ACTA ABP BIOCHIMICA POLONICA

Vol. 68, No 1/2021
29–31
https://doi.org/10.18388/abp.2020_5501

Communication

# Serendipitous crystallization of *E. coli* HPII catalase, a sequel to "the tale usually not told"*

Marta Grzechowiak[1], Bartosz Sekula[2], Mariusz Jaskolski[1,3] and Milosz Ruszkowski[1]✉

[1]Center for Biocrystallographic Research, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznań, Poland; [2]Synchrotron Radiation Research Section of MCL, National Cancer Institute, Argonne, U.S.A.; [3]Department of Crystallography, Faculty of Chemistry, A. Mickiewicz University, Poznań, Poland

**Protein crystallographers are well aware of the trap of crystallizing *E. coli* proteins instead of the macromolecule of interest if heterologous recombinant protein expression in *E. coli* was part of the experimental pipeline. Among the well-known culprits are YodA metal-binding lipocalin (25 kDa) and YadF carbonic anhydrase (a tetramer of 25 kDa subunits). We report a novel crystal form of another such culprit, *E. coli* HPII catalase, which is a tetrameric protein of ~340 kDa molecular weight. HPII is likely to contaminate recombinant protein samples, co-purify, and then co-crystallize with the target proteins, especially if their masses in size exclusion chromatography are ~300–400 kDa. What makes this case more interesting but also parlous, is the fact that HPII can crystallize from very low concentrations, even well below 1 mg/mL.**

**Keywords**: crystal growth, crystallization artifact; impurities; contaminations; *E. coli* proteins

**Received**: 23 September, 2020; revised: 23 October, 2020; accepted: 15 November, 2020; available on-line: 24 January, 2021

*Dedicated as a Birthday tribute to Prof. Wladek Minor who has blazed more than one trail in protein crystallography

✉e-mail: mruszkowski@ibch.poznan.pl

**Acknowledgments of Financial Support**: This work was supported by the National Science Centre grants SONATA 2018/31/D/NZ1/03630 to M. Ruszkowski and SYMFONIA 2016/20/W/ST5/00478 to M. Jaskolski, and by the Intramural Research Program of the NCI, Center for Cancer Research.

**Abbreviations**: AtDHS, *Arabidopsis thaliana* deoxyhypusine synthase; AtGDH2, *Arabidopsis thaliana* glutamate dehydrogenase 2; HPII, *Escherichia coli* catalase (alternative name hydroxyperoxidase II); MPD, 2-Methyl-2,4-pentanediol; MR, molecular replacement; PDB, Protein Data Bank; PEG, polyethylene glycol; Rmsd, root mean square deviation; SDS-PAGE, sodium dodecyl sulphate–polyacrylamide gel electrophoresis; SEC, size-exclusion chromatography; TEV, Tobacco Etch Virus

As of July 2020, there were ~166,000 macromolecular structures deposited in the Protein Data Bank (PDB) (Berman *et al.*, 2000). Most of them come from X-ray diffraction studies, as a result of meticulous procedures involving recombinant protein production, purification, crystallization, and X-ray diffraction that together form experimental basis for elucidation of three-dimensional atomic models. Heterologous protein expression is usually carried out in *Escherichia coli* cells. During purification, the protein of interest is separated from the host proteins, typically with the use of various chromatographic techniques. Unfortunately, some impurities are notoriously present in the protein samples that are used for crystallization. In exceptional instances these contaminant proteins happen to crystallize instead of the protein of interest. Such cases have been summarized by Niedzialkowska and others (Niedzialkowska *et al.*, 2016) in their paper "The tale usually not told", but other examples also exist in the literature (van Eerde *et al.*, 2006; Zaitseva *et al.*, 2009; Keegan *et al.*, 2016).

In our studies, carried out in several laboratories, our targets were the structures of two unrelated *Arabidopsis thaliana* proteins, glutamate dehydrogenase (*At*GDH2) and deoxyhypusine synthase (*At*DHS). Having obtained several morphologically different crystal forms in these projects, we collected high quality X-ray diffraction data (Table 1) for what appeared to be easy molecular replacement (MR) (Rossmann, 1990) problems. However, in both cases, despite the availability of very good models, all our numerous MR trials have failed. We then investigated the unit cell parameters of our crystal forms, to find that two of them (6ZTV, 6ZTX) were within a 3% margin of those reported for *E. coli* catalase HPII (Uniprot ID: P21179). A third crystal form (6ZTW) had different unit cell parameters but the structure could also be solved instantaneously with the model of *E. coli* HPII catalase (Table 1).

*E. coli* HPII is a homotetrameric enzyme with 222 symmetry, comprised of four 753-residue subunits (Bravo *et al.*, 1995). Each subunit contains a *cis*-heme *d* prosthetic group. Interestingly, there are 45 structures of *E. coli* HPII catalase in six crystal forms in the PDB (Table 2). It must be noted that our pipeline for protein purification in both projects involved Ni affinity chromatography, $His_6$-tag cleavage with His-tagged TEV protease, elimination of TEV protease and impurities by a second run of the Ni column, and finally size-exclusion chromatography (SEC). Despite this multistep procedure, a substantial residual amount of HPII remained in all samples, indicating that HPII may interact strongly enough with various proteins of interest to pass with them through all purification procedures.

