting models were taken from the relevant PDB depositions and the resolution limits, descriptions of the

Table 4
Selected parameters of the originally refined (4nsf, 4xan, 4mwk) and re-refined (4yem, 4yeo) lysozyme structures.

Values in parentheses are for the last resolution shell.

	HEWL–carboplatin complex			HEWL–cisplatin complex	
PDB code	4nsf	4xan	4yem	4mwk	4yeo
Resolution (Å)	55.56–1.47 (1.51–1.47)	55.56–1.47 (1.51–1.47)	39.29–1.47 (1.51–1.47)	29.28–0.98 (1.005–0.980)	29.26–0.98 (1.005–0.980)
No. of reflections (refinement/R_{free})	17672/933	17672/933	17672/933	48959/2646	48959/2646
Completeness (%)	91.0 (64.11)	91.0 (64.11)	91.0 (64.11)	90.56 (49.71)	90.56 (49.71)
Effective resolution† (Å)	1.52	1.52	1.52	1.01	1.01
R/R_{free}‡	0.179/0.214	0.126/0.181§, 0.163/0.202¶	0.105/0.156	0.119/0.145§, 0.144/0.165¶	0.113/0.131
No. of atoms					
Protein	1001	1001	1064	992	1045
Heterogens	35	57	36	92	56
Water	90	90	182	142	233
Ramachandran plot statistics (%)					
Favored/outliers	98.43	97.64	97.46	97.54	97.46
Outliers	0	0	0	0	0
R.m.s. deviations from target values					
Bond lengths (Å)	0.026	0.022	0.013	0.039	0.011
Bond angles (°)	1.94	2.59	1.49	2.91	1.71
LLDF > 2 [> 10]††	8 [2]	14 [4]	9 [1]	20 [7]	7 [1]

† As calculated by the formula $d_{ef} = d_{min}/C^{1/3}$, where d_{ef} is the effective resolution, d_{min} is the high-resolution limit of the data and C is the overall completeness of the data set (Weiss, 2001). ‡ $R = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively. R_{free} was defined by Brünger (1992). § As reported in the file deposited in the PDB. ¶ From the PDB validation report. †† LLDF > 2, as recommended by the PDB, seems to be too strict, as a very small change in occupancy changes the LLDF. An LLDF higher than 10 usually indicates problems that should be fixed.

B factors and reflections used to define R_{free} were the same as in the original files.

3.11.1. Atomic resolution structure of a cisplatin complex of triclinic lysozyme (PDB entry 4mwk). This structure of a cisplatin complex of triclinic lysozyme (PDB entry 4mwk; Tanley & Helliwell, 2014*b*) was selected as a candidate for re-refinement since its resolution was the highest among all of the structures discussed here and the quality of the diffraction data was reported to be very good [R_{merge} = 0.045, $\langle I/\sigma(I)\rangle$ = 15.6, redundancy of 3.5]. However, the reported resolution was questionable because the completeness in the highest resolution shell was only 49.7%. In addition, the criterion for treating reflections as 'observed' was reported as an $I/\sigma(I)$ of 2.0, which is a dubious practice if it were indeed used in data processing. Nevertheless, for consistency, we have re-refined the structure at the originally reported resolution of 0.98 Å.

The structure was re-refined using a full anisotropic ADP model. We were able to bring the overall geometry parameters to reasonable values, included a reasonable number of water molecules and significantly improved the R_{free} (Table 4). The coordinates were moved to within the standard unit cell, rather than outside (Kowiel *et al.*, 2014), as was the case in the original model. The original crystallographic model had an extremely high clashscore (29 according to the PDB validation report), mostly as a result of incorrectly identified and wrongly modeled ligand molecules. For example, the original model contained five MEB molecules, none of which were supported by electron density, and three of which were involved in intermolecular distances as short as 1.7 Å. We replaced three of those phony molecules with water molecules and two of them with ethylene glycol and a nitrate ion. Three DMSO molecules in the original model (DMS201, DMS205 and DMS207) had trigonal planar geometry. These molecules were replaced by nitrate ions, which were indeed present in the crystallization mixture at a concentration of 0.5 *M* and fitted the electron density perfectly (Fig. 4). However, several nitrate ions (NO3/210, NO3/213, NO3/214 and NO3/215) in the 4mwk structure had no electron density to support them and were therefore replaced by water molecules. In agreement with the observation that more dual-conformation residues become apparent with increased resolution (Addlagatta *et al.*, 2001), ten protein residues were modeled with alternate conformations, including the crucial His15, as opposed to the original model, in which only two residues had alternate positions. After re-refinement, the clashscore decreased to 3.

