STRUCTURAL BIOLOGY

ISSN: 2059-7983
journals.iucr.org/d

Towards a dependable data set of structures for L-asparaginase research

Alexander Wlodawer, Zbigniew Dauter, Jacek Lubkowski, Joanna I. Loch, Dariusz Brzezinski, Miroslaw Gilski and Mariusz Jaskolski

Acta Cryst. (2024). D**80**, 506–527

STRUCTURAL BIOLOGY

ISSN 2059-7983

Towards a dependable data set of structures for L-asparaginase research

Alexander Wlodawer,[a]* Zbigniew Dauter,[a] Jacek Lubkowski,[a] Joanna I. Loch,[b] Dariusz Brzezinski,[c] Miroslaw Gilski[d,e] and Mariusz Jaskolski[d,e]

[a]Center for Structural Biology, Center for Cancer Research, National Cancer Institute, Maryland, USA, [b]Department of Crystal Chemistry and Crystal Physics, Faculty of Chemistry, Jagiellonian University, Cracow, Poland, [c]Institute of Computing Science, Poznan University of Technology, Poznan, Poland, [d]Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland, and [e]Department of Crystallography, Faculty of Chemistry, Adam Mickiewicz University, Poznan, Poland. *Correspondence e-mail: wlodawer@nih.gov

Received 21 February 2024
Accepted 7 June 2024

Edited by B. Kobe, University of Queensland, Australia

Keywords: hydrolases; amidohydrolases; L-asparaginases; leukemia; Protein Data Bank.

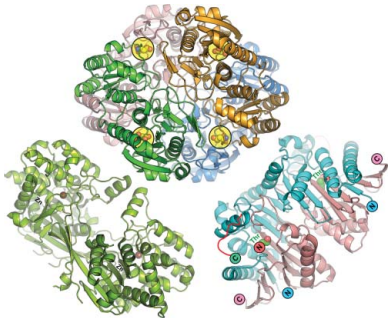
Supporting information: this article has supporting information at journals.iucr.org/d

The Protein Data Bank (PDB) includes a carefully curated treasury of experimentally derived structural data on biological macromolecules and their various complexes. Such information is fundamental for a multitude of projects that involve large-scale data mining and/or detailed evaluation of individual structures of importance to chemistry, biology and, most of all, to medicine, where it provides the foundation for structure-based drug discovery. However, despite extensive validation mechanisms, it is almost inevitable that among the ∼215 000 entries there will occasionally be suboptimal or incorrect structure models. It is thus vital to apply careful verification procedures to those segments of the PDB that are of direct medicinal interest. Here, such an analysis was carried out for crystallographic models of L-asparaginases, enzymes that include approved drugs for the treatment of certain types of leukemia. The focus was on the adherence of the atomic coordinates to the rules of stereochemistry and their agreement with the experimental electron-density maps. Whereas the current clinical application of L-asparaginases is limited to two bacterial proteins and their chemical modifications, the field of investigations of such enzymes has expanded tremendously in recent years with the discovery of three entirely different structural classes and with numerous reports, not always quite reliable, of the anticancer properties of L-asparaginases of different origins.

1. Introduction

L-Asparaginases hydrolyze L-asparagine to L-aspartate with the simultaneous release of ammonia. They are often referred to as ASNases and are assigned the EC number 3.5.1.1; if significant glutaminase co-activity is also present, the EC number is 3.5.1.38. Beyond pure academic curiosity, ASNases are also studied because of their potential application as first-line drugs for the treatment of acute lymphoblastic leukemia (ALL; Egler *et al.*, 2016; Beckett & Gervais, 2019; Ghasemian *et al.*, 2019; Chan *et al.*, 2019; Radadiya *et al.*, 2020). This use of ASNases dates back to the early experiments of Kidd (1953), who discovered the antitumor activity of guinea pig serum, and Broome (1961, 1963), who attributed this tumor inhibition to ASNase activity. Since the late 1970s, L-asparaginases of bacterial origin have been used in the clinical treatment of ALL. Moreover, there are recent reports that ASNases may also be effective against solid tumors (Tam *et al.*, 2023).

To date, three completely different structural classes of ASNases have been identified (Loch & Jaskolski, 2021), originally named according to the source organism of their isolation (Michalska & Jaskolski, 2006), namely class 1 (bacterial-type), class 2 (plant-type) and class 3 (*Rhizobium*



etli-type). This new classification intersects with an older convention which divided the known enzymes with L-asparaginase activity into five types, since both class 1 and class 3 contain two types that are distinguished according to their compartmentalization and expression profile. The prototypes of types I and II (in class 1) and type III (in class 2) are the *Escherichia coli* enzymes EcAI (cytosolic), EcAII (periplasmic) and EcAIII (also cytosolic), respectively. The prototypes of types IV and V (class 3) are the *R. etli* enzymes ReAIV (constitutive) and ReAV (inducible).

It is puzzling why nature has invented so many structural ways to catalyze a simple reaction that releases ammonia from a small molecule, L-asparagine (L-Asn). From the point of view of medicinal application, the most important aspect of the asparaginase reaction is the removal of circulating L-Asn, as starvation of this amino acid leads to the death of some cancer cells. However, to be effective in this role as injected drugs, the enzymes must have a sufficiently high substrate affinity (*i.e.* a sufficiently low K_m) to match the micromolar concentration of L-Asn circulating in the bloodstream (Ogawa *et al.*, 2017). Thus arises the main snag in the otherwise promising clinical use of ASNases: only enzymes of type II, *i.e.* EcAII-like, with a micromolar K_m are sufficiently potent for therapeutic application. Indeed, EcAII and ErA, a similar protein from *Erwinia chrysanthemi* (a bacterium now renamed *Dickeya dadantii*), are the only antileukemic drugs in this category that are currently approved. In their more recent formulations, these enzymes are manufactured with chemical modifications (Dinndorf *et al.*, 2007) to increase their lifetime in the blood and reduce their side effects, which can be very severe (Patel *et al.*, 2009; Koprivnikar *et al.*, 2017; Cecconello *et al.*, 2020).

In view of the negative clinical reactions accompanying the current administration protocols, it is not surprising that a significant research effort has been directed into engineering the other types of ASNases into useful drugs (Nguyen *et al.*, 2016*b*; Rigouin *et al.*, 2017; Loch & Jaskolski, 2021; Van Trimpont, Peeters *et al.*, 2022; Loch *et al.*, 2022, 2024; Janicki *et al.*, 2023). For such protein-engineering projects, it is important to have a reliable, well-validated database of protein structures. Naturally, for such a resource one turns to the Protein Data Bank (PDB; Protein Data Bank, 1971; wwPDB Consortium, 2018). The first L-asparaginase structure, of the EcAII enzyme, was published and deposited in the PDB in 1993 (Swain *et al.*, 1993). Over the last three decades, the number of L-asparaginase structures in the PDB has increased dramatically (Lubkowski & Wlodawer, 2021). However, a careful examination of these depositions reveals a number of problems of variable severity. In this work, we present our evaluation and validation of the crystallographic structures of ASNases found in the PDB.

We have carried out similar analyses of medicinally important protein models in the past, for instance for cisplatin and carboplatin complexes (Shabalin *et al.*, 2015), metallo-β-lactamases (Raczynska *et al.*, 2018) and SARS-CoV-2 proteins (Wlodawer *et al.*, 2020; Brzezinski *et al.*, 2021; Grabowski *et al.*, 2021; Jaskolski *et al.*, 2021). Our aim is to verify the structures deposited in the PDB and, if necessary, provide a better interpretation of the available crystallographic data.

2. Methods

2.1. PDB searches

In this study, ASNase structures were retrieved by the PDBe search command `Molecule name='asparaginase'` and manually verified. The results were then cross-checked with RCSB, PDBe and PDBj queries for EC Nos. 3.5.1.1, 3.4.19.5 and 3.5.1.38. Finally, 189 deposits were retained for further analysis. Since five of them were not accompanied by diffraction data, only 184 structures were analyzed more fully, and 30 of them were selected for re-refinement (see below).

2.2. Ranking of structures according to potential problems

To facilitate and prioritize the re-evaluation work, the structures within each class and type of ASNase were first ranked according to the severity of the perceived problems using the single-parameter statistic PQ1 (Brzezinski *et al.*, 2020). Briefly, PQ1 is a percentile score showing how the structural quality of a given structure compares with the entire PDB. The score combines information about the R_{free}, percentages of RSRZ (normalized real-space R value) outliers and Ramachandran outliers, rotamer outliers and clashscore of the structure, metrics that are described in detail in Wlodawer (2017) and are used in validation sliders at the top of the PDB report (Read *et al.*, 2011). The values of PQ1 range from 0 to 1, with 0 corresponding to the worst-quality structures in the PDB and 1 to the best-quality structures. The PQ1 calculations in this study were based on all PDB depositions based on X-ray crystal diffraction as of 22 November 2023 (180 418 structures). The resulting PQ1 scores for all of the analyzed ASNases can be found in the Excel files deposited as supporting information.

Ultimately, the PQ1 parameter was not fully correlated with the severity of the structural problems encountered in each structure. This was partly due to the fact that PQ1 focuses on what is wrong in the existing model, but not on its missing parts. The only component of PQ1 that contains some (indirect) information about uninterpreted electron density is the R_{free} factor.

2.3. Automatic analysis of difference electron-density maps

To aid the model-rebuilding work, a procedure was developed that automatically reports the most prominent (positive and negative) $F_o - F_c$ electron-density peaks for all structures under consideration. The procedure was written as a script for the *Coot* program (Emsley *et al.*, 2010), using its embedded Python interpreter. Positive and negative peaks in the $F_o - F_c$ difference maps were calculated for all 184 structures using the experimental structure factors that are available in the PDB as files labeled 'Map Coefficients (MTZ format)'. A list of peaks with values exceeding 5 root-mean-square deviations from an average was created for each structure (Excel files deposited as supporting information) using the internal *Coot*

Table 1
Structures of class 1 type I L-asparaginases deposited in the Protein Data Bank since 2021.

For structures deposited earlier, see Lubkowski & Wlodawer (2021).

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No. of missing residues‡	Problems identified	Reference
B. paralicheniformis	8h4f	T13A/P55F	1.9	Dimer	D-Asn	9/8	0 occ 4 Mg, 5 FMT, 2 DSG	Unpublished work
B. paralicheniformis	8h4g	T13A/M57N	1.81	Dimer	—	10/6	0 occ 3 Mg, 7 FMT	Unpublished work
B. paralicheniformis	8h4e	T13A/P55N	2.06	Dimer	D-Asn	10/7	0 occ 4 Mg, 8 FMT, 2 DSG	Unpublished work
B. paralicheniformis	8h4d	T13A/M57N	1.9	Dimer	—	9/2	0 occ 3 Mg, 7 FMT; 5 FMT, 46 Wat unrefined	Unpublished work
B. paralicheniformis	8h4c	T13A/M57P	1.6	Dimer	—	9/2	0 occ 7 Mg, 6 FMT	Unpublished work
B. paralicheniformis	8h4b	T13A/M57P	1.9	Dimer	L-Asn	7/8	0 occ 6 Mg, 2 FMT	Unpublished work
B. paralicheniformis	8h4a	T13A/M57P	1.9	Dimer	—	3/1	0 occ 7 Mg, 3 FMT; 4 FMT unrefined	Unpublished work
B. paralicheniformis	8h49	M57P	2.0	Dimer	—	7/6	0 occ 7 Mg, 3 FMT	Unpublished work
B. paralicheniformis	8h48	T13A/P55F	1.8	Dimer	L-Asn	10/1	0 occ 6 Mg, 3 FMT; 2 FMT, 1 Wat unrefined	Unpublished work
B. paralicheniformis	8h47	T13A/P55F	1.9	Dimer	—	8/7	0 occ 4 Mg, 8 FMT, 1 GOL; 1 FMT unrefined	Unpublished work
B. paralicheniformis	8h46	T13A/P55N	1.8	Dimer	L-Asn	1/1	0 occ 7 Mg, 3 FMT; 2 FMT, 4 Wat unrefined	Unpublished work
B. paralicheniformis	8h45	T13A/P55N	1.97	Dimer	—	10/6	0 occ 7 Mg, 6 FMT; 1 FMT unrefined	Unpublished work
B. paralicheniformis	8h44	P55N	1.8	Dimer	—	17/14	0 occ 7 Mg, 3 FMT; 2 FMT unrefined	Unpublished work
B. paralicheniformis	7cbw	Y278M	1.98	Dimer	—	1/1	Title in PDB wrong; lists T13A with D-Asn	Ran *et al.* (2021)
B. paralicheniformis	7cbu	T13A	2.25	Dimer	L-Asn	4/1	Missing atoms; 4 Mg, 3 FMT unrefined	Ran *et al.* (2021)
B. paralicheniformis	7cbr	T13A	1.8	Dimer	D-Asn	1/1	Missing atoms	Ran *et al.* (2021)
B. paralicheniformis	7cb4	WT	2.04	Dimer	—	1/4	Missing atoms, unrefined atoms	Ran *et al.* (2021)
B. paralicheniformis	7c91	WT	1.98	Dimer	D-Asn	1/2	Title in PDB wrong; lists T13A with D-Asn	Ran *et al.* (2021)
B. paralicheniformis	7c8x	T13A	1.99	Dimer	L-Asn	3/2	12 Wat, 1 residue unrefined	Ran *et al.* (2021)
B. paralicheniformis	7c8q	T13A	1.89	Dimer	D-Asn	1/1	5 Mg, 5 FMT, 7 Wat unrefined	Ran *et al.* (2021)
Y. pestis	7r6b	R43D/L124D/R125A/C273S	2.03	Dimer	—	36/37	One Ca OK, another doubtful	Strzelczyk *et al.* (2023)
Y. pestis	7r6a	L124D/R125A/C273S	1.89	2 dimers	—	9/16/21/15	—	Strzelczyk *et al.* (2023)
Y. pestis	7r69	R43D/C273S	1.8	2 dimers	—	19/26/30/35	—	Strzelczyk *et al.* (2023)

† ASU, asymmetric unit. ‡ Number of missing residues in each chain.

procedure `Validate → Difference Map Peak Analysis`.

2.4. Re-refinement of selected examples

Atomic coordinates of ASNases were downloaded from the PDB and displayed with *Coot*. Maps were calculated in *Coot* using structure factors downloaded from the PDB. The *ACHESYM* server (Kowiel *et al.*, 2014) was utilized in some cases to bring the coordinates of multiple files to a consistent location in the asymmetric unit of the unit cell. If we considered it advisable, models were adjusted in *Coot* and refined with *REFMAC*5 (Murshudov *et al.*, 2011). We used the default parameters of the latter program, with F_o and $\sigma(F_o)$ records (labeled FP and SIGFP) taken from the MTZ file. With one exception, we did not change the selection of reflections used to calculate R_{free} (Brünger, 1992). Altogether, 30 structures were re-refined in this way.

2.5. Availability of PDB files for the re-refined structures

The coordinates (in CIF and legacy PDB formats) of the structures that were re-refined can be found in the supporting information. They will also be accessible via the Proteopedia server (Prilusky *et al.*, 2011). Maps can be calculated by using the F_o and $\sigma(F_o)$ records in the original map coefficient (MTZ) files, which can be downloaded from the PDB, with the re-refined coordinates used to calculate the values of $|F_c|$ and the phases. These calculations can easily be performed in *Coot*.

3. Results and discussion

3.1. Class 1 L-asparaginases

The most recent review summarizing the crystal structures of all class 1 L-asparaginases determined by 2021 listed the PDB codes for 80 such depositions (Tables 1–3 in Lubkowski & Wlodawer, 2021). An additional 33 structures belonging to this category had been deposited by 15 November 2023, with 23 belonging to type I (Table 1) and ten belonging to type II (Table 2). The principal building block for these enzymes is a single polypeptide chain consisting of ∼320–350 amino acids folded into two domains, in some cases extended by another noncatalytic domain. Two such polypeptides extensively interacting with each other contain all of the machinery that is required for catalysis. Some class 1 ASNases (primarily the enzymes from extremophilic bacteria) exist as tight homodimers, although most of these enzymes form homotetramers, *i.e.* dimers of tight dimers, that include four equivalent active sites. Fig. 1 shows the tetramers, tight dimers and the details of the catalytic site for type II ASNases, exemplified by the first structure of an ASNase to be determined (PDB entry 3eca; Swain *et al.*, 1993). Tetramers, tight dimers and the allosteric

Table 2
Structures of class 1 type II L-asparaginases deposited in the Protein Data Bank since 2021.

Structures deposited earlier have been summarized by Lubkowski & Wlodawer (2021).