Both, *At*GDH2 (~45 kDa per subunit) and *At*DHS (~41 kDa subunits) form oligomers, with total molecular weight of ~270 and ~170 kDa, respectively. The molecular weight of the *E. coli* HPII tetramer is ~340 kDa. This indicates that special caution must be used when pooling SEC fractions corresponding to that mass range, as they may be contaminated with HPII. It is also important to note that when we attempted to crystalize *At*GDH2, the total protein concentration was ~4 mg/mL. Considering that *At*GDH2 was clearly a dominating band on SDS-PAGE (not shown), the concentration of HPII must have been well below 1 mg/mL. One must conclude, therefore, that *E. coli* HPII can be easily crystallized from very low concentrations, supporting the findings of Simpkin and others (Simpkin *et al.*, 2018).

**Table 1. Data collection and refinement statistics.**

| PDB ID | 6ZTV | 6ZTW | 6ZTX |
|---|---|---|---|
| Data collection | | | |
| Beamline | PETRA P13 | PETRA P13 | APS 22-ID |
| Wavelength (Å) | 1.03 | 1.03 | 1.00 |
| Space group | $P2_1$ | $P2_1$ | $P2_1$ |
| Unit cell parameters<br>*a, b, c* (Å)<br>β (°) | <br>73.78, 172.26, 123.18<br>104.52 | <br>121.19, 168.07, 137.98<br>105.24 | <br>93.21, 132.93, 121.65<br>109.65 |
| Resolution (Å) | 80-1.78 (1.89-1.78)[a] | 80-1.84 (1.95-1.84) | 50-1.30 (1.38-1.30) |
| Unique reflections | 282183 (45601) | 452364 (72060) | 670790 (106698) |
| Multiplicity | 4.6 (4.6) | 4.7 (4.5) | 4.8 (4.9) |
| Completeness (%) | 99.4 (99.5) | 98.4 (97.1) | 98.3 (96.9) |
| $R_{merge}$ (%) | 10.0 (80.2) | 12.4 (68.5) | 10.2 (110.1) |
| $<I/\sigma(I)>$ | 11.5 (1.8) | 9.9 (2.0) | 11.8 (2.0) |
| Program used for data processing | XDS (Kabsch, 2010) | | |
| Refinement | | | |
| Program used for MR | Phaser (McCoy *et al.*, 2007) | | |
| Program used for refinement | Phenix.refine (Afonine *et al.*, 2012) | | |
| $R_{free}$ reflections | 1411 | 2260 | 2638 |
| No. of atoms (non-H) | | | |
| protein | 23009 | 46122 | 23350 |
| ligands | 186 | 496 | 228 |
| solvent | 2077 | 4695 | 3309 |
| $R_{work}/R_{free}$ (%) | 18.9/22.9 | 14.4/18.4 | 13.5/16.2 |
| Average B-factor (Å$^2$) | 31.6 | 22.2 | 16.8 |
| Rmsd from ideal geometry | | | |
| bond lengths (Å) | 0.006 | 0.007 | 0.005 |
| bond angles (°) | 1.09 | 1.09 | 1.08 |
| Ramachandran statistics (%) | | | |
| favored | 97 | 98 | 98 |
| allowed | 3 | 2 | 2 |
| outliers | 0 | 0 | 0 |

[a]Values in parentheses refer to the highest-resolution shell

**Table 2. Crystal lattice parameters for *E. coli* HPII catalase.**

| Space group | Unit cell parameters (Å; °) | Number of PDB instances | PDB IDs |
|---|---|---|---|
| *P*1 | 69, 90, 115;<br>107, 106, 96 | 1 | 6BY0 |
| $P2_1$ | 94, 133, 122;<br>90, 109, 90 | 38 | 1CF9, 1GG9, 1GGE, 1GGF, 1GGH, 1GGJ, 1GGK, 1IPH, 1P7Y, 1P7Z, 1P80, 1P81, 1QF7, 1QWS, 3P9P, 3P9Q, 3P9R, 3P9S, 3PQ2, 3PQ3, 3PQ4, 3PQ5, 3PQ6, 3PQ7, 3PQ8, 3TTT, 3TTU, 3TTV, 3TTW, 3TTX, 4ENP, 4ENQ, 4ENR, 4ENS, 4ENT, 4ENU, 4ENV, 4ENW, **6ZTX**[a] |
| $P2_1$ | 73, 172, 123;<br>90, 105, 90 | 2 | 4BFL, **6ZTV**[b] |
| $P2_1$ | 121, 168, 138;<br>90, 105, 90 | 1 | **6ZTW**[c] |
| *C*2 | 162, 171, 122;<br>90, 122, 90 | 1 | 5BV2 |
| *I*222 | 136, 159, 167;<br>90, 90, 90 | 1 | 3VU3 |

Structures presented in this work are in bold. The crystallization conditions were as follows: [a]0.2 M $Li_2SO_4$, 0.1 M Tris pH 8.5, 25% w/v polyethylene glycol 3350; [b]0.2 M NaCl, 0.1 M Tris pH 7.5, 20% w/v PEG 4000, 10% MPD; [c]0.1 M Tris pH 7.5, 20% w/v PEG 4000.