The presence of cisplatin (or another Pt complex, possibly formed in the crystallization cocktail) binding to His15 of lysozyme is the main focus of the paper that reported the structure 4mwk (Tanley & Helliwell, 2014*b*). Unfortunately, interpretation of the electron density of this ligand is extremely difficult owing to structural disorder. His15 adopts two alternative conformations, and since cisplatin can be bound by the $N^{\delta1}$ and $N^{\varepsilon2}$ atoms of this residue, these alternative conformations result in at least three Pt-atom positions that are clearly visible in the anomalous map. Given the wavelength used for data collection (1.5418 Å), the composition of the crystallization cocktail, and the anomalous and electron-density maps, one can confidently assign these peaks to Pt atoms, which can be refined with an occupancy of 0.2–0.3. However, the overlapping of these modes of binding of cisplatin, together with other molecules possibly bound in the absence of cisplatin, results in very blurred electron density, making it virtually impossible to identify and locate all of the inner-sphere Pt ligands and to properly orient the cisplatin molecules with any degree of confidence. We placed one cisplatin molecule attached to the $N^{\varepsilon2}$ atom of the major conformation of His15. Its presence can be justified by the

Table 5
Selected parameters of the originally processed and reprocessed data as well as the originally refined and re-refined structure of tetragonal lysozyme with the original PDB code 4g4a (Tanley *et al.*, 2014).

Values in parentheses are for the last resolution shell. The data reprocessed at 2.4 Å resolution were only used for preliminary refinement and the final two columns are included solely to provide a more direct comparison with the statistics reported in the original deposition. Data beyond frame 3037 were very strongly affected by radiation damage and could not be processed in an acceptable manner.

Data processing	Original	Reprocessed		
PDB code	4g4a	4yen	—	—
Frame numbers	1–3366	1–347	1–347	2708–3037
Resolution (Å)	55.89–2.40 (2.44–2.40)	34.24–2.00 (2.03–2.00)	34.24–2.40 (2.44–2.40)	34.24–2.40 (2.44–2.40)
R_{merge} (%)	18.0 (94.7)	8.6 (57.1)	6.3 (24.6)	10.5 (>100)
No. of reflections (measured/unique; Bijvoet separate)	360972/9038	68208/15602	46742/9043	46516/9074
$\langle I/\sigma(I)\rangle$	23.6 (6.7)	19.4 (2.7)	30.4 (7.3)	12.4 (0.6)
Completeness (%)	99.9 (100)	100 (100)	100 (100)	99.9 (98.3)
Multiplicity	39.9 (30.5)	7.9 (5.6)	9.2 (7.1)	9.1 (5.0)
Refinement				
No. of reflections (refinement/R_{free})	4818/234	8107/450	4784/268	4565/239
R/R_{free}†	0.163/0.210	0.146/0.179	0.126/0.177	0.155/0.197
No. of atoms				
Protein	1001	1000	1000	1000
Heterogens	16	11	11	11
Water	23	75	75	75
Ramachandran plot statistics (%)				
Favored	96.06	97.62	97.62	97.62
Outliers	0	0	0	0
R.m.s. deviations from target values				
Bond lengths (Å)	0.014	0.013	0.018	0.022
Bond angles (°)	2.40	1.58	1.71	1.90
LLDF > 2 [> 10]	1 [1]	1 [0]		

† $R = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where F_{obs} and F_{calc} are the observed and calculated structure factors, respectively, calculated for all data. R_{free} was defined by Brünger (1992).

correct distance between Pt and $N^{\varepsilon 2}$ (2.0 Å) and by two electron-density peaks next to Pt which fit square-planar geometry, are located at a proper Pt—Cl distance (2.3 Å) and have reasonable electron density to accommodate Cl at the same occupancy as the Pt atom. We modeled two other anomalous peaks just as single Pt atoms because there are no clues as to how the putative cisplatin molecules could be oriented. In our opinion, one should not form any definitive conclusions about cisplatin binding based on this structure, or at least should be very careful about any statements except that there are indeed multiple anomalous density peaks indicating multiple modes of platinum binding. Thus, the five cisplatin molecules (or rather fragments of these molecules) reported in the original 4mwk deposition with significant violations of square-planar geometry and missing inner-sphere ligands may not represent a credible illustration of the mode of interaction of this compound with proteins.

3.11.2. High-resolution structure of tetragonal lysozyme with carboplatin (PDB entries 4nsf and 4xan). These two sets of coordinates resulted from refinement of the structure of tetragonal lysozyme crystals grown in the presence of 1 *M* NaBr as well as sodium acetate, DMSO and carboplatin. Whereas it was reported that the structure had been refined anisotropically using 1.47 Å resolution data collected at a wavelength of 0.9163 Å (Tanley *et al.*, 2014), the original PDB deposition (PDB entry 4nsf) included only isotropic displacement parameters, and the *R* factors in the PDB file were much higher than the published ones. These coordinates were later replaced in the PDB by data set 4xan, in which anisotropic *B* factors are present and the *R* factors are similar to those quoted in the original publication, suggesting that 4nsf did not correspond to the final state of the refinement. The numbers of protein atoms (1001) and water molecules (90) were the same in both models, whereas the number of heterogen atoms was 35 in PDB entry 4nsf and 57 in 4xan.