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No. of missing residues‡	Problems identified	Reference
E. coli	6v5f	WT	2.1	Tetramer	L-Asn	3/26/31/17	Many: see text	Unpublished work
E. coli	7a0u	T12S/T89S	2.26	Tetramer	—	11/1/22/21	Coordinate shift, minor corrections	Unpublished work
E. coli	7p9c	WT	1.6	Tetramer	L-Asn	5/15/0/16	Coordinate shift, deleted residues not in density	Maggi *et al.* (2021)
E. coli	7m11	WT	1.83	Tetramer	L-Asp	18/26/0/31	—	Unpublished work
E. coli	7r5q	N24S	1.9	Tetramer	L-Glu	20/20/15/23	Coordinate shift, minor adjustments	Maggi & Scotti (2022)
E. coli	7r57	N24S	1.4	Tetramer	—	16/16/17/10	Coordinate shift, added several residues, minor corrections	Maggi & Scotti (2022)
E. coli	8ecd	V27T	1.62	Tetramer	L-Asp	0/11/0/0	Coordinate shift, no other problems	Strzelczyk *et al.* (2022)
E. coli	8ece	V27T	2.5	Tetramer	L-Glu	21/21/23/18	—	Strzelczyk *et al.* (2022)
D. dadantii	7u6m	A31I, E63Q, S254Q; fused ABD§	1.75	Tetramer	L-Asp	19/19/5/19	Added 21 water molecules, deleted residues −14/−13 from chain *C* ABD	Van Trimpont, Schalk *et al.* (2022)
D. dadantii	8wgq	WT, commercial	2.75	Tetramer	—	15/2/14/17	Not significant, but more waters present	Unpublished work

† ASU, asymmetric unit. ‡ Number of missing residues in each chain. § ABD, albumin-binding domain.

sites of type I ASNases are shown in Fig. 2 based on the coordinates of PDB entry 6nxd (Lubkowski *et al.*, 2019). Whereas the fold of the individual protomers and the conformation of the tight dimers is similar for both types (Figs. 1*b* and 2*b*), the tetramers of type I enzymes, if present, are less compact (Fig. 2*a*) than the tetramers of type II ASNases (Fig. 1*a*).

Although the enzymatic mechanism of class 1 ASNases has been the subject of some controversy in the past, it is now generally agreed that it involves a double-displacement ('ping-pong') model. Reaction begins with nucleophilic attack on the side-chain amide C atom of the substrate by a threonine residue, activated with the help of a neighboring tyrosine residue through a series of water-mediated proton transfers. The catalytic threonine residue is located at the hinge of a long loop, designated as the active-site flexible loop (ASFL), that is usually disordered in the absence of a substrate. The resulting acyl-enzyme intermediate is then cleaved in the second step of the reaction by a water molecule activated by another threonine residue that belongs to a Thr–Lys–Asp triad, resembling the catalytic triad of serine proteases (Lubkowski *et al.*, 2020).

3.1.1. Class 1 type I (EcAI-type) enzymes

Five structures of type I ASNases were listed in the previously published review of structural data for L-asparaginases (Lubkowski & Wlodawer, 2021). An additional 14 structures were listed as 'unassigned', but these thermophilic and mammalian enzymes should also fall into this category (Loch & Jaskolski, 2021). This classification is based on three criteria: (i) higher sequence similarity to the type I *E. coli* ASNase (EcAI) than to EcAII, (ii) cytosolic localization and (iii) low substrate affinity (in the millimolar range). Whereas the enzymes from thermophilic archaea appear to function as homodimers (Yao *et al.*, 2005; Bansal *et al.*, 2012), those from mesophilic bacteria and enzymes of mammalian origin (the latter only represented in the PDB by structures of the guinea pig enzyme) form tetramers. However, it is possible that some

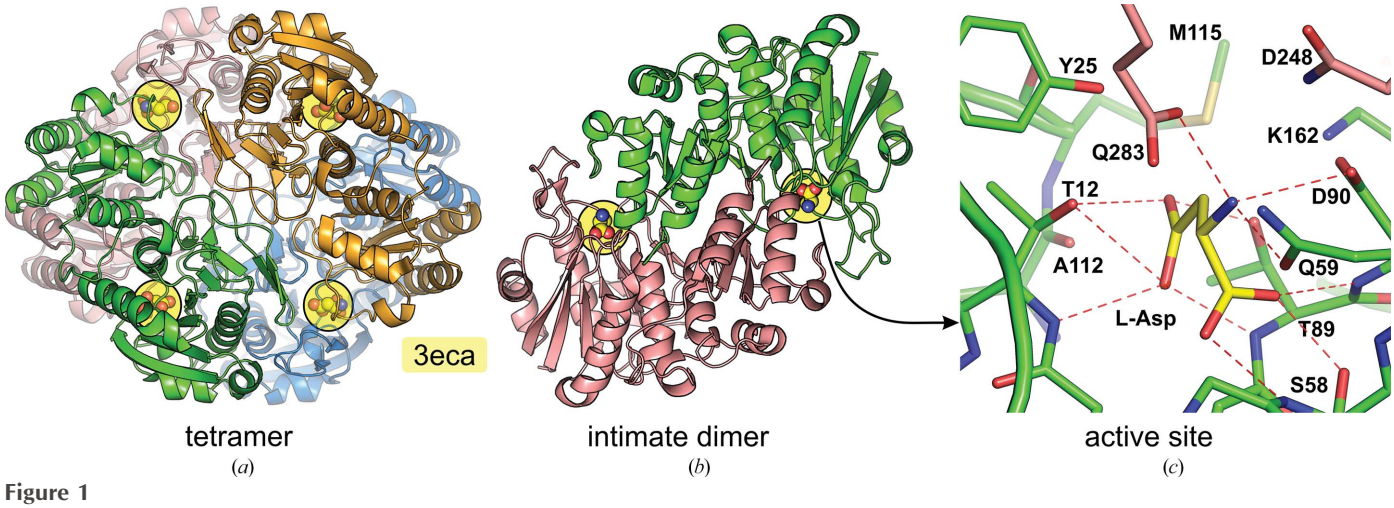


Figure 1
Crystal structure of the class 1 type II L-asparaginase from *E. coli* (PDB entry 3eca). (*a*) Two intimate dimers (*A*–*B*, green–pink; *C*–*D*, blue–orange) make up the EcAII homotetramer. The four active sites are marked by yellow circles with L-Asp molecules in CPK representation. (*b*) The intimate dimer (chain *A*, green; chain *B*, pink) of EcAII possesses two active sites [marked as in (*a*)]. (*c*) Detailed view of the active site of protomer *A* with a pattern of hydrogen bonds (red dashed lines) stabilizing the product molecule (L-Asp, yellow).

of them retain activity in a dimeric form (Strzelczyk *et al.*, 2023). A further 23 structures of type I ASNases have been published since 2021: three structures of *Yersinia pestis* ASNase (Strzelczyk *et al.*, 2023) and 20 structures of an enzyme from *Bacillus paralicheniformis* (Ran *et al.*, 2021; Zhou *et al.*, 2022).

3.1.1.1. Class 1 L-asparaginases from mesophilic bacteria

Three structures of EcAI were published at a resolution of 1.8–1.9 Å by Yun *et al.* (2007), either ligand-free (PDB entry 2p2d) or with a putative mixture of L-Asp and L-Asn in the active site (PDB entries 2p2n and 2him). The latter interpretation was subsequently questioned (Lubkowski *et al.*, 2019) and the active-site ligand was redefined as citrate (PDB entries 6nxd and 6nxc). Consequently, these cases will not be discussed any further here, but we note that none of them contested the presence of an L-Asn ligand in the presumed allosteric site (Fig. 2*c*). Despite their similarities to the EcAI structure (Fig. 2*a*), the related type I ASNases from *Y. pestis* (PDB entry 3ntx) and *Vibrio cholerae* (PDB entry 2ocd) did not contain ligands either in the active site or in the allosteric site. Investigations of *Y. pestis* ASNase were subsequently continued by another laboratory (Strzelczyk *et al.*, 2023), resulting in structures of acceptable quality (PDB entries 7r6a, 7r6b and 7r69) but without any ligands in the active or putative allosteric sites.

A set of 20 isomorphous structures of *B. paralicheniformis* ASNase provides an example of errors on the part of the depositors combined with limitations of the structure-validation process at the PDB. The structures have inconsistent descriptions in the PDB, with one batch (seven structures; Ran *et al.*, 2021) labeled as dimers with biological assembly evidence listed as 'gel filtration' and the other batch (13 structures, unpublished work) annotated as tetramers with the same 'gel filtration' evidence. Some of the titles of these entries in the PDB are not helpful to the user. They vary from the general and cryptic '*Crystal structure of BIAsnase*' (PDB entry 7cb4) to the misleading '*Blasnase-T13A with D-asn*' (PDB entries 7cbw and 7c8q), indicating a ligand that is not present in these particular structures.

Another problem on the examination of these 20 structures is that although the crystals were isomorphous, the protomers are scattered in different parts of the unit cell, making their comparison unnecessarily difficult (Fig. 3). This is unnecessary, especially since software helpful in the consistent placement of molecules in the asymmetric unit is available (*ACHESYM*; Kowiel *et al.*, 2014).

All 20 structures were refined at medium-to-high resolution (2.2–1.6 Å). Taking one of these structures (PDB entry 8h48) as an example, the values of the *R* factors (R = 0.180 and $R_{\rm free}$ = 0.204) would be acceptable at 1.8 Å resolution. Analysis of the stereochemistry with the *MolProbity* server (Williams *et al.*, 2018) assigns a score of 1.14, placing this structure in the top 1% of all PDB entries at a similar resolution. Thus, at face value it would appear that this structure is near the pinnacle in its category. However, a closer look at the deposited coordinates discloses obvious problems. Six Mg^{2+} ions and three formate (FMT) molecules have zero occupancy, although their temperature factors (or atomic displacement parameters, ADPs) are nonzero. Four additional FMT molecules with full occupancy have their ADPs set to 20.00 $Å^2$, suggesting that these parameters were never refined. Similarly, the ADPs of 27 water molecules are exactly 30.00 $Å^2$, again indicating that they were not refined. Some of these solvent/ligand molecules are located in clear $2F_o - F_c$ electron density, confirming their presence, whereas others are not supported by any difference electron density, thus questioning their modeling. Although atoms with zero occupancy are tagged in the PDB header, users of these coordinates would not be warned about this significant problem, *i.e.* about the fictitious status of the zero-occupancy atoms. The remaining 12 structures from the

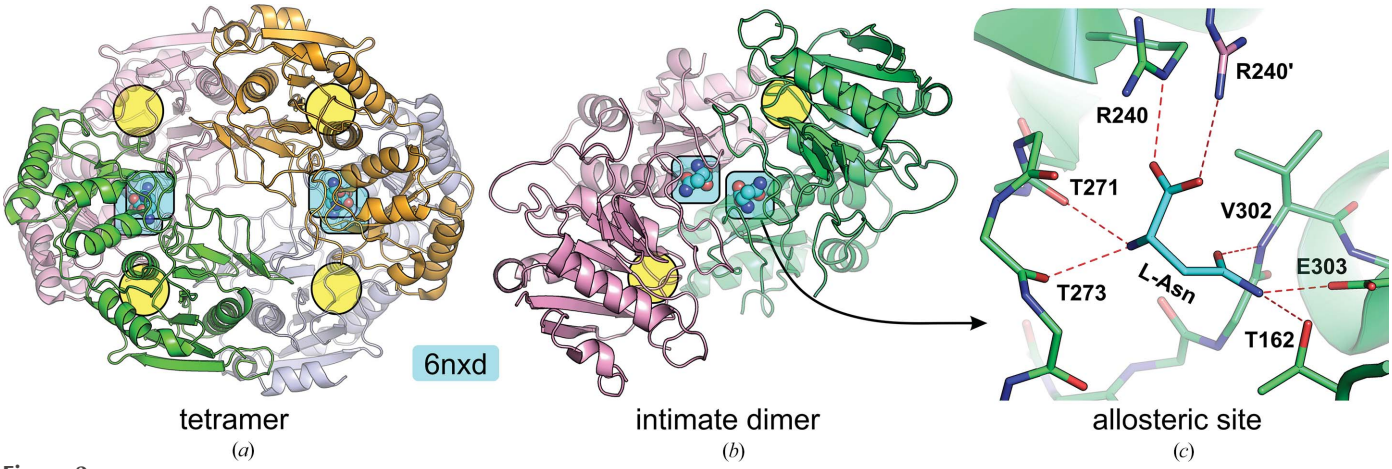


Figure 2
Crystal structure of the class 1 type I L-asparaginase from *E. coli* (PDB entry 6nxd). (*a*) Two intimate dimers (*A*–*B*, green–pink; *C*–*D*, violet–orange) make up the EcAI homotetramer. The four active sites are marked by yellow circles, while four allosteric sites are marked by cyan rectangles with L-Asn molecules in CPK representation. (*b*) The intimate dimer (chain *A*, green; chain *B*, pink) of EcAI possesses two active sites and two allosteric sites [marked as in (*a*)]. (*c*) Detailed view of the allosteric site of protomer *A* with a pattern of hydrogen bonds (red dashed lines) stabilizing the bound L-Asn (cyan) molecule; the primed residue Arg240′ belongs to the complementary (within the intimate dimer) subunit.

second series suffer from the same problems (Table 1), although no zero-occupancy atoms are present as in the earlier series. However, unrefined atoms are present in both batches.

The situation is only partially ameliorated by running PDB entry 8h48 through the automatic refinement procedure in *PDB-REDO* (Joosten *et al.*, 2011). Whereas the *R* factors reported by *PDB-REDO* are dramatically lower (0.135 and 0.165 for *R* and $R_{\rm free}$, respectively), this effect was achieved by utilizing anisotropic ADPs. Additionally, at the time of the original submission of this manuscript the *PDB-REDO* re-refinement did not correct the occupancy of the ligands, which remained at 0.0, although this has subsequently been rectified by *PDB-REDO* when the problem was brought to the attention of the developers (Joosten *et al.*, 2014). Our re-refinement, after setting all ligands to full occupancy and correcting the geometry around those Mg^{2+} ions that were clearly visible in the electron density, as well as the addition of ∼20 water molecules, resulted in *R* and $R_{\rm free}$ values of 0.161 and 0.188, respectively, with proper octahedral coordination for the four remaining Mg^{2+} ions. We did not re-evaluate the remaining 19 structures from this series, but we are confident that their refinement would yield similar results. In evaluating these type I structures we noticed that the tetramers corresponded closely to those of type II enzymes, as exemplified by EcAII (Fig. 2*a*), rather than those of type I, as seen for EcAI (Fig. 1*a*), further blurring the categorization of class 1 ASNases.

Although the ASNase from *Campylobacter jejuni* (PDB entry 3nxk; Center for Structural Genomics of Infectious Diseases, unpublished work) is annotated in the PDB as cytoplasmic (and thus presumably type I), its amino-acid sequence is much more similar to EcAII than to EcAI, and thus it should possibly be considered as type II (Lubkowski & Wlodawer, 2021). This structure itself was determined as part of a structural genomics project at medium resolution (2.4 Å) and does not appear to exhibit significant problems.

3.1.1.2. Structures of dimeric class 1 L-asparaginases from thermophilic archaea

The crystal structure of the *Pyrococcus horikoshii* ASNase was determined independently in two laboratories as part of a structural genomics project (PDB entry 1wnf; RIKEN Structural Genomics/Proteomics Initiative, unpublished work) and by Yao and coworkers (PDB entry 1wls; Yao *et al.*, 2005). The crystals used in both investigations were obtained under similar conditions and were isomorphous and, due to the high pH, there were no ligands bound in the active sites. Whereas the dimers were placed differently in the unit cell in the two models, superposition of their coordinates resulted in an r.m.s.d. of 0.38 Å for the C^{α} atoms. The only significant difference was found for residues 51–52 in protomer *A*, which were traced differently in the two models, but their conformation was supported by the respective electron-density maps. Since the diffraction data for PDB entry 1wls extended to higher resolution, this structure was re-refined by us. Adjustments mostly involved changing unlikely side-chain rotamers to their more energetically preferred orientations (as indicated by *Coot*) that still agreed with the electron-density map, and the deletion of residues *A*35–*A*39 which had been modeled with unacceptable geometry and for which there was no clear electron density. The resulting model has a considerably better overall geometry (r.m.s.d. for bonds of 0.009 versus 0.018 Å) and its $R_{\rm free}$ decreased from 0.253 to 0.240. The quaternary structure of this ASNase is clearly dimeric, since none of the tetramers seen in other type I enzymes are seen in these crystals.

