AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Bacteriology | Announcement

Structural genomics of bacterial drug targets: Application of a high-throughput pipeline to solve 58 protein structures from pathogenic and related bacteria

Nicole L. Inniss,[1,2] George Minasov,[1,2] Changsoo Chang,[2,3,4] Kemin Tan,[2,3,4] Youngchang Kim,[2,3,4] Natalia Maltseva,[2,3,4] Peter Stogios,[2,5] Ekaterina Filippova,[1,2] Karolina Michalska,[2,3,4] Jerzy Osipiuk,[2,3,4] Lukasz Jaroszewki,[2,6] Adam Godzik,[2,6] Alexei Savchenko,[2,7] Andrzej Joachimiak,[2,3,4] Wayne F. Anderson,[2,8] Karla J. F. Satchell,[1,2] the Center for Structural Biology of Infectious Diseases Team members

AUTHOR AFFILIATIONS See affiliation list on p. 4.

ABSTRACT Antibiotic resistance remains a leading cause of severe infections worldwide. Small changes in protein sequence can impact antibiotic efficacy. Here, we report deposition of 58 X-ray crystal structures of bacterial proteins that are known targets for antibiotics, which expands knowledge of structural variation to support future antibiotic discovery or modifications.

KEYWORDS X-ray, structure, antimicrobial agents, drug targets, bacteria, PDB

Antibiotic-resistant bacteria remain a global threat, with millions of deaths attributed to decreased drug efficacy (1, 2). Amino acid variation across different bacterial species can impact antimicrobials targeting essential biochemical pathways. To support antimicrobial discovery or chemical modification of current antibiotics, the Center for Structural Genomics of Infectious Diseases (now the Center for Structural Biology of Infectious Diseases [CSBID]) established a high-throughput (HTP) structural genomics pipeline to expand the diversity of structures available for proteins that are known drug targets. A list of proteins representing known antibiotic targets was curated using DrugBank (http://www.drugbank.ca/). The protein sequences were used as queries to identify homologs in bacterial species with genomic DNA available in the center repository. Proteins sharing at least 50% sequence identity across 75% of the protein sequence were selected. In total, 630 targets from 47 bacterial species entered the pipeline.

All targets were subjected to automated analyses supporting protein expression construct design. The genes encoding the selected proteins or protein domains were amplified by PCR using genomic DNA as a template. The PCR products were cloned into pMCSG53 (PSI:Biology-Materials Repository, http://psimr.asu.edu) according to published ligation-independent cloning procedures (3, 4). This vector introduced a protease-cleavable, N-terminal hexa-histidine purification tag. The clones were transformed into T7-polymerase expressing *Escherichia coli* strains and tested for expression and solubility. Soluble proteins were purified by nickel affinity chromatography according to published protocols (5, 6), and concentrated proteins were set up as 2-µL crystallization drops in 96-well plates using multiple screens. Resulting crystals were cryoprotected, cooled, and then screened for data collection at the Advanced Photon Source (APS) at Argonne National Laboratory.

In total, 24% of targets were purified, and 19% yielded protein preparations that entered HTP crystallization screens. Pipeline success rate from selection through structure determination was 7.6%. Forty-eight targets from 24 bacterial species produced high-quality crystals, yielding 58 structures (Fig. 1). The RCSB Protein Data

Editor Irene L. G. Newton, Indiana University, Bloomington, Bloomington, Indiana, USA

Address correspondence to Karla J. F. Satchell, k-satchell@northwestern.edu.

The authors declare no conflict of interest.

See the funding table on p. 5.

Received 26 February 2025
Accepted 3 April 2025
Published 20 May 2025

TABLE 1 Summary of bacterial drug targets with structures deposited to Protein Data Bank[a]