Selected parameters of the re-refined structure are shown in Table 4, together with the corresponding parameters for 4nsf and 4xan. We moved the coordinates within the unit cell rather than outside (Kowiel *et al.*, 2014) as in the original refinements (Tanley *et al.*, 2014). For consistency with the deposited 4xan model, the structure was re-refined with anisotropic ADP parameters. With the geometry restrained in a similar way, the final *R* factors are significantly lower than for both previously deposited models (Table 4). Since the resolution is rather low for full anisotropic refinement, we used tight restraints on the sphericity of individual anisotropic ellipsoids, and the distribution of atomic vibrational anisotropy was monitored using the *PARVATI* server (http://skuld.bmsc.washington.edu/parvati/). The number of solvent molecules in the current refinement is in line with what would be expected for a high-resolution structure. Only two of the six DMSO molecules found in 4xan were retained; one was changed to an acetate, two were changed to bromine ions and one was removed completely. Bromine ions were placed mostly on the basis of the presence of anomalous signal and a suitable chemical environment; more of them were located compared with the original model. All four acetate moieties present in the 4xan model were replaced by water molecules.

A strong anomalous peak 2.12 Å from the $N^{\delta 1}$ atom of His15 must unambiguously originate from the presence of a Pt

atom refined at 0.85 occupancy. In agreement with Tanley *et al.* (2014), the two strong anomalous peaks at an ∼90° angle to the Pt—$N^{\delta1}$ bond were interpreted as *trans* bromines and their occupancies match that of the Pt atom. The presence of these bromide ions is also supported by the Pt—Br distances of ∼2.5 Å. The site opposite $N^{\delta1}$ was modeled as Cl^-, also with a matching occupancy [note that chloride was present in the crystallization mixture at 0.005% as described in Tanley *et al.* (2014), which corresponds to 1.4 m*M*]. This site is difficult to model with confidence, but we found that placing a Cl^- ion to be reasonable because of a suitable distance to Pt (2.3 Å) and because placing a water or ammonia molecule there resulted in ∼6σ positive difference peak, strongly suggesting a heavier atom. The presence of a Br^- ion at this position is unlikely owing to the absence of a significant anomalous peak. However, we placed a bromide anion (Br204) in a small peak in the anomalous map located next to the Cl^- ion, 2.9 Å from the Pt site, and slightly out of the plane formed by Pt and its ligands. Such positioning makes it very unlikely that Br204 could be a Pt ligand; therefore, we assigned it an occupancy of 0.15 and set its conformation identifier to B, as an alternative to variant A presented by the Pt complex with one Cl^- and two Br^- ions. This interpretation is different from the 4nsf structure, where the site opposite to $N^{\delta1}$ was not modeled, or 4xan, where it contained several atoms belonging to a putative carboplatin but in an arrangement that could not be reconciled with the expected geometry and chemistry.

Two anomalous peaks are present in the vicinity of $N^{\varepsilon2}$ of His15. Our interpretation is quite similar to the original interpretation. The peak 2.68 Å from $N^{\varepsilon2}$ was interpreted as Pt (but not cisplatin as in the original deposition) with occupancy 0.15, even if this distance is considerably longer than expected for a Pt—N coordination bond (2.0 Å). The adjacent site, 3.73 Å from $N^{\varepsilon2}$ and 2.12 Å from the Pt atom, is in our interpretation a bromide ion with occupancy 0.6. This interpretation could be challenged by the absence of a complete coordination sphere for the square-planar platinum(II) atom and by most of the distances deviating from ideal Pt—*X* values. However, placement of these heavy atoms is supported by the anomalous map, by the characteristic 2.1 Å distance between Pt and the $N^{\eta2}$ atom of Arg14 and by a similar pattern of Pt binding in the 4g4a structure (see below). Nevertheless, owing to the low occupancies and weak electron density, this region gives no certain information about platinum/cisplatin binding to $N^{\varepsilon2}$ of His15 and should not be used to draw any conclusions in this matter.