The structure of the closely related ASNase from *P. furiosus* was determined for the wild-type enzyme (PDB entry 4q0m), as well as for different constructs (PDB entries 4nje, 4ra6, 4ra9, 5cbp, 5b74 and 5b5u) in which the linker between the N- and C-terminal domains was removed (Tomar *et al.*, 2014; Sharma *et al.*, 2020). The resolution of these structures varied between 2.05 and 2.6 Å and their quality was satisfactory in the PDB validation report. Nevertheless, we still re-refined the structure of the wild-type enzyme after improving a number of side-chain rotamers, adding 15 clearly defined water molecules and replacing the (4*S*)-2-methyl-1,4,5,6-tetrahydropyrimidine-4-carboxylic acid ligand (which was not listed as a component of the crystallization buffer and does not fit the electron density) by two water molecules. These changes led to a decrease in the *R* and $R_{\rm free}$ by about 0.02 each, without any major modifications of the protein model.

Although only a monomer was present in the asymmetric unit of all of these crystals, not all of them were isomorphous, but those that were (for instance PDB entries 5b74 and 5b5u) were placed differently in the unit cell without any regard for consistency. These structures provide a clear example of why such inconsistent placement leads to difficulties in comparing small differences that might be quite important. Both PDB entries 5b74 and 5b5u contained an additional four-residue peptide that was able to restore the activity of enzyme in which the linker residues 179–182 were removed. Two such peptides were visible in PDB entry 5b74: one in the location of the deleted sequence and another (also present in PDB entry 5b5u) forming intermolecular contacts. However, the latter tetrapeptide in PDB entry 5b74 included a very unlikely nonproline *cis*-peptide bond, most likely due to incorrect modeling of the side chain of the adjacent leucine residue. After bringing these structures to a consistent location in the unit cell with the *ACHESYM* server (Kowiel *et al.*, 2014), PDB entry 5b74 was re-refined and the resulting geometry of the peptide in question no longer violated the rules of stereochemistry.

Only a single 2.2 Å resolution structure (PDB entry 5ot0) is available for the related ASNase from *Thermococcus kodakarensis* (Guo *et al.*, 2017). Although the refinement statistics are acceptable for a structure at this resolution (*R* and $R_{\rm free}$ of 0.195 and 0.224, respectively), the large number of RSRZ and Ramachandran outliers put the quality of this structure into question. According to a remark on the *PDB-REDO* site, the $R_{\rm free}$ factor might be biased, and its correct, unbiased value was estimated to be 0.254. The main problem with PDB entry 5ot0 is the lack of electron density, indicating significant

disorder, in some of the six protomers in the asymmetric unit that were nevertheless modeled. The disordered regions in protomers *A*–*D* and *G* were mostly limited to the loop containing residues 37–40, whereas they were very extensive in protomer *F*, for which the electron density was very fragmented for a large part of the main chain. The parts of the structure that were in good electron density did not suffer from any problems, and neither *PDB-REDO* nor our re-refinement (including deletion of the 27–40 loop) changed them in any significant way. Thus, although the original structure was refined to the limit of the available data, the modeling of the disordered regions was over-optimistic.

3.1.1.3. Structures of eukaryotic class 1 L-asparaginases

The properties of the class 1 ASNase from guinea pig (*Cavia porcellus*) are rather unusual. Although this enzyme does not contain a signal peptide and thus is not secreted, a property associated with type I ASNases, its tight binding of a substrate (with a micromolar K_m) is mostly found in type II enzymes. It had been postulated that this was the agent responsible for the original discovery by Kidd (1953) of the anticancer properties of ASNases (Schalk *et al.*, 2014). Four isomorphous structures of the wild-type enzyme and its active-site mutants (PDB entries 4r8l, 5dnc, 5dnd and 5dne) contain a single tetramer in the asymmetric unit, whereas one (PDB entry 4r8k) contains two tetramers. All of these medium-resolution (2.2–2.4 Å) structures do not present any major concerns.

3.1.2. Class 1 type II (EcAII-type) enzymes

3.1.2.1. Structural studies of EcAII

The first structure of any L-asparaginase was that deposited in the PDB in 1993 for EcAII (PDB entry 3eca; Swain *et al.*, 1993; Fig. 1). The diffraction data for this medium-resolution (2.4 Å) structure were collected on an in-house system equipped with a multiwire area detector; thus, their quality was moderate by current standards and is not comparable to that of data collected subsequently using synchrotron light sources. The structure refinement was also inferior compared with current standards; for example, it lacked cross-validation by R_{free} (Brünger, 1992). The crystals used in these investigations were grown from the commercially available drug Elspar (Merck). In the absence of information about the amino-acid sequence of this variant of EcAII, it was assumed that the amino-acid sequence corresponded to that from the standard K12 strain of *E. coli*. However, it was found 25 years later that the actual sequence differs from that from the K12 strain at four posi-

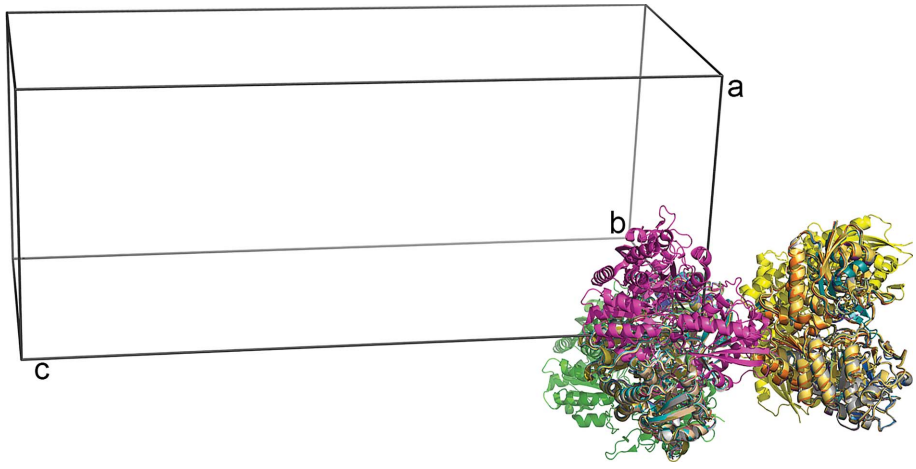


Figure 3
Tracing of the main chains of the 20 isomorphous depositions of *B. paralicheniformis* ASNase. The coordinates were placed in different locations (marked in different colors) in the tetragonal unit cell in space group $P4_12_12$. The asymmetric unit contains only a tight dimer of *B. paralicheniformis* ASNase.

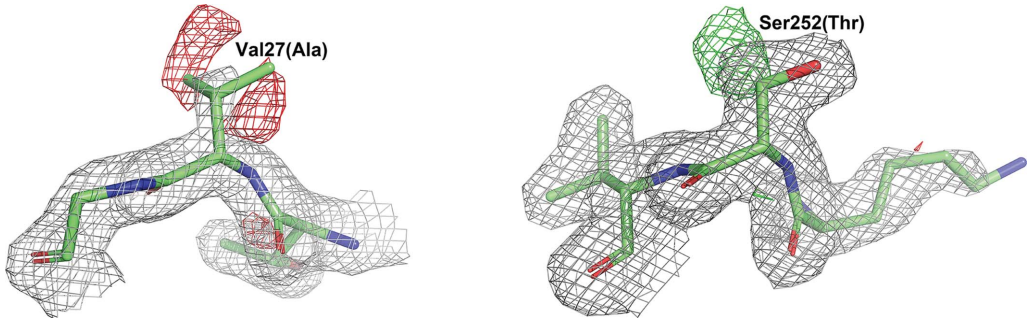


Figure 4
$2F_o - F_c$ (gray, contour 1.0σ) and $F_o - F_c$ (green/red, +/−3.0σ) electron-density maps for the originally deposited structure with PDB code 1nns (sticks), covering two residues that differ in the sequences of EcAII from the K-12 strain of *E. coli* and Elspar (the latter in parentheses).

tions (substitutions V27A, N64D, S252T and T263N). Ultimately, the coordinates of PDB entry 3eca were replaced with the version corresponding to the corrected sequence. These comments are meaningful since PDB entry 3eca is still used as a reference for type II ASNases.

The structure of Elspar was subsequently determined at a higher resolution (1.95 Å) using crystals in a different space group (Sanches *et al.*, 2003). This structure (PDB entry 1nns) is of generally high quality, although it could be further improved by correcting the rotamers of a significant number of side chains and making small corrections to the solvent structure. These corrections did not impact the values of the *R* factors, which were already very low. Importantly, as a result of the higher resolution and good quality of the diffraction data, the electron-density maps very clearly indicated substitutions at three of the four positions (V27A, S252T and T263N), two of which are shown in Fig. 4. The structure with the corrected sequence was subsequently re-refined by us.

Three more structures of commercial variants of *E. coli* ASNase were released in 2020 and described in a subsequent publication (de Araújo *et al.*, 2021). The structure of Aginasa was only determined in the absence of ligands in the active sites (PDB entry 6uoh), whereas structures of Leuginase were determined with L-Asn in the active sites (PDB entry 6uog) and for the ligand-free enzyme (PDB entry 6uod). At the time of the original submission of this manuscript the validation reports for PDB entries 6uog and 6uod on the RCSB PDB site mentioned that validation with *MolProbity* (Williams *et al.*, 2018) had failed, but this has subsequently been corrected. No similar problems were reported for PDB entry 6uoh. Although the geometry of all three models was highly idealized, some imperfections were still apparent. The asymmetric unit of PDB entry 6oug contains two homotetramers. Surprisingly, many side-chain rotamers of the eight protomers were modeled very differently, although the electron density did not support such variability. Unification of the conformation of most side chains, the imposition of noncrystallographic symmetry (NCS) restraints and the addition of 75 water molecules that were clearly indicated by the difference electron-density maps led to a decrease in R and $R_{\rm free}$ by about 0.02 each and to an overall better model. PDB entries 6uoh and 6uod were originally refined with *Phenix* using both $I(+)$ and $I(-)$ intensities, although no detectable anomalous signal was present in the diffraction data. However, the coordinates were in good agreement with the electron-density maps and these structures did not require any significant modifications. Generally, details of the amino-acid sequences of commercial preparations are not reported, and rarely are they experimentally determined prior to structural studies, as was the case for Elspar; thus, it is possible that the sequences in PDB entries 6uoh, 6uog and 6uod contain some discrepancies.

The structure of the *E. coli* type II ASNase released by the PDB in 2021 (PDB entry 6v5f) provides an example of a variety of problems resulting from actions by both the depositors and the PDB itself. There is still no publication associated with this deposition, thus it is not possible to assess some of the experimental details. The title of a putative publication '*L-asparaginase from Escherichia coli: Novel insights about the behavior of flexible part of the active site*' might suggest that the aim of this deposition was to study the wild-type enzyme with ligands in the active site. The asymmetric unit contains a homotetramer with four active sites. In two of them, the authors modeled the product L-aspartate (with unknown provenance, since it was not listed in the crystallization media), whereas the other two active sites were assigned by the authors as occupied by succinic acid (presumably contributed by the crystallization buffer). Although the claimed resolution is 2.1 Å, the outermost data shell has an $\langle I/\sigma(I)\rangle$ of 0.33 and an $R_{\rm r.i.m.}$ of 3.4 (*i.e.* 340%), indicating that this nominal resolution limit was quite optimistic. At the time of the original submission of this manuscript the validation report provided by the RCSB PDB indicated that the execution of both *MolProbity* and *EDS* (Kleywegt *et al.*, 2004) had failed, thus the structure was presented without proper validation. This problem has subsequently been remedied by the RCSB PDB. No such problems were present on the PDBe site, highlighting that in some cases the validation information on different PDB sites is not the same. We were able to re-refine this structure and, although the R factors were not changed significantly, we could correctly trace the C-terminus of protomer *D*, fix the orientation of the incorrectly placed L-Asp ligands, restore the side-chain atoms that were reported with zero occupancy and remove one succinic acid molecule that had no electron density. Analysis of the electron-density map resulting from our re-refinement strongly suggested that the amino-acid sequence of the crystallized protein is likely to correspond to the sequence of Elspar rather than the standard *E. coli* K12 strain (*i.e.* the sequence assumed by the authors). The structure of the N24S variant of EcAII (PDB entry 5mq5; Maggi *et al.*, 2017) has a rather unusual history. The original deposition reported that this high-resolution (1.6 Å) structure had been refined after dividing the model into 19 TLS groups, but the occupancies of the L-Asp ligands in all of the active sites were set to 0.0. The values of R and $R_{\rm free}$ were 0.156 and 0.190, respectively. These coordinates were later replaced by a file that claimed that TLS parameters were not used, and the R and $R_{\rm free}$ values were now 0.147 and 0.198, respectively. However, all coordinates and thermal parameters are identical in the two versions of the deposition, with the only difference being the occupancy of the L-Asp ligands, which was now set to 1.0. Interestingly, the PDB validation report shows the R factors calculated by the PDB to be almost identical to those in the original deposition, and not in the redeposition. The $F_{\rm o} - F_{\rm c}$ maps in the area of the ligands have both positive and negative density, indicating that the ADPs for the ligand atoms are incorrect. Our re-refinement of this structure without any rebuilding and using just isotropic ADPs (but after adjusting the placement of molecules in the cell with *ACHESYM*; see below), resulted in an R and $R_{\rm free}$ of 0.160 and 0.193, respectively. The $F_{\rm o} - F_{\rm c}$ maps near the ligands are now clean.

PDB entry 7p9c for wild-type EcAII (Maggi *et al.*, 2021) is isomorphous to PDB entry 5mq5, but the arrangement of the protomers in the asymmetric unit was different. This structure

is a replacement for the initially deposited PDB entry 6yzi. With a resolution of 1.6 Å and very good geometry, it did not require much change other than shifting the coordinates to the same location as in PDB entry 5mq5, deletion of a number of residues in the ASFL for which no convincing electron density was present and the addition of a few atoms at both the N- and C-termini. Although the diffraction data were comparable to those for PDB entry 5mq5, this structure was refined (and also re-refined) with isotropic ADPs and re-refinement did not change the *R* factors in a significant way.

The coordinates of the unpublished isomorphous structure with PDB entry 7a0u were also shifted to a common location and their refinement, after minor adjustments, decreased the *R* factors by ∼0.01. The coordinates of PDB entry 7r5q (Maggi & Scotti, 2022) were treated in an analogous manner and they were re-refined after the addition of several residues that could be modeled in the electron density. The same treatment was applied to PDB entry 7r57 (Maggi & Scotti, 2022) which, despite its high resolution (1.4 Å), had highly disordered ASFLs.

Some of the EcAII-type structures were not re-refined due to significant disorder of the ASFLs (PDB entry 7m11) or poor resolution and incomplete modeling (PDB entry 6eok; Cerofolini *et al.*, 2019). Similarly, the medium-resolution crystal structures of the D90E variant of EcAII (PDB entries 1ihd, 1jaz and 1jja; Borek *et al.*, 2014), structures of the inactive D90T/K162T mutant determined at different pH values (PDB entries 6nxa, 6nx6, 6nx7, 6nx8 and 6nx9) and the structure of the wild-type enzyme complexed with citrate (PDB entry 6nxb) (Lubkowski *et al.*, 2019) did not display any significant problems and were not analyzed here in more detail. Finally, two large series of structures of EcAII and its variants, one supporting the double-displacement catalytic mechanism of this enzyme (PDB entries 6pa2, 6pa3, 6pa4, 6pa5, 6pa6, 6pa8, 6pa9, 6paa, 6pab, 6pac and 6pae; Lubkowski & Wlodawer, 2019) and the other the catalytic mechanism with covalently bound acyl-enzyme and tetrahedral intermediates (PDB entries 6v23, 6v24, 6v25, 6v26, 6v27, 6v28, 6v29, 6v2a, 6v2b, 6v2c and 6v2g; Lubkowski *et al.*, 2020), posed no problems and were not analyzed further. As many of these structures contained clearly visible reaction intermediates bound covalently to Thr12, they made the previously determined structure of such a complex (PDB entry 4eca; Palm *et al.*, 1996) obsolete.

3.1.2.2. Structural studies of other bacterial type II ASNases

Structures of *Dickeya dadantii* ASNase (for historical reasons still called ErA here and in the literature) are of particular importance, since this enzyme is approved for therapeutic use as a second-line drug for ALL. Two isomorphous structures of ErA were determined at medium resolution with data collected at room temperature (PDB entry 1hfj) or using vitrified crystals (PDB entry 1hfk) (Jaskólski *et al.*, 2001). Neither structure exhibits significant problems and since both have been superseded by structures at much higher resolution, they have not been evaluated further.