PDB code	Csbid Ref #	Protein name	Organism	Resolution	Ligand[b]
6nbk	IDP07367	Arginase	*Bacillus cereus*	1.91 Å	–
6nfp	IDP07164	Arginase	*Bacillus subtilis*	1.70 Å	–
5us8	IDP07200	Argininosuccinate synthase	*Bordetella pertussis*	2.15 Å	Adenosine
6e5y	iDP07200	Argininosuccinate synthase	*Bordetella pertussis*	1.50 Å	AMP
6w2z	IDP07475	Beta-lactamase class A	*Bacillus subtilis*	1.50 Å	Avibactam
9bzn	IDP07519	Beta-lactamase class A	*Bordetella bronchiseptica*	1.05 Å	–
9bzq	IDP07519	Beta-lactamase class A	*Bordetella bronchiseptica*	1.47 Å	Avibactam
9bzr	IDP07519	Beta-lactamase class A	*Bordetella bronchiseptica*	1.40 Å	Clavulanate
6pua	IDP07511	Chloramphenicol acetyltransferase	*Vibrio cholerae*	2.00 Å	–
5ux9	IDP07301	Chloramphenicol acetyltransferase	*Vibrio fischeri*	2.70 Å	–
6pxa	IDP07301	Chloramphenicol acetyltransferase	*Vibrio fischeri*	1.82 Å	Taurocholic acid
6pu9	IDP07511	Chloramphenicol acetyltransferase	*Vibrio vulnificus*	1.70 Å	–
6b5f	IDP07570	CobT	*Yersinia enterocolitica*	1.95 Å	–
6azi	IDP07508	D-ala-D-ala-endopeptidase	*Enterobacter cloacae*	1.75 Å	–
6bz0	IDP07418	Dihydrolipoamide dehydrogenase	*Acinetobacter baumannii*	1.83 Å	FAD
6aon	IDP07182	Dihydrolipoamide dehydrogenase	*Bordetella pertussis*	1.72 Å	FAD
6cmz	IDP07673	Dihydrolipoamide dehydrogenase	*Burkholderia cenocepacia*	2.30 Å	FAD, NAD
6awa	IDP07540	Dihydrolipoamide dehydrogenase	*Pseudomonas aeruginosa*	1.83 Å	FAD, AMP
5tr3	IDP07540	Dihydrolipoamide dehydrogenase	*Pseudomonas putida*	2.50 Å	FAD
5umg	IDP07170	Dihydropteroate synthase	*Klebsiella pneumoniae*	2.60 Å	–
5usw	IDP07359	Dihydropteroate synthase	*Vibrio fischeri*	1.64 Å	–
6bq9	IDP07285	DNA Topoisomerase IV Subunit A	*Pseudomonas putida*	2.55 Å	–
–5vh6	IDP07716	Elongation factor G	*Bacillus subtilis*	2.61 Å	–
6bk7	IDP07555	Elongation factor G	*Enterococcus faecalis*	1.83 Å	–
6b8d	IDP07537	Elongation factor G	*Haemophilus influenzae*	1.78 Å	–
5ty0	IDP07381	Elongation factor G	*Legionella pneumophila*	2.22 Å	–
6n0i	IDP07336	Elongation factor G	*Pseudomonas putida*	2.60 Å	–
5tv2	IDP07581	Elongation factor G	*Vibrio vulnificus*	1.60 Å	–
6b4o	IDP07317	Glutathione reductase	*Enterococcus faecalis*	1.73 Å	FAD
5v36	IDP07311	Glutathione reductase	*Streptococcus mutans*	1.88 Å	FAD
6n7f	IDP07597	Glutathione reductase	*Streptococcus pyogenes*	1.90 Å	–
5u1o	IDP07224	Glutathione reductase	*Vibrio parahaemolyticus*	2.31 Å	FAD
5vdn	IDP07394	Glutathione reductase	*Yersinia pestis*	1.55 Å	FAD
6aoo	IDP07201	Malate dehydrogenase	*Haemophilus influenzae*	2.15 Å	–
6bal	IDP07201	Malate dehydrogenase	*Haemophilus influenzae*	2.10 Å	L-malate
5vfb	IDP07567	Malate synthase G	*Pseudomonas aeruginosa*	1.36 Å	Glycolytic acid
5ume	IDP07318	MetF	*Haemophilus influenzae*	2.70 Å	FAD
6po4	IDP07178	Methylthioadenosine/SAH nucleosidase	*Haemophilus influenzae*	2.10 Å	–
5ue1	IDP07462	Methylthioadenosine/SAH nucleosidase	*Vibrio fischeri*	1.14 Å	Adenine
6muq	IDP07205	Murein-DD-endopeptidase	*Yersinia enterocolitica*	1.67 Å	–
6c8q	IDP07348	NAD synthetase	*Enterococcus faecalis*	2.58 Å	NAD
5wp0	IDP07110	NAD synthetase	*Vibrio fischeri*	2.60 Å	–
5uu6	IDP07628	Nitroreductase A	*Vibrio parahaemolyticus*	1.95 Å	FMN
6czp	IDP07377	Nitroreductase A	*Vibrio vulnificus*	2.24 Å	FMN
6dll	IDP07306	p-Hydroxybenzoate Hydroxylase	*Pseudomonas putida*	2.20 Å	FAD
5u2g	IDP07344	Penicillin-binding protein 1A	*Haemophilus influenzae*	2.61 Å	–
5u47	IDP07211	Penicillin-binding protein 2X	*Streptococcus thermophilus*	1.95 Å	–
6blb	IDP07228	RuvB	*Pseudomonas aeruginosa*	1.88 Å	ADP
5u63	IDP07488	Thioredoxin reductase	*Haemophilus influenzae*	1.99 Å	–
5uwy	IDP07356	Thioredoxin reductase	*Streptococcus pyogenes*	2.72 Å	FAD
5utx	IDP07222	Thioredoxin reductase	*Vibrio vulnificus*	2.46 Å	–