We hope that this note will save time, effort, and resources when phasing X-ray data that do not correspond to the protein of interest. In cases of inexplicable MR difficulties, we recommend screening the PDB for unit cell parameters with a 5% margin, and using the hit protein models for MR. When the protein of interest has the molecular weight in its quaternary structure of ~300–400 kDa, it might be a good idea to try *E. coli* HPII first. In other cases, one might run a software pipeline, such as SIMBAD (Simpkin *et al.*, 2018) or ContaMiner (Hungler *et al.*, 2016) that can analyze unit cell parameters and suggest an isomorphous structure of a contaminant protein for MR. One of the structures presented in this work represents a new crystal form of *E. coli* HPII, not reported to date. It provides an important additional piece of information for data mining which will improve future lattice-parameter searches of isomorphous structures in the PDB, as models for MR trials.

## Acknowledgements

Diffraction data for the 6ZTX structure was collected at the Advanced Photon Source (APS), Argonne National Laboratory (ANL), at the SER-CAT beamline 22-ID supported by the U.S. Department of Energy, Office of Basic Energy Sciences, under contract W-31-109-Eng-38. For 6ZTV and 6ZTW structures, the synchrotron MX data was collected at beamline P13 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) (Cianci *et al.*, 2017).

## REFERENCES

Afonine PV, Grosse-Kunstleve RW, Echols N, Headd JJ, Moriarty NW, Mustyakimov M, Terwilliger TC, Urzhumtsev A, Zwart PH, Adams PD (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**: 352–367. https://doi.org/10.1107/S0907444912001308

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The Protein Data Bank. *Nucleic Acids Res* **28**: 235–242. https://doi.org/10.1093/nar/28.1.235

Bravo J, Verdaguer N, Tormo J, Betzel C, Switala J, Loewen PC, Fita I (1995) Crystal structure of catalase HPII from *Escherichia coli*. *Structure* 3: 491–502. https://doi.org/10.1016/S0969-2126(01)00182-4

Cianci M, Bourenkov G, Pompidor G, Karpics I, Kallio J, Bento I, Roessle M, Cipriani F, Fiedler S, Schneider TR (2017) P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. *J Synchrotron Radiat* **24**: 323–332. https://doi.org/10.1016/S0969-2126(01)00182-4

Hungler A, Momin A, Diederichs K, Arold ST (2016) ContaMiner and ContaBase: a webserver and database for early identification of unwantedly crystallized protein contaminants. *J Appl Crystallogr* **49**: 2252–2258. https://doi.org/10.1107/S1600576716014965

Kabsch W (2010) Xds. *Acta Crystallogr D Biol Crystallogr* **66**: 125–132. https://doi.org/10.1107/S0907444909047337

Keegan R, Waterman DG, Hopper DJ, Coates L, Taylor G, Guo J, Coker AR, Erskine PT, Wood SP, Cooper JB (2016) The 1.1 Å resolution structure of a periplasmic phosphate-binding protein from Stenotrophomonas maltophilia: a crystallization contaminant identified by molecular replacement using the entire Protein Data Bank. *Acta Cryst D* **72**: 933–943. https://doi.org/10.1107/S2059798316010433

Mccoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658–674. https://doi.org/10.1107/S0021889807021206

Niedzialkowska E, Gasiorowska O, Handing KB, Majorek KA, Porebski PJ, Shabalin IG, Zasadzinska E, Cymborowski M, Minor, W (2016) Protein purification and crystallization artifacts: The tale usually not told. *Protein Sci* **25**: 720–733. https://doi.org/10.1002/pro.2861

Rossmann MG (1990) The molecular replacement method. *Acta Cryst A* **46** (Pt 2): 73–82. https://doi.org/10.1107/S0108767389009815

Simpkin AJ, Simkovic F, Thomas JMH, Savko M, Lebedev A, Uski V, Ballard C, Wojdyr M, Wu R, Sanishvili R, Xu Y, Lisa MN, Buschiazzo A, Shepard W, Rigden DJ, Keegan RM (2018) SIMBAD: a sequence-independent molecular-replacement pipeline. *Acta Cryst D* **74**: 595–605. https://doi.org/10.1107/S2059798318005752

Van Eerde A, Wolterink-Van Loo S, Van Der Oost J, Dijkstra BW (2006) Fortuitous structure determination of 'as-isolated' Escherichia coli bacterioferritin in a novel crystal form. *Acta Cryst F* **62**: 1061–1066. https://doi.org/10.1107/S1744309106039583

Zaitseva J, Meneely KM, Lamb AL (2009) Structure of *Escherichia coli* malate dehydrogenase at 1.45 A resolution. *Acta Cryst F* **65**: 866–869. https://doi.org/10.1107/S1744309109032217