3.11.3. Moderate-resolution structure of tetragonal lysozyme with cisplatin (PDB entry 4g4a). Our initial attempts to reinterpret the PDB deposition 4g4a were difficult, as the map calculated from the deposited structure factors was quite noisy. Since the original data frames for this structure, refined at 2.4 Å resolution, are available at http://rawdata.chem.uu.nl, we decided to use them in order to evaluate whether the refinement results could be changed not only by re-refinement with the same data, but also by reprocessing (re-reduction) of the original diffraction images. The diffraction data had been collected with an in-house Bruker APEX II CCD detector system and consisted of 3366 0.5° frames measured in nine sweeps. The originally reported values of parameters such as $R_{\rm merge}$ were unusually high (Table 5), suggesting that significant radiation damage might be present. Indeed, when only the first sweep of data (347 frames, with an average redundancy of 7) was processed with *HKL*-3000 (Minor *et al.*, 2006), the resolution of the useable data could be extended to 2.0 Å and the value of $R_{\rm merge}$ became very considerably lower, and it was lower still at 2.4 Å resolution (Table 5). The re-reduction of frames 2708–3037 revealed that the crystal suffered a marked loss of diffraction power owing to radiation damage; statistically significant diffraction extended only to ∼2.65 Å resolution, using an $\langle I/\sigma(I)\rangle \simeq 2$ cutoff for the highest resolution shell. Clearly, merging too many frames strongly affected the results, producing a data set with much worse overall quality. Corrections for crystal decay might improve the quality somewhat, but even the best corrections cannot create data that are absent because of the resolution drop owing to significant crystal deterioration.

The re-refined structure not only has a much higher resolution, but also better *R* factors and other model-quality indicators (Table 5). Overall, reprocessing of the data and re-refinement resulted in electron-density maps with lower noise, which made their interpretation much easier. Several errors in the original model were clearly

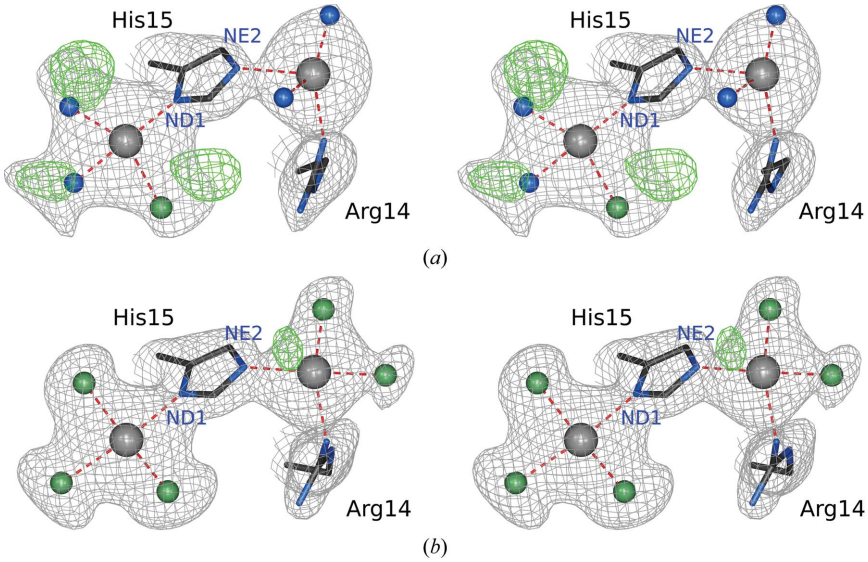


Figure 5
Stereoview showing the binding of platinum(II) adducts to lysozyme. $2mF_o - DF_c$ maps are displayed in gray contoured at the 1.2σ level. The $mF_o - DF_c$ difference map is contoured at the 4.2σ level using a green color for positive density; no negative density is present at this level. Chloride ions are depicted as green spheres and ammine groups of cisplatin are depicted as blue spheres. (*a*) The original PDB model 4g4a. (*b*) The same structure after reinterpretation and re-refinement.

revealed and successfully corrected. They include (i) an incorrect conformation of the His15 main chain, resulting in a severe clash between the O atoms of His15 and Met12 at 2.15 Å; (ii) an Arg14 side-chain conformation that is not supported by electron density; and (iii) a DMSO molecule, DMS201, that is not supported by electron density (replaced by two well fitting water molecules). As with the other structures re-refined in this work, the number of modeled water molecules is significantly higher than in the original model (75 *versus* 21).

Similarly to the structures 4nsf/4xan, there is a very strong anomalous peak in the vicinity of the $N^{\delta1}$ atom of His15, which unambiguously originates from the presence of a Pt atom refined at an occupancy of 0.8 both in the original deposition and in our model. However, this site was interpreted as $PtCl(NH_3)_2$ in the original deposition, which is not supported by the heights of the electron-density peaks and interatomic distances. The positive electron difference peaks present when an ammine group was placed and refined in the original positions were $\sim 7\sigma$, suggesting that the heavier Cl^- ion should be placed in all vacant positions around Pt (Fig. 5). The distances between these peaks and Pt (2.2–2.3 Å) also support this interpretation, even if, taken individually, they might not be significant at 2.0 Å resolution. Our interpretation therefore assumes the conversion of cisplatin to a chloride-substituted complex, which might be rather audacious to suggest. However, such a conversion is theoretically possible as the crystallization cocktail had been kept for 14 months with NaCl present at a concentration of 2.8 *M*.