(*Continued on next page*)

TABLE 1 Summary of bacterial drug targets with structures deposited to Protein Data Bank[a] (*Continued*)

PDB code	Csbid Ref #	Protein name	Organism	Resolution	Ligand[b]
5usx	IDP07222	Thioredoxin reductase	*Vibrio vulnificus*	2.60 Å	NADP, FAD
5vt3	IDP07222	Thioredoxin reductase	*Vibrio vulnificus*	1.98 Å	NADP, FAD
5v0i	IDP07325	Tryptophanyl-tRNA synthetase	*Escherichia coli*	1.90 Å	Tryptophan, AMP
6dfu	IDP07216	Tryptophanyl-tRNA synthetase	*Haemophilus influenzae*	2.05 Å	–
6cn1	IDP07215	UDP-GlcNAc 1-carboxyvinyltransferase	*Pseudomonas putida*	2.75 Å	UDP-GlcNAc
6nkj	IDP07236	UDP-GlcNAc 1-carboxyvinyltransferase	*Streptococcus pneumoniae*	1.30 Å	–
5wi5	IDP07236	UDP-GlcNAc 1-carboxyvinyltransferase	*Streptococcus pneumoniae*	2.00 Å	UDP-GlcNAc

[a]Access link for Data quality and refinement statistics 10.5281/zenodo.15224721.
[b]–, indicates no ligand.

Bank (PDB) deposition code, protein name, source DNA, and refinement statistics are listed in Table 1. Of the 58 structures determined, 55 are reported here for the first time, with three structures published previously (7, 8). Structures were derived from proteins involved in antibiotic modification, cell wall maintenance, oxidative stress, and metabolism.