Several other structures of ErA were determined using isomorphous crystals in space group *C*2. All of these structures were placed in the asymmetric unit in a consistent way. These structures represented complexes with D-Asp (PDB entry 1hg1), L-Glu (PDB entry 1hfw) and succinic acid (PDB entry 1hg0) (Aghaiypour *et al.*, 2001*a*), L-Asp (PDB entry 6pae; Lubkowski & Wlodawer, 2019), as well as with covalently bound 6-hydroxy-D-norleucine (PDB entry 1jsl) and 6-hydroxy-L-norleucine (PDB entry 1jsr) (Aghaiypour *et al.*, 2001*b*). These structures, determined at resolutions ranging from 1.6 to 1.9 Å, do not present significant problems and thus were not re-refined. An atomic resolution structure isomorphous to those above (PDB entry 1o7j) is discussed separately below.

Two isomorphous structures in space group $P2_12_12_1$ were determined of complexes with L-Asp (PDB entry 5f52) and L-Glu (PDB entry 5hw0) (Nguyen *et al.*, 2016*a*). The quality of these high-resolution structures is very good, although significant disorder of the ASFL is present in the latter structure. The same is true for two other structures isomorphous with those above (PDB entries 5i3z and 5i48) that describe the L-Asp complexes of the E63Q and A31I, E63Q variants of ErA with diminished L-Glu activity (Nguyen *et al.*, 2016*b*). Another structure in the latter series, of the L-Asp complex of the E63Q, S254N variant of ErA (PDB entry 5i4b), is not isomorphous with the structures above and also does not exhibit any quality problems.

In an interesting study (Van Trimpont *et al.*, 2022), a 19-residue albumin-binding domain (ABD) was attached to the N-terminus of a triple mutant of ErA with reduced glutaminase activity. The crystal structure of this construct at 1.75 Å resolution, with a homotetramer in the asymmetric unit, was described very briefly in the above publication and deposited as PDB entry 7u6m. A large number of water molecules that are clearly seen in the electron-density map were not modeled in the original deposition. Inclusion in our re-refinement of 21 water molecules found among the highest $F_o - F_c$ peaks improves *R* and R_{free} from 0.142 and 0.167 (0.144 and 0.166 in the PDB deposition) to 0.140 and 0.166, respectively. The ABD domain was only modeled in subunit *C* of the homotetramer. This modeling is quite convincing, except for the first 13–14 residues, which we removed in the re-refined structure. In the original deposition of PDB entry 7u6m, the L-Asp molecules in the four active sites were modeled in excellent electron density.

In the deposited medium-resolution structure of ErA (PDB entry 8wgq; F. Wang, W. Cheng, Z. Lv, C. Ju & J. Wang, unpublished work), there are no ligands present in the active site and the ASFL is disordered in three protomers, while it is fixed by crystal contacts in the fourth protomer. A number of unmodeled electron-density peaks are visible in the difference maps, but otherwise there are no concerns regarding this deposition and it was not re-refined.

Four structures of ASNase from *Erwinia carotovora* (now renamed *Pectobacterium carotovorum*) are available in the PDB. Two of them represent independent determination of the structure of the complex with L-Asp. The crystals used to obtain PDB entries 2gvn (Kravchenko *et al.*, 2008) and 2jk0

(Papageorgiou *et al.*, 2008) were isomorphous in space group $P2_12_12_1$, but the contents of the asymmetric unit (two tetramers) were described differently. PDB entry 2gvn (resolution 1.9 Å) was re-refined after shifting the coordinates with *ACHESYM*, renaming the protomers *ABCD* and *EFGH*, and removing the ASFLs (residues 21–32) from all subunits, with the exception of protomer *H*. These actions reduced R and R_{free} from 0.206 and 0.229 to 0.173 and 0.203, respectively. PDB entry 2jk0 was shifted in the cell to correspond to PDB entry 2gvn and was subjected to one refinement cycle without rebuilding. Even though no TLS parameters were used in the re-refinement, the R and R_{free} changed from 0.192 and 0.266 to 0.205 and 0.229, respectively. All eight protomers of these two structures can be superimposed with an r.m.s.d. of 0.55 Å. A low-resolution structure of the unliganded enzyme (PDB entry 1zcf; Kislitsyn *et al.*, 2006), isomorphous to the other two structure, was shifted to the same frame and re-refined without any rebuilding, resulting in R and R_{free} being lowered by ∼0.01 each.

The first structure of *Wolinella succinogenes* ASNase (PDB entry 1wsa; Lubkowski *et al.*, 1996) was deposited in 1996, but unfortunately no structure factors are available and thus this deposition cannot be properly evaluated. Four subsequently determined structures (PDB entries 5k3o, 5k45, 5k4g and 5k4h; Nguyen *et al.*, 2017) of the two natural variants of the enzyme, in which residue 121 is either Ser or Pro, were determined at a resolution of 1.6–2.0 Å. The first three are isomorphous in space group *C*2, but the description of the asymmetric unit was different in each of them. These structures were re-refined after they had been brought to a common placement in the unit cell with *ACHESYM*. We noticed that the 5% of free reflections (over 8400) in PDB entry 5k4g must not have corresponded to those used in the refinement since, despite the difference between R and R_{free} reported in the PDB file itself, their calculated values were identical (this is also obvious from the PDB validation report). Adjustment of the conformation of several side chains and a single round of refinement resulted in an R and R_{free} of 0.165 and 0.192, respectively, with better overall geometry. A single round of refinement of PDB entries 5k3o and 5k45 (without rebuilding) improved their geometry slightly but did not significantly affect the R factors. PDB entry 5k4h did not require any significant modification and was not re-refined.

Four structures of the V23Q/K24T/S121P variant of *W. succinogenes* ASNase (PDB entries 6rud, 6ruf, 6rue and 6syh; Timofeev *et al.*, 2020*a*,*b*) have been published. Two of them, PDB entries 6rud and 6ruf, were isomorphous, although the placement of the molecules in the unit cell had to be unified with *ACHESYM*, whereas the other two were in different crystal forms. PDB entry 6rud was re-refined after the adjustment of several rotamers, the addition of Ala2 at the N-terminus of each subunit, the addition of several residues in the ASFLs and the addition of an astonishing number of water molecules (over 350). This resulted in a decrease of both R and R_{free} from 0.162 and 0.194 to 0.143 and 0.174, respectively, and an improvement in the geometry of the model. PDB entry 6ruf was not re-refined but only shifted to a position corresponding to PDB entry 6rud. We previously found that PDB entry 6syh contained unambiguously visible molecules of HEPES in the vicinity of the active site of each protomer (Fig. 5*a*), although the use of this buffer was not mentioned either in the publications or in the PDB deposition (Lubkowski & Wlodawer, 2021). Re-refinement of this structure decreased R and R_{free} from 0.142 and 0.187 to 0.113 and 0.161, respectively. Interestingly, even after the addition of a few residues to the ASFL in these models, the two target mutations (residues 23 and 24) were still not visible.

Two isomorphous structures of ASNase from *Helicobacter pylori* were determined either at 100 K (PDB entry 2wlt, 1.4 Å resolution) or at room temperature (PDB entry 2wt4, 1.8 Å resolution) (Dhavala & Papageorgiou, 2009). Both structures are of high quality with low R factors (R and R_{free} of 0.131 and 0.169 for PDB entry 2wlt and 0.153 and 0.185 for PDB entry 2wt4, respectively) and both include L-Asp bound in the active site, although the electron density of the ASFL of the single protomer present in the asymmetric unit is weak and broken in places. PDB entry 2wlt was re-refined by us, with the main correction consisting of an adjustment of the amino-acid sequence from Glu137 to Asn and Val163 to Ala. The electron density unquestionably supported these changes in both structures. Strangely, Fig. 2 in the original manuscript (Dhavala & Papageorgiou, 2009) identified residue 137 as tyrosine, although another strain listed in this figure (HpA-CCUG) identified the two abovementioned sites (137 and 163) as in our reinterpretation. The ASNase gene used by the authors to prepare the expression construct was taken directly from the genomic DNA of *H. pylori* and thus there must have been some discrepancy in the identification of the specific strain, despite the claim that the expression constructs were

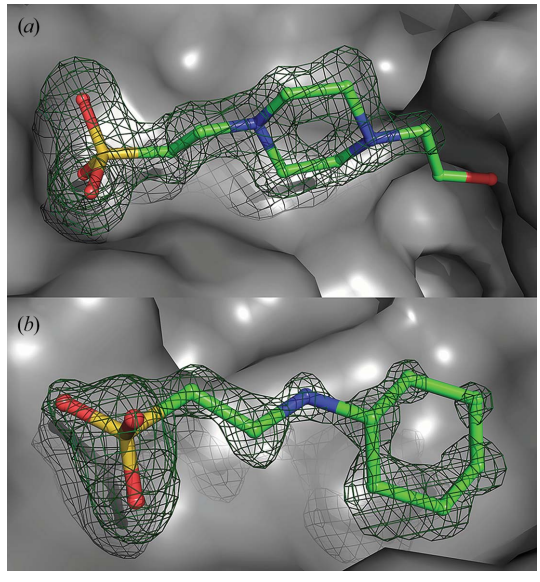


Figure 5
Missed buffer molecules rediscovered in the re-refined structures. (*a*) A HEPES molecule in PDB entry 6syh. (*b*) A CHES molecule in PDB entry 1o7j. In both panels ball-and-stick models of the buffer molecules are shown on the background of the protein surface (gray). The OMIT electron-density maps (dark green mesh) are contoured at the 2.5σ level.

sequenced. Refinement of the modified PDB entry 2wlt model resulted in an R and R_{free} of 0.112 and 0.141, respectively, which were significantly lower than in the original PDB deposition (0.131 and 0.169, respectively). This example emphasizes the importance of trusting the electron density in well refined structures and the need to question the identity of buried residues that do not agree with the electron-density maps.

Only a single, low-resolution (2.9 Å) structure of glutaminase–asparaginase from *Acinetobacter glutaminasificans* (PDB entry 1agx; Lubkowski, Wlodawer, Housset *et al.*, 1994) is present in the PDB. This structure is primarily of historical value, since it has not been refined to the current standards and its quality is low. Removal of the ASFL, for which there is no proper electron density, the reorientation of a number of side-chain rotamers and the addition of 11 water molecules decreased R from 0.171 to 0.133, with a vastly improved overall geometry. This large decrease may also be due to the use of individual ADPs rather than the group ADPs that were applied during the original refinement.

Another glutaminase–asparaginase that belongs to the type II class 1 enzymes is PGA from *Pseudomonas glutaminasificans*. Its first structure in a ligand-free form (PDB entry 3pga; Lubkowski, Wlodawer, Ammon *et al.*, 1994) is very much below the current standards, although the R factor is 0.165 at 2 Å resolution. The deposited data are only 47% complete, whereas the reported completeness is 86%, most likely due to truncation of the structure-factor file during deposition. The coordinates of the tetramer in the asymmetric unit include the ASFL modeled in both the open and closed conformations at zero occupancy. The model has very poor geometry, as shown by the validation diagram provided by the PDB. We re-refined the structure after removing the ASFLs from three protomers (the open conformation was still modeled in protomer *B*, although the electron density is fragmented), correcting almost 100 side chains to better rotameric conformations and removing many water molecules that were not supported by the electron density. The re-refinement led to a decrease in R to 0.122, but due to the questionable quality of the diffraction data even this model is not up to current standards. Three subsequently deposited structure files for this ASNase (PDB entries 4pga, 1djp and 1djo; Jakob *et al.*, 1997; Ortlund *et al.*, 2000) were not accompanied by structure factors, so they were not evaluated by us.

PDB entry 6wz6 represents the highest resolution structure from a series aiming to elucidate the catalytic mechanism of PGA (Strzelczyk *et al.*, 2020). This structure of the covalent acyl-enzyme intermediate between the PGA K173M mutant and L-glutamate was determined at 1.15 Å resolution, demonstrating that L-Asn and L-Gln are hydrolyzed by type II ASNases according to the same mechanism. The relatively high quality of the experimental data (R_{merge} = 0.060, $R_{p.i.m.}$ = 0.025, $CC_{1/2}$ = 0.997 at the highest resolution), as well as the low values of R and R_{free} (0.100 and 0.120, respectively), suggest that this should be a high-quality structure. Indeed, the electron density in the active-site area, *i.e.* the most interesting region of this structure, is excellent and the model fits it very well. However, in other parts of the structure we found several significant errors. For instance, Met137*B*, Arg153*B* and Glu180*C*, all modeled in multiple conformations, had total occupancies higher than 1.0. Asn284*D* had the main chain modeled in two conformations and the side chain in just one, a situation that we find highly unusual. The side chains of many residues were modeled as incorrect rotamers as identified by the *MolProbity* server. Over 100 water molecules were modeled with no clear electron density and many side chains at or near the molecular surface were severely disordered, thus preventing reliable modeling. We addressed these problems in a re-refinement that, despite clear improvement of the model, increased both R and R_{free} by 0.010 each. Other structures in this series (PDB entries 6wyw, 6wyx, 6wyy, 6wyz, 6wz4 and 6wz8) did not require re-refinement.

3.1.2.3. Re-refinement of the highest resolution ErA structure: PDB entry 1o7j

The structure of the clinically important *E. chrysanthemi* L-asparaginase ErA was solved early on and described in a publication (Miller *et al.*, 1993) but, for reasons that have now faded into oblivion, the 1.8 Å resolution model was never deposited in the PDB. Since the protein crystallized well (from CHES buffer with ammonium sulfate as precipitant) and the crystals diffracted X-rays to high resolution, about a decade later the structure was redetermined at the then record-breaking resolution of 1.00 Å, refined (with *SHELXL*; Sheldrick, 2015) to an R and R_{free} of 0.110 and 0.128, respectively, deposited in the PDB (PDB entry 1o7j) and published (Lubkowski *et al.*, 2003). The first issue that we encountered when inspecting PDB entry 1o7j was the erroneous crystallization conditions described in the PDB, specifying Tris instead of CHES (*N*-cyclohexyl-2-aminoethanesulfonate) buffer. This seemingly trivial (but misleading and frustrating reproducibility) mistake turned out to be essential when we noticed modeling problems at multiple sulfate sites present in the crystal structure. The four sulfate anions occupying the four active sites of the ErA homotetramer in the asymmetric unit had excellent electron density, but most of the remaining anions were either not supported by the electron density or were sitting, together with clusters of close water sites, in extended patches of electron density (both $F_o - F_c$ and $2F_o - F_c$), which could be easily interpreted as belonging to seven CHES molecules (Fig. 5*b*), some with partial occupancy. The original structure also included three PEG molecules [with a maximum length of five $-O(CH_2)_2-$ units, forming typical 'circles' around lysine NH_3^+ groups], with problematic (fragmented) modeling. Corrections of these molecules (consolidation and/or extension) was accompanied by the inclusion of six additional PEG molecules, with lengths ranging from three to ten units, for which sufficiently convincing electron density (sometimes interpreted as water in the original model) could be found. For the interpretation of ambiguous $2F_o - F_c$ maps we used $3F_o - 2F_c$ maps, which in many instances allowed a clearer interpretation.

Additional reinterpretation of the solvent area (sulfate ions, water molecules) was followed by re-refinement using the

Table 3
Structures of class 2 type III L-asparaginases deposited in the Protein Data Bank as of 22 November 2023.