In contrast to the structures 4nsf/4xan, only a single anomalous peak is present in the vicinity of the $N^{\varepsilon2}$ atom of His15. This peak has an acceptable distance (2.3 Å) to the $N^{\varepsilon2}$ atom, supporting the presence of a Pt atom modeled by us with an occupancy of 0.4. Similarly to the $N^{\delta1}$ site, the $N^{\varepsilon2}$ site was also modeled as $PtCl(NH_3)_2$ in the original structure, but (i) it had tetrahedral instead of square-planar geometry, (ii) one NH_3 ligand was out of density and (iii) both NH_3 ligands had full occupancy despite forming one molecule with Pt at an occupancy of 0.5. We interpreted this site as Pt coordinated by His15 $N^{\varepsilon2}$, Arg14 $N^{\eta1}$ and two chloride ions, in congruence with the $N^{\delta1}$ site. This configuration was further supported by the interatomic distances and heights of the electron-density peaks. The possibility of transition-metal coordination by a guanidine group (which must not be protonated in this role) has been conclusively demonstrated by Legin *et al.* (2014).

4. Discussion and conclusions

Crystallographic data, once released as part of the literature and deposited in the PDB, tend to be accepted by the nonstructural community as definitive, their correctness unquestioned. The 'ripple effect' of structural results (good and bad!) is very significant, given that many other fields of science use these structural models. Thus, for example, Fig. 6(*a*) of a review by Pinato *et al.* (2014) shows an atomic model of two cisplatin complexes bound equivalently at the two sides of His15 in HEWL. As discussed above, this is a grossly oversimplified picture without solid experimental verification, yet it has already been referenced in this review article. Similarly, PDB deposition 3fj0, describing the structure of β-glucosidase, was downloaded from the PDB over 30 000 times before it was replaced by the reinterpreted model 4hz8. These are just examples of how questionable or unproven results can be propagated in the scientific literature and into structural databases if there is no mechanism to correct them and to follow the replacement models, or the eradication of errors, for example in the PDB.

It is generally accepted that chemical crystallography is critical for drug discovery. Indeed, most of the publications dealing with the complexes of the anticancer platinum compounds with proteins cited here stressed health-related motivation as the impetus for such work. This particular aspect might not be so critical with regard to protein complexes of cisplatin and carboplatin, since the intended biological targets of these drugs are nucleic acids. Still, understanding the chemical nature of the protein adducts might be of importance in the search for new molecules with better therapeutic performance and fewer side effects. However, in other cases, where crystallographic models of direct targets for drug discovery might have similar problems to those discussed here, the concerns would be not only of a purely academic nature but could have very unpleasant consequences for translational research based on structural data, including retardation of human health-related research.

How can such problems be avoided in the future, or at least mitigated? Several generalizations can be made based on the experiences gained from the analysis and re-refinement of the structures discussed here. First of all, the onus is squarely on the authors, who need to pay enough attention to both the technical aspects of their work and adherence to established principles of chemistry. The PDB validation reports are important tools for maintaining the high quality of deposited crystallographic models, but they can be helpful only if authors choose to analyze (and correct) the reported abnormalities. The validation reports for some of the crystallographic models discussed here indicated, for example, large differences between the values of the *R* factors quoted in the coordinate files and those recalculated during the validation process. Such discrepancies must not be ignored by the authors, and the source of the differences should be identified. Similarly, the LLDF parameters (if used by PDB at the time of deposition) should have indicated that placement of some ligands was not supported by the electron density, and ignoring these alerts is another missed opportunity to improve the models. LLDF and even RSRZ parameters are not perfect and require additional testing on a large number of cases, but they are very helpful to show areas that require special attention (see the footnote to Table 4).

Some of the problems that were due to incorrect stereochemical restraints are very clearly noted in the validation reports as abnormalities in bond lengths and angles, but they were clearly disregarded. In this case, an aphorism (coined by Dr Theodore Woodward in the 1940s) is particularly apt: 'when you hear hoofbeats, think of horses, not

zebras'. The probability that outliers are a consequence of 'new chemistry', rather than modeling errors, is usually very low, especially when other validation problems are present as well. On the other hand, we note that setting unrealistically stringent validation criteria is not the best way to obtain better models. For example, flagging ligand bonds and angles with $Z > 2$ (compared with $Z > 5$ for proteins and nucleic acids) and an electron-density fit criterion of RMSZ > 2 (still a 5% chance of being correct) only lead to lengthy error reports which might not be taken seriously.