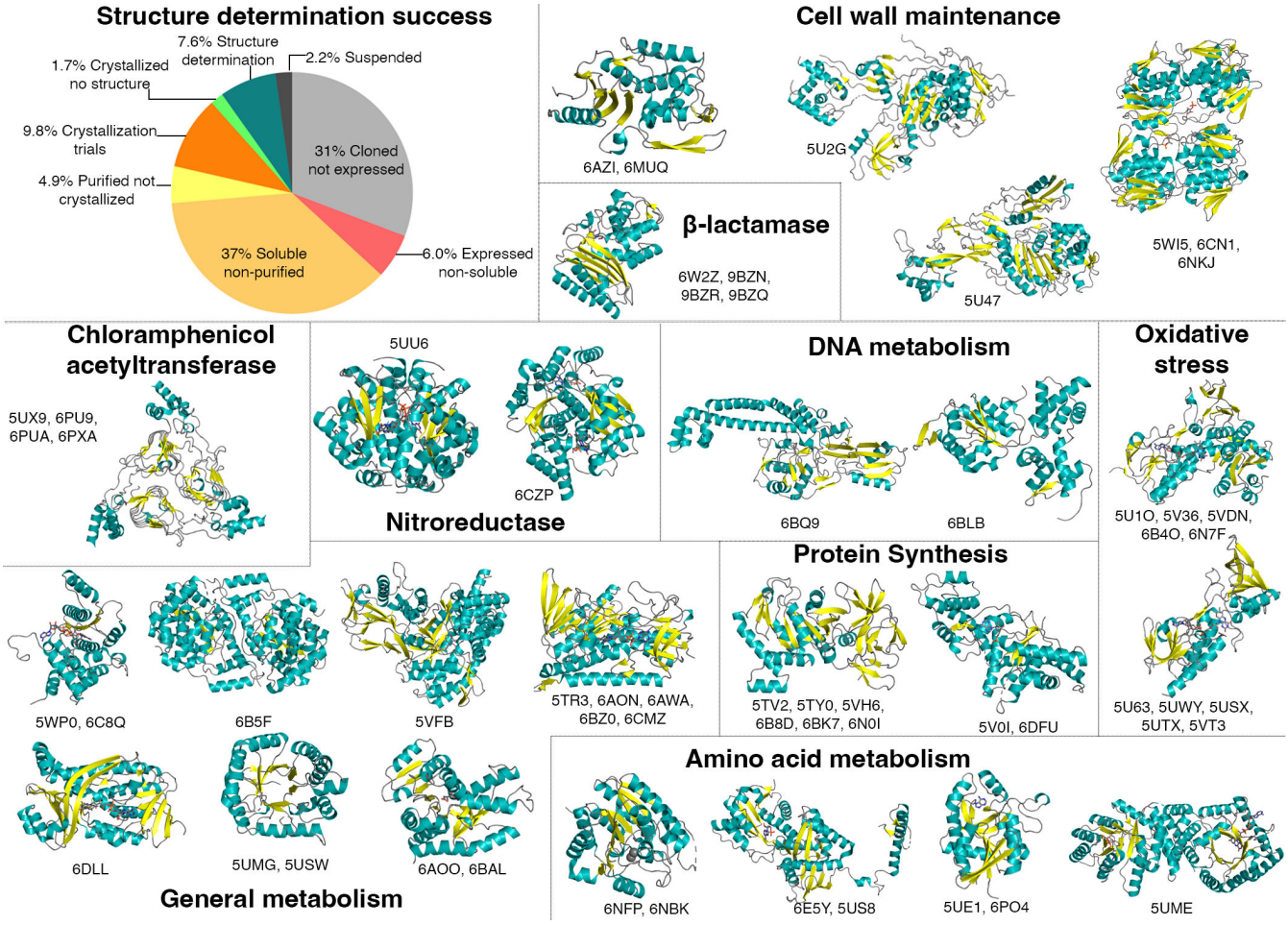

FIG 1 Percentage of approved bacterial drug targets at each stage in the structure determination pipeline and representative X-ray structures. The pie chart shows the overall success rate of proteins in the structure determination pipeline from a total of 630 targets. Work was completed between 2016 and 2024. Twenty-five representative structures are depicted as cartoons: β-sheets are colored yellow, α-helices are teal, and loops are gray. Associated crystal variants, complexes with ligands, and homologous structures are annotated below each image, totaling 58 structures solved. The proteins were sorted according to their known function in bacteria. Associated ligands and crystallographic details are described in Table 1.

Data collection and data quality information are available on the PDB. Structures of proteins grown in selenomethionine medium were solved by single-wavelength anomalous diffraction method, using the Automatic Structure Solution from HKL-3000 (9) and Auto-build package from PHENIX (10). Structures of native proteins were solved by molecular replacement in the CCP4 suite (11). Diffraction data were used for structure solution using either the structure of the closest sequence homolog in the PDB in PHASER or the target protein sequence using MORDA and MRBUMP. Structures were refined using REFMAC5 (12) or PHENIX and visually corrected in Coot (13). Water molecules were generated using ARP/wARP (14), and ligands were fit into electron density maps in Coot. Translation–Libration–Screw groups were generated by the TLSMD server (15), and corrections were applied during refinement finalization. Models were validated using MolProbity (16).

ACKNOWLEDGMENTS

This project has been funded in whole or in part with Federal funds from the Department of Health and Human Services, NIH, and National Institute of Allergy and Infectious Diseases under contract no. HHSN272201200026C (to W.F.A), HHSN272201700060C, and 75N93022C00035 (to K.J.F.S). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02- 06CH11357. The use of Structural Biology Center beamlines is supported by U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817).

N.L.I. analyzed structures, drafted the manuscript, and prepared figures. G.M., K.T., C.C., Y.K., P.S., E.F., N.M., K.M., and J.O. are the first authors on the reported PDB deposits and are responsible for the quality of the structures. Structure author order is determined first by the number of structures deposited and then by alphabetical order. L.J. and A.G. curated the target list. A.S., A.J., W.A., and K.S. supervised projects and provided funding. Senior authors are listed in reverse order of the total number of structures in their group. K.S. edited the manuscript.