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No. of missing residues‡	Problems identified	Reference
E. coli	1t3m	WT	1.65	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	NO_3^-/nitrate ion (NO3)	20*A*/20*B*/12*C*/12*D*	Incorrect numbering of residues (shift by −1)	Prahl *et al.* (2004)
E. coli	2zal	WT	1.90	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	L-Asp (ASP)	3*C*/1*B*/1*D*	Historically, PDB entry 2zal replaces PDB entry 1seo§	Michalska *et al.* (2005)
L. luteus	2gez	WT	2.60	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	Cl^-/chloride ion (CL)	36*A*/28*C*/29*E*/36*G*/0*B*/0*D*/0*F*/0*H*	—	Michalska *et al.* (2006)
E. coli	1jn9	WT	2.30	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	18*A*/19*C*/8*B*/7*D*	—	Michalska, Borek *et al.* (2008)
E. coli	1k2x	WT	1.65	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	20*A*/20*C*/7*B*/7*D*	—	Michalska, Borek *et al.* (2008)
E. coli	2zak	T179A	2.01	Homodimer (uncleaved precursor)	—	20*A*/25*B*	—	Michalska, Borek *et al.* (2008)
E. coli	3c17	T179A	1.95	Homodimer (uncleaved precursor)	Cl^-/chloride ion (CL)	15*A*/12*B*	—	Michalska, Borek, Hernández-Santoyo & Jaskolski (2008)
H. sapiens	3tkj	T168A	2.30	Homodimer (uncleaved precursor)	SO_4^{2-} (SO4)	21*A*/22*B*	—	Li *et al.* (2012)
H. sapiens	4et0	WT	3.30	Homodimer (circularly permuted)	—	30*A*/27*B*	Low resolution, lots of unmodeled blobs	Li *et al.* (2012)
H. sapiens	4pvp	WT	1.85	Mixture of mature and immature protein	Malonate (MLI)	11*A*/13*B*	Zero occupancy of Phe236; no information about partial cleavage in the structure title	Nomme *et al.* (2012)
H. sapiens	4pvq	WT	2.13	Mixture of mature and immature protein	SO_4^{2-} (SO4), I^- (I)	14*A*/12*B*	Ramachandran outliers; no information about partial cleavage in the structure title	Nomme *et al.* (2012)
H. sapiens	4pvr	WT	1.75	Mixture of mature and immature protein	L-Asp (ASP)	15*A*/13*B*	Some Ramachandran outliers	Nomme *et al.* (2012)
H. sapiens	4pvs	WT	1.84	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	L-Asp (ASP)	13*A*/13*B*	—	Nomme *et al.* (2012)
H. sapiens	4osx	WT	1.95	Homodimer (uncleaved precursor)	Gly (GLY)	8*A*/8*B*	—	Su *et al.* (2013)
H. sapiens	4osy	WT	1.91	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	Gly (GLY)	13*A*/13*B*	—	Su *et al.* (2013)
H. sapiens	4o0c	WT	1.50	Homodimer (uncleaved precursor)	—	12*A*/12*B*	Incorrectly modeled Na^+	Nomme *et al.* (2014)
H. sapiens	4o0d	T168S	1.95	Homodimer (uncleaved precursor)	Gly (GLY)	9*A*/10*B*	Some Ramachandran outliers	Nomme *et al.* (2014)
H. sapiens	4o0e	T186V	1.71	Homodimer (uncleaved precursor)	—	11*A*/11*B*	—	Nomme *et al.* (2014)
H. sapiens	4o0f	T219A	1.92	Homodimer (uncleaved precursor)	Gly (GLY)	12*A*/13*B*	Incorrectly modeled Na^+	Nomme *et al.* (2014)
H. sapiens	4o0g	T219V	2.10	Mixture of mature and immature protein		13*A*/12*B*	—	Nomme *et al.* (2014)
H. sapiens	4o0h	WT	1.97	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	IAS (L-Asp covalently bound)	13*A*/13*B*	Poor density for covalently bound ligand	Nomme *et al.* (2014)
C. porcellus	4o47	WT	1.90	Homodimer (uncleaved precursor)	—	36*A*/40*B*	Unmodeled density in the active site; high Matthews coefficient of 3.83 $Å^3$ Da^{-1}	Schalk & Lavie (2014)
C. porcellus	4o48	WT	2.29	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	L-Asp (ASP)	33*A*/33*B*	High Matthews coefficient of 3.83 $Å^3$ Da^{-1}	Schalk & Lavie (2014)
P. vulgaris	4pu6	WT	2.30	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	L-Asp (ASP)	39*A*/39*C*/0*B*/0*D*	—	Bejger *et al.* (2014)
P. vulgaris	4pv2	WT	1.79	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	38*A*/41*C*/0*B*/0*D*	—	Bejger *et al.* (2014)
P. vulgaris	4pv3	WT	2.09	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	39*A*/41*C*/0*B*/0*D*	—	Bejger *et al.* (2014)
H. sapiens	4zm9	WT	2.51	Homodimer (circularly permuted)	Citrate (FLC), dithiothreitol (DTT), Gly (GLY)	19*A*/19*C*/19*B*/19*D*	Poor density for DDT and FLC; incorrectly modeled Na^+; several unmodelled waters and empty density blobs; some geometric issues in the side chains	Li *et al.* (2016)
E. coli	7qsf	G206C/R207T/D210A/S211A	1.60	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	18*A*/24*C*/7*B*/8*D*	—	Loch *et al.* (2022)

Table 3 (continued)

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No. of missing residues‡	Problems identified	Reference
E. coli	7qtc	G206H/R207T/D210P/S211Q	2.55	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	33*A*/35*C*/7*B*/8*D*	—	Loch *et al.* (2022)
E. coli	7qvr	G206S/R207T/D210S	1.90	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	16*A*/22*C*/7*B*/7*D*	—	Loch *et al.* (2022)
E. coli	7qy6	WT	1.65	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	23*A*/22*C*/7*B*/7*D*	—	Loch *et al.* (2022)
E. coli	7qym	R207V/D210P/S211W	1.20	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	21*A*/18*C*/7*B*/7*D*	—	Loch *et al.* (2022)
E. coli	7qyx	R207A/D210S/S211T	1.85	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	28*A*/33*C*/8*B*/7*D*	—	Loch *et al.* (2022)
E. coli	7r1g	R207C/D210S/S211V	1.95	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	19*A*/26*C*/7*B*/7*D*	—	Loch *et al.* (2022)
E. coli	7r5c	G206C/R207S/D210L/S211V	2.20	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	23*A*/23*C*/7*B*/7*D*	—	Janicki *et al.* (2023)
E. coli	8bi3	M200W	1.45	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	19*A*/20*C*/7*B*/7*D*	—	Janicki *et al.* (2023)
E. coli	8bkf	M200T	1.22	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	20*A*/20*C*/7*B*/7*D*	—	Janicki *et al.* (2023)
E. coli	8bp9	M200W	1.70	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	22*A*/18*C*/7*B*/8*D*	—	Janicki *et al.* (2023)
E. coli	8bqo	M200I	2.10	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	16*A*/22*C*/9*B*/9*D*	—	Janicki *et al.* (2023)
E. coli	8c0i	M200L	1.90	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	IAE (L-Asp covalently bound)	21*A*/25*C*/7*B*/7*D*	—	Janicki *et al.* (2023)
E. coli	8c23	M200T	1.84	$2 \times (\alpha\beta)_2$ heterotetramer (fully cleaved)	—	19*A*/20*C*/22*E*/20*G*/8*B*/8*D*/8*F*/9*H*	—	Janicki *et al.* (2023)
R. halodurans	8dqm	WT	2.70	$(\alpha\beta)_2$ heterotetramer (fully cleaved)	—	21*A*/20*C*/5*B*/6*D*	—	Sharon & Schmeing (2023)

† The number of separate protein chains. The uncleaved precursor is a homodimer. Upon maturation, each immature chain is cleaved into α and β subunits, and the mature protein is an $(\alpha\beta)_2$ heterotetramer. ‡ Number of missing residues in each chain. § PDB entry 2zal was created by including a missing 5 Ca + 3 L-Asp cluster in PDB entry 1seo.

diffraction data from PDB entry 1o7j. Since the last cycle of the original *SHELXL* refinement combined the work and test reflections (a practice recommended at very high resolution), the original R_{free} set (which is excessively large with 5592 reflections) was not properly marked in the deposition (a practice that we do not recommend). For this reason, we had to select a new R_{free} set (of 1500 reflections) and anneal the structure first (using the *SHELXL* STIR command) to wipe the memory of the model in the test set.

With a newly generated *SHELXL* input file (including correctly defined restraints), our refinement of the original PDB entry 1o7j model converged at an R and R_{free} of 0.111 and 0.128, respectively (as in the original refinement), while the residual R factors for the corrected model were 0.109 and 0.127, respectively. This example shows that even the best structures in the PDB (in terms of resolution and refinement statistics) are not guaranteed to be flawless.

3.2. Class 2 L-asparaginases

Currently, there are 43 structures of class 2 (or type III) L-asparaginases deposited in the PDB (Table 3): 20 structures of the enzyme from *E. coli* (EcAIII) and its variants (Prahl *et al.*, 2004; Michalska, Hernandez-Santoyo & Jaskolski, 2008; Loch *et al.*, 2022; Janicki *et al.*, 2023), 15 structures of the human protein HsAIII, WT and its variants (Li *et al.*, 2012; Su *et al.*, 2013; Nomme *et al.*, 2014), two structures of the guinea pig (*C. porcellus*) protein CpAIII (Schalk & Lavie, 2014), three structures of common bean (*Phaseolus vulgaris*) L-asparaginase (Bejger *et al.*, 2014) and one structure of the yellow lupin (*Lupinus luteus*) enzyme (Michalska *et al.*, 2006). Recently, the structure of a class 2 L-asparaginase from the bacteriochlorophyll-containing bacterium *Roseivivax halodurans* has also been published (Sharon & Schmeing, 2023) and deposited in the PDB (PDB entry 8dqm). Most of the structures of class 2 L-asparaginases were correctly refined, and only minor refinement/validation issues (listed in Table 3) were detected in the PDB depositions, for example some Ramachandran outliers (PDB entries 4o0d, 4pvr and 4pvq) or alternative conformers with zero occupancy (for example PDB entry 4pvp). However, in several structures some more important problems related to the correct assignment of the number of protein chains, the location of the metal ions in the structures, the positioning of ligands or the composition of the asymmetric unit were found, as discussed below.

3.2.1. Enzymatic mechanism of class 2 L-asparaginases in brief

Class 2 L-asparaginases belong to the family of Ntn-hydrolases, which means that they are expressed as inactive precursors (Fig. 6) that must undergo autoproteolytic cleavage into α and β subunits to achieve maturation (Linhorst & Lübke, 2022). Maturation leads to the liberation of the catalytic threonine residue at the N-terminus of subunit β (Thr179 in EcAIII) and leaves the ∼20-residue flexible linker at the C-terminus of subunit α at a different stage of partial digestion (in all structures of the mature enzymes there is no electron density

for most of the residues in the linker region; Table 3). It is interesting that the proteolytic autoactivation is catalyzed by the same threonine residue in front of its peptide bond. The enzymatic reaction is a two-step process, in which an acyl-enzyme is first formed at Thr179, with the release of ammonia, and an activated water molecule then hydrolyzes this ester intermediate. A characteristic Na^+-binding loop in subunit α stabilizes the position of the threonine nucleophile. The oxyanion hole for the substrate C═O group during the nucleophilic attacks is formed by the OH group of another threonine and the NH group of the following glycine. A key residue for the presentation of the L-Asn substrate is Arg207 (EcAIII numbering), which forms a salt bridge with the α-carboxylate group of the substrate. In a subgroup of potassium-dependent class 2 ASNases there is another alkali metal-binding loop, also in subunit α. It activates the enzyme by positioning Arg207 correctly when K^+ is bound and switches it off when Na^+ is bound. Some class 2 L-asparaginases also have isoaspartyl dipeptidase coactivity (EC 3.4.19.5), which in some cases may even be the dominant activity (Borek *et al.*, 2004).

3.2.2. Assignment of separate protein chains to the α and β subunits of mature class 2 enzymes

The immature polyproteins (composed of uncleaved α–β chains) form homodimers, and the maturation process leaves this architecture largely intact. Mature class 2 L-asparaginases are therefore dimers of $\alpha\beta$ heterodimers, or $(\alpha\beta)_2$ hetero-tetramers, composed of four independent protein chains, typically labeled *A*, *B*, *C* and *D*, where chains *A* and *C* correspond to subunit α and chains *B* and *D* correspond to subunit β (Fig. 6; Michalska, Hernandez-Santoyo & Jaskolski, 2008; Su *et al.*, 2013). However, in some PDB depositions containing structures of mature class 2 L-asparaginases only two protein chains are present, instead of the four independent chains that define the mature, *i.e.* fully processed into subunits α and β, form of Ntn-amidohydrolase. Examples of such structures can be found among structures of human class 2 L-asparaginase (HsAIII): PDB entries 4o0h, 4pvs and 4osy. Moreover, some PDB depositions for HsAIII contain a 'heterodimer' which is a mixture of fully cleaved (mature) and uncleaved (immature) subunits. Such a situation is present in PDB entries 4o0g, 4pvp and 4pvr. However, this is properly mentioned (as 'partially cleaved') only in the title of PDB entry 4pvr. We note that the lack of information about the maturation state of the protein might be misleading to PDB users.

Several PDB depositions (PDB entries 3tkj, 4o0d, 4o0e, 4o0f and 4pvq) contain structures of uncleaved Ntn-hydrolase precursors but their titles are unclear. We encourage all structure authors to include in the titles of future depositions the relevant information about the maturation state of the Ntn-amidohydrolases, especially corresponding to precursors, by using phrases such as 'uncleaved precursor', 'immature

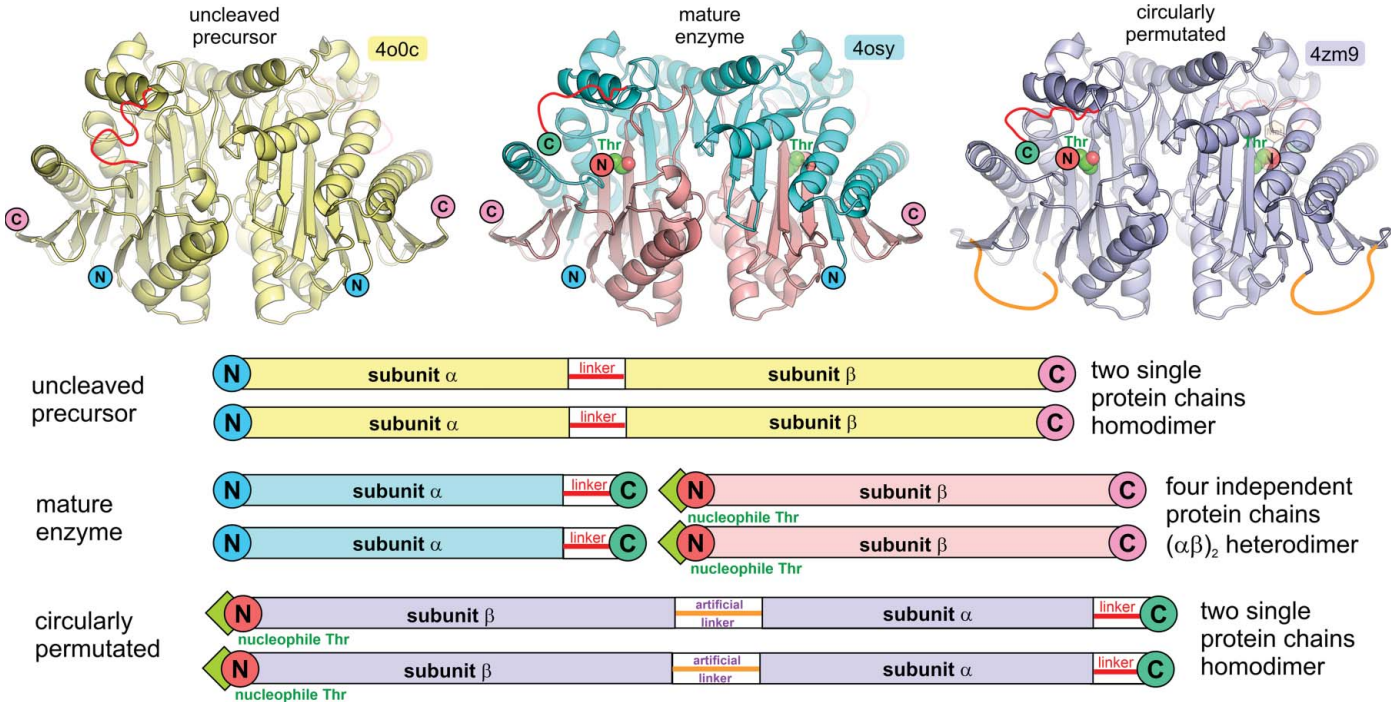


Figure 6
Assignment of the number of protein chains to class 2 L-asparaginases. In the PDB, three possible 'forms' of class 2 L-asparaginases are deposited. In the figure, these 'forms' are presented using the human enzyme HsAIII as an example: the uncleaved precursor (PDB entry 4o0c), the mature enzyme (PDB entry 4osy) and a circularly permuted protein (PDB entry 4zm9). Uncleaved precursors and circularly permuted enzymes should be treated as homodimers, while the mature enzymes should be treated as $(\alpha\beta)_2$ heterotetramers with four independent protein chains.

protein', 'partially cleaved structure' *etc*. Although sequence databases only provide sequences of Ntn-amidohydrolase precursors, we recommend modeling (and refining) the structures of mature Ntn-hydrolases as $(\alpha\beta)_2$ heterotetramers, *i.e.* as four separate chains (two for subunits α and two for subunits β), with the use of the original residue numbers. For example, in the mature human class 2 enzyme HsAIII (UniProt ID Q7L266), the first chain ID, assigned to subunit α, should cover residues 1–167, whereas the second chain ID, assigned to subunit β, should include residues 168–308. Such an approach to residue numbering and structure organization is consistent with the biochemistry of Ntn-hydrolases: in the mature enzymes, the nucleophilic threonine located at the N-terminus of subunit β possesses a free amino group, which is necessary for the catalytic activity of the enzyme.