The key point that requires emphasizing is that whereas validation of the correctness of the protein model in a crystal structure is comparatively easy, it is much harder to make sure that the same is true of the ligands. Metals, in particular, seem to present constant problems with verification of their nature and the correctness of their interactions with the macromolecules. It is our experience that the fraction of suspicious metals in the PDB files does not change over time, as opposed to the overall structure quality, which is constantly improving (Domagalski *et al.*, 2014). If indeed the fraction of dubious models in the PDB is roughly constant, then this would mean that the total volume of incorrect information is actually increasing with time at an increasing rate, which would be a very dangerous and grim outlook for the future validity and integrity of structural databases.

In the present analysis of deposited structures, we encountered three options for the availability of experimental data, namely (i) structure factors are not available, (ii) structure factors are available but raw diffraction images are not and (iii) raw diffraction images are available. Of the three possibilities, it is clearly option (iii) that allows the most desirable approach to further analysis. If errors or other flaws have been introduced during the processing of diffraction images, the structure factors may be contaminated with errors that cannot be corrected.

Since February 2008, all depositions in the PDB must include the associated experimental data in the form of structure factors. These additional data allow re-refinement using automated approaches such as *PDB_REDO* (Joosten *et al.*, 2012), flagging of suspicious models (Janssen *et al.*, 2007) and validation by their 'consumers'. However, the structure-factor files may not always contain all of the information that is necessary for the full reinterpretation and improvement of a particular PDB model. For example, the PDB currently contains 13 483 depositions that report the presence of anomalous signal in the diffraction data, but only around 20% of the structure-factor files associated with those depositions have actual data for the calculation and use of the anomalous signal. This affected the work described here, since some of the data sets lacked this crucial information for the independent identification of ligands. Many of these problems could have been solved if public access to raw diffraction images was available. In addition, as we have demonstrated in the example of the reinterpretation of PDB entry 4g4a, the availability of the raw diffraction data may allow the re-refinement of previously deposited structures to higher resolution and may facilitate the use of more sophisticated corrections for radiation decay and/or anisotropic diffraction.

What should happen to the reinterpreted and corrected depositions? Clearly, they should be made available to the scientific community, but there seems to be no agreement as to the best strategy for accomplishing such a task. We have deposited the re-refined structures in the PDB and they were issued with new accession codes, while being linked to the original depositions. However, we are not sure that this way of archiving the data would be most useful to the community. Ideally, the PDB should transition from a static repository towards a dynamic one (Terwilliger & Bricogne, 2014) that contains endlessly improving models obtained by using the continuously progressing methods and software, according to the idea that the originally measured data are eternal but any interpretation is subject to evolution (McNutt, 2015). The new crystallographic models should be made generally available together with all the precursor entries, and our suggestion would be that they should all be easily traced to the original deposition. Whereas the authors of the original work may not have participated in the reinterpretation of their data, and might not even agree with the new conclusions, their contribution needs to be very clear to the users of such potentially improved data. We hope that the results and divagations presented here will provide a stimulus for moving in this direction.

Acknowledgements

We would like to thank Nicholas P. Farrell and Przemek Porebski for valuable discussions and Joanna Raczynska for reading the manuscript. This project was supported in part by the Intramural Research Program of the National Cancer Institute, Center for Cancer Research. The collaboration of MJ and ZD was supported in part by a grant (2013/10/M/NZ1/00251) from the National Science Center (Poland). IS and WM were supported by federal funds from the NIAID, NIH, Department of Health and Human Services under Contract No. HHSN272200700058C and by NIGMS grants GM094585, GM094662 and GM093342.

References

Addlagatta, A., Krzywda, S., Czapinska, H., Otlewski, J. & Jaskolski, M. (2001). *Acta Cryst.* D**57**, 649–663.
Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Andrejašič, M., Pražnikar, J. & Turk, D. (2008). *Acta Cryst.* D**64**, 1093–1109.
Appleton, T. G., Berry, R. D., Davis, C. A., Hall, J. R. & Kimlin, H. A. (1984). *Inorg. Chem.* **23**, 3514–3521.
Banci, L., Bertini, I., Blaževitš, O., Calderone, V., Cantini, F., Mao, J., Trapananti, A., Vieru, M., Amori, I., Cozzolino, M. & Carrì, M. T. (2012). *J. Am. Chem. Soc.* **134**, 7009–7014.
Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucleic Acids Res.* **28**, 235–242.
Boal, A. K. & Rosenzweig, A. C. (2009). *J. Am. Chem. Soc.* **131**, 14196–14197.
Bragg, W. L. (1913). *Proc. R. Soc. A Lond. A*, **89**, 248–277.
Brünger, A. T. (1992). *Nature (London)*, **355**, 472–475.