Members of the CSBID team involved in the technical execution of the project include personnel (in alphabetical order) from J. Craig Venter Institute (Sarah Grimshaw, Keewhan Kwon, Jason Stam); Northwestern University (Alvin Cardona-Correa, Ievgenia Dubrovska, Olga Kiryukhina, Misty Kuhn, Amanda Olphie, Ludmilla Shuvalova, Zdzislaw Wawrzak, James Winsor); University of Chicago/Argonne National Laboratory (Michael Endres, Robert Jedrzejczak, Rory Mulligan, Min Zhou), University of Toronto (Christopher McChesney, Elena Evdokimova, Tatiana Skarina), and University of Virginia (Marek Grabowski, Wladek Minor, Ivan G. Shabalin).

AUTHOR AFFILIATIONS

[1]Department of Microbiology-Immunology, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA
[2]Center for Structural Biology of Infectious Diseases, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA
[3]Consortium for Advanced Science and Engineering, University of Chicago, Chicago, Illinois, USA
[4]Structural Biology Center, X-ray Science Division, Argonne National Laboratory, Lemont, Illinois, USA
[5]Biozone, Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario, Canada
[6]Biosciences Division, University of California, Riverside, School of Medicine, Riverside, California, USA

[7]Department of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Calgary, Alberta, Canada
[8]Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA

AUTHOR ORCIDs

Nicole L. Inniss http://orcid.org/0000-0002-1510-7412
Kemin Tan http://orcid.org/0000-0002-4003-7903
Peter Stogios http://orcid.org/0000-0001-8663-1425
Alexei Savchenko http://orcid.org/0000-0002-5256-9237
Andrzej Joachimiak http://orcid.org/0000-0003-2535-6209
Karla J. F. Satchell http://orcid.org/0000-0003-3274-7611

FUNDING

Funder	Grant(s)	Author(s)
National Institute of Allergy and Infectious Diseases	75N93022C00035,HHSN272201700060C,HHSN272201200026C	Andrzej Joachimiak Wayne F. Anderson Karla J. F. Satchell
U.S. Department of Energy	DE-AC02- 06CH11357,DE-AC02-06CH11357	Andrzej Joachimiak Wayne F. Anderson Karla J. F. Satchell
Michigan Economic Development Corporation	085P1000817	Andrzej Joachimiak Wayne F. Anderson Karla J. F. Satchell
Michigan Technology Tri-Corridor	085P1000817	Andrzej Joachimiak Wayne F. Anderson Karla J. F. Satchell

AUTHOR CONTRIBUTIONS

Nicole L. Inniss, Data curation, Formal analysis, Project administration, Visualization, Writing – original draft | George Minasov, Data curation, Formal analysis | Changsoo Chang, Data curation, Formal analysis | Kemin Tan, Data curation, Formal analysis | Youngchang Kim, Data curation, Formal analysis | Natalia Maltseva, Data curation, Formal analysis | Peter Stogios, Data curation, Formal analysis | Ekaterina Filippova, Data curation, Formal analysis | Karolina Michalska, Data curation, Formal analysis | Jerzy Osipiuk, Data curation, Formal analysis | Lukasz Jaroszewki, Conceptualization, Data curation | Adam Godzik, Conceptualization, Data curation | Alexei Savchenko, Funding acquisition, Project administration, Supervision | Andrzej Joachimiak, Funding acquisition, Project administration | Wayne F. Anderson, Conceptualization, Funding acquisition, Project administration | Karla J. F. Satchell, Funding acquisition, Project administration, Writing – review and editing

DATA AVAILABILITY

The list of protein targets with solved structures is available at http://targets.csbid.org/targets organized as batch "set296" with the internal tracking number (IDP) listed in Table 1. The entire list of 630 protein targets including primer sequences and minor changes to protein purification can be found on the legacy database at csgid.org or by request from the Center for Structural Biology of Infectious Diseases. All coordinates for all final models and experimental data have been deposited to the Protein Data Bank (https://www.rcsb.org/), and can be found using PDB codes 5TR3, 5TV2, 5TY0, 5U1O, 5U2G, 5U47, 5U63, 5UE1, 5UME, 5UMG, 5US8, 5USW, 5USX, 5UTX, 5UU6, 5UWY, 5UX9, 5V0I, 5V36, 5VDN, 5VFB, 5VH6, 5VT3, 5WI5, 5WP0, 6AON, 6AOO, 6AWA, 6AZI, 6B4O, 6B5F, 6B8D, 6BAL, 6BK7, 6BLB, 6BQ9, 6BZ0, 6C8Q, 6CMZ, 6CN1, 6CZP, 6DFU, 6DLL, 6E5Y, 6MUQ, 6N0I, 6N7F, 6NBK, 6NFP, 6NKJ, 6PO4, 6PU9, 6PUA, 6PXA, 6W2Z, 9BZN, 9BZR, and 9BZQ.