Uncleaved precursors of Ntn-hydrolases should be modeled and refined as homodimers comprised of two chains. The same approach can be used for the refinement of circularly permutated Ntn-amidohydrolases (Fig. 6). For example, HsAIII engineered by circular permutation has a linker between the original N- and C-termini of the precursor, and the newly designed protein is already expressed in the mature form with the nucleophilic Thr168 located at the new genuine N-terminus, just behind the initiator methionine, which is cleaved after translation. Two structures of circularly permuted HsAIII are currently available in the PDB: PDB entries 4et0 and 4zm9 (Li *et al.*, 2012, 2016). These structures were refined at relatively low resolutions of 3.30 and 2.51 Å, respectively, which is too poor for the visualization of many important details, the most important being the position of the 12-residue artificial linker that connects Gly1 and Pro308. The electron-density map of PDB entry 4et0 (3.30 Å resolution) also has several large unmodeled blobs, especially in the region of Thr168. The same density, in the neighborhood of Thr168, was observed in PDB entry 4zm9 (2.51 Å resolution), but in this case it was interpreted as a citrate or DTT molecule, although the fitting of these ligands to the density is far from satisfactory.

3.2.3. Metal ions in the structures of class 2 L-asparaginases

The structures of all class 2 L-asparaginases contain a stabilization loop (SL), which usually coordinates (exclusively through main-chain C=O groups) a sodium cation (Fig. 7), although coordination of K^+ ions is also possible, as shown in PDB entry 4pu6. This loop is responsible for maintaining the correct conformation of the free amino group of the nucleophilic threonine residue (Loch & Jaskolski, 2021). A subgroup of potassium-dependent class 2 L-asparaginases have a second metal-binding loop, called the activation loop (AL), which also has the ability to coordinate monovalent alkali-metal cations (K^+ or Na^+). The AL is part of a 'catalytic switch', which activates the enzyme when potassium is bound (Bejger *et al.*, 2014). The alkali-metal cations coordinated in these structural loops usually have the typical octahedral geometry (Figs. 7*b* and 7*h*) and are predominantly coordinated by O atoms (Lev *et al.*, 2013). As an exception, the Na^+ cation in the SL loop of the yellow lupin protein LlAIII (PDB entry 2gez) is coordinated by only five C=O ligands, the first Leu59 element of the loop being too far (~3.5 Å) for proper coordination (Figs. 7*m* and 7*n*). This is not a modeling error but a genuine fact, as mentioned in the original publication (Michalska *et al.*, 2006) and confirmed by the multiple copies (four) of the protein in the asymmetric unit.

Apart from the stabilization and activation loops, where monovalent alkali-metal cations are usually modeled in the structures of class 2 L-asparaginases, in some PDB depositions there are also some unusual Na^+ sites. For example, in the crystal structure of the HsAIII enzyme (PDB entry 4o0c), sodium ions were modeled near Phe61 (chain *A*; Fig. 7*c*) or near the genuine Na^+ cation in the AL of chain *B* (see below), with a coordination sphere (consisting of only two water molecules!) and $Na^+\cdots$Wat distances of 2.52–2.97 Å, clearly suggesting that these 'sodium ions' should be replaced by water molecules.

In the same structure, PDB entry 4o0c, a sodium ion was also modeled in the loop formed by Met109–His114 of chain *B* (Fig. 7*d*), while the same site in chain *A* was interpreted as a water molecule. This region of the HsAIII protein corresponds to the activation loop of potassium-dependent L-asparaginases, although this fact escaped the notice of the authors of this entry, as it had been deposited shortly before the publication by Bejger *et al.* (2014) in which the structural grounds of potassium-dependence were clarified for the first time. It is, therefore, very exciting to hypothesize that the human HsAIII protein also has an AL loop that could influence the enzymatic activity depending on the alkali metal coordinated in it. The bond lengths (2.33–2.80 Å), the coordination number of five and the appearance of the electron-density peak all confirm the identity of the Na^+ cation in the AL loop of molecule *B*. Using the same argument, the corresponding water molecule in chain *A* should be converted to Na^+.

Incorrectly interpreted sodium cations are also present in the active sites of both chains (*A* and *B*) of PDB entry 4o0f, close to the nucleophilic Thr168 (Figs. 7*f* and 7*g*). These very high $2F_o - F_c$ peaks (~6σ), with coordination patterns that are not suitable for a metal cation and long N—H$\cdots X$ hydrogen bonds, in all probability represent Cl^- anions. An interesting situation is found in the active sites of the structure of EcAIII with PDB code 3c17. At both sites there is an obvious Cl^- anion and, much less typically for this site, an Na^+ cation (Michalska, Borek *et al.*, 2008). The sodium cations have an incomplete coordination sphere, since on one side they form a short ionic contact with the chloride anions (Fig. 7*j*). As the crystals used to obtain PDB entry 3c17 were obtained at a high NaCl concentration (4.3 *M*), this EcAIII structure is saturated with Na^+ ions, and apart from the active site, Na^+ cations were identified at the dimer interface (Fig. 7*l*), at the loop near Phe290 (Fig. 7*o*), at the helix located at the beginning of the linker (Fig. 7*p*) and in the region corresponding to the activation loop in potassium-dependent enzymes (Fig. 7*k*). Therefore, the EcAIII structure (PDB

entry 3c17), together with PDB entry 4o0c (HsAIII, see above), indicates that potassium-independent enzymes can also coordinate alkali metals in the activation loop, but such coordination does not activate the catalytic switch.

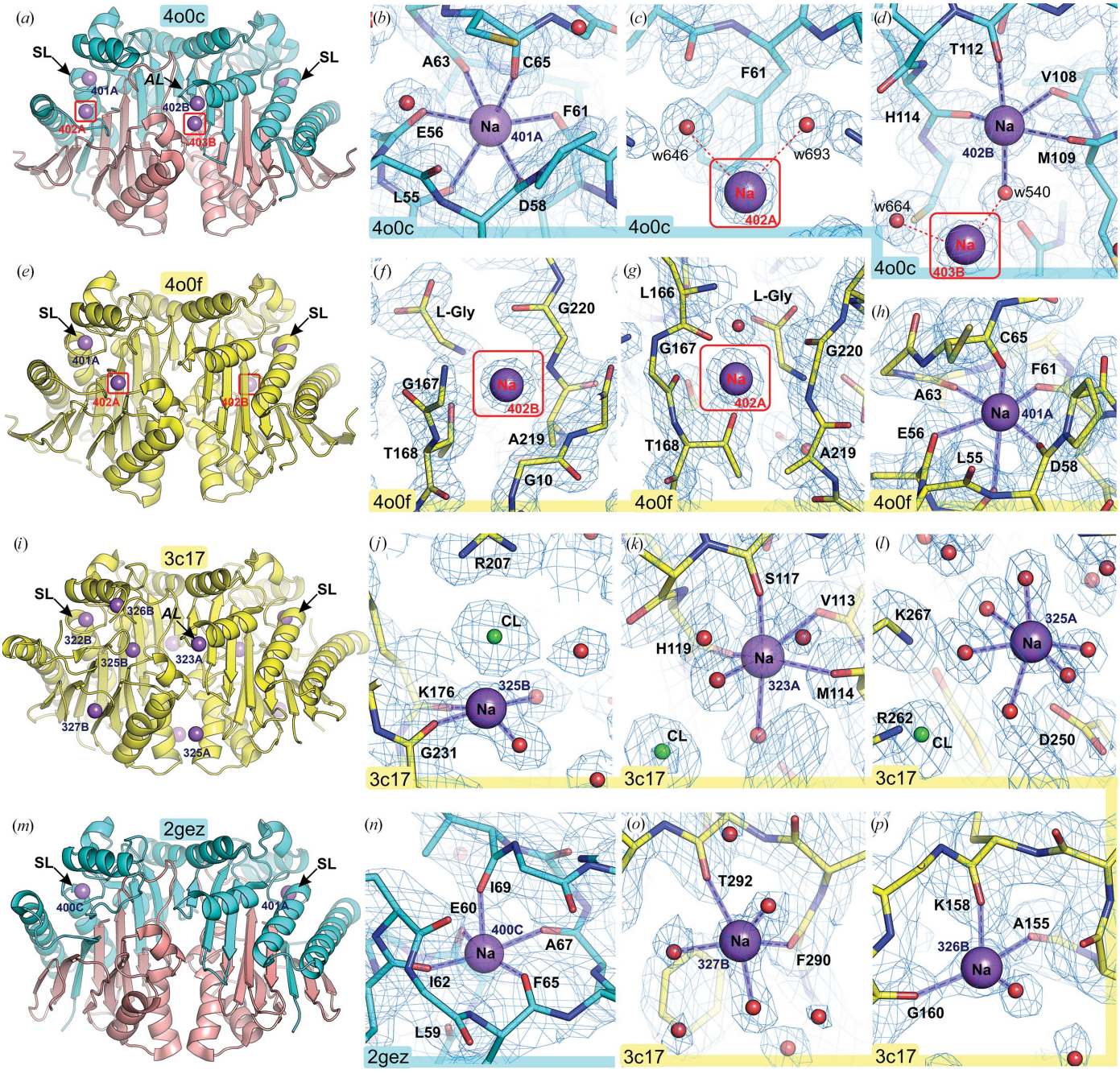


Figure 7
An assortment of Na^+ cations (violet spheres) in class 2 L-asparaginases. Incorrectly modeled Na^+ ions are marked by red frames, while properly identified Na^+ ions are shown with coordination bonds marked as violet solid lines. Numbering of residues is the same as in the PDB depositions. $2F_o - F_c$ maps (blue mesh) are presented at the 1.50σ or 1.00σ level depending on the resolution and structure quality. (*a*) PDB entry 4o0c for mature HsAIII (1.50 Å) with the positions of Na^+ ions. (*b*) Typical octahedral geometry of the Na^+ cation (401*A*) coordinated in the stabilization loop (SL). (*c*) Incorrectly located Na^+ ion 402*A*, which should be replaced by a water molecule. (*d*) Na^+ 402*B* coordinated in the region corresponding to the activation loop (AL) and the incorrectly located Na^+ 403*B* in its close neighborhood. (*e*) PDB entry 4o0f for the uncleaved precursor of HsAIII (1.92 Å) with the positions of Na^+ ions. (*f*, *g*) Incorrectly located Na^+ ions 402*B* and 402*A* without a coordination sphere. (*h*) Properly modeled Na^+ 401*A* in the stabilization loop. (*i*) PDB entry 3c17 for the uncleaved precursor of EcAIII (1.95 Å resolution) with the positions of Na^+ ions. (*j*) Na^+ 325*B* coordinated in the active site in an ion pair with a Cl^- anion. (*k*) Na^+ coordinated in the region corresponding to the activation loop. (*l*) Na^+ 325*A* located at the dimer interface. (*o*) Na^+ 327*B* bound near Phe290. (*p*) Na^+ 326*B* located at the beginning of the linker region; the geometry and O···Na distances around 326*B* indicate a sodium cation with an incomplete coordination sphere rather than a water molecule. (*m*) One of two dimers of mature LlAIII present in the asymmetric unit of PDB entry 2gez (2.60 Å resolution). (*n*) Na^+ 400*C* coordinated in the stabilization loop of subunit C of entry 2gez; in this crystal structure all four Na^+ cations bound in the stabilization loops have an atypical pentadentate coordination.

3.2.4. Ligands and different conformation of residues in class 2 L-asparaginases

Many structures of class 2 L-asparaginases contain ligands bound in the active site (Fig. 8). The most common is glycine (PDB entries 4o0d, 4o0f, 4osx, 4osy, 4zm9 and 8c23; Fig. 8*c*), followed by L-Asp (PDB entries 2zal, 4pvr, 4pvs, 4o48 and 4pu6; Fig. 8*a*) and inorganic anions, such as sulfate (PDB entries 1t3m and 4pvq; Fig. 8*g*), nitrate (PDB entry 1t3m; Fig. 8*f*), chloride (PDB entry 3c17; Fig. 8*e*) or iodide (PDB entry 4pvq). In two structures organic ions are present, malonate (PDB entry 4pvp; Fig. 8*d*) and citrate (PDB entry 4zm9), although the electron density for the citrate anion in PDB entry 4zm9 is not very convincing, as already mentioned above. Two structures represent the covalent acyl-enzyme intermediate of the L-asparagine hydrolysis reaction: residue IAS in the HsAIII structure with PDB code 4o0h and residue AEI in the EcAIII structure with PDB code 8c0i (Fig. 8*b*). In most of the structures the ligands are correctly modeled in the electron density, but interestingly in some cases the ligand was found only in one enzyme subunit, for example the malonate ion in PDB entry 4pvp (Fig. 7*d*), while in other structures both subunits are occupied, for example by L-Asp in PDB entry 4pvs. In the latter case, the position and orientation of the L-Asp ligands in the two active sites of the dimeric enzyme are slightly different; these two active sites also differ in the conformation of the residues surrounding the ligand. These observations suggest that the two ($\alpha\beta$) units of class 2 L-asparaginases act rather independently (*i.e.* not synchronously) and that Ntn-amidohydrolases can have different conformations not only at residues in the active site but also of entire segments of the structure (Loch *et al.*, 2022). The hypothesis about the independent action of the two catalytic units of class 2 L-asparaginases seems to find confirmation in several crystal structures that possess the substrate/product ligand in only one active site. The maturation process also proceeds independently in the two subunits, as found in the HsAIII structure with PDB code 4pvr. These observations should be taken into account during structure refinement, especially if one uses NCS restraints in the refinement procedure.

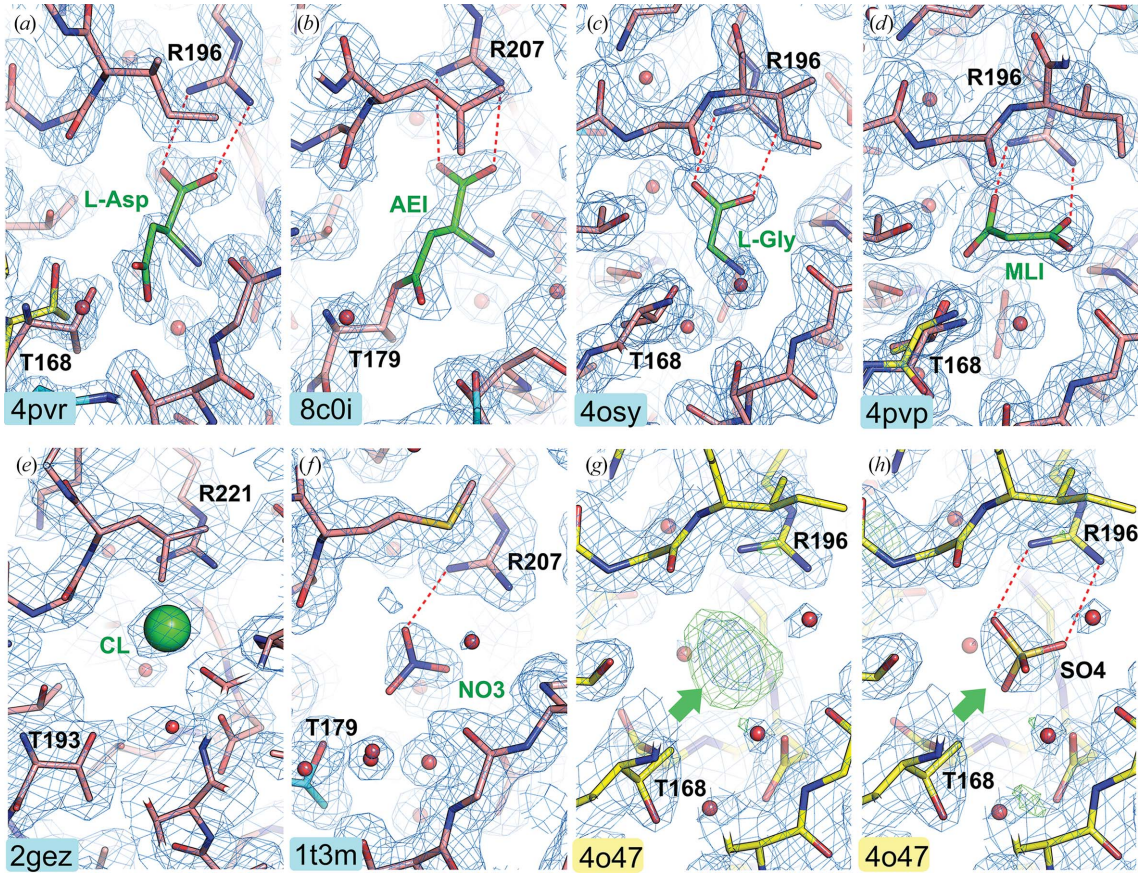


Figure 8
Examples of ligands identified in the active site of class 2 L-asparaginases deposited in the PDB. (*a*) L-Asp in the HsAIII structure with PDB code 4prv. (*b*) L-Asp covalently bound to Thr179 in an acyl-enzyme intermediate of EcAIII mutant M200L (PDB entry 8c0i). (*c*) Gly in the HsAIII structure with PDB code 4osy. (*d*) Malonate (MLI) in the HsAIII structure with PDB code 4pvp. (*e*) Chloride ion in the LlAIII structure with PDB code 2gez. (*f*) Nitrate anion in the EcAIII structure with PDB code 1t3m. (*g*) Unmodeled electron density (marked by an arrow) in the active site of the uncleaved precursor of CpAIII (PDB entry 4o47). (*h*) Re-refinement of PDB entry 4o47 allowed the modeling of a sulfate anion in the empty density. The electron-density maps are $2F_o - F_c$ (blue) at 1.5σ or 1.0σ and $F_o - F_c$ [green in (*g*)] at 3.0σ. Red dashed lines mark hydrogen bonds between ligands and an arginine residue present in the active site.