Calderone, V., Casini, A., Mangani, S., Messori, L. & Orioli, P. L. (2006). *Angew. Chem. Int. Ed.* **45**, 1267–1269.
Casini, A., Mastrobuoni, G., Temperini, C., Gabbiani, C., Francese, S., Moneti, G., Supuran, C. T., Scozzafava, A. & Messori, L. (2007). *Chem. Commun.* **2007**, 156–158.
Cech, T. R. (2000). *Science*, **289**, 878–879.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A. (1949). *Chemistry of Penicillin*, edited by H. Clarke, T. Johnson & J. R. Robinson, pp. 310–367. Princeton University Press.
Dauter, Z., Wlodawer, A., Minor, W., Jaskolski, M. & Rupp, B. (2014). *IUCrJ*, **1**, 179–193.
Deller, M. C. & Rupp, B. R. (2015). *J. Comput. Aided Mol. Des.*, doi:10.1007/s10822-015-9833-8.
Domagalski, M. J., Zheng, H., Zimmerman, M. D., Dauter, Z., Wlodawer, A. & Minor, W. (2014). *Methods Mol. Biol.* **1091**, 297–314.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gust, R. & Schnurr, B. (1999). *Monatsh. Chem.* **130**, 637–644.
Helliwell, J. R. & Tanley, S. W. M. (2013). *Acta Cryst.* D**69**, 121–125.
Hooft, R. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
Ivanov, A. I., Christodoulou, J., Parkinson, J. A., Barnham, K. J., Tucker, A., Woodrow, J. & Sadler, P. J. (1998). *J. Biol. Chem.* **273**, 14721–14730.
Janssen, B. J. C., Read, R. J., Brünger, A. T. & Gros, P. (2007). *Nature (London)*, **448**, E1–E2.
Jaskolski, M. (2013). *Acta Cryst.* D**69**, 1865–1866.
Joosten, R. P., Joosten, K., Murshudov, G. N. & Perrakis, A. (2012). *Acta Cryst.* D**68**, 484–496.
Kane, S. A. & Lippard, S. J. (1996). *Biochemistry*, **35**, 2180–2188.
Kleywegt, G. J. & Harris, M. R. (2007). *Acta Cryst.* D**63**, 935–938.
Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wählby, A. & Jones, T. A. (2004). *Acta Cryst.* D**60**, 2240–2249.
Kowiel, M., Jaskolski, M. & Dauter, Z. (2014). *Acta Cryst.* D**70**, 3290–3298.
Lamb, A. L., Kappock, T. J. & Silvaggi, N. R. (2015). *Biochim. Biophys. Acta*, **1854**, 258–268.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Legin, A. A., Jakupec, M. A., Bokach, N. A., Tyan, M. R., Kukushkin, V. Y. & Keppler, B. K. (2014). *J. Inorg. Biochem.* **133**, 33–39.
Lippard, S. J. (1982). *Science*, **218**, 1075–1082.
McNutt, M. (2015). *Science*, **347**, 7.
Messori, L. & Merlino, A. (2014). *Inorg. Chem.* **53**, 3929–3931.
Milne, A. A. (1928). *The House at Pooh Corner*, p. 1. London: Methuen.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perutz, M. F. (1970). *Nature (London)*, **228**, 726–734.
Phillips, D. C. (1967). *Proc. Natl Acad. Sci. USA*, **57**, 483–495.
Pinato, O., Musetti, C. & Sissi, C. (2014). *Metallomics*, **6**, 380–394.
Read, R. J. *et al.* (2011). *Structure*, **19**, 1395–1412.
Sen, S. *et al.* (2014). *Database (Oxford)*, **2014**, bau116.
Shakespeare, W. (1603). *The Tragicall Historie of Hamlet Prince of Denmarke*, Act III, Scene II. London: Nicholas Ling & John Trundell.
Spek, A. L. (2009). *Acta Cryst.* D**65**, 148–155.
Tanley, S. W. M., Diederichs, K., Kroon-Batenburg, L. M. J., Schreurs, A. M. M. & Helliwell, J. R. (2013). *J. Synchrotron Rad.* **20**, 880–883.
Tanley, S. W. M. & Helliwell, J. R. (2014*a*). *Acta Cryst.* F**70**, 1127–1131.
Tanley, S. W. M. & Helliwell, J. R. (2014*b*). *Struct. Dyn.* **1**, 034701.
Tanley, S. W. M., Schreurs, A. M. M., Helliwell, J. R. & Kroon-Batenburg, L. M. J. (2013). *J. Appl. Cryst.* **46**, 108–119.
Tanley, S. W. M., Schreurs, A. M. M., Kroon-Batenburg, L. M. J. & Helliwell, J. R. (2012*a*). *Acta Cryst.* F**68**, 1300–1306.
Tanley, S. W. M., Schreurs, A. M. M., Kroon-Batenburg, L. M. J. & Helliwell, J. R. (2012*b*). *Acta Cryst.* A**68**, s155.
Tanley, S. W. M., Schreurs, A. M. M., Kroon-Batenburg, L. M. J., Meredith, J., Prendergast, R., Walsh, D., Bryant, P., Levy, C. & Helliwell, J. R. (2012). *Acta Cryst.* D**68**, 601–612.
Tanley, S. W. M., Starkey, L.-V., Lamplough, L., Kaenket, S. & Helliwell, J. R. (2014). *Acta Cryst.* F**70**, 1132–1134.
Terwilliger, T. C. & Bricogne, G. (2014). *Acta Cryst.* D**70**, 2533–2543.
Tickle, I. J. (2012). *Acta Cryst.* D**68**, 454–467.
Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* D**55**, 191–205.
Wang, D. & Lippard, S. J. (2005). *Nature Rev. Drug Discov.* **4**, 307–320.
Weichenberger, C. X., Pozharski, E. & Rupp, B. (2013). *Acta Cryst.* F**69**, 195–200.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Werner, A. (1893). *Z. Anorg. Chem.* **3**, 267–330.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Zheng, H., Chordia, M. D., Cooper, D. R., Chruszcz, M., Müller, P., Sheldrick, G. M. & Minor, W. (2014). *Nature Protoc.* **9**, 156–170.