REFERENCES

1. Antimicrobial Resistance Collaborators. 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. Lancet 399:629–655. https://doi.org/10.1016/S0140-6736(21)02724-0
2. CDC. 2019. Antibiotic resistance threats in the united states. Department of Health and Human Services, CDC, Atlanta, GA.
3. Eschenfeldt WH, Lucy S, Millard CS, Joachimiak A, Mark ID. 2009. A family of LIC vectors for high-throughput cloning and purification of proteins. Methods Mol Biol 498:105–115. https://doi.org/10.1007/978-1-59745-196-3_7
4. Stols L, Gu M, Dieckman L, Raffen R, Collart FR, Donnelly MI. 2002. A new vector for high-throughput, ligation-independent cloning encoding a tobacco etch virus protease cleavage site. Protein Expr Purif 25:8–15. https://doi.org/10.1006/prep.2001.1603
5. Shuvalova L. 2014. Parallel protein purification. Methods Mol Biol 1140:137–143. https://doi.org/10.1007/978-1-4939-0354-2_10
6. Makowska-Grzyska M, Kim Y, Maltseva N, Li H, Zhou M, Joachimiak G, Babnigg G, Joachimiak A. 2014. Protein production for structural genomics using *E. coli* expression. Methods Mol Biol 1140:89–105. https://doi.org/10.1007/978-1-4939-0354-2_7
7. Lazar JT, Shuvalova L, Rosas-Lemus M, Kiryukhina O, Satchell KJF, Minasov G. 2019. Structural comparison of p-hydroxybenzoate hydroxylase (PobA) from *Pseudomonas putida* with PobA from other *Pseudomonas* spp. and other monooxygenases. Acta Crystallogr F Struct Biol Commun 75:507–514. https://doi.org/10.1107/S2053230X19008653
8. Alcala A, Ramirez G, Solis A, Kim Y, Tan K, Luna O, Nguyen K, Vazquez D, Ward M, Zhou M, Mulligan R, Maltseva N, Kuhn ML. 2020. Structural and functional characterization of three Type B and C chloramphenicol acetyltransferases from *Vibrio* species. Protein Sci 29:695–710. https://doi.org/10.1002/pro.3793
9. Minor W, Cymborowski M, Otwinowski Z, Chruszcz M. 2006. HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta Crystallogr D Biol Crystallogr 62:859–866. https://doi.org/10.1107/S0907444906019949
10. Liebschner D, Afonine PV, Baker ML, Bunkóczi G, Chen VB, Croll TI, Hintze B, Hung LW, Jain S, McCoy AJ, Moriarty NW, Oeffner RD, Poon BK, Prisant MG, Read RJ, Richardson JS, Richardson DC, Sammito MD, Sobolev OV, Stockwell DH, Terwilliger TC, Urzhumtsev AG, Videau LL, Williams CJ, Adams PD. 2019. Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. Acta Crystallogr D Struct Biol 75:861–877. https://doi.org/10.1107/S2059798319011471
11. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AGW, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. 2011. Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr 67:235–242. https://doi.org/10.1107/S0907444910045749
12. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA. 2011. REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67:355–367. https://doi.org/10.1107/S0907444911001314
13. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of coot. Acta Crystallogr D Biol Crystallogr 66:486–501. https://doi.org/10.1107/S0907444910007493
14. Cohen SX, Ben Jelloul M, Long F, Vagin A, Knipscheer P, Lebbink J, Sixma TK, Lamzin VS, Murshudov GN, Perrakis A. 2008. ARP/wARP and molecular replacement: the next generation. Acta Crystallogr D Biol Crystallogr 64:49–60. https://doi.org/10.1107/S0907444907047580
15. Painter J, Merritt EA. 2006. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. Acta Crystallogr D Biol Crystallogr 62:439–450. https://doi.org/10.1107/S0907444906005270
16. Chen VB, Arendall WB 3rd, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr D Biol Crystallogr 66:12–21. https://doi.org/10.1107/S0907444909042073