3.2.5. Lessons from the example of PDB entry 2zal

In 2005, a 1.9 Å resolution structure of EcAIII with L-Asp in the active site was described in a publication (Michalska *et al.*, 2005) and deposited in the PDB as entry 1seo. A few years later, when preparing this PDB deposition for a student tutorial, the authors noticed that a large cluster of electron density had been overlooked in the original electron-density map. Careful inspection led to the conclusion that this electron density corresponded to a 5 Ca^{2+} + 3 Asp cluster, which in PDB entry 1seo had been incorrectly marked as just two Na^{+} ions. The calcium–aspartate cluster had an excellent experimental justification, as the protein had been crystallized after incubation with an ∼200-fold molar excess (100 m*M*) of L-Asp (which was mistakenly utilized instead of the intended ∼20-fold excess) from a precipitating buffer that contained 80 m*M* Ca^{2+} cations. After correct reinterpretation of the structure with a calcium–aspartate clump sitting between the protein molecules in the crystal lattice, the corrected structure was redeposited in the PDB in 2007 with accession code 2zal, superseding the original PDB entry 1seo. It has to be noted though that due to the current PDB rules of allocation of chain IDs to the various structural components according to their proximity to crystallographic copies of the macromolecule, the 5 Ca^{2+} + 3 Asp cluster in PDB entry 2zal has been scattered all over the standard unit cell (and outside it), so that it is again very difficult to make sense of it. We also note that the *PDB-REDO* server claimed that it was able to fix a number of further mistakes in PDB entry 2zal such as 30 corrected side-chain rotamers or the removal of six suspicious water molecules. However, inspection of these corrections leads us to the conclusions that some of them have no merit.

The PDB entry 1seo/2zal example is a case in point illustrating the importance of inspecting macromolecular crystal structures not only within the electron density for which there is already some sort of (correct or suspicious) interpretation (model), but indeed also at electron-density peaks for which there is no current interpretation ('hidden' in intermolecular spaces).

3.2.6. Unmodeled electron density in structures of the guinea pig class 2 enzyme

Analysis of the structures of the guinea pig enzyme CpAIII (PDB entries 4o47 and 4o48) leads to several interesting observations. In the structure of the uncleaved precursor, PDB entry 4o47, there is a large spherical density in the active site near Arg196 of chain *A*. As the crystallization solution contained ammonium sulfate, it is quite probable that a sulfate ion can be modeled in this density (Figs. 8*g* and 8*h*). The second structure, PDB entry 4o48, of the mature form of CpAIII contains L-Asp (with slightly different conformations) in the two active sites. Although PDB entries 4o47 and 4o48 have good refinement statistics (R_{free} of 0.213 and 0.221, respectively; Schalk & Lavie, 2014), careful inspection of the electron-density maps revealed that in these structures vast areas of unmodeled electron density located between the protein molecules are present in large empty channels with a diameter of ∼80 Å running along the trigonal 3_2 screw axis. These structures also have a relatively high Matthews coeffi-

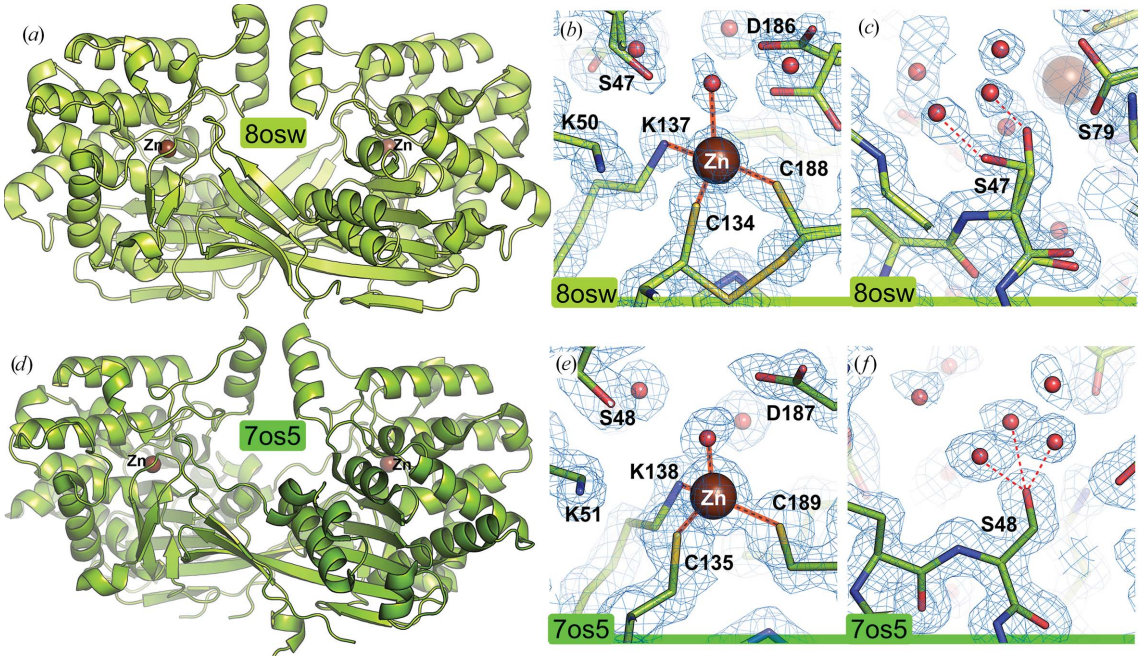


Figure 9
Class 3 L-asparaginases. (*a*) Structure of ReAIV (PDB entry 8osw). (*b*) Zn^{2+} coordinated in the active site of ReAIV; the active site is not fully saturated with Zn^{2+} and the alternative conformers of Cys134 and Cys188 form a disulfide bond. (*c*) The nucleophilic Ser47 in the active site of ReAIV, occupying two conformations. (*d*) Structure of ReAV (PDB entry 7os5). (*e*) Zn^{2+} coordinated in the active site of ReAV. (*f*) The nucleophilic Ser48 in the active site of ReAIV, with three tightly bound water molecules. The $2F_o - F_c$ electron-density maps (blue mesh) are contoured at 1.5σ.

Table 4
Structures of class 3 type IV L-asparaginases deposited in the Protein Data Bank as of 22 November 2023.

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No. of missing residues‡	Problems identified	Reference
R. etli	8cly	WT	2.50	Dimer	—	1/1	—	Loch *et al.* (2023)
R. etli	8clz	WT	1.50	Dimer	—	1/2	—	Loch *et al.* (2023)
R. etli	8col	WT	2.00	2 dimers	—	2/0/1/0	—	Loch *et al.* (2023)
R. etli	8ori	WT	1.35	2 dimers	—	1/0/1/0	—	Loch *et al.* (2023)
R. etli	8osw	WT	1.30	Dimer	—	0/1	—	Loch *et al.* (2023)

† ASU, asymmetric unit. ‡ Number of missing residues in each chain.

Table 5
Structures of class 3 type V L-asparaginases deposited in the Protein Data Bank as of 22 November 2023.

Organism	PDB code	Variant	Resolution (Å)	Content of ASU†	Active-site ligand	No of missing residues‡	Problems identified	Reference
R. etli	7os3	WT	2.18	Dimer	—	2/18	—	Loch *et al.* (2021)
R. etli	7os5	WT	1.29	Dimer	—	5/18	—	Loch *et al.* (2021)
R. etli	7os6	WT	1.43	2 dimers	—	20/17/17/17	—	Loch *et al.* (2021)
R. etli	7ou1	WT	1.65	2 dimers	—	17/17/17/17	—	Loch *et al.* (2021)
R. etli	7oz6	WT	1.76	Dimer	—	6/5	—	Loch *et al.* (2021)

† ASU, asymmetric unit. ‡ Number of missing residues in each chain.

cient (3.83 Å^3 Da^{-1}), suggesting that something might be missing in the structure model. However, different attempts to model 'orphan' protein molecules in this density have all failed, and we must conclude that this electron density is the result of artificial accumulation, which is sometimes observed across symmetry elements in X-ray crystal structures.

3.3. Class 3 L-asparaginases

While the existence of an alien type of L-asparaginase in the symbiotic nitrogen-fixing bacterium *Rhizobium etli* had been recognized long ago (Borek & Jaskólski, 2001), the structure of the inducible and thermolabile prototype ReAV has only been recently solved and deposited in the PDB (PDB entries 7os3, 7os5, 7os6, 7ou1 and 7oz6; Loch *et al.*, 2021; Fig. 9*b*). Soon thereafter, structures (PDB entries 8cly, 8clz, 8col, 8ori and 8osw) of the constitutive and thermostable isoform ReAIV were published (Loch *et al.*, 2023; Fig. 9*a*). The structures are remarkably similar, despite very low (∼30%) sequence identity. The class 3 enzymes are homodimeric, with one subunit built around an active site containing two Ser–Lys tandems, of which Ser48 (in ReAV numbering) is the tentative catalytic nucleophile, and a Zn^{2+} cation coordinated by two cysteine residues, a lysine residue and a water molecule (Figs. 9*b* and 9*e*). The metal cation does not play a catalytic role, but it affects the substrate-binding potential (K_m). The enzymes are most active at high pH, at which the tandem lysine residues, as well as the lysine from the Zn^{2+} coordination sphere, are unprotonated (neutral). In a number of structures, the hydroxyl group of Ser48 (or Ser47 in ReAIV) is surrounded by three very tightly hydrogen-bonded water molecules (Fig. 9*f*). An alternative interpretation of the electron density of these clusters as a phosphorylation modification could be excluded on the grounds of structure refinement and mass-spectrometric analysis (Loch *et al.*, 2021). This unusual hydration pattern of Ser48/47 provides indirect evidence of the nucleophilic character of this residue. At present, there are five PDB depositions (WT and mutant proteins) for both ReAIV (1.3–2.5 Å resolution) and ReAV (1.29–2.18 Å resolution) (Supplementary Table S1). We did not find any major issues with these depositions, noting only the extremely short O···O distances (some <2.4 Å) in the hydration patterns of the Ser48/47 residues.

4. Conclusions and outlook

In this work, we have compiled a list of 189 crystal structures of different L-asparaginases or their variants that had been deposited and released in the PDB as of 22 November 2023 (Excel files deposited as supporting information). The enzymes come from three structural classes and five types, depending on their mechanism of L-asparagine hydrolysis, biophysical properties, phylogenetic classification and compartmentalization in the living cell. Since some of the depositions represented multiple structures from a single publication, in these cases we usually looked at only one representative data set in detail. We found that the majority of the depositions are without any serious errors, oversights or misinterpretations. The remaining models, however, posed various kinds of problems, from trivial but annoying inconsistent placement in the asymmetric unit, to misrepresentations of the solvent area (too many or too few water molecules, missed buffer molecules *etc.*), to inconsistencies between the deposited and published data or the eternal 'to be published' declaration. Ultimately, and these were the most serious cases, we found crystal structures in which parts of the protein were modeled without any support from the electron density or even in stark defiance against such evidence, or where, on the contrary, stretches of interpretable protein electron density were left unmodeled. In between were very frequent cases of the incorrect modeling of side-chain rotamers, of disallowed interatomic contacts or misidentification of metal cations. We even found errors in the reference (PDB entry 2zal for class 2)

or highest resolution (PDB entry 1o7j for class 1) structures, some of our own provenance. As a positive observation, we note that the usually very sensitive issue of incorrect/fanciful modeling of protein ligand molecules with regard to the active-site or allosteric site ligands was not particularly disturbing in this project.

In all inspected cases, we corrected the noticed errors, mistakes and omissions, and as a minimum listed them in our summary Tables 1–5 (in the case of class 1 ASNases only for structures deposited since 2021; Lubkowski & Wlodawer, 2021). When the problems were particularly severe and the depositions were accompanied by experimental diffraction data, we strove to carry out independent structure reinterpretation and re-refinement. In most such cases this procedure resulted in improved refinement statistics and better models. When structure models were re-refined by us, we kept the new models (30 in total) in a dedicated repository, which is available as supporting information to this publication and in Proteopedia.

Some of the problems (for example failure of the validation or electron-density map calculation) could have been rectified during deposition. Structures with zero-occupancy ligands should not be accepted. We also note that the PDB continues to use, despite previous alerts (see, for example, Jaskolski, 2013), incorrect templates, standard files and diagrams, of which a 'sulfate' (SO_4^{2-}) anion with two different sets of bonds (two single S—O bonds and two double S═O bonds) is an example. Despite these critical comments, we emphasize that the PDB is valued for not being the data police, but is regarded as an example of data sharing and open science. Moreover, as mentioned throughout the paper, many issues reported to the PDB and PDB-REDO have been rectified during this study. This demonstrates that identifying and communicating inconsistencies can be beneficial and that providing direct feedback to databanks can enhance their quality.

In view of the great interest in L-asparaginase structures as potential starting points for the engineering of new antileukemic drug candidates, researchers in this area may find our collection of re-refined structures to be a better source of data than the PDB itself. The presentation of a comprehensive collection of all crystal structures of enzymes with L-asparaginase activity should also be of value to the wider scientific community. However, we would like to postulate that the type of analysis presented here could also be applied to other important classes of proteins for which multiple structures are available. Finally, we may hope that by systematic validation of its holdings, this will make the PDB a better database in general and will promote a more conscientious approach to PDB deposition.

Acknowledgements

Author contributions were as follows. AW, ZD and JL analyzed the structures of class 1 asparaginases. JIL analyzed the structures of class 2 asparaginases. MJ inspected class 3 asparaginases. MG and MJ inspected the most difficult cases. DB and MG provided tools for comparative and statistical analyses. All authors wrote the manuscript.

Funding information

This work was supported by National Science Centre (NCN, Poland) grants 2020/37/B/NZ1/03250 to MJ and 2020/38/E/NZ1/00035 to JIL and by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research (ZD, JL and AW).