scientific comment

BIOLOGICAL CRYSTALLOGRAPHY

ISSN 1399-0047

Responses to *Crystallography and chemistry should always go together: a cautionary tale of protein complexes with cisplatin and carboplatin*

Todd O. Yeates*

Department of Chemistry and Biochemistry, University of California, Box 951569, Los Angeles, CA 90095-1569, USA. *Correspondence e-mail: yeates@mbi.ucla.edu

Received 16 July 2015
Accepted 29 July 2015

Keywords: cisplatin; carboplatin; response.

In this issue of *Acta Cryst. D*, Shabalin *et al.* (2015) critique and re-evaluate dozens of crystal structures in the PDB in which proteins are bound to the platinum compounds cisplatin or carboplatin. Investigators whose structures were critiqued were contacted and several wrote comments in response. Through all those comments there was much agreement with certain sentiments expressed by Shabalin *et al.*, including the idea that identification and refinement of metal ligands (and other unanticipated molecules in a crystal) is often extremely difficult. There was a shared opinion that greater vigilance and further tools for validation are needed. Shabalin *et al.* offer challenges to previous structural interpretations that vary in their severity. One end of the spectrum concerns cases where difficult decisions were required about whether or not to model a ligand into relatively weak features in an electron-density map. For example, in one protein that was re-examined (SOD), Shabalin *et al.* conclude that a missing fourth ligand to one of the Pt atoms should have been included but was not, and that a ligand to another Pt atom was included where the electron density was too weak for accurate modeling. They express similar opinions in their re-examination of RNase. Responses to Shabalin *et al.* in these two cases by A. Merlino, L. Messori, V. Calderone and S. Mangani include concessions on at least one point, that platinum is four-coordinated in the SOD structure; its omission by the original authors was a modeling decision or oversight not intended to convey that platinum was actually three-coordinated. In addressing other challenges such as whether reported ligands to platinum were reliably modeled, the responders maintained that their original assignments reflect plausible interpretations of electron density. At the other end of the critique spectrum, Shabalin *et al.* identify specific errors and arrive at alternate interpretations after re-refining the crystal structures of some other proteins bound to platinum compounds. These include ATOX1 and hen egg-white lysozyme. Here, Shabalin *et al.* offer challenges to the assignment of electron-density features to platinum atoms *versus* other metals, the identification of the ligands to platinum, and the modeling of various buffer constituents in the crystal structures. Owing to the specific re-interpretations offered by Shabalin *et al.* in these two cases, responses from the original investigators are published in this issue (Tanley *et al.*, 2015; Boal & Rosenzweig, 2015). Irrespective of the

particular findings and conclusions in these studies, the take-home lesson is clear that we must be ever more vigilant in protecting and improving the veracity of the structural and chemical information in the PDB so that it will stand as one of the most valuable (and hard-won) repositories in modern science.

References

Shabalin, I., Dauter, Z., Jaskolski, M., Minor, W. & Wlodawer, A. (2015). *Acta Cryst.* D**71**, 1965–1979.

Tanley, S. W. M., Diederichs, K., Kroon-Batenburg, L. M. J., Levy, C., Schreurs, A. M. M. & Helliwell, J. R. (2015). *Acta Cryst.* D**71**, 1982–1983.

Boal, A. K. & Rosenzweig, A. C. (2015). *Acta Cryst.* D**71**, 1984–1986.