References

Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*a*). *Biochemistry*, **40**, 5655–5664.
Aghaiypour, K., Wlodawer, A. & Lubkowski, J. (2001*b*). *Biochim. Biophys. Acta*, **1550**, 117–128.
Araújo, T. S. de, Scapin, S. M. N., de Andrade, W., Fasciotti, M., de Magalhães, M. T. Q., Almeida, M. S. & Lima, L. (2021). *Biophys. Chem.* **271**, 106554.
Bansal, S., Srivastava, A., Mukherjee, G., Pandey, R., Verma, A. K., Mishra, P. & Kundu, B. (2012). *FASEB J.* **26**, 1161–1171.
Beckett, A. & Gervais, D. (2019). *World J. Microbiol. Biotechnol.* **35**, 152.
Bejger, M., Imiolczyk, B., Clavel, D., Gilski, M., Pajak, A., Marsolais, F. & Jaskolski, M. (2014). *Acta Cryst.* D**70**, 1854–1872.
Borek, D. & Jaskólski, M. (2001). *Acta Biochim. Pol.* **48**, 893–902.
Borek, D., Kozak, M., Pei, J. & Jaskolski, M. (2014). *FEBS J.* **281**, 4097–4111.
Borek, D., Michalska, K., Brzezinski, K., Kisiel, A., Podkowinski, J., Bonthron, D. T., Krowarsch, D., Otlewski, J. & Jaskolski, M. (2004). *Eur. J. Biochem.* **271**, 3215–3226.
Broome, J. D. (1961). *Nature*, **191**, 1114–1115.
Broome, J. D. (1963). *J. Exp. Med.* **118**, 99–120.
Brünger, A. T. (1992). *Nature*, **355**, 472–475.
Brzezinski, D., Dauter, Z., Minor, W. & Jaskolski, M. (2020). *FEBS J.* **287**, 2685–2698.
Brzezinski, D., Kowiel, M., Cooper, D. R., Cymborowski, M., Grabowski, M., Wlodawer, A., Dauter, Z., Shabalin, I. G., Gilski, M., Rupp, B., Jaskolski, M. & Minor, W. (2021). *Protein Sci.* **30**, 115–124.
Cecconello, D. K., Magalhães, M. R., Werlang, I. C. R., Lee, M. L. M., Michalowski, M. B. & Daudt, L. E. (2020). *Hematol. Transfus. Cell. Ther.* **42**, 275–282.
Cerofolini, L., Giuntini, S., Carlon, A., Ravera, E., Calderone, V., Fragai, M., Parigi, G. & Luchinat, C. (2019). *Chem. A Eur. J.* **25**, 1984–1991.
Chan, W. K., Horvath, T. D., Tan, L., Link, T., Harutyunyan, K. G., Pontikos, M. A., Anishkin, A., Du, D., Martin, L. A., Yin, E., Rempe, S. B., Sukharev, S., Konopleva, M., Weinstein, J. N. & Lorenzi, P. L. (2019). *Mol. Cancer Ther.* **18**, 1587–1592.
Dhavala, P. & Papageorgiou, A. C. (2009). *Acta Cryst.* D**65**, 1253–1261.
Dinndorf, P. A., Gootenberg, J., Cohen, M. H., Keegan, P. & Pazdur, R. (2007). *Oncologist*, **12**, 991–998.
Egler, R. A., Ahuja, S. P. & Matloub, Y. (2016). *J. Pharmacol. Pharmacotherapeutics*, **7**, 62–71.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Ghasemian, A., Al-marzoqi, A., Al-abodi, H. R., Alghanimi, Y. K., Kadhum, S. A., Shokouhi Mostafavi, S. K. & Fattahi, A. (2019). *J. Cell. Physiol.* **234**, 19271–19279.
Grabowski, M., Macnar, J. M., Cymborowski, M., Cooper, D. R., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., Dauter, Z., Rupp, B., Wlodawer, A., Jaskolski, M. & Minor, W. (2021). *IUCrJ*, **8**, 395–407.

Guo, J., Coker, A. R., Wood, S. P., Cooper, J. B., Chohan, S. M., Rashid, N. & Akhtar, M. (2017). *Acta Cryst.* D**73**, 889–895.
Jakob, C. G., Lewinski, K., LaCount, M. W., Roberts, J. & Lebioda, L. (1997). *Biochemistry*, **36**, 923–931.
Janicki, M., Ściuk, A., Zielezinski, A., Ruszkowski, M., Ludwików, A., Karlowski, W. M., Jaskolski, M. & Loch, J. I. (2023). *Protein Sci.* **32**, e4647.
Jaskolski, M. (2013). *Acta Cryst.* D**69**, 1865–1866.
Jaskolski, M., Dauter, Z., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., Rupp, B. & Wlodawer, A. (2021). *IUCrJ*, **8**, 238–256.
Jaskólski, M., Kozak, M., Lubkowski, J., Palm, G. & Wlodawer, A. (2001). *Acta Cryst.* D**57**, 369–377.
Joosten, R. P., Joosten, K., Cohen, S. X., Vriend, G. & Perrakis, A. (2011). *Bioinformatics*, **27**, 3392–3398.
Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. (2014). *IUCrJ*, **1**, 213–220.
Kidd, J. G. (1953). *J. Exp. Med.* **98**, 565–582.
Kislitsyn, Y. A., Kravchenko, O. V., Nikonov, S. V. & Kuranova, I. P. (2006). *Crystallogr. Rep.* **51**, 811–816.
Kleywegt, G. J., Harris, M. R., Zou, J., Taylor, T. C., Wählby, A. & Jones, T. A. (2004). *Acta Cryst.* D**60**, 2240–2249.
Koprivnikar, J., McCloskey, J. & Faderl, S. H. (2017). *Onco Targets Ther.* **10**, 1413–1422.
Kowiel, M., Jaskolski, M. & Dauter, Z. (2014). *Acta Cryst.* D**70**, 3290–3298.
Kravchenko, O. V., Kislitsin, Y. A., Popov, A. N., Nikonov, S. V. & Kuranova, I. P. (2008). *Acta Cryst.* D**64**, 248–256.
Lev, B., Roux, B. & Noskov, S. Y. (2013). *Encyclopedia of Metalloproteins*, edited by R. H. Kretsinger, V. N. Uversky & E. A. Permyakov, pp. 2112–2118. New York: Springer.
Li, W., Cantor, J. R., Yogesha, S. D., Yang, S., Chantranupong, L., Liu, J. Q., Agnello, G., Georgiou, G., Stone, E. M. & Zhang, Y. (2012). *ACS Chem. Biol.* **7**, 1840–1847.
Li, W., Irani, S., Crutchfield, A., Hodge, K., Matthews, W., Patel, P., Zhang, Y. J. & Stone, E. (2016). *Biochemistry*, **55**, 960–969.
Linhorst, A. & Lübke, T. (2022). *Cells*, **11**, 1592.
Loch, J. I., Imiolczyk, B., Sliwiak, J., Wantuch, A., Bejger, M., Gilski, M. & Jaskolski, M. (2021). *Nat. Commun.* **12**, 6717.
Loch, J. I. & Jaskolski, M. (2021). *IUCrJ*, **8**, 514–531.
Loch, J. I., Klonecka, A., Kądziołka, K., Bonarek, P., Barciszewski, J., Imiolczyk, B., Brzezinski, K., Gilski, M. & Jaskolski, M. (2022). *Acta Cryst.* D**78**, 911–926.
Loch, J. I., Ściuk, A., Kilichowska, M., Pieróg, I., Łukaszczyk, W., Zimowska, K. & Jaskolski, M. (2024). *Acta Biochim. Pol.* **71**, 12299.
Loch, J. I., Worsztynowicz, P., Sliwiak, J., Grzechowiak, M., Imiolczyk, B., Pokrywka, K., Chwastyk, M., Gilski, M. & Jaskolski, M. (2023). *Acta Cryst.* D**79**, 775–791.
Lubkowski, J., Chan, W. & Wlodawer, A. (2019). *Sci. Rep.* **9**, 11070.
Lubkowski, J., Dauter, M., Aghaiypour, K., Wlodawer, A. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 84–92.
Lubkowski, J., Palm, G. J., Gilliland, G. L., Derst, C., Röhm, K. H. & Wlodawer, A. (1996). *Eur. J. Biochem.* **241**, 201–207.
Lubkowski, J., Vanegas, J. M., Chan, W. K., Lorenzi, P. L., Weinstein, J. N., Sukharev, S., Fushman, D., Rempe, S., Anishkin, A. & Wlodawer, A. (2020). *Biochemistry*, **59**, 1927–1945.
Lubkowski, J. & Wlodawer, A. (2019). *Protein Sci.* **28**, 1850–1864.
Lubkowski, J. & Wlodawer, A. (2021). *FEBS J.* **288**, 4183–4209.
Lubkowski, J., Wlodawer, A., Ammon, H. L., Copeland, T. D. & Swain, A. L. (1994). *Biochemistry*, **33**, 10257–10265.
Lubkowski, J., Wlodawer, A., Housset, D., Weber, I. T., Ammon, H. L., Murphy, K. C. & Swain, A. L. (1994). *Acta Cryst.* D**50**, 826–832.
Maggi, M., Meli, M., Colombo, G. & Scotti, C. (2021). *Sci. Rep.* **11**, 18885.
Maggi, M., Mittelman, S. D., Parmentier, J. H., Colombo, G., Meli, M., Whitmire, J. M., Merrell, D. S., Whitelegge, J. & Scotti, C. (2017). *Sci. Rep.* **7**, 14479.
Maggi, M. & Scotti, C. (2022). *Int. J. Mol. Sci.* **23**, 5942.
Michalska, K., Borek, D., Hernández-Santoyo, A. & Jaskolski, M. (2008). *Acta Cryst.* D**64**, 309–320.
Michalska, K., Brzezinski, K. & Jaskolski, M. (2005). *J. Biol. Chem.* **280**, 28484–28491.
Michalska, K., Bujacz, G. & Jaskolski, M. (2006). *J. Mol. Biol.* **360**, 105–116.
Michalska, K., Hernandez-Santoyo, A. & Jaskolski, M. (2008). *J. Biol. Chem.* **283**, 13388–13397.
Michalska, K. & Jaskolski, M. (2006). *Acta Biochim. Pol.* **53**, 627–640.
Miller, M., Rao, J. K. M., Wlodawer, A. & Gribskov, M. R. (1993). *FEBS Lett.* **328**, 275–279.
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.
Nguyen, H. A., Durden, D. L. & Lavie, A. (2017). *Sci. Rep.* **7**, 41643.
Nguyen, H. A., Su, Y. & Lavie, A. (2016*a*). *Biochemistry*, **55**, 1246–1253.
Nguyen, H. A., Su, Y. & Lavie, A. (2016*b*). *J. Biol. Chem.* **291**, 17664–17676.
Nomme, J., Su, Y., Konrad, M. & Lavie, A. (2012). *Biochemistry*, **51**, 6816–6826.
Nomme, J., Su, Y. & Lavie, A. (2014). *J. Mol. Biol.* **426**, 2471–2485.
Ogawa, C., Taguchi, F., Goto, H., Koh, K., Tomizawa, D., Ohara, A. & Manabe, A. (2017). *Pediatr. Blood Cancer*, **64**, e26475.
Ortlund, E., Lacount, M. W., Lewinski, K. & Lebioda, L. (2000). *Biochemistry*, **39**, 1199–1204.
Palm, G. J., Lubkowski, J., Derst, C., Schleper, S., Röhm, K. H. & Wlodawer, A. (1996). *FEBS Lett.* **390**, 211–216.
Papageorgiou, A. C., Posypanova, G. A., Andersson, C. S., Sokolov, N. N. & Krasotkina, J. (2008). *FEBS J.* **275**, 4306–4316.
Patel, N., Krishnan, S., Offman, M. N., Krol, M., Moss, C. X., Leighton, C., van Delft, F. W., Holland, M., Liu, J., Alexander, S., Dempsey, C., Ariffin, H., Essink, M., Eden, T. O., Watts, C., Bates, P. A. & Saha, V. (2009). *J. Clin. Invest.* **119**, 1964–1973.
Prahl, A., Pazgier, M., Hejazi, M., Lockau, W. & Lubkowski, J. (2004). *Acta Cryst.* D**60**, 1173–1176.
Prilusky, J., Hodis, E., Canner, D., Decatur, W. A., Oberholser, K., Martz, E., Berchanski, A., Harel, M. & Sussman, J. L. (2011). *J. Struct. Biol.* **175**, 244–252.
Protein Data Bank (1971). *Nat. New Biol.* **233**, 223.
Raczynska, J. E., Shabalin, I. G., Minor, W., Wlodawer, A. & Jaskolski, M. (2018). *Drug Resist. Updat.* **40**, 1–12.
Radadiya, A., Zhu, W., Coricello, A., Alcaro, S. & Richards, N. G. J. (2020). *Biochemistry*, **59**, 3193–3200.
Ran, T., Jiao, L., Wang, W., Chen, J., Chi, H., Lu, Z., Zhang, C., Xu, D. & Lu, F. (2021). *J. Agric. Food Chem.* **69**, 223–231.
Read, R. J., Adams, P. D., Arendall, W. B. III, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütteke, T., Otwinowski, Z., Perrakis, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. J., Vriend, G. & Zwart, P. H. (2011). *Structure*, **19**, 1395–1412.
Rigouin, C., Nguyen, H. A., Schalk, A. M. & Lavie, A. (2017). *Sci. Rep.* **7**, 10224.
Sanches, M., Barbosa, J. A. R. G., de Oliveira, R. T., Abrahão Neto, J. A. & Polikarpov, I. (2003). *Acta Cryst.* D**59**, 416–422.
Schalk, A. M. & Lavie, A. (2014). *Biochemistry*, **53**, 2318–2328.
Schalk, A. M., Nguyen, H. A., Rigouin, C. & Lavie, A. (2014). *J. Biol. Chem.* **289**, 33175–33186.
Shabalin, I., Dauter, Z., Jaskolski, M., Minor, W. & Wlodawer, A. (2015). *Acta Cryst.* D**71**, 1965–1979.
Sharma, P., Tomar, R., Yadav, S. S., Badmalia, M. D., Nath, S. K., Ashish & Kundu, B. (2020). *Sci. Rep.* **10**, 21702.
Sharon, I. & Schmeing, T. M. (2023). *Sci. Rep.* **13**, 8314.
Sheldrick, G. M. (2015). *Acta Cryst.* C**71**, 3–8.
Strzelczyk, P., Zhang, D., Alexandratos, J., Piszczek, G., Wlodawer, A. & Lubkowski, J. (2023). *FEBS J.* **290**, 780–795.
Strzelczyk, P., Zhang, D., Dyba, M., Wlodawer, A. & Lubkowski, J. (2020). *Sci. Rep.* **10**, 17516.

Strzelczyk, P., Zhang, D., Wlodawer, A. & Lubkowski, J. (2022). *FEBS Lett.* **596**, 3060–3068.
Su, Y., Karamitros, C. S., Nomme, J., McSorley, T., Konrad, M. & Lavie, A. (2013). *Chem. Biol.* **20**, 533–540.
Swain, A. L., Jaskólski, M., Housset, D., Rao, J. K. & Wlodawer, A. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 1474–1478.
Tam, S.-Y., Chung, S.-F., Kim, C.-F., To, J. C., So, P.-K., Cheung, K.-K., Chung, W.-H., Wong, K.-Y. & Leung, Y.-C. (2023). *Int. J. Biol. Macromol.* **253**, 127742.
Timofeev, V. I., Zhukhlistova, N. E. & Kuranova, I. P. (2020*a*). *Crystallogr. Rep.* **65**, 586–592.
Timofeev, V. I., Zhukhlistova, N. E. & Kuranova, I. P. (2020*b*). *Russ. J. Bioorg. Chem.* **46**, 171–180.
Tomar, R., Sharma, P., Srivastava, A., Bansal, S., Ashish & Kundu, B. (2014). *Acta Cryst.* D**70**, 3187–3197.
Van Trimpont, M., Peeters, E., De Visser, Y., Schalk, A. M., Mondelaers, V., De Moerloose, B., Lavie, A., Lammens, T., Goossens, S. & Van Vlierberghe, P. (2022). *Cancers*, **14**, 902.
Van Trimpont, M., Schalk, A. M., De Visser, Y., Nguyen, H. A., Reunes, L., Vandemeulebroecke, K., Peeters, E., Su, Y., Lee, H., Lorenzi, P. L., Chan, W.-K., Mondelaers, V., De Moerloose, B., Lammens, T., Goossens, S., Van Vlierberghe, P. & Lavie, A. (2022). *Haematol*, **108**, 409–419.
Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., Verma, V., Keedy, D. A., Hintze, B. J., Chen, V. B., Jain, S., Lewis, S. M., Arendall, W. B., Snoeyink, J., Adams, P. D., Lovell, S. C., Richardson, J. S. & Richardson, J. S. (2018). *Protein Sci.* **27**, 293–315.
Wlodawer, A. (2017). *Methods Mol. Biol.* **1607**, 595–610.
Wlodawer, A., Dauter, Z., Shabalin, I. G., Gilski, M., Brzezinski, D., Kowiel, M., Minor, W., Rupp, B. & Jaskolski, M. (2020). *FEBS J.* **287**, 3703–3718.
wwPDB Consortium (2018). *Nucleic Acids Res.* **47**, D520–D528.
Yao, M., Yasutake, Y., Morita, H. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 294–301.
Yun, M. K., Nourse, A., White, S. W., Rock, C. O. & Heath, R. J. (2007). *J. Mol. Biol.* **369**, 794–811.
Zhou, Y., Jiao, L., Shen, J., Chi, H., Lu, Z., Liu, H., Lu, F. & Zhu, P. (2022). *Int. J. Mol. Sci.* **23**, 9663.